



ÉVORA UNIVERSITY

DEPARTMENT OF MATHEMATICS

Automated Acoustic Identification of Bat Species

Bruno Miguel Santos Antunes Silva

Supervisor: Doctor Gonçalo João Costa Jacinto

Master in Statistical Modeling and Data Analysis

Thesis

Évora, 2013

Acronyms

ANN – Artificial Neural Networks

ASR – Automatic Speech Recognition

BP – Backpropagation Training Algorithm

CF – Constant Frequency

DFA – Discriminant Function Analysis

ENN – Ensembles of Artificial Neural Networks

FFT – Fast Fourier Transform

FM – Frequency Modulated

FMAXE – Frequency of Maximum Energy

GLM – Generalized Linear Models

GMM – Gaussian Mixture Models

GRPROP – Globally Convergent Resilient Backpropagation Training Algorithm

GUI – Graphical User Interface

HMM – Hidden Markov Models

LR – Logistic Regression

OR – Odds Ratio

RGW – Robust Generalized Weights

ROC – Receiver Operating Characteristic

RPROP – Resilient Backpropagation Training Algorithm

SSE – Sum of Squared Errors

SVM – Support Vector Machines

Abstract

Automated acoustic identification of bat species

Recent improvements in bat survey methods in Portugal, especially automatic recording stations, have led to an analysis problem due to the amount of data obtained. In this thesis we propose to develop an automated analysis and classification method for bat echolocation calls by developing a computer program based on statistical models and using a reference database of bat calls recorded in Portugal to quickly analyze and classify large amounts of recordings.

We recorded 2968 calls from 748 bats of 20 (of the 25) bat species known in mainland Portugal and coded a program in R that automatically detects bat calls in a recording, isolates the calls from the background noise and measures 19 parameters from each call.

A two stage hierarchical classification bat call scheme was implemented based on logistic regression models and ensembles of artificial neural networks. In the first stage calls were classified in six major groups with individual correct classification rates that varied between 93% and 100%. In the second stage calls were classified in species or groups of species with classification rates that varied between 50% and 100%.

Keywords: bat, echolocation, ensemble of artificial neural network, logistic regression model, species identification, R project

Resumo

Identificação acústica automatizada de espécies de morcegos

Desenvolvimentos recentes nas metodologias de monitorização de morcegos utilizadas em Portugal, especialmente estações de gravação automáticas, conduziram a um problema de análise devido à quantidade de dados obtida. Nesta tese 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 continental para rapidamente analisar e classificar grandes quantidades de gravações.

Gravámos 2968 pulsos de 748 morcegos de 20 (das 25) espécies de morcegos conhecidas em Portugal continental e codificámos em R um programa para automaticamente detectar pulsos de morcego numa gravação, isolar os pulsos do ruído de fundo e medir 19 parâmetros de cada pulso.

Foi implementado um esquema hierárquico de classificação de pulsos em duas etapas baseado em modelos de regressão logística e conjuntos de redes neuronais artificiais. Numa primeira etapa os pulsos foram classificados em seis grupos com taxas individuais de classificações correctas que variaram entre 93% e 100%. Numa segunda fase os pulsos foram classificados em espécies ou grupos de espécies com taxas de classificação correctas que variaram entre 50% e 100%.

Palavras-chave: morcego, ecolocalização, conjunto de redes neuronais artificiais, modelo de regressão logística, identificação de espécies, R project

ACKNOWLEDGMENTS

To Gonçalo Jacinto, my supervisor, for his unwavering support and encouragement along this long and difficult path

To Ana Rainho for her comments and clarifications that greatly improved this project and my own knowledge

To Bat Conservation International, Inc for funding this project through a Student Research Scholarship

To Plecotus – Estudos Ambientais Unip, Lda for providing the technical equipment that made this project a reality

To Pettersson Elektronik AB for providing a D1000x bat detector

To Russell Alpizar Jara and Anabela Afonso for their support with the sampling design

To Paulo Infante for his willingness to help with the statistical models

To Luisa Rodrigues for her help and support with the sampling methods and field work

To Sílvia Barreiro for embracing this adventure as her own and keeping me sane along the way

To Francisco Amorim, Hugo Rebelo and Pedro Alves that contributed greatly to the acquisition of field recordings

To Marta Borges for providing a photographic record of field work

To Jens Rydell and Stephan Pettersson for their encouragement and help in the initial stage of this project

To all the innumerable people that helped and assisted in the field that weren't mentioned before

To my grandmother and my parents who are my safety net and gave me the opportunity to learn the skills that were so important throughout my life

To Sílvia Barreiro my best friend and life companion who always supported me even in the darkest hours

To Nenas and Cuca that gave me the unconditional support I needed when things got really difficult along this journey

To all my friends and family without whom life would make a lot less sense

Index

1.	INTRODUCTION.....	11
1.1	BATS IN PORTUGAL	12
1.2	ECHOLOCATION CALLS: CHARACTERISTICS.....	14
1.3	ECHOLOCATION CALLS: SPECIES IDENTIFICATION.....	15
1.4	MOTIVATION.....	16
1.5	OBJECTIVES.....	17
1.6	RELATED WORK.....	18
1.7	CLAIM OF CONTRIBUTIONS	22
1.8	OVERVIEW OF THE DISSERTATION.....	23
2.	STATISTICAL BACKGROUND.....	24
2.1	GENERALIZED LINEAR MODELS	24
2.1.1	LOGISTIC REGRESSION.....	24
2.1.2	MODEL SELECTION.....	25
2.1.3	MODEL VALIDATION	27
2.1.4	MODEL INTERPRETATION.....	28
2.2	ARTIFICIAL NEURAL NETWORKS	29
2.2.1	MULTILAYER FEEDFORWARD PERCEPTRON NETWORKS	30
2.2.2	TRAINING	31
2.2.3	ARCHITECTURE SELECTION AND VALIDATION.....	34
2.2.4	MODEL INTERPRETATION.....	34
3.	METHODS.....	36
3.1	SAMPLING.....	36
3.2	DETECTING AND ISOLATING BAT CALLS	42
3.3	PARAMETER MEASUREMENT.....	45
3.4	BAT CALL CLASSIFICATION.....	49
3.5	STATISTICAL MODELING WITH R.....	51
3.5.1	LOGISTIC REGRESSION.....	51
3.5.2	MULTILAYER FEEDFORWARD PERCEPTRON NETWORKS	51
4.	RESULTS & DISCUSSION	54
4.1	CALL DATABASE	54
4.2	CALL CLASSIFICATION	59
4.2.1	CLASSIFICATION OF MAJOR GROUPS OF SPECIES	59
4.2.2	CLASSIFICATION OF <i>RHINOLOPHUS</i> SPECIES.....	62

4.2.3	CLASSIFICATION OF <i>MYOTIS</i> SPECIES	63
4.2.4	CLASSIFICATION OF <i>PIPISTRELOID</i> SPECIES	70
4.2.5	CLASSIFICATION OF <i>NYCTALOID</i> SPECIES	78
5	CONCLUSIONS & FUTURE WORK	85
	BIBLIOGRAPHY	88
	APPENDIX 1 – IMPLEMENTED FUNCTIONS IN R	94

TABLE INDEX

Table 1 – Bat species present in mainland Portugal and respective conservation status: CR – “Critically Endangered”, EN – “Endangered”, VU – “Vulnerable”, NT – “Near Threatened”, LC – “Least Concern”, DD – “Data Deficient” (Cabral <i>et al.</i> , 2006).	13
Table 2 – Interpretation of AUC values	28
Table 3 – Maximum and minimum values of f_{max} in literature for each species and the respective calculated reference sample size for each species (Russ, 1999; Russo & Jones, 2002; Zingg, 1990).....	37
Table 4 – Parameters measured from bat calls for classification purposes	46
Table 5 – Description of the measured parameters from bat calls for classification purposes ...	47
Table 6 – Number of cases in each dataset used for ENN training and validation.....	52
Table 7 – Number of recordings and respective number of calls, number of recording sites and number of recordings obtained in each recording condition (Hr – hand release; Re – roost entrance; Ff – free flying)	55
Table 8 – Descriptive statistics of the five commonly used parameters in manual bat call identification. Parameter mean and the 95% confidence interval is presented (Fini – initial frequency, F_{MAXE} – frequency of maximum energy, F_{fin} – final frequency, Bw – bandwidth of the call, Dur – duration of the call, N rec – number of recordings, N call – number of individual calls). * indicates values that differ considerably between our study and previously published studies	57
Table 9 – Classification results of the ENN for major groups classification (Myo – <i>Myotis</i> ; Pip – <i>Pipistrelloid</i> ; Nyc – <i>Nyctaloid</i> ; Bbar – <i>Barbastella</i> ; Plec – <i>Plecotus</i> ; Sens – Sensitivity; 1-Spec – 1-Specificity). Overall correct rate highlighted on the bottom right cell	62
Table 10 – Main effects model for <i>Myotis</i> species.....	64
Table 11 – Robust generalized weights for each covariate and output category of the ENN; means are presented in each cell (M35 – <i>M. myotis</i> / <i>M. blythii</i> ; M45 – <i>M. escalerae</i> / <i>M. daubentonii</i> / <i>M. emarginatus</i>). Most influent covariates on the log-odds are highlighted in bold	67
Table 12 – Classification results of the ENN for <i>Myotis</i> species classification (M35 – <i>M. myotis</i> / <i>M. blythii</i> ; M45 – <i>M. escalerae</i> / <i>M. daubentonii</i> / <i>M. emarginatus</i>). Overall correct rate highlighted on the bottom right cell.....	68
Table 13 – Main effects model for <i>Pipistrelloid</i> species.....	72
Table 14 – First binomial logistic model to compare <i>M. schreibersii</i> and <i>P. pipistrellus</i>	72
Table 15 – Second binomial logistic model to compare <i>M. schreibersii</i> and <i>P. pygmaeus</i>	73
Table 16 – Classification results of the multinomial logistic regression model for <i>Pipistrelloid</i> species (Msch – <i>M. schreibersii</i> ; Ppyg – <i>P. pygmaeus</i> ; Ppip – <i>P. pipistrellus</i>).....	75
Table 17 – Classification results of the ENN for <i>Pipistrelloid</i> species classification (Msch – <i>M. schreibersii</i> ; Pkuh – <i>P. kuhlii</i> ; Ppip – <i>P. pipistrellus</i> ; Ppyg – <i>P. pygmaeus</i>); Sens – Sensitivity; 1-Spec – 1-Specificity). Overall correct rate highlighted on the bottom right cell	78
Table 18 – Main effects model for <i>Nyctaloid</i> species.....	79
Table 19 – Classification results of the ENN for <i>Nyctaloid</i> species (Tten – <i>T. teniotis</i> ; NN – <i>N. leisleri</i> / <i>N. lasiopterus</i> ; EE – <i>E. serotinus</i> / <i>E. isabellinus</i>).....	84

FIGURE INDEX

Figure 1 – Diagram of a network with i covariates, 1 hidden layer with j neurons, k output neurons and bias on the hidden and output neurons.....	31
Figure 2 – Basic concept of the backpropagation algorithm illustrated for a univariate error function $E(w)$, from Günther & Fritsch (2010).....	32
Figure 3 – Location of the 39 recording sites in mainland Portugal with capital city (Lisbon) highlighted (Google Earth Image)	39
Figure 4 – Capture techniques used. Top left – harp trap at the entrance of a cave; top right – capturing bats inside a cave with hand-net; bottom left – mist nets surrounding a water tank; bottom right – mist nets covering a bridge over a stream of water.....	40
Figure 5 – Measuring bats after capture. Top left – releasing a captured bat from a mist net; top right – one of the camp sites; bottom – measurements of captured bats.....	41
Figure 6 – Recording of bat calls. Top left – bat taking off a tree surface; top right – recording of a hand released bat; bottom left – hand release of a bat with light tag; bottom right – recording of free flying bats outside a known roost	42
Figure 7 – Spectrogram (top) and energy envelope (100 points rollmean) of an <i>E. serotinus</i> call. Raw signal (left) and Teager energy transformed signal (right).	44
Figure 8 – Identification of the start point of the call (left) and end point of the call (right) with the binary segmentation method.....	45
Figure 9 – Schematic representation of the measured parameters that were used for classification purposes.....	47
Figure 10 – Screenshots of the implemented analysis GUI. Top – after importing and filtering a recording, a spectrogram and oscilogram is drawn and the calls are automatically identified (blue line). Bottom – after the parameter measurements, a spectrogram representation of the automatically extracted calls with the start (green line) and end (red line) identified and a table with the parameters measured for all calls, with their respective mean and median presented.	48
Figure 11 – Two stage identification scheme	50
Figure 12 – Histogram of f_{maxe} for <i>P. pygmaeus</i>	58
Figure 13 – Histogram of f_{maxe} for <i>M. schreibersii</i>	59
Figure 14 – Boxplot of call f_{maxe} for recorded bats with the cutpoint indicated. Results presented for groups of species for better visualization.....	60
Figure 15 – Sum of squared errors of the training and validation data by the different ENN architectures tested for major groups classification.....	61
Figure 16 – Boxplot of call f_{maxe} for <i>Rhinolophus</i> species (Rfer – <i>R. ferrumequinum</i> , Reur – <i>R. euryale</i> , Rmeh – <i>R. mehelyi</i> , Rhip – <i>R. hipposideros</i>).....	63
Figure 17 – Lowess representation of covariate ff_{in15}	64
Figure 18 – ROC curve of the final model for <i>Myotis</i> species	65
Figure 19 – Probabilities change of a call belonging to M35 or M45 (M35 – <i>M. myotis</i> / <i>M. blythii</i> ; M45 – <i>M. escalerae</i> / <i>M. daubentonii</i> / <i>M. emarginatus</i>). Left – $ff_{in15} = 30$ kHz and $bw_{fini15} = 26$ kHz: center – $posF_{MAXE} = 3$ ms and $bw_{fini15} = 26$ kHz; right – $posF_{MAXE} = 3$ ms and $ff_{in15} = 30$ kHz	66
Figure 20 – Sum of squared errors of the training and validation data by the different ENN architectures tested for <i>Myotis</i> species classification	66
Figure 21 – Lowess representation of the distribution of the RGW along the input space of the most significant covariates for the classification of <i>M. escalerae</i> / <i>M. daubentonii</i> / <i>M. emarginatus</i> calls.....	68

Figure 22 – LR weights and ENN RGW for the reference group M45 (M45 – <i>M. escalerae</i> / <i>M. daubentonii</i> / <i>M. emarginatus</i>), with respective boxplot of the significant covariates identified. Left – LR weights; right – ENN RGW	70
Figure 23 – Boxplot of call fmaxe for the Pipistrelloid species with the cutpoint indicated (Msch – <i>M. schreibersii</i> ; Pkuhl – <i>P. kuhlii</i> ; Pip – <i>P. pipistrellus</i> ; Ppyg – <i>P. pygmaeus</i>).....	71
Figure 24 – Pearson residual analysis of the first model	73
Figure 25 – Probabilities of a call belonging to Msch – <i>M. schreibersii</i> ; Ppyg – <i>P. pygmaeus</i> or Ppip – <i>P. pipistrellus</i> . Top left: posFMAXE = 3.0 kHz, ffin15 = 50 kHz and bwffin15 = 1.2 kHz; top right: duration = 5.5 ms, ffin15 = 50 kHz and bwffin15 = 1.2 kHz; bottom left: duration = 5.5 ms, posFMAXE = 3.0 kHz and bwffin15 = 1.2 kHz; bottom right duration = 5.5 ms, posFMAXE = 3.0 kHz and ffin15 = 50 kHz.....	76
Figure 26 – Sum of squared errors of the training and validation data by the different ENN architectures tested for <i>Pipistrelloid</i> species classification	77
Figure 27 – Boxplot of covariate ffin10 for Nyctaloid species with the cutpoint value indicated (EE – <i>E. serotinus</i> / <i>E. isabellinus</i> ; NN – <i>N. leisleri</i> / <i>N. lasiopterus</i> ; Tten – <i>T. teniotis</i>).....	79
Figure 28 – Lowess representation of covariate bw10	80
Figure 29 – ROC curve of the final model for <i>Nyctaloid</i> species.....	81
Figure 30 – Probabilities of a call belonging to <i>E. serotinus</i> / <i>E. isabellinus</i> or <i>N. leisleri</i> / <i>N. lasiopterus</i> . On the left bw10 is constant at 10 kHz and on the right ffin15 is fixed at 20 kHz....	82
Figure 31 – Sum of squared errors of the training and validation data by the different ENN architectures tested for <i>Nyctaloid</i> species classification	83

1. INTRODUCTION

This work is focused on the study and protection of bats, a vulnerable and extraordinary mammal group. They are the only mammals with the ability of true flight and many species have an advanced echolocation system (Airas, 2003). They belong to the second largest placental mammals order (*Chiroptera*) with more than 1100 known species, and are widespread all over the world, except in Antarctica (Altringham, 2011; Jones, 2002). They also play a major role on the maintenance of world ecosystems: fruit and pollen eating bats are very important for seed dispersal and plant pollination; insectivorous bats, around 70% of all known bat species, are of critical importance in pest control for agriculture and in human disease vectors as they can eat up to half their weight in insects per night (Boyles *et al.*, 2011; Nowak, 1994).

Bats are threatened worldwide (Mickleburgh, Hutson, & Racey, 2002; Racey, 2009). Their low birth rate, late sexual maturity, and the fact that many species form large colonies makes them highly vulnerable to habitat destruction and roost disturbance (Palmeirim & Rodrigues, 1992; Racey, 2009) and besides the intrinsic biological and ecological characteristics, Sherwin, Montgomery, & Lundy (2012) also showed that climate change is also becoming a major threat by influencing bat biogeography, access to food, timing of hibernation, reproduction and development, frequency and duration of torpor and rate of energy expenditure.

Although nowadays the ecological importance and conservation threats of bats are becoming common knowledge for the general population, the predominantly nocturnal habits combined with the use of ultrasound (inaudible to humans) have always constrained their study in the wild and delayed for many years the recognition of their ecological importance (O'Shea, Ellison, & Stanley, 2004). It wasn't before the development of specialized equipment (ultrasound bat detectors) and the discovery of species-specific echolocation call structures in bats (Fenton & Bell, 1981), which allowed to identify bats species from echolocation calls, that the knowledge on these mammal boosted.

Identifying bat species from echolocation calls revealed more difficult than initially expected. In fact, bat echolocation calls may be similar between species and calls show a large variability due to sex, age, body weight (Jones, Gordon, & Nightingale, 1992), behavior, geographical location (Parsons, 1997; Rydell, 1993), habitat structure, flight height, distance to obstacles (Kalko & Schnitzler, 1993; Rydell, 1993), presence of conspecifics (Obrist, 1995) weather patterns and other environmental factors (Larom *et al.*, 1997).

Despite the difficulties, the analysis of ultrasound recordings of bats in flight has proved that identifying bat species is possible by measuring some parameters from bat calls and comparing them with species databases. Yet, due to the variability present in bat calls, this type of identification has a high degree of subjectivity (Ahlén & Baagøe, 1999; Parsons & Jones, 2000; Russ, 1999; Russo & Jones, 2002; Zingg, 1990).

1.1 BATS IN PORTUGAL

Four bat families with 27 species are known in Portugal, 25 of which are present in the mainland, one, *Nyctalus azoreum*, is only present in Azores archipelago and other, *Pipistrellus maderensis*, in Madeira and Azores archipelagos (Cabral *et al.*, 2006). They are mostly insectivorous, although some species may occasionally feed on small birds, mammals and fish (Ibáñez *et al.*, 2001; Siemers *et al.*, 2001).

All bat species known in Portugal and most of their roosts are protected by International legislation such as the Berne Convention (Convention on the Conservation of European Wildlife and Natural Habitats - annex II and III), the Bonn Convention (Convention on the Conservation of Migratory Species of Wild Animals), the Habitat Directive (annex II and IV) and EUROBATS (Agreement on the Conservation of Populations of European Bats) transposed to the Portuguese law. Three species are classified as “Critically Endangered”, one as “Endangered” and five as “Vulnerable”. Adding to this nine other species are classified as “Data Deficient”, which means that the existing information is too scarce to know for sure what is the real conservation status of these species (Cabral *et al.*, 2006). In Table 1 are presented bat species known in mainland Portugal.

Table 1 – Bat species present in mainland Portugal and respective conservation status: CR – “Critically Endangered”, EN – “Endangered”, VU – “Vulnerable”, NT – “Near Threatened”, LC – “Least Concern”, DD – “Data Deficient” (Cabral *et al.*, 2006).

Family	Species	Conservation status
RHINOLOPHIDAE	<i>Rhinolophus ferrumequinum</i>	VU
	<i>Rhinolophus hipposideros</i>	VU
	<i>Rhinolophus euryale</i>	CR
	<i>Rhinolophus mehelyi</i>	CR
VESPERTILIONIDAE	<i>Myotis bechsteinii</i>	EN
	<i>Myotis myotis</i>	VU
	<i>Myotis blythii</i>	CR
	<i>Myotis escaleraei</i>	VU
	<i>Myotis emarginatus</i>	DD
	<i>Myotis mystacinus</i>	DD
	<i>Myotis daubentonii</i>	LC
	<i>Pipistrellus pipistrellus</i>	LC
	<i>Pipistrellus kuhlii</i>	LC
	<i>Pipistrellus pygmaeus</i>	LC
	<i>Hipsugo savii</i>	DD
	<i>Nyctalus leisleri</i>	DD
	<i>Nyctalus noctula</i>	DD
	<i>Nyctalus lasiopterus</i>	DD
	<i>Eptesicus serotinus</i>	LC
	<i>Eptesicus isabellinus</i>	-
	<i>Barbastella barbastellus</i>	DD
	<i>Plecotus austriacus</i>	DD
	<i>Plecotus auritus</i>	LC
MINIOPTERIDAE	<i>Miniopterus schreibersii</i>	VU
MOLOSSIDAE	<i>Tadarida teniotis</i>	DD

1.2 ECHOLOCATION CALLS: CHARACTERISTICS

Echolocation is a biological sonar used for navigation and hunting, being a common characteristic of many bat species (Airas, 2003; Moss & Sinha, 2003). Echolocating bats emit calls (usually ultrasonic) out to the environment and listen to the returning echoes to detect, classify and spatially locate objects. The echoes allow the bat to detect a possible target or prey and evaluate its size, material, shape and texture. The bat can also estimate the vertical and horizontal angles to the target (spatial position) from echo clues and estimate its range by the time delay between the emitted signal and the returning echo (Schnitzler & Kalko, 2001).

Echolocation calls can be composed by one or a combination of two fundamental components: constant frequency (CF) tones and frequency modulated (FM) sweeps (Moss & Sinha, 2003; Schnitzler & Kalko, 2001).

CF tones have its energy concentrated around a particular frequency, more or less constant over time and are mostly used for target detection. They can also be used for range estimation and classification, especially long duration calls with Doppler shift compensation, but they are rather imprecise for this task. On the other hand a FM sweep is a broadband signal, having its energy distributed along a broad range of frequencies over time. This type of signal is very precise for estimating range and angles to targets and to infer about texture and depth structure, being often used for target location and classification (Schnitzler & Kalko, 2001).

The typical structure of echolocation calls in bats can be broadly categorized into three groups (adapted from Airas (2004)) and can be correlated with different hunting strategies and foraging areas (Schnitzler & Kalko, 2001):

- FM bats – these bats use short, downward broadband signals and usually forage in background cluttered spaces which include edges or vegetation gaps, near ground and water surfaces;
- Long FM/CF bats – these bats use a long CF component preceded and followed by a quick FM sweep, using Doppler shift information to forage in background cluttered or even high-cluttered spaces, where prey detection is harder for pure FM bats;
- Short FM/CF bats – these bats employ an intermediate pulse design with both FM and CF components and are able to adapt the call structure to the surrounding environment. They use quasi-CF signals while foraging in open environments and almost pure FM signals in more cluttered spaces.

The temporal structure of echolocation signals of bats can also be categorized into two duty cycles, i.e. the percentage of time that signals are being produced (Airas, 2004; Jones, 1999):

- Low-duty cycle bats – these bats emit short bursts and as they are intolerant to call-echo overlap their duty cycles usually do not exceed 20%. All FM and short FM/CF bats belong to this category;
- High-duty cycle bats – these bats can tolerate call-echo overlap and their duty cycles can vary from 30% to 90%. All long FM/CF bats belong to this category.

Despite the variability due to behavior and the involving environment, there are certain characteristics of echolocation that can be related to wing morphology and body size (Jones, 1999; Norberg & Rayner, 1987). Insectivorous bats that forage in open spaces tend to have longer wings with pointed tips giving good agility even at high flight speeds and insectivorous bats that hunt among vegetation usually have very short and broad wings with rounded tips, offering very good maneuverability at low flight speeds (Norberg & Rayner, 1987). These evolutionary trends of wing morphology are reflected in echolocation call structure and body size owing to the joint constraints of flying and foraging strategies (Norberg & Rayner, 1987). For instance, call frequency is negatively correlated with body mass and positively correlated with call duration (Jones, 1999).

1.3 ECHOLOCATION CALLS: SPECIES IDENTIFICATION

Bat detectors probably remote to the 1940's (Allen, Romeling, & Robbins, 2011) but more serious bat studies using detectors only became popular in mid 1970's (Ahlèn & Baagoe, 1999). These detectors did a conversion of ultrasonic frequencies to audible ones by means of heterodyne (Hogan, 1921) or frequency division techniques (Miller, 1939). These techniques performed a transformation of the original signal and although they could be used to record calls, the recordings were not adequate for posterior analysis and study of echolocation calls. The usage of expanded time recording devices combined with heterodyne or frequency division bat detectors (Fenton & Bell, 1981) finally allowed recording bat calls without significant loss of information from the original signal and boosted the knowledge of bat echolocation calls.

Nowadays, with the use of a bat detector to record bat echolocation calls, and posterior measurement of call parameters and comparison with species parameters databases, it is possible to classify an emitting bat as belonging to a species or a group of species with some degree of certainty (Fenton & Bell, 1981). Nevertheless some species are extremely difficult to discriminate due to an overlap in their echolocation call parameters and some mismatch classifications may occur (Parsons & Jones, 2000; Russo & Jones, 2002; Vaughan, Jones, & Harris, 1997). Besides, as there is no reference database of bat calls recorded in Portugal, Portuguese bat researchers have been using, among others, databases compiled in England (Russ, 1999), Italy (Russo & Jones, 2002) and Switzerland (Zingg, 1990), which can introduce

additional bias in species classification due to geographic variation of echolocation calls (Parsons, 1997; Rydell, 1993).

Traditionally six parameters have been used to describe and classify echolocation calls: initial frequency, final frequency, bandwidth, frequency of maximum energy, frequency at half the duration and call duration (Parsons & Jones, 2000; Redgwell *et al.*, 2009). The identification process is normally done by a human operator, and even with only these parameters it's a very time consuming procedure. Nevertheless, until the introduction of automatic recording stations in late 1990's (Chick & Lumsden, 1999) only manual bat detectors were available, which require a human operator to scan a range of frequencies at ground level until an echolocation call is detected and recorded to posterior analysis. The amount of recordings obtained this way is relatively small, usually no more than a hundred per night, and the identification method employed, although slow, was sufficient to analyze it.

The development and increased usage of automatic recording stations, which can be left in the field at ground level or in height, with no human intervention, and designed to record bats during a time span of days or weeks, exponentially increased the amount of data collected (sometimes more than a thousand recordings per night) and the existing identification method (conducted by a human operator) revealed inappropriate to respond to this volume of data. This means that bat monitoring over large periods of time is possible but unrealistic due to the huge amount of time needed for the classification of all the recorded calls, unless an automatic identification system is used (Allen *et al.*, 2011).

1.4 MOTIVATION

Recent studies have showed that the accuracy of bat identification from their echolocation calls can be highly improved with the use of statistical tools and more detailed measurements to describe the echolocation calls (Jennings, Parsons, & Pocock, 2008; Parsons & Jones, 2000; Walters *et al.*, 2012).

With this in mind our main motivation is to develop an automated method to analyze and classify bat calls recorded in real bat-work situations with high accuracy. We intend to develop an automatic analysis and classification computer program, using statistical models based on recordings of bats collected in mainland Portugal. In this way we intend to decrease the amount of time needed to analyze and identify bat calls while increasing the objectivity and reliability of the identification.

Having a tool with the ability to integrate the analysis of a large amount of bat calls and a statistical evaluation of the species recorded will permit to drastically reduce the time and

financial costs of bat data analysis which could be a major advantage in environmental impact assessment studies or in wildlife management programs.

By using an automated analysis and statistical classification of bat calls, a continuous bat monitoring over an indefinite period of time using automatic recording stations would then be feasible, resulting also in a better accuracy in discriminating similar calls of distinct species, which is particularly relevant when those species have different conservation status. Moreover, the compilation of a database of bat calls collected in mainland Portugal could be very helpful for future bat related work in this country.

1.5 OBJECTIVES

As seen before, manually identifying bats from their echolocation calls is a slow and subjective process which can lead to some uncertainty or mismatches in classifications and is unable to respond effectively to the volume of data that recording stations can produce. Adding to this, we don't have a reference database of bats calls recorded in Portugal.

To overcome these gaps in knowledge we established three main objectives: compilation of a database of bat calls; development of a computer program for automatic detection and analysis of bat calls; classification of bat calls using statistical models.

Compilation of a recordings database

The first objective is to compile a collection of previously identified bat calls obtained in mainland Portugal by recording emerging bats from known roosts at sunset, hand released bats or free flying bats. These recordings will be the foundation for the first Portuguese bat call database and could be quite valuable in future studies. We intend to record bats in a real bat-work environment to obtain recordings with high and low signal-to-noise ratios, from cluttered and open environments, with rustling foliage, sympatric fauna and other sources of noise or atmospheric attenuation, etc. covering this way the call variability that bat researchers face in the field and therefore incorporate these factors in the statistical models.

Development of a computer program for automatic detection and analysis of bat calls

Having a compiled database of recordings, the second objective is to develop a program to automatically filter a sound file to reduce background noise, identify individual bat calls, determine the start and end positions of each call and measure parameters from the isolated calls.

Bats in Portugal cover a wide range of echolocating frequencies, approximately from 9 kHz to 115 kHz (see Table 3) and if we take into account that insects emit in frequency ranges of 4 kHz to 30 kHz (Pollack & Imaizumi, 1999) it's easy to realize that there is much insect noise overlapping echolocation calls, besides other sources of environmental noise such as rustling foliage for instance.

Most of previous related studies (see Section 1.6 for further details) used amplitude thresholds or a visual inspection for detecting and isolating bat calls from background noise but we wanted to go further by developing and implementing a fully automatic computer program for this end.

Bat call classification using statistical models

The third objective is to fit both binomial or multinomial logistic regression models and artificial neural networks for bat call classification, to identify the most important parameters for the classification of certain species (or groups of species) and compare the classification accuracy and results of both statistical approaches.

The usage of artificial neural networks allows to handle large data sets with colinearity between variables and to detect complex nonlinear interactions between them but cannot be accurately trained with small datasets (Rojas, 1996). On the other hand, logistic regression models can give new insights about the importance of certain parameters for bat calls classification and can be used to predict the outcome of a categorical response variable even on small data samples (Hosmer & Lemeshow, 2000), for example when bats are rare or difficult to capture/record and sample recordings are scarce, which could give logistic regression models an unquestionable statistical value over artificial neural networks.

1.6 RELATED WORK

Many studies have been written over the years on bat species identification from echolocation calls but our primary interest is the identification using statistical models. In this sense the review presented here only covers papers that make use of statistical models for classification or detection of bat calls.

Parsons & Jones (2000) compared the results of a multilayer perceptron artificial network and discriminant function analysis (DFA) to identify 14 species of bats from Britain. Recordings were made outside known roosts, after hand release or at foraging sites where individual bats and species could be identified unambiguously and almost every species was recorded at different geographic locations.

The recording apparatus was static, with the microphone mounted on a tripod 1.2 m above ground, and a 2 s recording was manually triggered. The majority of the recordings were made with the bat flying at less than 2 m from the microphone. After recording, the call with the best signal-to-noise ratio was chosen by a human operator, extracted from the background noise by using an arbitrary threshold (equal to all calls) and saved for further analysis.

Five parameters (temporal and spectral) were then directly measured from the call: duration, frequency of maximum energy, starting frequency, ending frequency, center frequency, plus the frequency-time course of the call, that was obtained by fitting eight mathematical functions and using the mean squared error of each of the fits.

The multilayer perceptrons were trained using the backpropagation algorithm on half the database and the remaining half was only used for validating the results. Several architectures were tested and the one that gave best classification results on the validation dataset was selected. The discriminant function analysis used quadratic functions and the same 13 input variables. The relative importance of each of the inputs on the discriminant analysis was accessed via the Wilk's λ . Two separate analyses were tested, one just to discriminate genus and the other to discriminate species.

The overall correct call classification rate for genus classification was, for ANN, 96% with individual rates ranging from 89% (*Eptesicus*) to 100% (*Pipistrellus* and *Myotis*) and for DFA was 94%, with individual rates ranging from 70% (*Plecotus*) to 100% (*Myotis*).

As for the classification of individual species the overall correct rate, for ANN, was 87% with individual species correctly classified from 56% (*M. mystacinus*) to 100% (*B. barbastellus*) and for DFA was 79%, with individual rates ranging from 49% (*M. mystacinus*) to 96% (*M. nattereri*).

Obrist *et al.* (2004) used a synergetic algorithm approach to classify all the 26 Swiss bat species. These algorithms are commonly used in product control and it's the first time they were used to classify bat calls.

Recordings were obtained after hand release of captured bats, near the entrance of known roosts or from free flying bats. After recording, a high-pass filter (7.5 kHz) was applied and single calls were cut with the help of a simple integrating detector algorithm and saved for further analysis. The automated extracted calls were submitted to a visual inspection and almost 3/4 were considered just noise or otherwise inadequate for analysis.

From the original database three random subsets of 520 calls (20 calls per species) were selected for training and other three random subsets were used for validation. The classification was achieved with the SC-MELT algorithm using as input the entire spectrogram data. The global correct classification rate was 77% with individual species identified correctly with 27% (*M. brandtii*) to 100% (*N. leisleri* and *T. teniotis*) accuracy.

Skowronski & Harris (2006) compared a Gaussian Mixture Model (GMM) and a Hidden Markov Model (HMM) for the classification of 5 bat species and tested a GMM for detecting bat calls. These models are commonly used in automatic speech recognition (ASR). Recordings were obtained from free flying bats or outside know roosts. The recording equipment was set on previously chosen locations and recorded continuously (only limited by battery life or storage space).

The calls were transferred to Matlab where a custom script measured two parameters (frequency of peak energy and peak energy) and their respective first and second temporal derivatives. These six measures were the inputs used for the GMM and HMM for detection and classification of bat species.

The detection model showed a 96% rate of correct identification of bat calls but only accounted for “typical” background noise (electronic noise from the recording equipment, rustling foliage or sympatric fauna). If novel sources of noise were introduced the author refers that the detector would probably indicate a bat call instead of background noise.

Both models had a correct classification rate of nearly 100% and identified correctly 97% (*Molossus molossus*) to 100% (*Lasiurus borealis*, *Lasiurus cinereus semotus* and *Tadarida brasilienses*) of the individual calls. These results seem very good but are limited by the fact that each species was recorded at a single location, separate from all other species and only five species were used. Furthermore, characteristics of each recording site could influence the measured parameters.

Redgwell *et al.* (2009) compared ensembles of neural networks (ENN), discriminant function analysis (DFA) and support vector machines (SVM) for the classification of 14 species of bats. The recordings used came from a previously published study (Parsons & Jones, 2000) and only the call with best signal-to-noise ratio was selected from each recording. Calls were extracted from the background noise using two thresholds.

Twelve parameters were measured on the extracted call: duration, frequency of maximum energy, starting frequency, ending frequency and center frequency, rate of change of the frequency-time course, standardized bandwidth (at 80% of the maximum energy), proportion of total energy in each quartile (4 parameters, 1 for each quartile) and call type. These measurements were all made on the power spectrum.

The discriminant function analysis used quadratic functions and the relative importance of each of the inputs on the discriminant analysis was accessed via the Wilk’s λ .

The neural networks (multilayer perceptrons) used a sigmoidal transfer function and were trained using a backpropagation algorithm on Matlab. Half of the database was used for training and the remaining half was only used for validating the results. Several architectures were tested and the best was considered the one that had the highest minimum accuracy. Two sets

of ENN were generated: one to discriminate species and other to discriminate genus and species.

The SVM classifier was trained and tested using Matlab as well. The kernel functions of the support vector machines used a radial basis function with a gamma parameter. For each value of gamma, the support vector machine was reinitialized 20-times. A support vector machine was trained for every target case in the dataset (each genus, or species) to classify an instance as either belonging to that case, or not. All classifiers were then combined and categorized each echolocation call as either belonging to a specific class (genus or species), or not. A call was considered correctly classified only if a single support vector machine classified it.

Ensembles of networks show less prediction variability and a better accuracy than a single network and obtained the best results of the study, 98% correct classification against the 87% of SVM and 73% of DFA. As for the correct individual species classification rate, ENN ranged from 91% (*N. leisleri*) to 100% (*B. barbastellus*, *M. bechsteinii*, *M. daubentonii*, *P. pipistrellus*, *P. pygmaeus*, *R. ferrumequinum* and *R. hipposideros*), SVM from 64% (*M. mystacinus*) to 100% (*R. ferrumequinum*) and DFA from 42% (*M. mystacinus*) to 100% (*R. ferrumequinum* and *R. hipposideros*). The use of a hierarchical ENN classification didn't seem to improve the species classification results.

More recently Walters *et al.* (2012) used ensembles of neural networks to classify 34 European bat species. Recordings for this study were obtained from EchoBank (Collen, 2012) and the commercially available sound analysis software SonoBat was used to automatically find and measure calls in recorded sequences. Calls were detected using amplitude thresholds, were visually inspected to ensure a correct selection and 24 parameters describing the frequency and time course of the call were measured.

The neural networks (multilayer perceptrons) used a sigmoidal transfer function and were trained using a backpropagation algorithm on a custom Java program. Half of the database was used for training and the remaining half was only used for validating the results. Several architectures were tested and the best was considered the one that had the highest minimum accuracy and the highest classification rate on the validation dataset. After selection, the top 50 networks were refined and the best-performing 21 of these networks were selected.

The ENN (multilayer perceptrons) were assembled in a hierarchical structure, classifying into genus or subgroup level first and only then to the species (i.e. one ensemble just discriminates genus/subgroup and then an ensemble discriminates species belonging to a particular genus/subgroup).

The first classification into genus/subgroups shows a 98% correct classification rate and species classifications of 84% with individual species correct classification rates of 49% (*M.*

bechsteinii, *M. daubentonii* and *M. mystacinus*) to 100% (*R. blasii*, *R. ferrumequinum* and *R. euryale*).

Although the results achieved by the studies described above seem very promising, some cautionary notes must be made on the differences between echolocation calls of the species used in each study, the recording conditions and the decisions of what calls were suited to analysis.

Some species are naturally easier to classify than others and sometimes the inclusion of one extra species in the analysis could transform completely the classification rates. This fact also means that the performance of the techniques used in studies involving different sets of species cannot be directly compared, since a given study which classifies species with great differences in call characteristics will perform better than more ambitious models that include species with similar calls. In this sense, only the work by Obrist *et al.* (2004) used calls of all bat species, that would probably be presented in the future to the fitted model (if used in national surveys).

As for the recording conditions and call selection, Parsons & Jones (2000) and Redgwell *et al.*, (2009) shared the same dataset, which was obtained with a static microphone and most of the bats were recorded at less than 2 m from it. As the distance from the microphone was very similar for most recordings the energy of the ultrasonic calls suffered a similar (and small) attenuation permitting the usage of a pre-defined threshold for call detection and extraction. Besides that only the call with the best signal-to-noise ratio was selected for analysis, by a human operator. Obrist *et al.* (2004) also enhanced their database by manually discarding calls that seemed inappropriate for analysis.

1.7 CLAIM OF CONTRIBUTIONS

In field recordings there are some variability factors that cannot be controlled or even accounted for. For instance, the distance of the flying bat from the microphone is not fixed, implying that the call energy has a great variability and a pre-defined threshold to identify or extract calls could reveal inappropriate. Furthermore the calls in some recordings may have very small signal-to-noise ratios and if the database that was used for fitting the statistical models doesn't include a high variability of calls (with better and worse quality) it will not be able to classify them accurately. Adding to this, the attenuation of the ultrasonic wave is much more intense in higher frequencies than in lower ones (Denny, 2004) implying that some variables used for classification can sometimes depend more on the distance of the bat to the microphone than the emitting species itself, and should be used with caution, especially initial frequency.

Since the main motivation of our work is to analyze and classify bat calls obtained in real bat-work situations the compilation of recordings aimed to maximize call variability. We used several detectors from different models to obtain recordings in different clutter environments while encompassing several source of noise in different locations across mainland Portugal. We believe that the higher variability introduced in the measured parameters could substantially improve classification results.

This higher variability led us to develop our own method for detecting and isolating calls in a recording avoiding amplitude thresholds which revealed inadequate to handle recordings with low amplitude calls or with much noise. Because of these fainter calls some of the parameters used in previous studies were also avoided (for instance, initial frequency or peak energy) and substituted by others, measured around the point of maximum energy of the call, which is less prone to atmospheric attenuation. Also we did not enhance our database by excluding weaker calls and, as expected, the higher variability led to an increased difficulty in fitting the statistical models.

We also propose to model bat calls using logistic regression models which, to the best of our knowledge, were never done. Although most studies used artificial neural networks with good results, we believe that logistic regression models could also provide good classification results, giving extra information about the relevance of the different parameters.

1.8 OVERVIEW OF THE DISSERTATION

In chapter 2 is presented the theoretical background for the statistical models used: logistic regression models and multilayer feedforward perceptrons.

In chapter 3 are presented the methods used for capturing bats and obtaining recordings; the method we developed for detecting and isolating bat calls in a recording; the method used for parameter measurements and call classification; and the modeling strategies followed for both statistical approaches used.

In chapter 4 is presented the descriptive analysis of the calls recorded and the classification results of both logistic regression models and artificial neural networks.

In chapter 5 are presented the conclusions of this thesis and future work planned.

2. STATISTICAL BACKGROUND

In this chapter we will do a brief introduction to the broader families of models used in our study, generalized linear models and artificial neural networks and then particularize for the specific models used, logistic regression models and multilayer feedforward networks.

2.1 GENERALIZED LINEAR MODELS

Generalized Linear Models (GLMs) were formulated by Nelder & Wedderburn (1972) as a unified theory of statistical modeling. The linear model is generalized by the introduction of a function linking the linear part of the model and the response variable. This way the response variable may be non-linear but the transformed response variable is. GLMs are characterized by having three components (Agresti, 2007; Hosmer & Lemeshow, 2000):

- the random component – identifies the response variable which follows a probability distribution of the exponential family;
- The systematic component – specifies the explanatory variables (covariates);
- The link function – connects the random and systematic components.

The modeling strategy used in GLM can be resumed in three main stages:

- Model selection – the characteristics of the candidate models are defined by selecting an appropriate probabilistic distribution for the response variable, identifying possible explanatory variables and the correct linking function;
- Model assessment – the models previously selected are fitted to the data, i.e., the model parameters are estimated, and the existence of outliers or influential observations is analyzed;
- Model validation – the most appropriate model is identified and the fit of the model to the data is tested.

After the modeling phase the parameters of the model are interpreted by evaluating the significant covariates and their quantitative contribute to the explanation of the response variable.

2.1.1 Logistic regression

Logistic regression models (LR) are a particular class of GLMs that make use of the logit¹ link function and are employed for predicting the outcome of a categorical response variable Y (Hosmer & Lemeshow, 2000). If there are only two categories in the response variable (the event happened *versus* the event didn't happened), it's called binary or binomial otherwise is

¹ The logit function is: $\text{logit}(\pi) = \text{Ln}\left(\frac{\pi}{1-\pi}\right)$. The inverse is the logistic function: if $\text{logit}(\pi) = z$, then $\pi = \frac{1}{(1+e^{-z})}$

called multinomial. For the binomial logistic regression model, let the conditional probability that the event happened be denoted by $P(Y = 1|x) = \hat{\pi}$, then the logistic regression model that represents the probability of the outcome by equal to 1 can be written as (Hosmer & Lemeshow, 2000):

$$\hat{\pi} = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \quad (\text{Eq. 1})$$

where β_0 is the intercept and β_1 to β_p are the coefficients of the x_1 to x_p explanatory variables. By applying the logit function to equation 3.1 we obtain the linearized form of the model, expressed as log-odds (Hosmer & Lemeshow, 2000):

$$\text{logit}(\hat{\pi}) = \text{Ln}\left(\frac{\hat{\pi}}{1 - \hat{\pi}}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (\text{Eq. 2})$$

When the response variable Y has more than two levels, we have a multinomial logistic regression model. The logistic model can be easily modified to handle a multinomial response variable and the same methods used to fit and assess the model assumptions used in binomial logistic regression models could be used in the multinomial case. However, not all methods have been fully implemented in software packages. For that reason, to assess the model assumptions is often used the independence of irrelevant alternatives (IIA), as suggested in Hosmer & Lemeshow (2000). Basically, IIA states that the odds of preferring one class over another do not depend on the presence or absence of other "irrelevant" alternatives, which in fact means that separate multiple binomial models can be used as an approximation of a full multinomial model, being the estimates of the logistic regression coefficients obtained in this manner consistent with the multinomial model.

To fit the logistic regression model to a specific dataset the unknown parameters β are estimated by the maximum likelihood method, which yields values for the parameters that maximize the probability of obtaining the observed data. As the ML equation system doesn't have an analytical solution, the parameters are fitted through an iterative computational algorithm. Usually the iteratively reweighted least squares method is employed (McCullagh & Nelder, 1989).

2.1.2 Model selection

When selecting a logistic regression model for a given data set there has to be a compromise between the complexity of the model and its interpretability. More complex models tend to fit better to the data but as complexity increases the interpretability of the model decreases. So, if there are too many explanatory variables or if the sample size is too small or unbalanced, one should reduce the number of explanatory variables entering the model (Agresti, 2007). Note that some covariates should be based on others in the model to allow for interaction between them.

The significance of the coefficients in the model can be assessed by the Wald statistic by testing the null hypotheses, $H_0: \beta_j = 0$ (Hosmer & Lemeshow, 2000):

$$W_j = \frac{\hat{\beta}_j}{\widehat{SE}(\hat{\beta}_j)} \sim N(0, 1) \quad (\text{Eq. 3})$$

where $\hat{\beta}_j$ is the estimate of β_j and $\widehat{SE}(\hat{\beta}_j)$ is the estimated standard error of β_j ; or by testing the null hypotheses by the likelihood ratio test. The maximized value of the likelihood function under $H_0: \beta = 0, L_0$, is compared with the maximized value of the likelihood function when $\beta \neq 0, L_s$, by means of the statistic:

$$D = -2 \ln \left(\frac{L_0}{L_s} \right) \quad (\text{Eq. 4})$$

which, under the null hypothesis has χ^2 distribution with one degree of freedom (Hosmer & Lemeshow, 2000).

For model selection Hosmer & Lemeshow (2000) suggests the following approach:

- an exploratory univariate assessment of the significance of each covariate, where each univariate model is compared with the null model using the likelihood ratio test;
- all the covariates significant at a p-value of 0.25 are included in an initial multivariate model;
- using the likelihood ratio test, a backward stepwise approach refines the initial multivariate model by removing (one by one) every covariate not significant at a p-value of 0.05 (less conservative p-values of 0.1 or even 0.15 can also be used, depending on the theoretical knowledge of the phenomenon under analysis), starting with the one whose p-value is greater;
- covariates not included in the initial multivariate model are now included in the model to assess their significance. At this stage the main effects model is obtained;
- the linearity in the logit for continuous covariates in the main effects model is tested. Three (complementary) methods are proposed: (1) a lowess smoothed scatterplot on the logit scale for each covariate, (2) design variables based on the quartiles of the distribution and (3) the method of fractional polynomials. Methods (1) and (2) just require a visual inspection of the plots produced whereas method (3) is an analytical approach. Multiple transformations on the covariate are tested and compared (via the Likelihood ratio test) with the model without the covariate transformation as a way of testing if a particular transformation could improve the linearity of the logit. Further explanation of this methods can be found in Hosmer & Lemeshow (2000);

- the significance of interactions between all pairs of covariates is then tested with the likelihood ratio test, and even if significant at a p-value of 0.05 or 0.01 the interaction should only be included in the model if it makes sense in a clinical point of view, i.e. if there is any scientific basis for supporting that particular interaction;
- the preliminary final model is then obtained and should include all the covariates and (logical) interaction pairs significant at a 0.05 p-value;
- the model's adequacy and fit must be assessed, i.e., it must be validated.

2.1.3 Model Validation

After the model selection with the significant covariates and interactions, it's important to investigate how effectively a model describes the outcome variable, its goodness-of-fit. This starts by an evaluation of the sample for outliers (observations that have a great deviation from the remaining sample) and influent observations (that have an exaggerated effect on the coefficients estimations). This assessment is often done graphically by the analysis of the residuals (summary measures of the distance between observed and estimated values. Several types of residuals are commonly used for this end: Pearson, Deviance, Leverage, Cook's distance and the Dfbeta. Further explanation of this methods can be found in Hosmer & Lemeshow (2000) and Mccullagh & Nelder (1989).

For complementing the residuals inspection there are three diagnose tests commonly used for assessing the goodness-of-fit: the Hosmer and Lemeshow test, the Nagelkerke's pseudo R^2 and the ROC curve.

The Hosmer and Lemeshow test, used for binomial LR, groups the data in deciles based on the values of the estimated probabilities. The goodness-of-fit of the model can be assessed by the statistic, \hat{C} , which is obtained by calculating the Pearson chi-square statistic from the $g \times 2$ table of observed and estimated expected frequencies, by testing the null hypotheses, H_0 : *observed frequencies = estimated frequencies*:

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} \quad (\text{Eq. 5})$$

n'_k is the total number of observations in the k^{th} group, o_k is the number of responses per covariate pattern (c_k) in the k^{th} group and $\bar{\pi}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n'_k}$ is the average estimated probability. If there is at least one continuous covariate the number of covariate patterns is equal to the number of observations ($c_k = n'_k$) and the statistic \hat{C} is well approximated by the chi-square distribution with $g - 2$ degrees of freedom. Further details can be found in Hosmer & Lemeshow (2000).

The Nagelkerke's pseudo R^2 gives an idea of the effect dimension of the covariates on the outcome variable but, differently from R^2 from linear regression, its value is not a measure of the percentage of variability of the outcome variable explained by the model but instead is based on a comparison between predicted values from the fitted model and the null model. This measure assumes values in the range of [0,1] and is expressed as (Nagelkerke, 1991):

$$R_N^2 = \frac{1 - e^{(D_1 - D_0)/n}}{1 - e^{D_0/n}} \quad (\text{Eq. 6})$$

being D_1 the deviance of the fitted model and D_0 the deviance of the null model.

Another way of evaluating the general performance of a model built for classification is by looking at the classification results. After a cutpoint has been specified, estimated probability values of the model equal or over the cutpoint can be classified as belonging to the reference category of the response variable. Three summary statistics based on the classification performance can then be used to evaluate the model: sensitivity, specificity and overall correct rate. Sensitivity indicates the correct rate of classifications of the reference group ($Y=1$), specificity the correct classification rate of the other category ($Y=0$) and overall correct rate indicates the correct classification rate of the entire model (Hosmer & Lemeshow, 2000).

The ROC curve is a graphical representation of the positives (sensitivity) versus the false positives (1-specificity) and is a robust way of analyzing the sensitivity and specificity for different cutpoints by measuring the area under the curve (AUC). The value of the AUC ranges from 0 to 1 and is an indication of the model's ability to correctly discriminate between the two groups. In Table 2 is presented a possible interpretation of the AUC values as suggested by Hosmer & Lemeshow (2000):

Table 2 – Interpretation of AUC values

AUC < 0.5	No discrimination
0.5 < AUC < 0.7	Poor discrimination
0.7 < AUC < 0.8	Acceptable discrimination
0.8 < AUC < 0.9	Excellent discrimination
AUC > 0.9	Outstanding discrimination

After the preliminary final model has been validated, we obtain the final model which can be used for inference.

2.1.4 Model interpretation

The interpretation of the fitted model will allow practical inference from the estimated coefficients. The focus is the phenomenon under study and the underlying relationships between the variables in the model. For most models this involves only the estimated

coefficients for the covariates in the model but in rare occasions the estimated intercept coefficient is also of interest.

The coefficients for the covariates represent the rate of change (slope) of the logit of the response variable per unit of change in the covariate. Thus, interpretation of a model involves (1) the determination of the functional relationship between the response variable and the covariate(s) and (2) defining the unit of change for the covariate(s). Normally this is done through one measure of association, the odds ratio (OR), which represent the ratio between the odds of the event happening ($Y=1$) against the odds of not happening ($Y=0$) and can be expressed as:

$$OR = \frac{\pi(1)((1 - \pi(1)))}{\pi(0)((1 - \pi(0)))} = e^{\beta_1} \quad (\text{Eq. 7})$$

This simple relation between the coefficients and the odds ratio is one of the reasons why logistic regression has proven to be such a powerful research tool. The value of the OR indicates the odds change when the variable x is present or changes one unit (if is measured on a continuous scale) when all other covariates remain constant². The endpoints of a 100*(1- α)% confidence interval for the coefficients are:

$$\hat{\beta} \pm z_{1-\alpha/2} \widehat{SE}(\hat{\beta}) \quad (\text{Eq. 8})$$

and for the logit are:

$$\hat{g}(x) \pm z_{1-\alpha/2} \widehat{SE}[\hat{g}(x)] \quad (\text{Eq. 9})$$

2.2 ARTIFICIAL NEURAL NETWORKS

Artificial neurons and artificial neural networks (ANN) were initially proposed as a new computing paradigm roughly based on biological neural networks (which are able to acquire knowledge after a learning phase) but nowadays they are mostly seen as powerful statistical tools for regression and classification as they are capable of approximating non-linear functions (Rojas, 1996).

The artificial neurons are simple computing units that may have multiple inputs but only one output and process information in a very simple way: they sum all their inputs (after being multiplied by a specific weight), apply an activation function and send the result to the next layer (Rojas, 1996). The output of a neuron t is determined from inputs x_i and can be expressed as:

$$t(x) = f_0 \left(w_0 + \sum_i w_i x_i \right) \quad (\text{Eq. 10})$$

where f_0 is the activation function, w_i are the weights and w_0 is the bias.

² For computation of OR in models with interaction please refer to Hosmer & Lemeshow (2000)

An artificial neural network is a network of interconnected artificial neurons (or in other words, functions) disposed in layers (Rojas, 1996).

2.2.1 Multilayer feedforward perceptron networks

Multilayer feedforward perceptron networks are a class of artificial neural networks that are characterized by having neurons disposed in multiple layers and the flow of information is one way only, from the inputs to the outputs (hence the designation “feedforward”). In the context of this thesis, artificial neural networks always refer to this particular class of network.

This class of network has one output layer and at least one hidden layer³. The composite function calculated by a multilayer perceptron (with one hidden layer) can be expressed as (Günther & Fritsch, 2010; Intrator & Intrator, 2001):

$$o(x) = f_0 \left(w_{0k} + \sum_j w_{jk} * f_0 \left(w_{0j} + \sum_i x_i * w_{ij} \right) \right) \quad (\text{Eq. 11})$$

where o is the output vector (with k dimensions), k indexes the neurons in the output layer, j indexes the neurons in the hidden layer, i indexes the covariates, f_0 is the logistic activation function, x are the covariates, w_{ij} are the weights attached to the hidden layer neurons, w_{0j} are the intercepts (or bias) of the hidden layer neurons, w_{jk} are the weights of the output neurons and w_{0k} are the intercepts of the output neurons. In Figure 1 we can observe a diagram of this type of network with one hidden layer.

The activation function is necessary to assure that the composite function (and the error) resulting from individual interconnected neurons is continuous and differentiable. Two common choices are the logistic function $f_0(x) = \frac{1}{1+e^{-x}}$ and the hyperbolic tangent $f_0(x) = \frac{1-e^{-x}}{1+e^{-x}}$ (Rojas, 1996). The parameters of the network (weights and intercepts) are estimated from the data during the training phase.

³ A hidden layer designates a layer of neurons located between the input vector and the output layer.

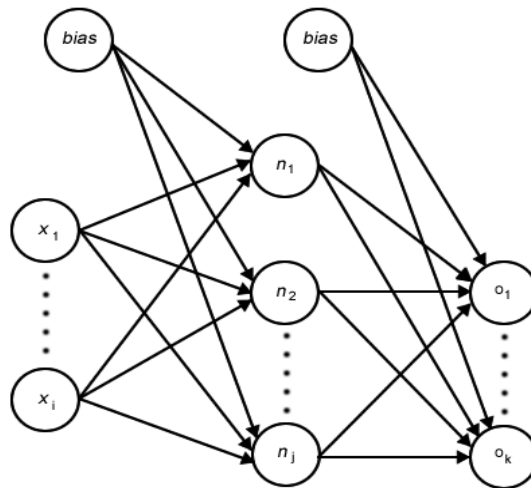


Figure 1 – Diagram of a network with i covariates, 1 hidden layer with j neurons, k output neurons and bias on the hidden and output neurons.

This type of network is well suited for classification problems and tends to perform well in large covariate spaces with multicollinearity (Hastie, Tibshirani, & Friedman, 2009; Intrator & Intrator, 2001; Kotsiantis, 2007) and has already obtained very good results in bat classification in the past (Jennings *et al.*, 2008; Parsons & Jones, 2000; Redgwell *et al.*, 2009; Walters *et al.*, 2012)

2.2.2 Training

Multilayer perceptrons used for classification and regression are usually trained using supervised learning methods, where the correct output is known for every input vector. The network computes the input vectors (initially with random weights and intercepts), compares the output result with the correct output using an error function and adjusts the weights iteratively, using a set of previously defined rules – the learning algorithm – trying to minimize the error function, by adding a user defined learning rate (Rojas, 1996). The error function is a summary measure of the total error of the network and a common choice is the sum of squared errors:

$$E = \frac{1}{2} \sum_{l=1}^L \sum_{h=1}^H (o_{lh} - y_{lh})^2 \quad (\text{Eq. 12})$$

where l indexes the observations and h the output neurons (Günther & Fritsch, 2010).

The learning algorithm then uses information from the error function to update the weights. Although there are several different algorithms we will focus on the globally convergent resilient backpropagation (GRPROP) which is a gradient descent method that uses first derivative related information for the update of the weights, based on the traditional backpropagation algorithm (Anastasiadis *et al.*, 2005).

The backpropagation algorithm became very popular in the 1980's due to the work of Rumelhart & McClelland (1986). In this algorithm the gradient of the error function is calculated with respect to the weights $\left(\frac{\partial E}{\partial w}\right)$, and the weights of the $t + 1$ iteration are calculated according to the delta rule:

$$w_k^{(t+1)} = w_k^{(t)} - \eta \left(\frac{\partial E^{(t)}}{\partial w_k^{(t)}} \right) \quad (\text{Eq. 13})$$

where k indexes the weights and η is the user defined learning rate (Günther & Fritsch, 2010; Rojas, 1996). The weights are modified in the opposite direction of the gradient of the error function until a local minimum is reached. In Figure 2 we can see a graphical representation of the backpropagation algorithm for a univariate error function. In the left side of the figure the gradient of the error function is negative and the weight is increased, in the right side is positive and the weight is decreased.

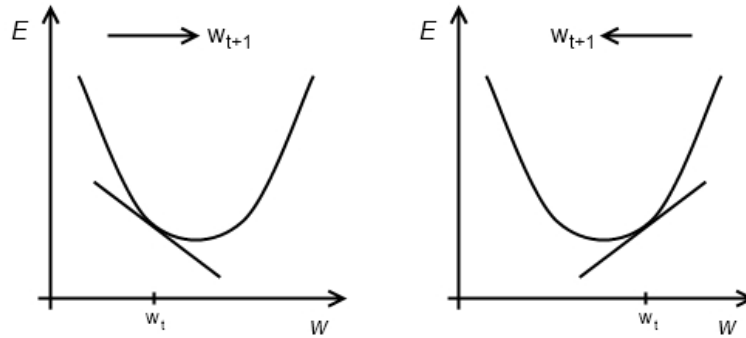


Figure 2 – Basic concept of the backpropagation algorithm illustrated for a univariate error function $E(w)$, from Günther & Fritsch (2010)

Because this algorithm just has one learning rate for the update of the entire set of weights it is very slow to converge. Additionally the influence of the learning rate is higher for some weights because it depends on the magnitude of the gradient of the error function which can sometimes difficult convergence (Günther & Fritsch, 2010; Rojas, 1996).

Riedmiller & Braun (1993) introduced some major modifications to the BP and developed the resilient backpropagation algorithm (RPROP). In RPROP there is a learning rate for each weight that changes during training and an equal influence of the learning rate over the entire network is guaranteed because only the sign of the partial derivative is used and not its magnitude. In this algorithm the weights are updated according to:

$$w_k^{(t+1)} = \begin{cases} w_k^{(t)} - \eta_k^{(t)}, & \text{if } \left(\frac{\partial E^{(t)}}{\partial w_k^{(t)}} \right) > 0 \\ w_k^{(t)} + \eta_k^{(t)}, & \text{if } \left(\frac{\partial E^{(t)}}{\partial w_k^{(t)}} \right) < 0 \\ 0, & \text{otherwise} \end{cases} \quad (\text{Eq. 14})$$

and the learning rates are updated according to:

$$\eta_k^{(t+1)} = \begin{cases} \eta^+ * \eta_k^{(t)}, & \text{if } \left(\frac{\partial E^{(t)}}{\partial w_k^{(t)}} * \frac{\partial E^{(t+1)}}{\partial w_k^{(t+1)}} \right) > 0 \\ \eta^- * \eta_k^{(t)}, & \text{if } \left(\frac{\partial E^{(t)}}{\partial w_k^{(t)}} * \frac{\partial E^{(t+1)}}{\partial w_k^{(t+1)}} \right) < 0 \\ 0, & \text{otherwise} \end{cases} \quad (\text{Eq. 15})$$

where η^- denotes the decrease factor and η^+ the increase factor for the learning rate where $0 < \eta^- < 1 < \eta^+$ (Günther & Fritsch, 2010; Riedmiller, 1994; Rojas, 1996).

Every time the gradient of the error function for a specific weight changes sign over two consecutive iterations it's an indication that the last learning update was too large causing the algorithm to jump over a local minimum. In this case the learning rate is decreased by a factor η^- and the weights aren't updated at this iteration to enforce returning to the local minimum. If the gradient retains its sign it's an indication of a shallow region of the error function and the learning rate is increased by a factor η^+ to accelerate convergence. Constraints for the minimum and maximum values that the learning rates may assume can be added to the algorithm to avoid too large or too small weight updates.

Although RPROP outperforms the traditional backpropagation algorithm, especially in large covariate spaces with multicollinearity, the assumption that a change of sign of the partial derivatives implies a jump over a local minimum does not take into account whether the weight update increased or decreased the error. In fact the change in signs could also mean a jump over a local maximum. Over the years several modifications for RPROP have been proposed to overcome this drawback one of which is the globally convergent resilient backpropagation algorithm (GRPROP) introduced by Anastasiadis *et al.* (2005).

The difference between RPROP and GRPROP is that in the latter an additional modification on the smallest learning rate or on the learning rate associated with the smallest gradient of the error function is performed at each iteration. This single learning rate is modified by:

$$\eta_i^{(t)} = - \frac{\sum_{k; k \neq i} \eta_k^{(t)} * \left(\frac{\partial E^{(t)}}{\partial w_k^{(t)}} \right)}{\frac{\partial E^{(t)}}{\partial w_i^{(t)}}}, \left(\frac{\partial E^{(t)}}{\partial w_i^{(t)}} \right) \neq 0 \quad (\text{Eq. 16})$$

Anastasiadis *et al.* (2005) proved that GRPROP is able to converge to a local minimum of a non-linear function from almost any starting point and showed by benchmark simulation that the algorithm is faster and has more accurate predictions than backpropagation and RPROP.

As a final remark it should be noted that regardless the training algorithm used, standardization of the input vectors for values that can be mapped by the activation function is very advisable and sometimes even required for the convergence of the network parameters (Heaton, 2005).

2.2.3 Architecture selection and validation

When applying an ANN to solve a specific problem the best architecture (number of hidden layers and number of neurons) is most often unknown. There are some guidelines to help with this decision but the usual way to search for the best solution is by testing several architectures and comparing their performances. One hidden layer is considered enough for the majority of classification problems and the number of neurons in this layer should be between a minimum of 2/3 and a maximum of twice the number of neurons in the input layer. With less than 2/3 the network won't probably fit and with more than the double there is a great probability of overfitting which would lead to poor generalization capabilities (Heaton, 2005).

As ANN don't have a unique solution, that is for a given data set and any given architecture the weight matrix is not uniquely determined because ANN in general find many local minima when starting from different random initial weights. Thus, averaging the predictions of several models can reduce variance and lead to better prediction (Intrator & Intrator, 2001). This way using an ensemble of networks (ENN) for prediction instead of a single network will probably lead to better classification results. The number of experiments, e , with different random starting weights, needed to ensure at least one solution within a desirable lower percentile of all possible experiments is given by (Iyer & Rhinehart, 1999):

$$e = \frac{\ln(1 - \alpha)}{\ln(1 - p)} \quad (\text{Eq. 17})$$

where $1 - \alpha$ is the confidence level and p is the percentile.

For validating trained ANN models a cross-validation method is most often used. For a given dataset, only some of the observations are used to train the network, normally around 70-80%, leaving the remaining of the observations to validate the performance of the network. This way the generalization capability of the model is tested because the final results come from observations that were not entered in the training process. The average error over the validation set of observations is taken as an approximation of the true error of the model over the input space, especially if the sample effectively reflects the population variability (Rojas, 1996).

2.2.4 Model interpretation

Most of ANN configurations are black-boxes, i.e., they output a result but do not give any information on how the covariates are related to that result, and besides the result no other information is available from the network. However, in a multilayer feed-forward network using logistic activation function on the hidden and output layers it's possible to know the relative contribution of each covariate to the result and if the effect is linear or not by calculating the generalized weights, which are similar to weights in logistic regression (Intrator & Intrator, 2001).

In terms of the log odds, the ANN can be written as:

$$\log \left(\frac{o(x)}{1-o(x)} \right) = w_{0k} + \sum_j w_{jk} * f_0 \left(w_{0j} + \sum_i x_i * w_{ij} \right) \quad (\text{Eq. 18})$$

and by taking the derivative, the effect of an infinitesimal change in the variable x_i on the logit transform of the output function can be observed (Intrator & Intrator, 2001). Since the odds are expressed as a linear combination of the inputs, the effect of each covariate x_i on the log odds of each neuron is given by the individual weights w_{ij} , and the effect of each covariate over the entire set of neurons, i.e. over the entire ANN, is given by the generalized weights.

$$\bar{w}_{ik}(x) = w_{0k} + \sum_j w_{jk} * f_0' \left(w_{0j} + \sum_i x_i * w_{ij} \right) w_{ij} \quad (\text{Eq. 19})$$

Thus, in neural network modeling, the generalized weights have similar interpretation as weights have in logistic regression, i.e., the contribution to the log odds. However, unlike logistic regression where weights are unique for each covariate, in ANN there is a generalized weight for each specific point of a covariate and the variance of the generalized weights for a particular covariate can be visually inspected for an indication of the linearity of its effect. A small variance of the distribution of the generalized weight indicates that the effect of the covariate is linear (Intrator & Intrator, 2001).

A robust generalized weight (RGW) for each covariate is an average of the generalized weights obtained for predictions of an ensemble of estimates. A scatterplot of the RGW of each variable with respect to its values provide a mean for examining the possibility of nonlinearity, although not necessarily its form (Intrator & Intrator, 2001).

Although the RGW approach could, in theory, be used for interpreting all multilayer feed-forward networks using logistic activation function on the hidden and output layers Intrator & Intrator (2001) only demonstrated its validity on networks trained for binomial classification or regression with one hidden layer.

3. METHODS

In this chapter are presented the methods used in this study for recording bat calls, development of the program for automatic analysis of those recordings and the modeling strategies used for bat call classification.

3.1 SAMPLING

As mentioned before bat echolocation calls show great variability, which differs among species. Thus, different sample sizes will be needed according to the species. To optimize resources, as field work is logistically very demanding, a reference sample size was calculated for each species using the random sampling method. The reference sample number for each species was obtained using the expression (Cochran, 1977):

$$n_i = s_i^2 \left(\frac{z_{1-\alpha/2}}{e} \right)^2 \quad (\text{Eq. 20})$$

where n_i is the sample size for species i ; s_i is the estimated standard deviation for the parameter of species i ; $z_{1-\alpha/2}$ is the critical value for the confidence level $1 - \alpha$ and e is the maximum desired error.

The calculations were based on a call parameter, frequency of maximum energy (fmaxe), which is one of the most widely used parameters for characterizing echolocation calls and is sometimes referred as the characteristic frequency of the species. The standard deviation of fmaxe is unknown for most species or is different between studies from different authors. To overcome this we estimated the standard deviation for each species based on the minimum and maximum values obtained from previously published studies (Parsons & Jones, 2000; Russ, 1999; Russo & Jones, 2002; Zingg, 1990). For skewed distributions a good estimate of the standard deviation s is given by Hozo *et al.* (2005):

$$s = \frac{\max - \min}{4} \quad (\text{Eq. 21})$$

For a confidence level of 95% and a maximum error of 0.8 kHz regarding the real mean of fmaxe of the species population we obtained the reference sample sizes for each species. Our calculations show that a total of 1263 recordings of these 24 species⁴ are a sufficient number for our analysis. The sample size ranged from six recordings for *Rhinolophus ferrumequinum* to 247 recordings for *Myotis escalerai* (Table 3). Although the sampling scheme was planned for all species⁵ we could not capture/record three of them: *M. mystacinus*, *M. bechsteinii* and *H. savii*. Furthermore at this point we still did not reach the reference samples for most species. Nevertheless, compilation of bat call recordings is a demanding process and is still undergoing,

⁴ Even though there are 25 species in mainland Portugal, *N. noctula* is extremely rare and was intentionally left out of this study due to the very low capture probability.

as the work presented in this thesis is part of a research project that begun in 2011 and is still in active research state.

Table 3 – Maximum and minimum values of f_{max} in literature for each species and the respective calculated reference sample size for each species (Russ, 1999; Russo & Jones, 2002; Zingg, 1990)

Species	F _{max} minimum	F _{max} maximum	Reference sample size
<i>Rhinolophus ferrumequinum</i>	80.0	84.0	6
<i>R. hipposideros</i>	106.4	113.3	18
<i>R. euryale</i>	102.0	109.0	18
<i>R. mehelyi</i>	104.0	112.0	24
<i>Myotis myotis</i>	27.0	33.0	14
<i>M. blythii</i>	27.0	33.0	14
<i>M. mystacinus</i>	41.7	69.4	128
<i>M. emarginatus</i>	50.0	60.0	38
<i>M. escaleraei</i>	27.9	66.4	247
<i>M. bechsteinii</i>	45.6	55.2	35
<i>M. daubentonii</i>	39.6	56.7	110
<i>Pipistrellus pipistrellus</i>	41.7	51.8	38
<i>P. pygmaeus</i>	48.8	61.6	61
<i>P. kuhlii</i>	35.2	44.8	35
<i>Hypsugo savii</i>	30.4	37.6	19
<i>Nyctalus leisleri</i>	20.8	28.8	24
<i>N. lasiopterus</i>	16.8	26.8	38
<i>Eptesicus serotinus</i>	23.2	35.2	54
<i>E. isabellinus</i>	23.2	35.2	54
<i>Barbastella barbastellus</i>	28.0	40.0	24
<i>Plecotus auritus</i>	31.7	45.0	66
<i>P. austriacus</i>	31.7	45.0	66
<i>Miniopterus schreibersii</i>	48.8	70.7	80
<i>Tadarida teniotis</i>	8.8	15.2	15

Although a random choice of capture/recording sites would be advisable, constraints related to bat activity, distribution status and capturing probability made it impossible. Sites had to be chosen to maximize capture/recording probability.

Recordings were made at 39 distinct locations all over the country (Figure 3) after hand release of captured bats, outside known roosts or at foraging sites where the identification of the species by echolocation calls was unambiguously. This was possible to achieve with *B. barbastellus* as this species alternates between two very peculiar call types at distinct frequencies (Denzinger *et al.*, 2001) and with *P. pipistrellus* and *P. kuhlii* at foraging sites. Although this two species can be confused one with the other in high to medium cluttered environments, in situations of low clutter the shape of their calls and *f_{maxe}* are different and the distinction can be done with some degree of certainty (Russo & Jones, 2002). We are aware that the inclusion of recordings of free flying *P. pipistrellus* and *P. kuhlii* was risky and probably introduced some mismatch classifications in the database, but the sample sizes resulting from recording outside roosts or after hand release, for these species, were too low for analysis and thus leading to this decision.

For *Plecotus* species, as they have very distinctive multi-harmonic echolocation calls (Russo & Jones, 2002; Waters & Jones, 1995), the discrimination from all other species is quite easy, even though the distinction between *P. auritus* and *P. austriacus* is not. As these species are very hard to capture we decided to consider only the genus *Plecotus* for classification purposes to allow for the use of free flying recordings.

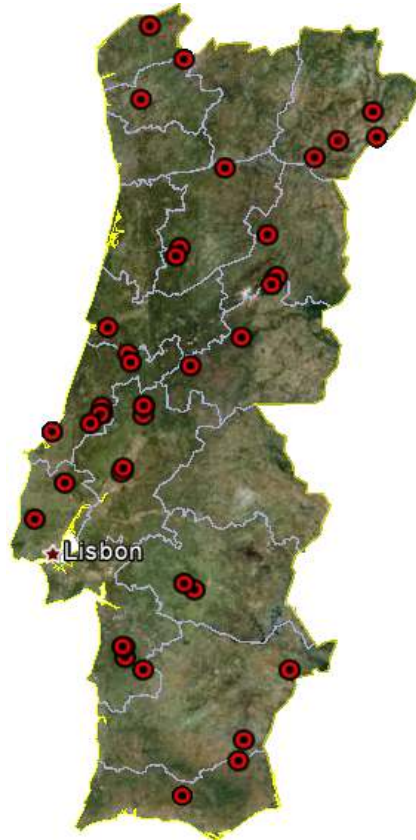


Figure 3 – Location of the 39 recording sites in mainland Portugal with capital city (Lisbon) highlighted (Google Earth Image)

In order to increase capture probabilities and to speed up recordings compilation, a database of known roosts was assembled from previously published work (Palmeirim & Rodrigues, 1992; Rainho, 1995, 1996; Rodrigues & Palmeirim, 2008; Rodrigues, 1996) and by direct contact with Portuguese bat researchers.

Bats were captured with hand-nets, harp traps and/or mist nets and in some cases by hand, according to the place of capture. Mist nets were usually mounted around water tanks, over water streams or small ponds and along the riverside (Kunz & Parsons, 2009) as can be seen in Figure 4.



Figure 4 – Capture techniques used. Top left – harp trap at the entrance of a cave; top right – capturing bats inside a cave with hand-net; bottom left – mist nets surrounding a water tank; bottom right – mist nets covering a bridge over a stream of water

The captured bats were identified by a bat expert to species level along with gender and approximate age. Weight and forearm were measured and occasionally, when the morphological identification was very inconclusive, a genetic sample was taken (Figure 5).



Figure 5 – Measuring bats after capture. Top left – releasing a captured bat from a mist net; top right – one of the camp sites; bottom – measurements of captured bats

Search phase calls, as defined by Griffin, Webster, & Michael (1958), are one of the most species-specific call variety (Fenton & Bell, 1981) and they are emitted continuously throughout the flight (Pfalzer & Kusch, 2003), probably making them the most suited calls for bat species identification. Social calls, another very species-specific call variety are significantly more prone to variation than search-phase calls (Fenton, 1994) and much less known, besides being emitted only occasionally during bat activity (Pfalzer & Kusch, 2003). This lead us to the decision of only using search phase calls for classification purposes. Nevertheless, immediately after taking flight from a roost or after hand release, bats do not usually use search phase calls as they are accelerating and orientating and not in a steady flight. To compensate for this, all the recording were made as far as possible from the release point or roost (Parsons & Jones, 2000).

When hand releasing, the use of light tags (Kunz & Parsons, 2009) helped to increase the distance between release and recording points to at least 8 to 10 meters and sometimes more because it was possible to track the bat for some time (Figure 6). Increasing the distance (and consequently time) between release and record points improves the likelihood that the bat is already in a search phase flight upon recording. Reflective tape was also tested instead of the light tag but with little success. Outside known roosts a night vision monocular with head mount

(Pulsar Challenger GS 1x20) was used to maximize the distance to the entrance point while still being sure that the bats recorded emerged from the roost.



Figure 6 – Recording of bat calls. Top left – bat taking off a tree surface; top right – recording of a hand released bat; bottom left – hand release of a bat with light tag; bottom right – recording of free flying bats outside a known roost

Recordings were made using D240x bat detectors (Pettersson Elektronik AB) with 10x time-expanded mode and sample rate of 44.1 kHz connected to a digital recorder (Archos Gmini 402) and D1000 bat detectors (Pettersson Elektronik AB) recording in real-time at a sample rate of 300.0 kHz.

3.2 DETECTING AND ISOLATING BAT CALLS

A program implemented in R environment was used to filter the recordings, detect and isolate echolocation calls, and measure parameters.

After importing the recording and prior to analysis, a standard high-pass 10th order Butterworth filter at 8 kHz was applied. Then, a Fast Fourier Transform (FFT) was computed for the entire recording producing a two-dimensional array of amplitude values with X and Y dimensions representing time and frequency (this array can be visually presented as a power spectrum) to obtain the frequency with maximum amplitude over the entire recording. Again a

10th order Butterworth filter was applied, this time with a high-pass at 1/3 of the frequency of maximum amplitude. This procedure excluded most of the insect noise from the recordings without interfering with the bat calls.

There were problems though when insect noise was in the ultrasonic range of frequencies and louder than every bat call on the recording (after the first Butterworth filter was applied) if bat calls f_{max} was at lower frequencies than the insect noise. In these recordings the second Butterworth filter could affect bat call energy, introducing errors in posterior measurements, and we decided to discard them. The R code which enabled to import, filter and obtain f_{max} over the entire recording is presented in Appendix 1.

After filtering, and for detecting and isolating bat calls the raw signal was transformed by the Teager energy algorithm and the energy envelop was obtain using a roll mean of 100-points on the absolute energy values. Teager algorithm is one way of computing the energy of the signal. Given a sinusoidal signal, the instantaneous energy E_i of the sample x_i is given by (Gu, 2002):

$$E_i = x_i^2 - x_{i+1}x_{i-1} \cong A^2\Omega^2 \quad (\text{Eq. 22})$$

where A is the amplitude and Ω is the discrete time frequency. From Equation 22 it's easy to observe that the Teager energy expression is based not only on the signal amplitude, but also on the corresponding frequency component. In reality what happens to the signal is that the amplitude of the higher frequency components with higher energy is inflated and the amplitude of the lower frequency components with lower energy is deflated. The R function for computation of the Teager algorithm is presented in Appendix 1.

Figure 7 shows the energy envelope of a call from *E. serotinus* with some undesired noise near the end (an echo) before and after the transformation with the Teager algorithm. With the transformation is evident that the call is more detached from the background noise and the beginning and the end of the call are easier to detect. One of the drawbacks of this technique is that it shortens the call. Normally no more than 50-100 μs , but in the case of very faint calls the shortening can be up to 500 μs .

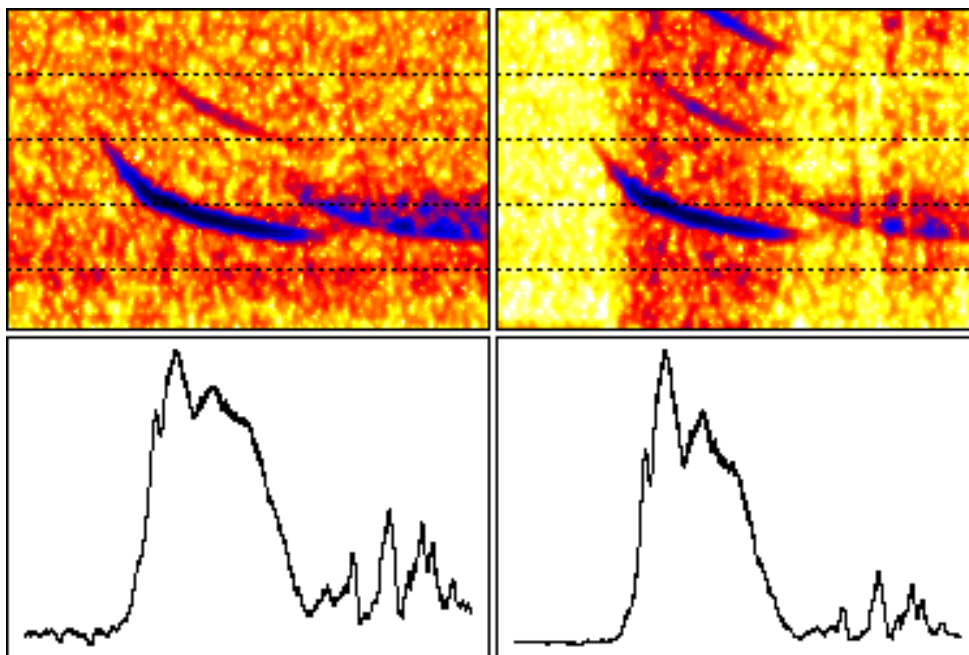


Figure 7 – Spectrogram (top) and energy envelope (100 points rollmean) of an *E. serotinus* call. Raw signal (left) and Teager energy transformed signal (right).

The signal was then shortened 200 times through linear interpolation using the R base function “approx” and the amplitude peaks were detected using the function “fpeaks” from package “seewave” (Sueur, Aubin, & Simonis, 2008). The shortening through linear interpolation was used to (1) smooth the energy envelope making the “fpeaks” function more reliable and (2) to increase the speed of peak detection (otherwise it could take an unreasonable amount time to process longer sound files). The maximum number of peaks detected on each recording was 10, starting by the one with greater amplitude with a minimum interval of 25 ms imposed between two consecutive peaks.

For isolating/extracting bat calls the start and end positions of the call have to be determined. First a segment of 40ms centered at the point of maximum energy of the call was selected (signal with original length and with Teager energy). Then this segment was split in half, and the function “binseg.meanvar.exp” was applied to each of the halves to detect the major change point (in each half) of the amplitude variance using a binary segmentation method (Killick & Eckley, 2011). The major change point detected in the first half is the start of the call and the major change point detected in the second half is the end of the call (see Figure 8).

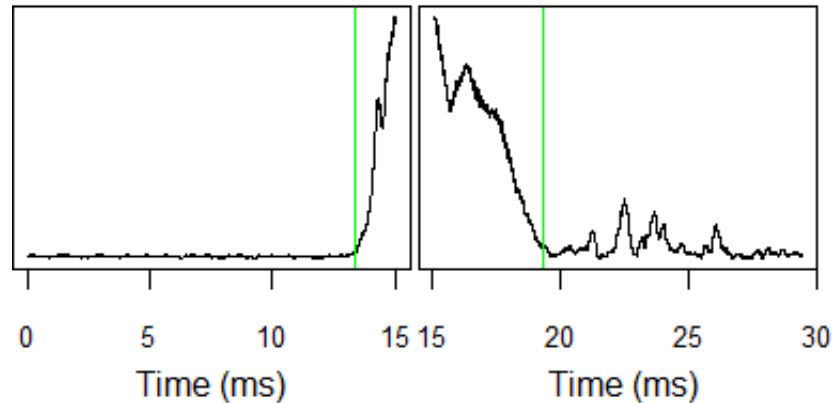


Figure 8 – Identification of the start point of the call (left) and end point of the call (right) with the binary segmentation method.

This method for isolating bat calls revealed very adequate for most bat species, even in noisy conditions and weak amplitude calls. However, for high-duty cycle bats, of which *Rhinolophus* genus is the example in Portugal, it did not work (*Rhinolophidae* is one of the three high-duty families of bats in the world (Airas, 2003) and the only one known in Portugal). As this method only seemed adequate to handle low-duty cycle bats another strategy had to be used for *Rhinolophus* species. As these species echolocate at very high frequencies, higher than any other species in Portugal, we could use a cutpoint to discriminate *Rhinolophus spp.* from all other species. After importing the recording and applying the first of the two Butterworth filters used, f_{maxe} was obtained over the entire recording and if the value was over 75 kHz (for further details please refer to Section 4.2.1) the algorithm would not try to detect and isolate individual calls and just outputted the f_{maxe} value of the entire recording. The R function to detect and isolate bat calls from a recording is presented in Appendix 1.

3.3 PARAMETER MEASUREMENT

For classification of echolocation bat calls we measured 19 variables which resume information from the time and frequency domains (Table 4). As referred in Section 1.7 we decided not to use some of the variables commonly used for describing bat calls in previous studies (namely initial frequency, final frequency and call bandwidth) but they were still measured for future reference and comparison with previously published studies.

Table 4 – Parameters measured from bat calls for classification purposes

Parameter	Parameter description
fmaxe	Frequency with maximum energy
dur	Call duration
fini15	Initial frequency standardized at -15dB of fmaxe intensity
fini10	Initial frequency standardized at -10dB of fmaxe intensity
fini5	Initial frequency standardized at -5dB of fmaxe intensity
ffin15	Final frequency standardized at -15dB of fmaxe intensity
ffin10	Final frequency standardized at -10dB of fmaxe intensity
ffin5	Final frequency standardized at -5dB of fmaxe intensity
bw15	Bandwidth between fini15 and ffin15
bw10	Bandwidth between fini10 and ffin10
bw5	Bandwidth between fini5 and ffin5
posFMAXE	Time position of fmaxe
ptFMAXE	Time position of fmaxe as a proportion of call duration
bwffini15	Bandwidth between fmaxe and fini15
bwffini10	Bandwidth between fmaxe and fini10
bwffini5	Bandwidth between fmaxe and fini5
bwffin5	Bandwidth between fmaxe and Ffin5
bwffin10	Bandwidth between fmaxe and Ffin10
bwffin15	Bandwidth between fmaxe and Ffin15

The main function used for measuring parameters was called “freq_maxe”. This function normalizes the length of a selected portion of a signal to 32768 points (through zero padding) and computes the FFT over the normalized segment. It returns a list of the frequency range (kHz) versus amplitude (dB) and the frequency with greater amplitude. For computing freq_maxe two auxiliary functions were also coded, one for zero padding a signal to a desired length and another to compute FFT (based on the R base function fft). The R functions “freq_maxe”, “zero_pad” and “fft_new” are presented in Appendix 1.

After the measurement of all the isolated calls, a control algorithm eliminates a particular call from the final results if its fmaxe value was more than 8 kHz apart from the fmaxe median value of all calls. This was implemented to eliminate misidentified calls that could derive from the call detection algorithm (especially due to insect noise) and as a consequence, the higher frequency calls of *B. barbastella* (this species alternates between two very peculiar call types at distinct frequencies) were always removed from the analysis because they were always detected in smaller number. This happened because the detection algorithm starts by the more intense calls in the recording and these higher frequency ones were always fainter.

The parameters used for classification were measured as indicated in Table 5. Initial frequency and final frequency were measured by applying function “freq_maxe” to the first and last 0.25 ms of the isolated calls, being call bandwidth the difference between the two.

Table 5 – Description of the measured parameters from bat calls for classification purposes

Parameter	Measurements description
fmaxe	Obtained by applying “freq_maxe” to the entire call
dur	Duration of the extracted call
posFmaxe, ptFmaxe	Measured on the energy envelope of the call. A 10 point rollmean on the absolute value of amplitude is applied to the raw signal to obtain the energy envelope. The energy is then normalized between 0 and 1 dividing by maximum energy. The position of maximum energy is obtained (posFmaxe) and divided by total duration of the call to obtain ptFmaxe
fini15, fini10, fini5, ffin15, ffin10, ffin5	“freq_maxe” is applied to the entire call. Further analysis is conducted on the resulting power spectrum. For the initial frequencies an algorithm iterates from the frequency with most intensity to the frequency at the beginning of the call registering the frequencies where the intensity drops 15dB, 10dB and 5dB. For the final frequencies an algorithm iterates from the frequency with most intensity to the frequency at the end of the call registering the frequencies where the intensity drops 15dB, 10dB and 5dB
bw5, bw10, bw15	Difference between the respective initial and final frequencies. The difference between ffin5 and fini5 is the bw5, and so on
bwffini15, bwffini10, bwffini5, bwffin15, bwffin10, bwffin5	Difference between fmaxe and specific call positions, bwffini15 is the frequency difference between fmaxe and fini15, bwffin15 is the frequency difference between fmaxe and ffin15 and so on

A schematic representation of the measured parameters can be seen in Figure 9.

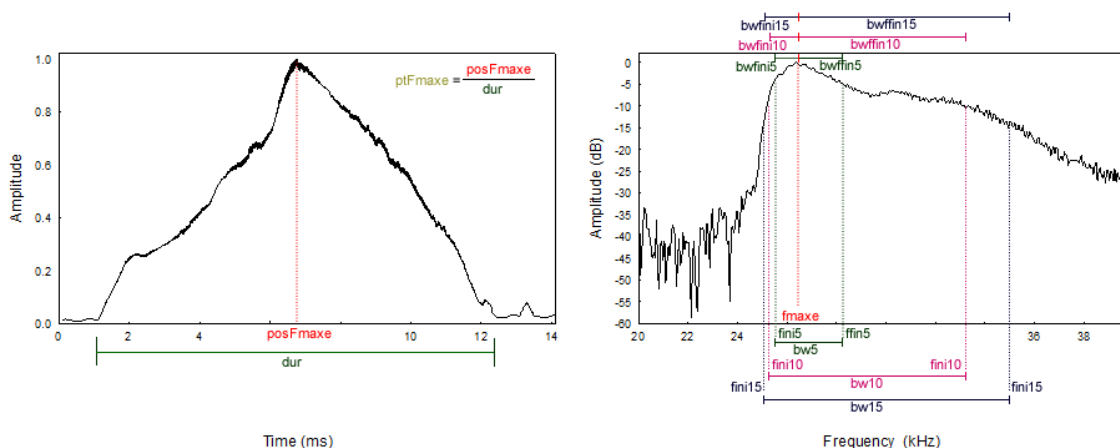


Figure 9 – Schematic representation of the measured parameters that were used for classification purposes

The program was implemented with a GUI in Tcl/Tk. It uses the R graphics engine, has custom table outputs and a custom menu. Screenshots can be seen in Figure 10.

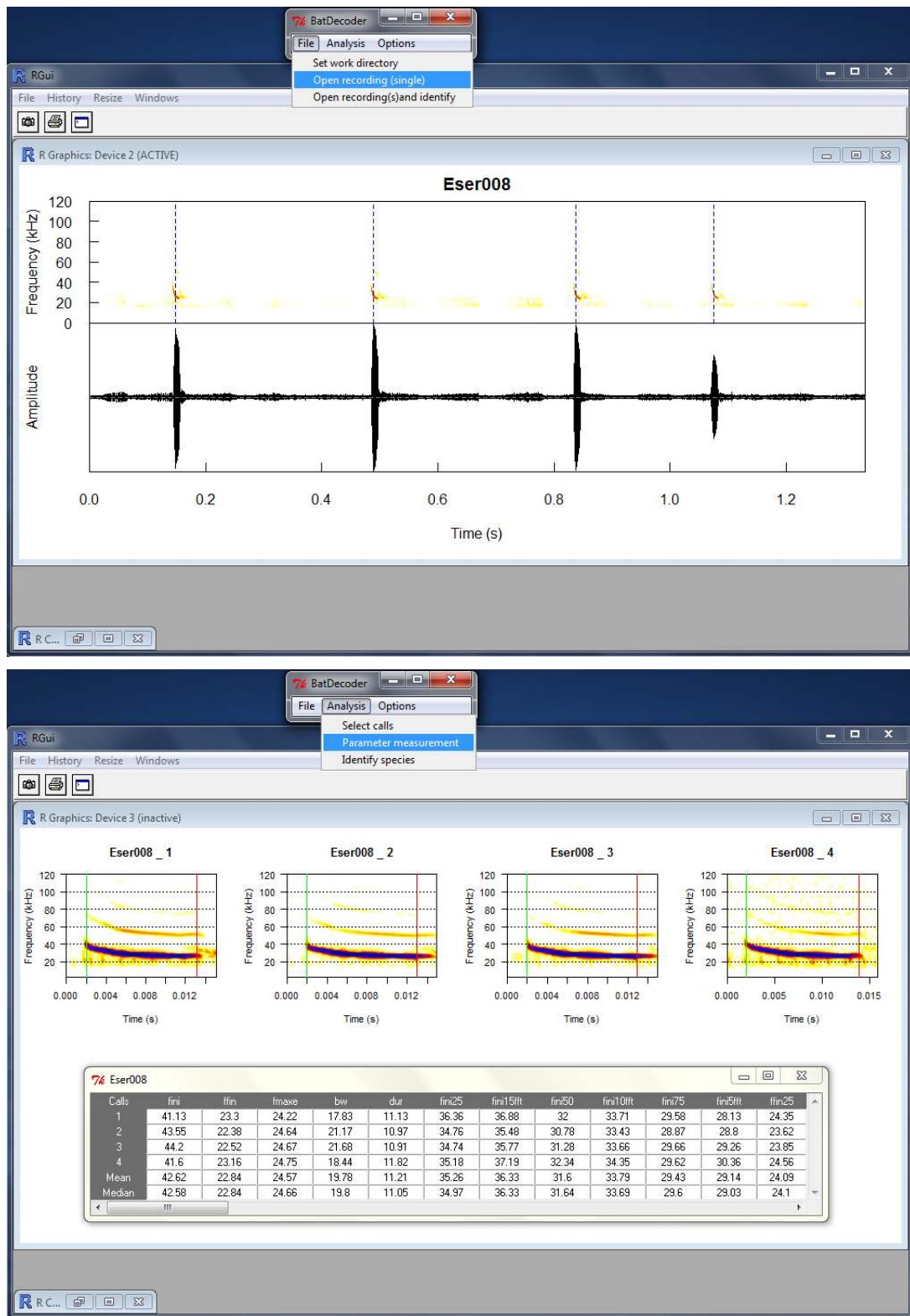


Figure 10 – Screenshots of the implemented analysis GUI. Top – after importing and filtering a recording, a spectrogram and oscillogram is drawn and the calls are automatically identified (blue line). Bottom – after the parameter measurements, a spectrogram representation of the automatically extracted calls with the start (green line) and end (red line) identified and a table with the parameters measured for all calls, with their respective mean and median presented

3.4 BAT CALL CLASSIFICATION

For the classification of bat calls we used a hierarchical classification scheme to enable classification to genus or subgroup level for those calls that cannot be confidently classified to species level (Obrist *et al.*, 2004; Walters *et al.*, 2012).

We divided the classification process in two stages. In the first stage calls are classified into one of six major groups of species: *Rhinolophus* (including all species of genus *Rhinolophus*), *Myotis* (including all species of genus *Myotis*), *Pipistrelloid* (including all species of genera *Pipistrellus* and *Miniopterus*), *Nyctaloid* (including all species of genera *Nyctalus*, *Eptesicus* and *Tadarida*), *Plecotus* (including all species of genus *Plecotus*) and *Barbastella* (including only the species *B. barbastella*). After a call was classified as belonging to one of these major groups a second classification into species (or groups of species) was performed inside each major group except for *Barbastella*, as this group is composed of only one species and *Plecotus*, as the two species of this group weren't discriminated in this study (refer to Section 3.1 for further details). The classification scheme can be observed in the Figure 11.

As some of the species had very small sample sizes they were grouped, based on biological and echolocation characteristics closeness, for classification purposes. The groups formed are as follows:

- *Eptesicus* spp. (including species *E. serotinus* and *E. isabellinus*);
- *Nyctalus* spp. (including species *N. leisleri* and *N. lasiopterus*);
- *R. euryale* / *R. mehelyi*;
- *M. myotis* / *M. blythii*;
- *M. emarginatus* / *M. escalerae* / *M. daubentonii*.

We used two different approaches for classification of calls, ensembles of neural networks and logistic regression models. Before any models were made, an exploratory analysis was conducted to identify parameters that by themselves alone could discriminate some groups with great accuracy. This was done not only to reduce some categories in certain classifications, and therefore facilitate fitting logistic models (for ENN this wasn't an issue), but also to provide a better insight on the differences between echolocation calls of the species.

For the classification we fitted four different sets of statistical models, the first classified calls into major groups, the second classified the *Myotis* species, the third the *Pipistrelloid* species and the fourth the *Nyctaloid* species.

In the neural network approach we trained four sets of ENN, three multinomial: for major groups of species, for *Pipistrelloid* species and for *Nyctaloid* species. And one binomial for *Myotis* species. We used one output neuron for each category being classified, even for the

binomial ENN. Although this was not strictly necessary it allowed for a null classification, i.e., the ENN is not forced to classify every case to a category.

For the logistic regression approach, we only used three different sets. Two binomial logistic regression models: for *Nyctaloid* species and for *Myotis* species; and a multinomial one for *Pipistrelloid* species.

Classification of *Rhinolophus* species was done by defined cutpoints using the only variable that was possible to measure for these species (see Section 4.2.2).

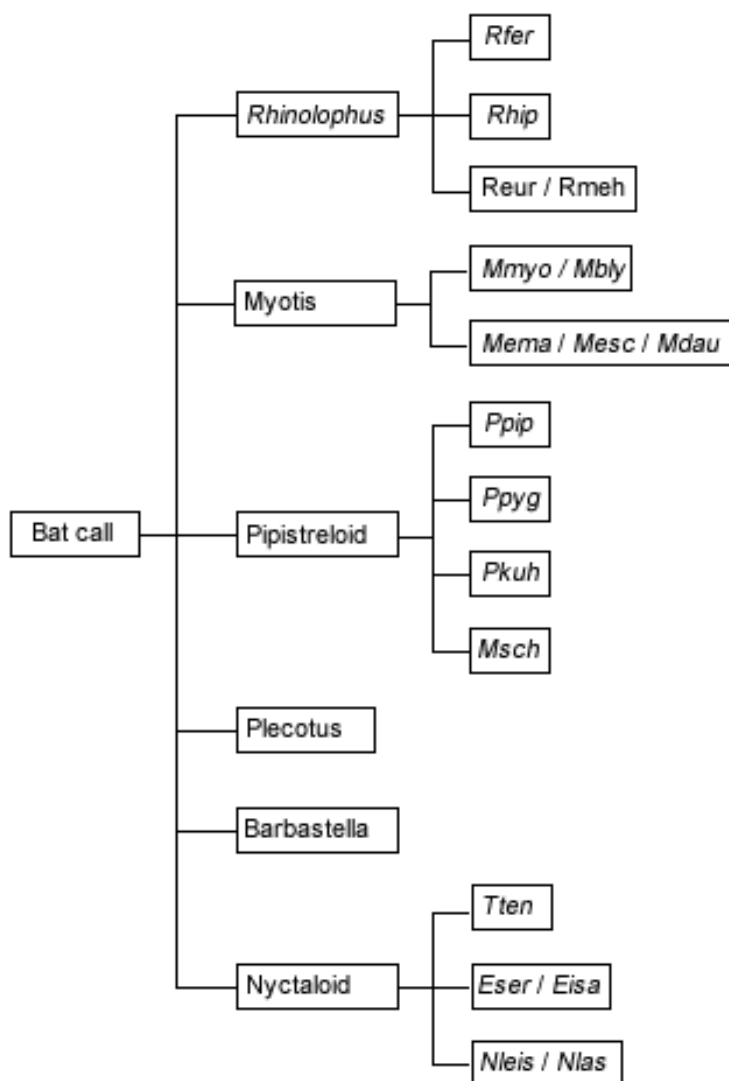


Figure 11 – Two stage identification scheme

3.5 STATISTICAL MODELING WITH R

3.5.1 Logistic regression

The binomial logistic regression models were fitted in R using the package “brglm”. This package allows fitting binomial GLMs using either a modified-score approach to bias reduction or maximum penalized likelihood where penalization is by Jeffrey’s invariant prior. Fitting takes place by iteratively fitting a local GLM on a pseudo-data representation. More flexibility is provided in this package by the fact that custom pseudo-data representations can be specified and used for model fitting and functions for the construction of confidence intervals for the bias-reduced estimates are provided (Kosmidis, 2007).

For fitting multinomial models the package “mlogit” was used but we used the strategy of fitting multiple binomial models as suggested by Hosmer & Lemeshow (2000) to validate the assumptions.

Using all individual measured calls revealed extremely difficult for fitting and validating the models. There were many outliers and the assumption of linearity was nearly impossible to verify. To reduce variability in the data and facilitate LR modeling we opted for using just the median of each recording.

Unlike neural networks, where every variable could be used despite the colinearity between them, in logistic regression models we had to choose an initial set of covariates with less colinearity as possible. As some covariates are simply linear functions of others they were immediately discarded. For the remaining covariates we evaluated the Spearman rank correlation and selected those that presented the lowest ρ between them, but there was always some degree of correlation in the initial set of covariates used.

The linearity of the covariates was accessed by the smoothed lowess, quartiles and fractional polynomials methods; outliers were accessed by the Pearson and Deviance residuals and influential observations by Cook’s distance and Dfbeta. The goodness of fit was accessed by the Hosmer and Lemeshow test (function presented in Appendix 1) and the Nagelkerke’s R^2 .

3.5.2 Multilayer feedforward perceptron networks

Throughout this thesis we used a particular class of artificial neural networks, the multilayer feedforward perceptron. The networks were designed and trained using the R package “neuralnet” (Günther & Fritsch, 2010).

The networks had one hidden layer and used the logistic activation function in the hidden and output layers. They were trained using a supervised learning algorithm, the GRPROP, with a weight increase factor of 1.2, a weight decrease factor of 0.5 and an initial learning rate of 0.1,

without any boundaries for the learning rate. The error function used in the training was the sum of squared errors and the observations were presented in batch mode to the network.

Instead of single networks we used ensembles of networks. Equation 17 was used to estimate the ensembles size and number of repetitions needed. We wanted to assure at least one solution with a confidence, $1 - \alpha = 99\%$ amongst the best $p = 20\%$ performing networks, so we trained 21 networks and used an ensemble of the top performing 5 for prediction. The individual predictions were averaged to obtain the final prediction.

The original dataset was split in two, with 75% of the data destined for training purposes and the other 25% just for validating the trained networks. A custom function was implemented to randomize observations and split the data in the two desired sets while maintaining the original strata proportions on the new datasets. This function made use of the package “sampling” and is presented in Appendix 1.

All individual measured calls were used for the ENN modeling in order to include as much variability as possible in these models and to have as many training and validation cases as possible. On Table 6 is possible to observe the number of cases of each of the new datasets created. The values of the datasets were standardized by dividing each covariate by its absolute maximum to map the input values between 0 and 1.

Table 6 – Number of cases in each dataset used for ENN training and validation

Species / Groups of species	ENN Training cases	ENN Validation cases
<i>M. myotis</i> / <i>M. blythii</i>	119	35
<i>M. emarginatus</i> / <i>M. escalerae</i> / <i>M. daubentonii</i>	248	82
<i>P. pipistrellus</i>	238	75
<i>P. pygmaeus</i>	483	159
<i>P. kuhlii</i>	174	58
<i>N. leisleri</i> / <i>N. lasiopterus</i>	20	8
<i>E. serotinus</i> / <i>E. isabellinus</i>	118	45
<i>B. barbastellus</i>	158	58
<i>Plecotus spp.</i>	91	28
<i>M. schreibersii</i>	515	186
<i>T. teniotis</i>	55	15

For selecting the best architecture for the desired classification task we tested the performance of several ENN designs, changing each time the number of neurons in the hidden layer. We started by 15 neurons, approximately $2/3$ the size of the input vector as suggested by Heaton (2005) and added neurons in steps of 5 until we observed that the error of the validation data increased when compared with the previous (less complex) design. The increasing of the validation data error with a higher number of neurons is a clear indication of overfitting to the training data and loss of generalization capabilities (Heaton, 2005). The selected architecture was the one that presented a lower SSE on the validation data.

4. RESULTS & DISCUSSION

In this chapter we present the results obtained from the descriptive analysis of the measured parameters of bat calls and the classification results obtained by the statistical models.

4.1 CALL DATABASE

The compiled database included 748 recordings (from approximately the same number of individual bats) with 2968 individual calls, of (at least) 20 different species. In 723 recordings the species was identified and in 25 only the genus (*Plecotus*) was identified (report to Section 3.1 for further explanation). The recordings were obtained after hand release of captured bats (374 recordings), near the entrance of known roosts (228 recordings) or from free flying bats (146), as shown in Table 7. As stated before, the work presented in this thesis is part of an ongoing research project and the sampling process is still in progress. Nevertheless as we can see in Table 8, for nine of the species the sample error is already below the defined maximum error (0.8 kHz) of our reference sampling scheme (see Table 3). To know: *R. ferrumequinum*, *R. mehelyi*, *P. pipistrellus*, *P. pygmaeus*, *P. kuhlii*, *E. serotinus*, *B. barbastellus*, *M. schreibersii* and *T. teniotis*.

Table 7 – Number of recordings and respective number of calls, number of recording sites and number of recordings obtained in each recording condition (Hr – hand release; Re – roost entrance; Ff – free flying)

Species	Number of recordings	Number of calls	Number of recording sites	Recording conditions
<i>R. ferrumequinum</i>	24	N/A	3	20 Hr, 4 Re
<i>R. hipposideros</i>	19	N/A	2	19 Re
<i>R. euryale</i>	3	N/A	2	3 Hr
<i>R. mehelyi</i>	32	N/A	3	5 Hr, 27 Re
<i>M. myotis</i>	47	140	4	26 Hr, 21 Re
<i>M. blythii</i>	6	14	1	6 Hr
<i>M. emarginatus</i>	25	78	3	25 Hr
<i>M. escalerae</i>	27	102	4	27 Hr
<i>M. daubentonii</i>	30	150	4	30 Hr
<i>P. pipistrellus</i>	65	313	13	7 Hr, 21 Re, 37 Ff
<i>P. pygmaeus</i>	118	642	8	22 Hr, 96 Re
<i>P. kuhlii</i>	43	232	8	4 Hr, 39 Ff
<i>N. leisleri</i>	7	21	2	7 Hr
<i>N. lasiopterus</i>	4	7	2	4 Hr
<i>E. serotinus</i>	35	141	6	29 Hr, 6 Re
<i>E. isabellinus</i>	7	22	3	7 Hr
<i>B. barbastellus</i>	45	216	8	2 Hr, 43 Ff
<i>Plecotus spp.</i>	25	119	8	9 Hr, 16 Ff
<i>M. schreibersii</i>	150	701	6	116 Hr, 34 Re
<i>T. teniotis</i>	36	70	5	25 Hr, 11 Ff

By observing our mean values and confidence intervals (Table 8) in contrast with some mean reference values commonly used for bat call identification some differences are noted, especially in two parameters: initial frequency and fmaxe. For initial frequency the differences are quite evident and mean values for this parameter even differ greatly amongst published studies, especially for some species (values presented below are means in kHz):

- *P. pipistrellus* – 54.4 (Zingg, 1990), 54.7 (Russ, 1999) and 68.8 (Russo & Jones, 2002) as opposed to our own of 67.2;
- *P. pygmaeus* – 76.2 (Zingg, 1990), 64.7 (Russ, 1999) and 79.6 (Russo & Jones, 2002) as opposed to our own of 77.9;

- *P. kuhlii* – 46.4 (Zingg, 1990) and 72.0 (Russo & Jones, 2002) as opposed to our own of 77.9;
- *E. serotinus* – 43.2 (Zingg, 1990), 57.4 (Russ, 1999) and 50.4 (Russo & Jones, 2002) as opposed to our own of 49.2;
- *N. leisleri* – 24.8 (Zingg, 1990) and 31.4 (Russ, 1999) as opposed to our own of 47.7.

We believe that these differences are probably reflecting variable atmospheric attenuation due to variations in the distance of the emitting bat to the recording device or even to different humidity, wind or temperature conditions (Larom *et al.*, 1997) reinforcing our decision of not using this parameter in our classification models as it seems to have a great deal of variability depending on external factors.

As for *f_{maxe}* the values amongst several published studies (Russ, 1999; Russo & Jones, 2002; Zingg, 1990) are generally quite similar, but we observed some species on our sample that show considerable differences from these previous studies (Table 8) (values presented below are means in kHz):

- *M. myotis* – 39.1 (Russo & Jones, 2002) as opposed to our own of 34.2;
- *M. daubentonii* – 47.8 (Russ, 1999) and 47.0 (Russo & Jones, 2002) as opposed to our own of 44.9;
- *P. pygmaeus* – 55.5 (Russ, 1999) and 57.7 (Russo & Jones, 2002) as opposed to our own of 52.7;
- *E. serotinus* – 32.0 (Russ, 1999) and 29.9 (Russo & Jones, 2002) as opposed to our own of 27.4;
- *M. schreibersii* – 54.4 (Zingg, 1990) and 54.2 (Russo & Jones, 2002) as opposed to our own of 52.7;
- *T. teniotis* – 10.4 (Zingg, 1990) and 13.0 (Russo & Jones, 2002) as opposed to our own of 14.7.

For *M. myotis*, *M. daubentonii* and *E. serotinus* our sample sizes are quite small for definite conclusions but some of these differences could be related to different sampling methods, recording devices or even to geographic variation.

For *T. teniotis* the majority of our recordings was made outside known roosts or at foraging sites that were located in the vicinity of scarps. As this species can be considered a short FM/CF (see Section 1.2) it can change its signal depending on the surrounding environment (Airas, 2003) and our higher *f_{maxe}* could be related with the proximity of the scarps, the foraging sites themselves or with flight altitude (Rydell, 1993).

Table 8 – Descriptive statistics of the five commonly used parameters in manual bat call identification. Parameter mean and the 95% confidence interval is presented (Fini – initial frequency, FMAXE – frequency of maximum energy, Ffin – final frequency, Bw – bandwidth of the call, Dur – duration of the call, N rec – number of recordings, N call – number of individual calls). * indicates values that differ considerably between our study and previously published studies

Species	N rec	N call	Fini	FMAXE	Ffin	Bw	Dur
<i>R. ferrumequinum</i>	24	N/A	n/a	82.0 [81.7, 82.4]	n/a	n/a]	n/a
<i>R. hipposideros</i>	19	N/A	n/a	110.4 [109.4, 111.4]	n/a	n/a	n/a
<i>R. euryale</i>	3	N/A	n/a	103.3 [94.4, 112.2]	n/a	n/a	n/a
<i>R. mehelyi</i>	32	N/A	n/a	104.0 [103.5, 104.5]	n/a	n/a	n/a
<i>M. myotis</i>	47	140	69.3 [64.6, 74.1]	34.2* [33.4, 35.0]	27.0 [26.2, 27.8]	41.6 [37.3, 46.0]	4.0 [3.5, 4.4]
<i>M. blythii</i>	6	14	88.4 [67.0, 109.8]	42.1 [34.6, 49.7]	28.9 [26.9, 30.8]	59.9 [39.3, 80.4]	3.3 [2.5, 4.1]
<i>M. emarginatus</i>	25	78	106.8 [102.0, 111.7]	58.5 [55.9, 61.2]	46.0 [43.9, 48.0]	60.7 [55.4, 66.0]	2.7 [2.4, 3.0]
<i>M. escaleraei</i>	27	102	96.0 [85.0, 106.9]	32.1 [20.1, 36.1]	25.3 [23.5, 27.2]	70.7 [60.2, 81.1]	2.8 [2., 3.7]
<i>M. daubentonii</i>	30	150	82.2 [78.3, 86.0]	44.9* [44.0, 45.8]	32.4 [31.4, 33.5]	49.6 [45.8, 53.4]	3.5 [3.2, 3.7]
<i>P. pipistrellus</i>	65	313	67.2* [64.4, 70.1]	46.3 [45.9, 46.7]	45.1 [44.7, 45.5]	22.0 [19.4, 24.7]	5.4 [5.1, 5.7]
<i>P. pygmaeus</i>	118	642	77.9* [74.8, 80.9]	52.7* [52.3, 53.2]	51.5 [51.1, 52.0]	26.3 [23.5, 29.2]	6.3 [6.1, 6.6]
<i>P. kuhlii</i>	43	232	60.6* [56.1, 65.1]	38.2 [37.7, 38.8]	36.0 [35.4, 36.5]	24.6 [20.4, 28.9]	5.9 [5.3, 6.5]
<i>N. leisleri</i>	7	21	47.7* [40.4, 55.0]	28.3 [25.8, 30.8]	25.9 [24.1, 27.6]	21.5 [14.6, 28.5]	7.0 [3.4, 10.7]
<i>N. lasiopterus</i>	4	7	32.7 [16.8, 48.6]	19.6 [16.1, 23.1]	17.0 [13.7, 20.3]	15.7 [1.8, 29.6]	4.6 [2.9, 6.3]
<i>E. serotinus</i>	35	141	49.2* [46.6, 51.9]	27.4* [26.6, 28.1]	24.9 [24.4, 25.3]	24.4 [21.9, 26.8]	7.5 [6.6, 8.4]
<i>E. isabellinus</i>	7	22	44.8 [33.7, 55.8]	26.0 [24.8, 27.2]	23.7 [22.6, 24.8]	21.1 [10.9, 31.3]	7.4 [5.2, 9.6]
<i>B. barbastellus</i>	45	216	38.4 [37.8, 39.0]	33.1 [32.7, 33.5]	29.1 [28.8, 29.5]	9.2 [8.5, 10.0]	2.0 [1.8, 2.2]
<i>Plecotus spp.</i>	25	119	43.3 [41.6, 44.0]	27.5 [26.2, 28.9]	22.1 [20.7, 23.5]	21.1 [19.4, 22.9]	3.9 [3.0, 4.8]
<i>M. schreibersii</i>	150	701	93.1 [90.8, 95.3]	52.7* [52.4, 53.0]	49.9 [49.7, 50.2]	43.1 [40.9, 45.3]	4.9 [4.7, 5.1]
<i>T. teniotis</i>	36	70	24.2 [22.2, 26.2]	14.7* [14.3, 15.1]	12.3 [11.9, 12.7]	11.9 [15.1, 14.4]	12.2 [11.2, 13.2]

For *P. pygmaeus* the observation of Figure 12 indicates that fmaxe has a high variability, with a peak between 52-53 kHz and also a large number of calls between 55-56 kHz. This latter interval is consistent with the values observed by Russ (1999) and Russo & Jones (2002). In fact this species has individuals with different echolocation characteristics it could be an

indication that in different regions different proportions of these individuals could exist and the usage of a local recorded database could be an advantage.

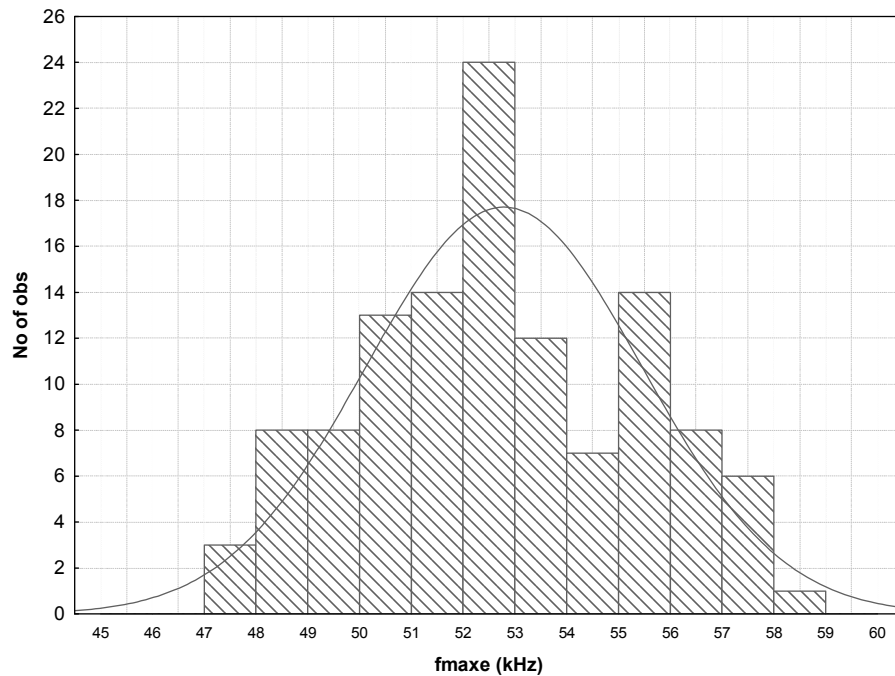


Figure 12 – Histogram of fmaxe for *P. pygmaeus*

For *M. schreibersii*, by observing Figure 13, we can see that fmaxe values of our sample are effectively concentrated around 53 kHz and the differences from previous studies could in fact reflect geographic differences reinforcing the usage of a local recorded database for classification purposes.

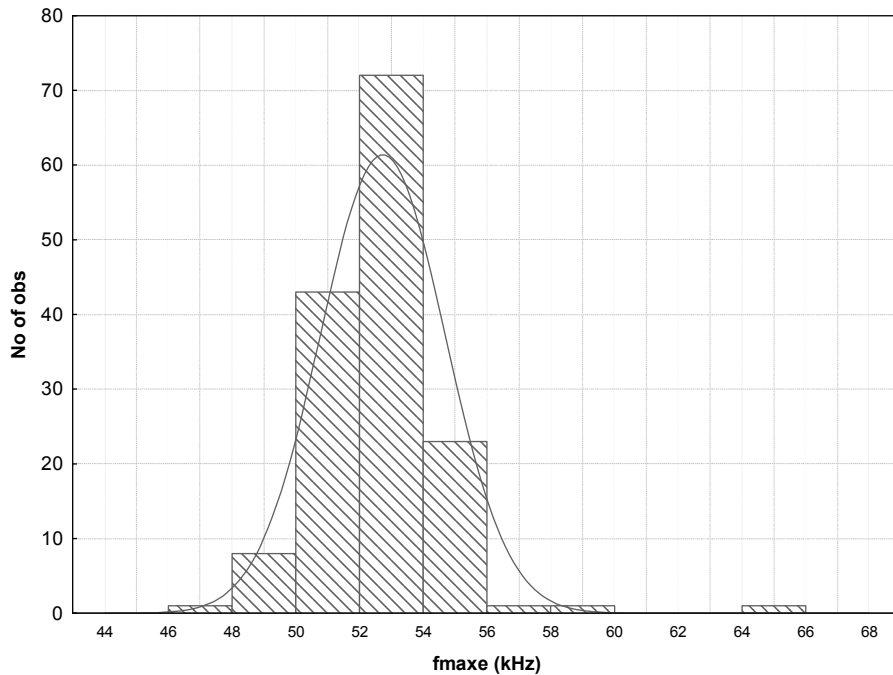


Figure 13 – Histogram of fmaxe for *M. schreibersii*

4.2 CALL CLASSIFICATION

Next are presented the several sets of statistical models used for classification of bat calls and their classification results. For binomial models are presented the sensitivity, specificity and overall correct classification rate. For multinomial models are presented the sensibility, 1-specificity (i.e., the false-positive rate) and overall correct classification rate. For multinomial regression models and ENN models the final classification was considered the category that had the highest predicted probability, whereas for binomial regression models a cutpoint that maximizes the sensibility and the specificity is chosen from the ROC curve.

The classification results presented for the ENN are based on the validation dataset and as this calls were not presented during training to the network they can be interpreted as the generalization capabilities of the ENN and can be seen as a fairly good approximation of the true error of the model as suggested by Rojas (1996).

4.2.1 Classification of major groups of species

In this first classification stage we classified bat calls as belonging to one of six major groups of species: *Rhinolophus*, *Myotis*, *Pipistrelloid*, *Nyctaloid*, *Plecotus* and *Barbastella*.

The classification into the major group *Rhinolophus* was done by a simple fmaxe cutpoint, defined at 75 kHz (see Figure 14). All calls with fmaxe equal or higher than 75 kHz were considered as belonging to the *Rhinolophus* major group. This cutpoint allowed us to correctly classify 100% of these calls.

For call classification of the remaining five major groups we were only able to use ENN. We were able to fit a multinomial logistic regression model, but the absence of appropriate R packages to evaluate the multinomial model assumptions prevented us to verify those assumptions, as they couldn't be verified using a set of independent binomial logistic regression models, as suggested by Hosmer & Lemeshow (2000).

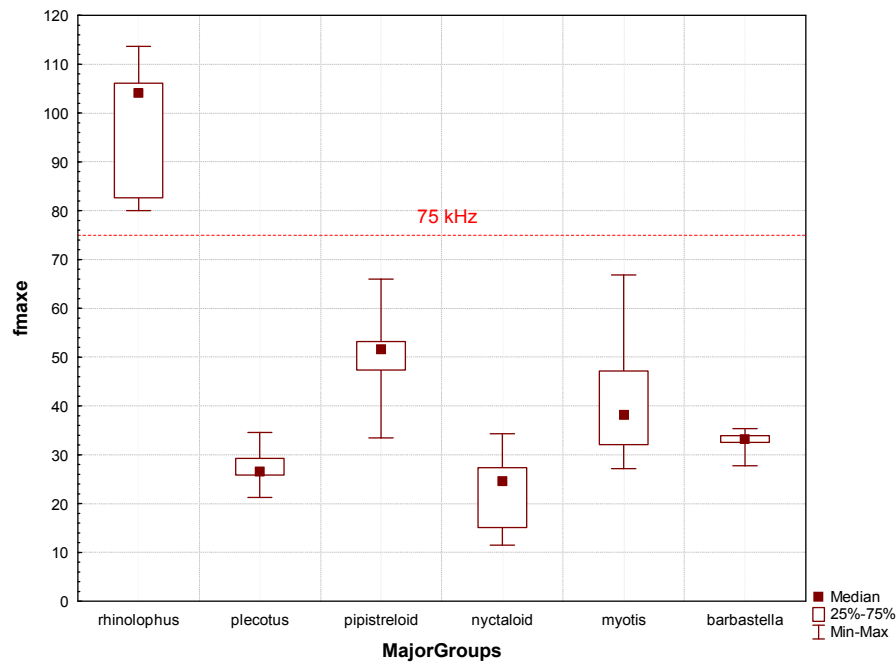


Figure 14 – Boxplot of call fmaxe for recorded bats with the cutpoint indicated. Results presented for groups of species for better visualization

As we can see in Figure 15, ensembles of networks with 15 or 20 neurons have the smallest SSE for the validation data clearly indicating better generalization capabilities. The ENN with 20 neurons was selected because it was the one that showed a smaller SSE for the validation data.

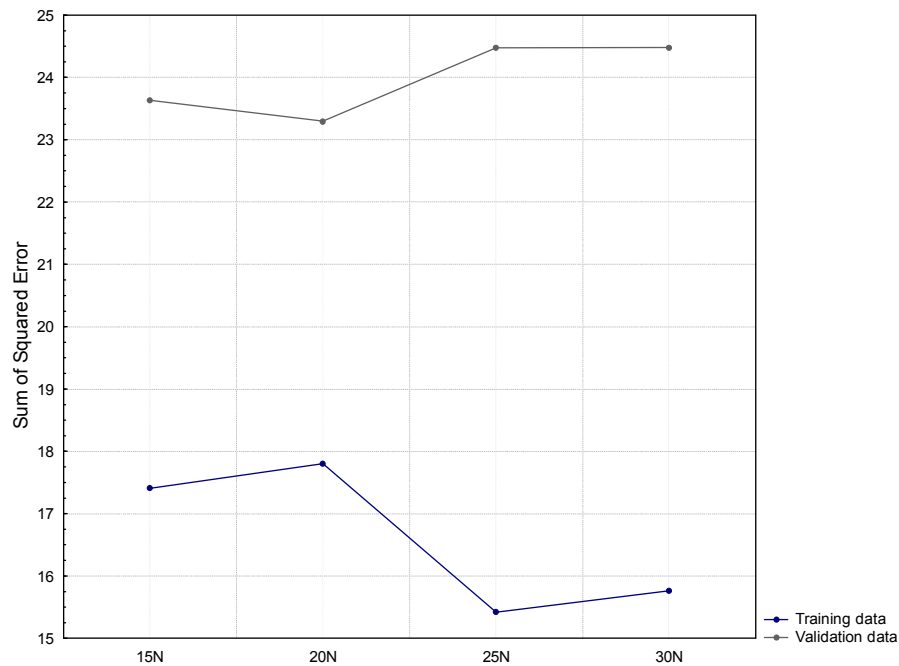


Figure 15 – Sum of squared errors of the training and validation data by the different ENN architectures tested for major groups classification

In Table 9 are presented the results of classification of the 20 neurons ENN on the validation data. The ENN correctly classified 98.3% of *Myotis* calls, 99.0% of *Pipistrelloid* calls, 97.1% of *Nyctaloid* calls, 96.6% of *Barbastella* calls and 92.9% of the *Plecotus* calls, having an overall correct classification rate of 98.3%, as can be observed in Table 9.

By observing the misclassifications rate (1-specificity) of the model we observe that only 0.2% of the calls were incorrectly classified as *Pipistrelloid* calls. For *Myotis* and *Nyctaloid* calls the model is also very precise with misclassification rates of approximately 3%. For *Barbastella* and *Plecotus* we observe that about 7% and 10% of the calls (respectively) were incorrectly classified as belonging to these major groups. By a close analysis of the misclassifications results we observe that for most of the groups the errors were unsystematic. For instance the ENN classified calls belonging either to *Pipistrelloid*, *Nyctaloid* and *Plecotus* as *Myotis* calls. On the other hand, the ENN tends to classify *Pipistrelloid* calls as *Barbastella* and *Barbastella* calls as *Plecotus*. However, as *Barbastella* and *Plecotus* are also the groups with the lower sample sizes it was somehow expected that their accuracy would be the worst because even though the ENN does not work with *a priori* probabilities the groups with larger samples will always contribute more to the error function that regulates the weight update and thus the networks will fit more accurately to categories with larger samples (Rojas, 1996).

From these results we can note that when this ENN classifies a call as *Pipistrelloid*, *Myotis* or *Nyctaloid* we can strongly rely on the result, but if a call is classified as *Plecotus* or *Barbastella* a more close inspection would be advisable.

Table 9 – Classification results of the ENN for major groups classification (Myo – *Myotis*; Pip – *Pipistrelloid*; Nyc – *Nyctaloid*; Bbar – *Barbastella*; Plec – *Plecotus*; Sens – Sensitivity; 1-Spec – 1-Specificity). Overall correct rate highlighted on the bottom right cell

		Observed					Classification results		
		Myo	Pip	Nyc	Barb	Plec	Total pred	Sens (%)	1-Spec (%)
Predicted	Myo	115	1	1	0	1	118	98.3	2.5
	Pip	1	473	0	0	0	474	99.0	0.2
	Nyc	1	0	66	0	1	68	97.1	2.9
	Bbar	0	4	0	56	0	60	96.6	6.7
	Plec	0	0	1	2	26	29	92.9	10.3
	Total obs	117	478	68	58	28	749	98.3	

4.2.2 Classification of *Rhinolophus* species

An exploratory analysis showed that by using variable *fmaxe* it's possible to discriminate between *R. ferrumequinum*, *R. euryale* / *R. mehelyi* and *R. hipposideros* with 100% accuracy, as can be seen in Figure 16.

By using a 92.0 kHz cutpoint, *R. ferrumequinum*, with a maximum sample value of 83.4 kHz, is clearly separated from the rest of *Rhinolophus* species, which have a minimum sample value of 99.2 kHz. As for the remaining species, *R. hipposideros* has a minimum sample value of 106.6 kHz and *R. euryale* / *R. mehelyi* a maximum sample value of 106.1 kHz and a cutpoint of 106.3 kHz can be used to discriminate them. A *Kruskal-Wallis* test comparing *fmaxe* rank values for these three groups, complemented by a multiple comparison test, had p-values <0.01, which asserts the robustness of using *fmaxe* cutpoints for discriminating between the species *R. ferrumequinum*, *R. hipposideros* and the group of species *R. euryale* / *R. mehelyi*.

As *Rhinolophus spp.* have some auditory specializations, which include neural centers devoted to analysis of a narrow range of frequencies for heightened sensitivity in the spectral region of the their signals, *fmaxe* tends to be very constant within each *Rhinolophus* species (Airas, 2003; Moss & Sinha, 2003). This fact suggests that for the discrimination of *R. ferrumequinum* the 92.0 kHz cutpoint can be generalized for *Rhinolophus* populations with great certainty.

The cutpoint of 106.3 kHz on the other hand, worked very well on our sample, but if generalized will probably lead to some classification errors. A more robust sample size, especially of *R. hipposideros* and *R. euryale* is needed before a definite conclusion can be made.

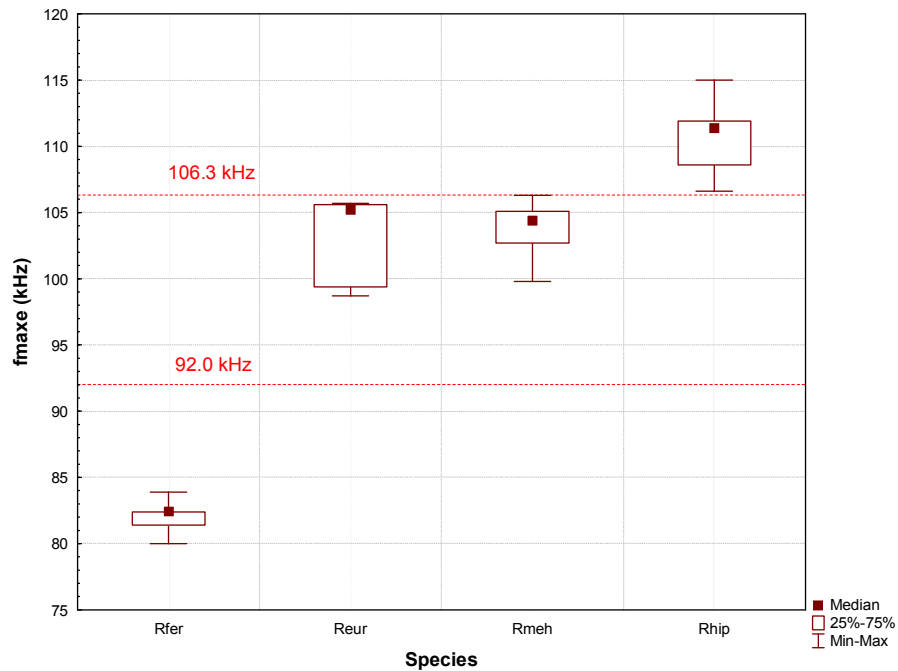


Figure 16 – Boxplot of call fmaxe for *Rhinolophus* species (Rfer – *R. ferrumequinum*, Reur – *R. euryale*, Rmeh – *R. mehelyi*, Rhip – *R. hipposideros*)

4.2.3 Classification of *Myotis* species

For this classification we were able to fit both a binomial logistic model and a binomial ENN. This particular ENN, as it is binomial is suited for a RGW analysis which, besides providing an insight on the ENN classification process, allowed for a comparison with the logistic model results. We will start by presenting the logistic approach, then the ENN approach and afterwards a comparison of the results of both methods. Recall that by the reasons already appointed in Section 3.5, the sample sizes and initial set of covariates is different in both approaches and some conclusion can differ due to this.

For the logistic regression model the response variable was coded as: 0 – *M. myotis* / *M. blythii* and 1 – *M. escalerai* / *M. daubentonii* / *M. emarginatus*. The initial subset of covariates (refer to Section 3.5.1 for further details) was composed by duration, posFMAXE, ffin15, bw10, bwfini15 and bwffin15. All these variables are considered continuous. After selecting the main effects model as described in Section 2.1.2, the significant variables were identified and are presented in Table 10.

Table 10 – Main effects model for *Myotis* species

Coefficients	Estimate	S.E.	p-value	OR	OR IC _{95%}
Intercept	-7.593	1.685	0.000		
posFMAXE	-0.511	0.224	0.023	0.60	(0.39; 0.93)
ffin15	0.254	0.055	0.000	1.29	(1.6; 1.44)
bwfini15	0.083	0.022	0.000	1.09	(1.04; 1.13)

The linearity assumption of each significant covariate was evaluated by the quartile method, the lowess representation and the fractional polynomials method. All covariates were considered linear despite some doubts regarding ffin15. As can be seen in Figure 17 the lowess representation for this covariate shows some non-linearity for values below 25 kHz and the fractional polynomials method suggests a transformation: $\text{ffin15T1} = 10/\text{ffin15}$. Nevertheless when we compared the model using the transformed covariate with the original model by the likelihood ratio test the p-value was 1, indicating that the model with the transformed covariate did not differ from the original model and thus we decided not to transform this covariate.

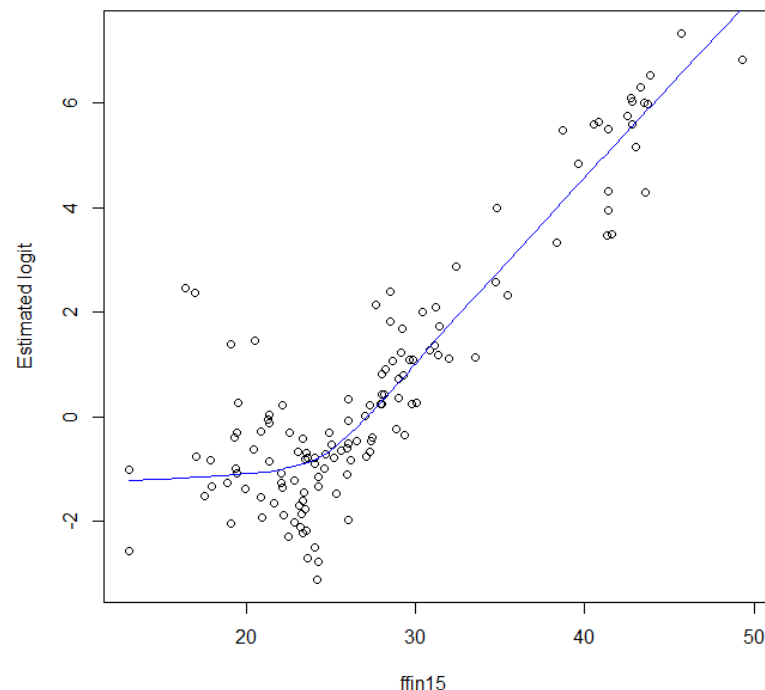


Figure 17 – Lowess representation of covariate ffin15

We then tested for interactions, but none were detected at a 1% significance level, and the analysis of the residuals didn't show any outlier or influential observation.

The final model is the one presented in Table 10. This model presented a good fit to the data, with a p-value for the Hosmer-Lemeshow statistic of 0.46, and a Nagelkerke's R^2 value of 0.42. The model also presented an excellent discriminative capability as demonstrated by an AUC of 0.88 and, for a cutpoint of 0.49 it correctly classified 85.9% of the *M. myotis* / *M. blythii* calls and 77.5% of the *M. escalerae* / *M. daubentonii* / *M. emarginatus* calls, as can be seen in Figure 18, with an overall correct classification rate of 80.7%.

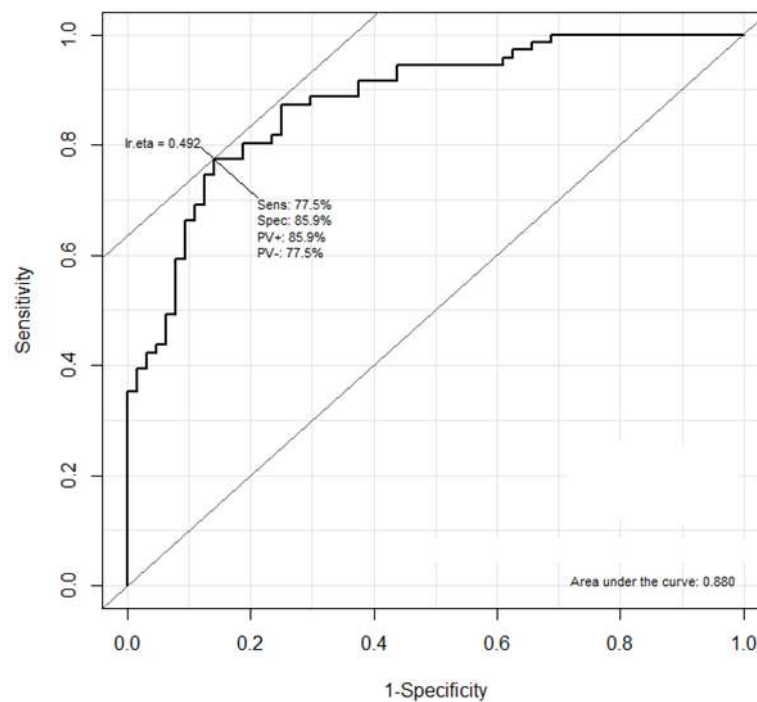


Figure 18 – ROC curve of the final model for *Myotis* species

To evaluate how a given parameter is related to call classification, in Figure 19 we plotted the probability of a given call belonging to *M. myotis* / *M. blythii* or *M. escalerae* / *M. daubentonii* / *M. emarginatus*, by varying one of the covariates while maintaining the remaining constant. For the fixed values of the covariates we used $\text{posFMAXE} = 3.0$ ms, $\text{ffin15} = 30$ kHz and $\text{bwfini15} = 26$ kHz, as these values can be found in the calls of both groups. We can observe that an increase of the values of ffin15 and bwfini15 will increase the probability of a call being classified in the *M. escalerae* / *M. daubentonii* / *M. emarginatus* group.

This results can also be presented by calculating the odds ratio (see Table 10) where we can analyze the odds change by varying one covariate while maintaining all the others constant: by each unit increase (1 kHz) of covariate ffin15 or bwfini15 , the odds of a call being classified as *M. escalerae* / *M. daubentonii* / *M. emarginatus* increase 1.29 and 1.09 times, respectively; and by decreasing posFMAXE by one unit (1 ms) the odds increase 1.67 times.

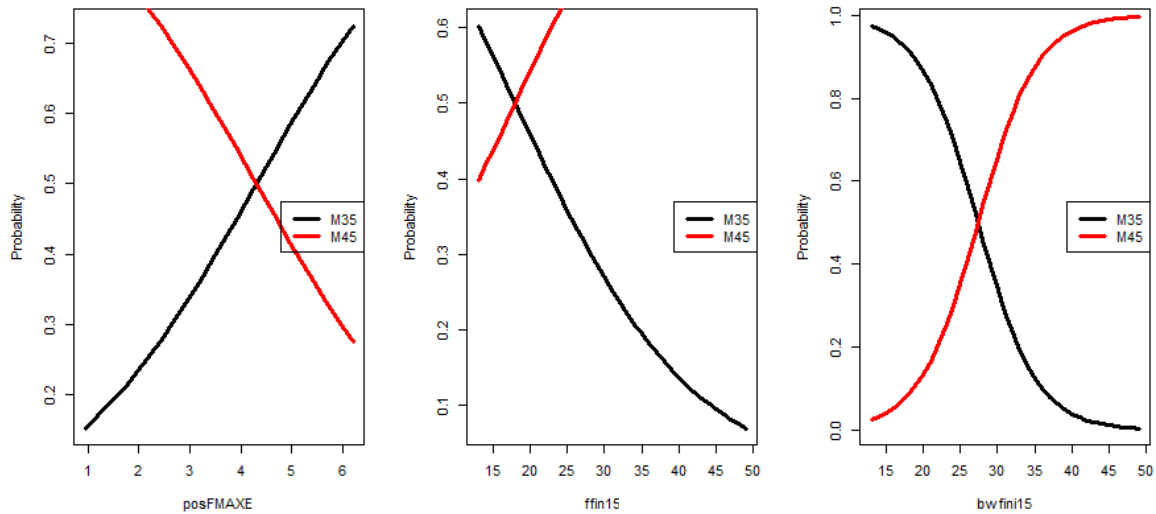


Figure 19 – Probabilities change of a call belonging to M35 or M45 (M35 – *M. myotis* / *M. blythii*; M45 – *M. escaleraei* / *M. daubentonii* / *M. emarginatus*). Left – ffin15 = 30 kHz and bwfini15 = 26 kHz; center – posFMAXE = 3 ms and bwfini15 = 26 kHz; right – posFMAXE = 3 ms and ffin15 = 30 kHz

For the ENN approach, we can see in Figure 20 that ENN with 20 or 25 neurons have the smallest SSE for the validation data. These results show that ENN with 20 to 25 neurons are those that present better results for the classification of *Myotis* species and the 25 neurons architecture was chosen because it has a smaller SSE for the validation data.



Figure 20 – Sum of squared errors of the training and validation data by the different ENN architectures tested for *Myotis* species classification

The RGW analysis revealed that 4 of the 19 covariates used have a very strong effect on the log-odds of this model but we can see that every covariate has some influence (Table 11). It

can be also observed that the contribution of the covariates to the log-odds of one category is almost symmetrical to the other. This was expected because the output of the network is an accurate approximation to the Bayesian *a posteriori* probability of the class membership given the input (Saerens, Latinne, & Decaestecker, 2002)

Table 11 – Robust generalized weights for each covariate and output category of the ENN; means are presented in each cell (M35 – *M. myotis* / *M. blythii*; M45 – *M. escalerae* / *M. daubentonii* / *M. emarginatus*). Most influential covariates on the log-odds are highlighted in bold

	M35	M45
fmaxe	-3.505	3.999
duration	19.375	-19.967
fini15	-6.023	5.9427
fini10	-1.459	3.362
fini5	0.550	0.308
ffin15	6.439	-6.286
ffin10	2.572	-2.316
ffin5	-5.933	6.279
bw15	-9.307	9.240
bw10	-4.945	5.566
bw5	0.549	0.146
posFMAXE	-5.598	5.558
ptFMAXE	-13.303	12.680
bwfini15	5.922	-6.756
bwfini10	3.631	-4.681
bwfini5	-5.360	5.242
bwffin5	8.704	-8.090
bwffin10	-20.596	22.189
bwffin15	-2.474	3.617

An analysis of the lowest representation of the distribution of the RGW along the input space of the most significant covariates identified on the analysis of Table 11 for the classification of *M. escalerae* / *M. daubentonii* / *M. emarginatus* is presented in Figure 21. We can observe that these covariates have a non-linear behavior (of different magnitudes though) on the log-odds and the non-linearity is more evident for covariates duration and bwffin10 as revealed by the larger variation of the RGW along the range of these covariates.

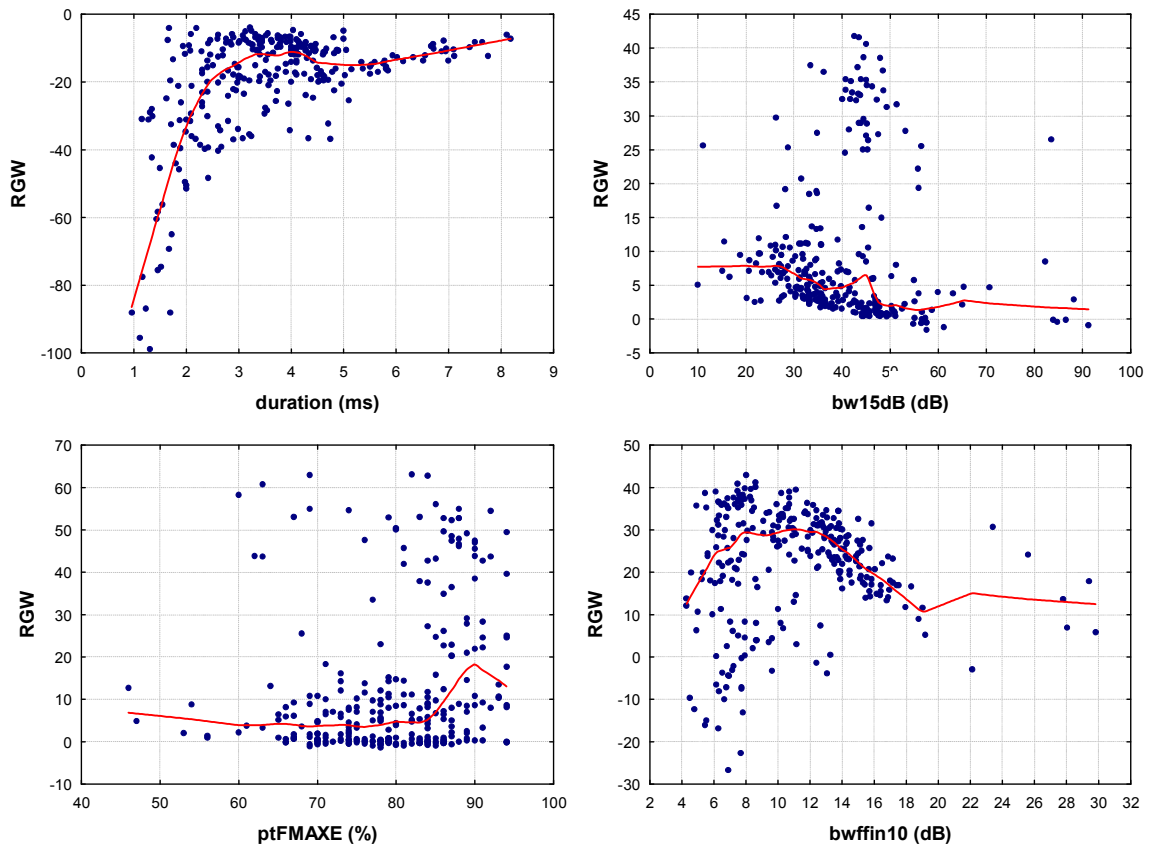


Figure 21 – Lowess representation of the distribution of the RGW along the input space of the most significant covariates for the classification of *M. escaleraei* / *M. daubentonii* / *M. emarginatus* calls

The discriminative capability of this ENN is excellent as demonstrated by a sensibility of 88.9% and a specificity of 96.3%, as can be seen in Table 12, that is, the model correctly classified 88.9% of *M. myotis* / *M. blythii* calls and 96.3% of *M. escaleraei* / *M. daubentonii* / *M. emarginatus* calls. The overall correct classification rate was 94.0%.

Table 12 – Classification results of the ENN for *Myotis* species classification (M35 – *M. myotis* / *M. blythii*; M45 – *M. escaleraei* / *M. daubentonii* / *M. emarginatus*). Overall correct rate highlighted on the bottom right cell

		Observed			Classification results	
		M35	M45	Total pred	Sens (%)	Spec (%)
Predicted	M35	32	4	36	88.9	96.3
	M45	3	78	81		
	Total obs	35	82	117	94.0	

A comparative analysis of both the classification results and significant covariates identified in the logistic regression and ENN models revealed some differences. The most important covariates identified by LR and ENN are different but this was expected because for LR most of the covariates were left out from the model selection stage. Nevertheless both approaches identified significant covariates for discriminating *Myotis* calls that make sense from a biological point of view. In LR we can see that *M. escalerae* / *M. daubentonii* / *M. emarginatus* calls tend to have higher *ffin15* and *bwfini15* values and lower *posFMAXE* which agree with the weight signs for these covariates (Figure 22). For ENN we can see that *M. escalerae* / *M. daubentonii* / *M. emarginatus* calls have lower duration and higher *bw15* and *bwffin10* values which agree with the signs of the RGW for these covariates. For *ptFMAXE* though, *M. escalerae* / *M. daubentonii* / *M. emarginatus* tends to have slightly lower values than *M. myotis* / *M. blythii*, which contradicts the sign of the respective RGW (Figure 22), nevertheless the RGW for this covariate are non linear (Figure 21) and adding the fact that the ENN uses highly correlated covariates could justify this apparent inconsistency.

In terms of accuracy the overall results obtained by ENN were better, probably because the ENN could take into account all the covariates, even the collinear ones and was able to find complex nonlinear interactions between them. In contrast, the logistic regression models provided simpler and more comprehensible insights about the covariates and thus about the differences in echolocation calls.

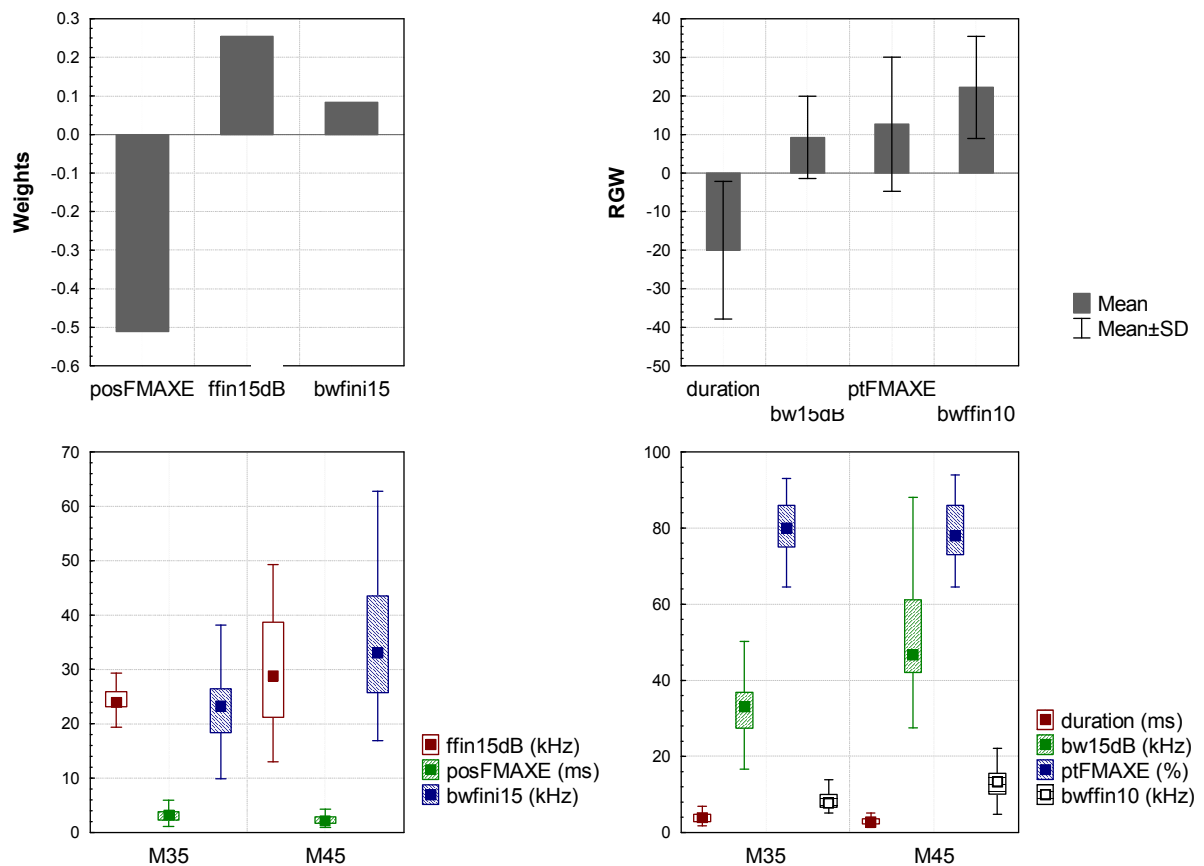


Figure 22 – LR weights and ENN RGW for the reference group M45 (M45 – *M. escalerae* / *M. daubentonii* / *M. emarginatus*), with respective boxplot of the significant covariates identified. Left – LR weights; right – ENN RGW

4.2.4 Classification of *Pipistrelloid* species

For this classification we were able to fit both a multinomial logistic model and a multinomial ENN. An exploratory analysis enabled to find a covariate (fmaxe) that can discriminate *P. kuhlii* from the remaining species of this major group. By observing Figure 23, we can realize that values of fmaxe below 43.0 kHz will indicate, with 100% accuracy (at least on our sample), a *P. kuhlii* call and consequently fmaxe values over 43.0 kHz will indicate a *M. schreibersii*, *P. pipistrellus* or *P. pygmaeus* call. However this cutpoint cannot be trusted for generalization as the majority of *P. kuhlii* recordings were obtained from free flying bats and identified to species level by using pre-defined echolocation parameters. A larger sample size for this species, recorded under different circumstances, has to be obtained before definite conclusions can be made.

For the logistic model we used this first discrimination and only classified three species. For the ENN we did not use this previous discrimination and opted for a model to classify the four species. We will start by presenting the logistic approach, then the ENN approach.

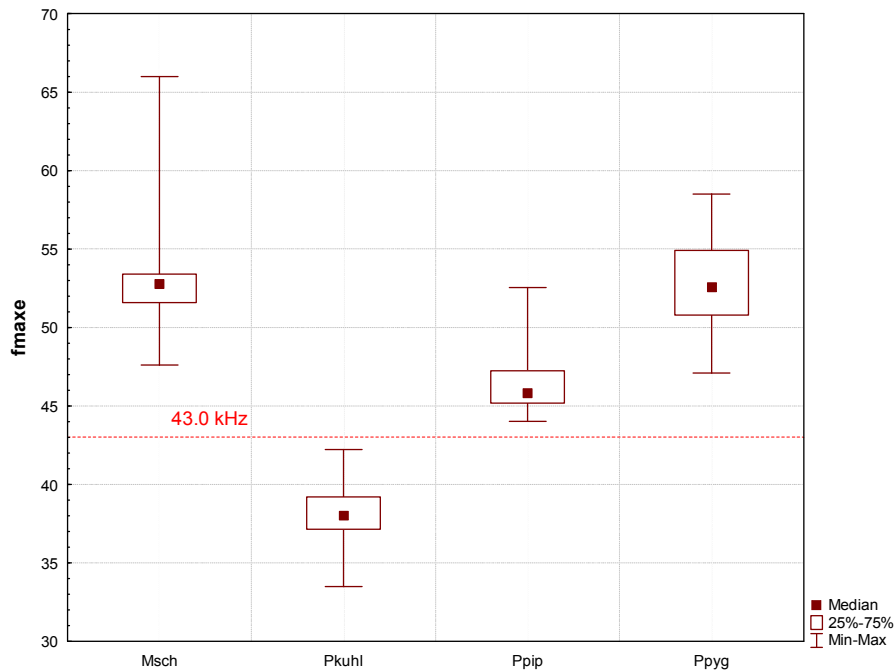


Figure 23 – Boxplot of call fmaxe for the Pipistrellid species with the cutpoint indicated (Msch – *M. schreibersii*; Pkuhl – *P. kuhlii*; Ppip – *P. pipistrellus*; Ppyg – *P. pygmaeus*)

For the multinomial logistic model the response variable was coded as: 0 – *M. schreibersii*; 1 – *P. pipistrellus*; 2 – *P. pygmaeus*. The initial subset of covariates (refer to Section 3.5.1. for further details) was composed by duration, posFMAXE, ffin15, bw10, bwfin15 and bwffin15. All these variables are considered continuous. After selecting the main effects model, using the same approach used for the binomial logistic models as described in Section 2.1.2, the significant variables obtained are presented in Table 13.

Table 13 – Main effects model for *Pipistrelloid* species

Coefficients	Estimate	S.E.	p-value	OR	OR IC _{95%}
1:(intercept)	111.269	26.772	0.000	-	-
2:(intercept)	-22.487	7.149	0.002	-	-
1:duration	-1.597	0.668	0.017	0.20	(0.05; 0.75)
2:duration	0.534	0.245	0.029	1.71	(1.06; 3.75)
1:posFMAXE	-3.168	1.063	0.003	0.04	(0.01; 0.34)
2:posFMAXE	-0.916	0.317	0.004	0.40	(0.21; 0.74)
1:ffin15	-1.879	0.461	0.000	0.15	(0.06; 0.37)
2:ffin15	0.528	0.131	0.000	1.70	(1.31; 2.19)
1:bwffin15	-3.610	1.008	0.000	0.03	(0.01; 0.19)
2:bwffin15	-3.127	0.555	0.000	0.04	(0.01; 0.13)

To verify the model assumptions, we followed Hosmer & Lemeshow (2000) suggestion to build two separate binomial models, the first to compare *M. schreibersii* and *P. pipistrellus*, and the second to compare *M. schreibersii* and *P. pygmaeus*. The linearity assumption, interactions, influential observation and outlier evaluation were done in the two separate logistic regression models. For the first model the response variable was coded as: 0 – *M. schreibersii*; 1 – *P. pipistrellus*. This model is presented in Table 14.

Table 14 – First binomial logistic model to compare *M. schreibersii* and *P. pipistrellus*

Coefficients	Estimate	S.E.	p-value
Intercept	136.812	44.486	0.002
duration	-2.218	0.944	0.019
posFMAXE	-4.930	2.101	0.019
ffin15	-2.200	0.703	0.002
bwffin15	-4.488	1.497	0.002

For the second model the response variable was coded as: 0 – *M. schreibersii*; 1 – *P. pygmaeus*. This model is presented in Table 15.

Table 15 – Second binomial logistic model to compare *M. schreibersii* and *P. pygmaeus*

Coefficients	Estimate	S.E.	p-value
Intercept	-21.810	6.987	0.002
duration	0.507	0.236	0.032
posFMAXE	-0.847	0.309	0.006
ffin15	0.509	0.129	0.000
bwffin15	-2.961	0.532	0.000

The relations between the coefficients of these two binomial logistic models and the multinomial model can be verified by looking at Table 13 and comparing the coefficients with both the coefficients in Tables 14 and 15.

For the first binomial model all covariates verified the linearity assumption and no significant interactions were found. As for outliers and influential observations, case 170 could be a possible outlier (Figure 24). However we decided not to remove it because a close inspection of this observation didn't indicate any data measurement errors and the analysis of Cook's distance and Dfbeta didn't revealed it to be influential.

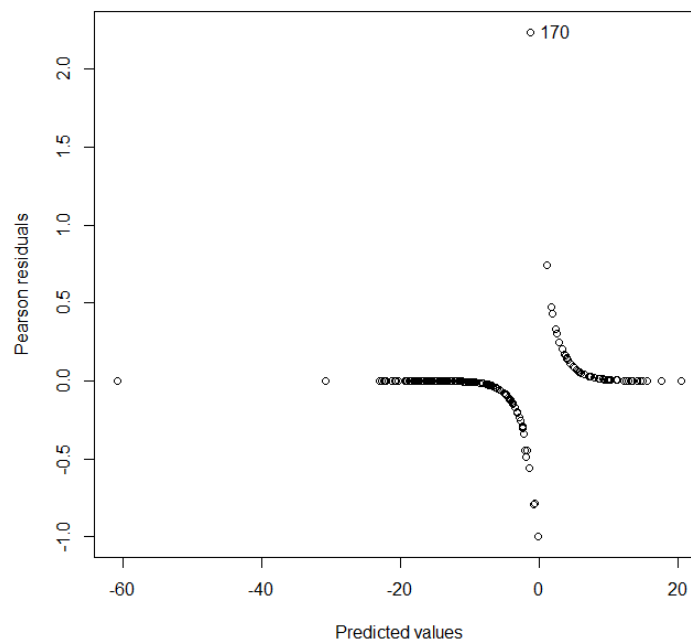


Figure 24 – Pearson residual analysis of the first model

This model as a good fit to the data with a Nagelkerke's R^2 of 0.98 and a Hosmer and Lemeshow test p-value of 0.94. With this analysis we can conclude that covariates duration, posFMAXE, ffin15 and bwfini15 are relevant to discriminate between *M. schreibersii* and *P.*

pipistrellus calls (Table 14). This model has a sensibility of 100%, a specificity of 96.7% and an AUC of 100%.

For the second binomial model, the linearity assumption was also verified for all covariates, no significant interactions were detected but three observations were identified as possible outliers. However we decided not to remove them because a close inspection of these observations didn't indicate any data measurement errors and the analysis of Cook's distance and Dfbeta didn't revealed them to be influential. This model presents a good fit to the data, with an R^2 of 0.80 and a Hosmer and Lemeshow test p-value of 0.09. With this analysis we can conclude that covariates duration, posFMAXE, ffin15 and bwfini15 are relevant to discriminate between *M. schreibersii* and *P. pygmaeus* calls. This model has a sensibility of 90.7%, a specificity of 90.7% and an AUC of 96%.

Having verified the model assumptions for the two independent binary logistic regression model, we can conclude that the multinomial logistic regression model also presents a good fit to the data, as indicated by Hosmer & Lemeshow (2000). The final model is the one presented in Table 13.

The classification results of the multinomial model can be seen in Table 16. The model correctly classified 96.9% of *P. pipistrellus* calls, 89.0% of *P. pygmaeus* calls and 89.3% of *M. schreibersii* calls with an overall correct classification rate of 90.7%.

By observing the misclassifications rate (1-specificity) of the model we observe that around 6% of *P. pipistrellus* calls and approximately 14% of *P. pygmaeus* and 7% of *M. schreibersii* were misclassifications. By a close analysis of the misclassifications results we observe that the errors seem unsystematic although most misclassifications were *P. pygmaeus* calls classified as *M. schreibersii* and vice-versa. Since we have used a fmaxe cutpoint to classify *P. kuhlii*, we observe that using that approach we were able to correctly classify all the calls in our sample.

Table 16 – Classification results of the multinomial logistic regression model for *Pipistrelloid* species (Msch – *M. schreibersii*; Ppyg – *P. pygmaeus*; Ppip – *P. pipistrellus*)

		Observed				Classification results	
		Ppip	Ppyg	Msch	Total pred	Sens (%)	1-Spec (%)
Predicted	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	

To understand how the parameters influence call classification, in Figure 25 we plotted the probability of a given dataset belonging to each of the species, by varying one of the variables while maintaining the others constant. To exemplify, we used the fixed values of the variables: duration = 5.5 ms, posFMAXE = 3.0 kHz, ffin15 = 50 kHz and bwffin15 = 1.2 kHz. For instance, we can observe that higher call durations along with higher values of ffin15, will increase the probability that a call belongs to *P. pygmaeus* and higher values of posFMAXE along with higher values of bwffin15 will increase the probability that a call belongs to *M. schreibersii*.

These results can also be presented by calculating the odds ratio (see Table 13) where we can analyze the odds change by varying one covariate while maintaining all the others constant. For instance, by each unit increase (1 ms) of covariate duration, the odds of a call being classified as *P. pygmaeus* increase 1.7 times.

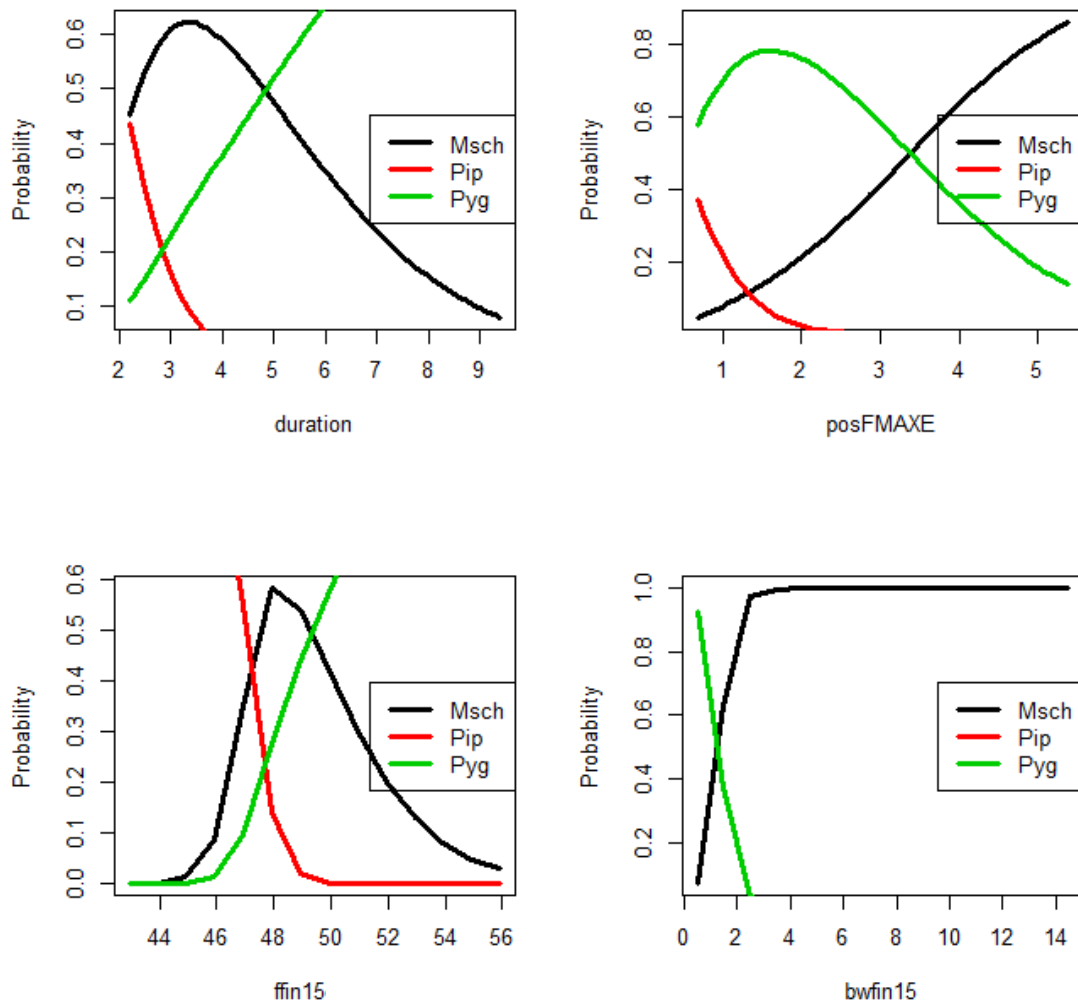


Figure 25 – Probabilities of a call belonging to Msch – *M. schreibersii*; Ppyg – *P. pygmaeus* or Ppip – *P. pipistrellus*. Top left: posFMAXE = 3.0 kHz, ffin15 = 50 kHz and bwffin15 = 1.2 kHz; top right: duration = 5.5 ms, ffin15 = 50 kHz and bwffin15 = 1.2 kHz; bottom left: duration = 5.5 ms, posFMAXE = 3.0 kHz and bwffin15 = 1.2 kHz; bottom right duration = 5.5 ms, posFMAXE = 3.0 kHz and ffin15 = 50 kHz

For the ENN approach, we can see in Figure 26 that ENN with 15 neurons have the smallest SSE for the validation data clearly indicating better generalization capabilities.

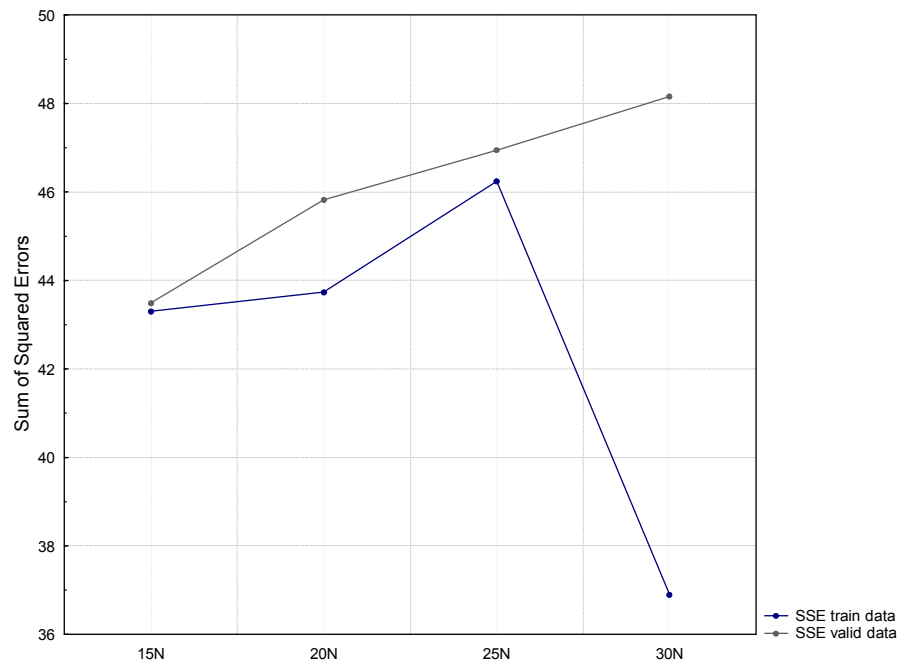


Figure 26 – Sum of squared errors of the training and validation data by the different ENN architectures tested for *Pipistrelloid* species classification

In Table 17 are presented the classification results of the 15 neurons ENN on the validation data. The ENN correctly classified 93.3% of *P. pipistrellus* calls, 91.8% of *P. pygmaeus* calls, 100.0% of *P. kuhlii* calls and 96.2% of *M. schreibersii* calls, having an overall correct classification rate of 94.8%.

By observing the misclassifications rate (1-specificity) of the model we observe that it is very accurate for *P. kuhlii*, presenting a null error for this particular species. For the remaining species we observe that around 4% of *P. pipistrellus* calls and approximately 6% of *P. pygmaeus* and *M. schreibersii* calls were misclassifications. By a close analysis of the misclassifications results we observe that the errors seem unsystematic although most misclassifications were *P. pygmaeus* calls classified as *M. schreibersii* and vice-versa.

Table 17 – Classification results of the ENN for *Pipistrelloid* species classification (Msch – *M. schreibersii*; Pkuh – *P. kuhlii*; Ppip – *P. pipistrellus*; Ppyg – *P. pygmaeus*); Sens – Sensitivity; 1-Spec – 1-Specificity). Overall correct rate highlighted on the bottom right cell

		Observed				Classification results		
		Ppip	Ppyg	Pkuhl	Msch	Total pred	Sens (%)	1-Spec (%)
Predicted	Ppip	70	3	0	0	73	93.3	4.1
	Ppyg	3	146	0	7	156	91.8	6.4
	Pkuh	0	0	58	0	58	100.0	0.0
	Msch	2	10	0	179	191	96.2	6.3
	Total obs	75	159	58	186	478	94.8	

Once again we could conclude that ENN gives better classification results than logistic regression models, however logistic models gives insights about the importance of certain parameters that could be very helpful in future classification studies.

4.2.5 Classification of *Nyctaloid* species

For this classification we were able to fit a binomial logistic model and a multinomial ENN. An exploratory analysis enabled to find a covariate (ffin10) that can discriminate *T. teniotis* calls from both *E. serotinus* / *E. isabellinus* and *N. leisleri* / *N. lasiopterus* calls and thus allowing to fit a binomial logistic model. This was needed because our sample size was too small for a multinomial model (Agresti, 2007). For the ENN we didn't use this previous discrimination and fitted a multinomial model.

In the logistic approach, as observed in Figure 27, we realize that values of ffin10 below 15.2 kHz will indicate, with 100% accuracy (at least on our sample), a *T. teniotis* call and consequently ffin10 values over 15.2 kHz will indicate an *E. serotinus* / *E. isabellinus* or *N. leisleri* / *N. lasiopterus* call. Nevertheless the small sample sizes prevent definite conclusions about this cutpoint.

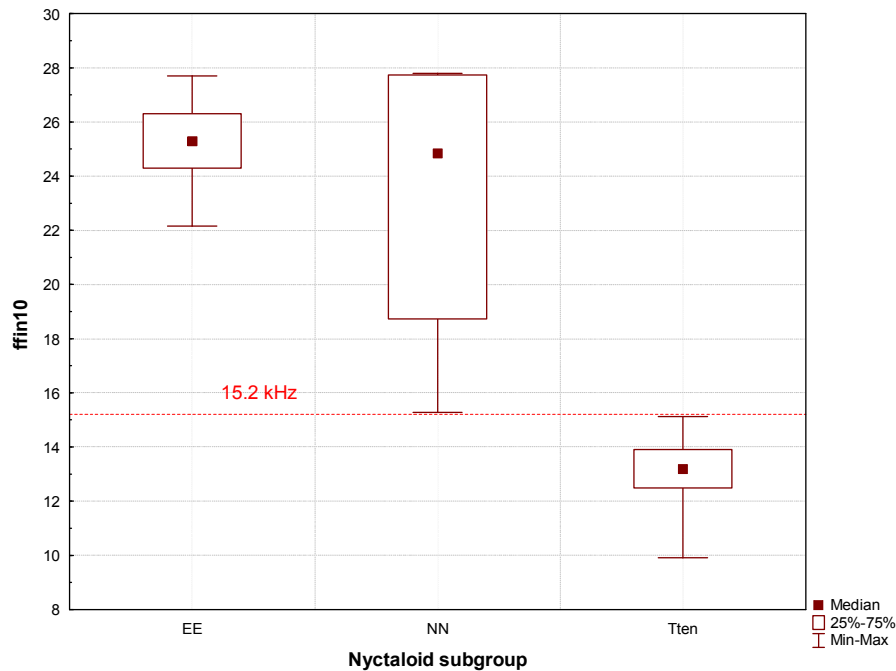


Figure 27 – Boxplot of covariate ffin10 for Nyctaloid species with the cutpoint value indicated (EE – *E. serotinus* / *E. isabellinus*; NN – *N. leisleri* / *N. lasiopterus*; Tten – *T. teniotis*)

For the binomial logistic model the response variable was coded as: 0 – *E. serotinus* / *E. isabellinus* and 1 – *N. leisleri* / *N. lasiopterus*. The initial subset of covariates (refer to Section 3.5.1. for further details) was composed by duration, posFMAXE, fini15, ffin15, bw10, bwfini15 and bwffin15. All these variables are considered continuous. A preliminary main effects model that contained four significant covariates was fitted but, as Agresti (2007) suggested regarding the number of covariates, for the sample size we have available they should not be more than two. Based on the Spearman rank correlation tests applied (refer to Section 3.5.1. for further details) we decided to discard two of them and a new, final main effects model was fitted with only two covariates, this model is presented in Table 18.

Table 18 – Main effects model for *Nyctaloid* species

Coefficients	Estimate	S.E.	p-value	OR (variation of -1 unit)	OR IC95%
Intercept	7.851	4.018	0.069		
ffin15	-0.280	0.164	0.111	1.28	(0.94; 1.75)
bw10	-0.234	0.119	0.073	1.22	(0.98; 1.53)

All covariates verified the linearity assumption. As an example, the lowess representation for bw10 is presented in Figure 28 where the linearity of this covariate can be easily verified.

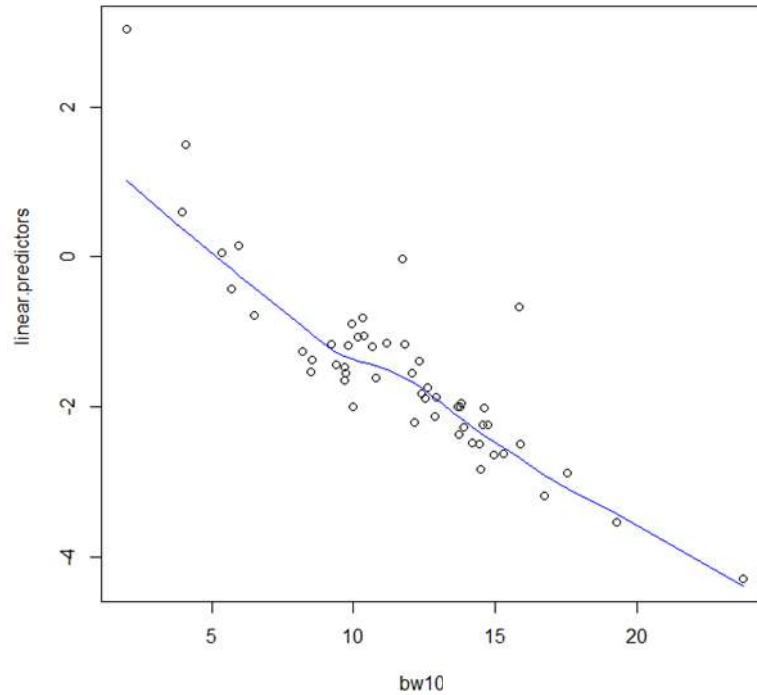


Figure 28 – Lowess representation of covariate bw10

Also, no significant interaction was found and no outliers or influential observations were detected. The final model is the one presented above, in Table 18.

This model presented a good fit to the data, with a p-value of the Hosmer-Lemeshow test of 0.54 and a Nagelkerke's R^2 value of 0.13. The model has an acceptable discriminative capacity with an AUC of 0.79. For a cutpoint of 0.32 it correctly classified 97.6% of *E. serotinus* / *E. isabellinus* calls and 63.6% of *N. leisleri* / *N. lasiopterus* calls (Figure 29). The overall correct classification rate was 90.6%.

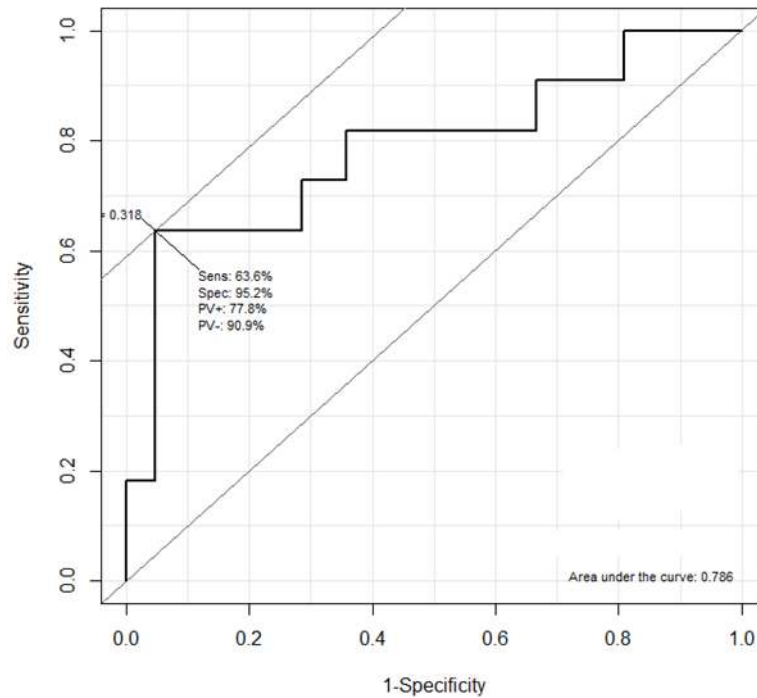


Figure 29 – ROC curve of the final model for *Nyctaloid* species

In Figure 30 we plotted the probability of a given data belonging to each of the groups, by varying one of the variables while maintaining the other constant. For the fixed values of the variables we used $ffin15 = 20$ kHz and $bw10 = 10$ kHz, as these values can be found in *E. serotinus* / *E. isabellinus* and *N. leisleri* / *N. lasiopterus* calls. We can observe that smaller values of $ffin5$ and $bw10$ increase the probability of a call being classified as *N. leisleri* / *N. lasiopterus*.

This results can also be presented by calculating the odds ratio (see Table 18) where we can analyze the odds change by varying one covariate while maintaining all the others constant: by each unit decrease (1 kHz) of covariate $ffin15$ or $bw10$, the odds of a call being classified as *N. leisleri* / *N. lasiopterus* increase 1.29 and 1.22 times, respectively.

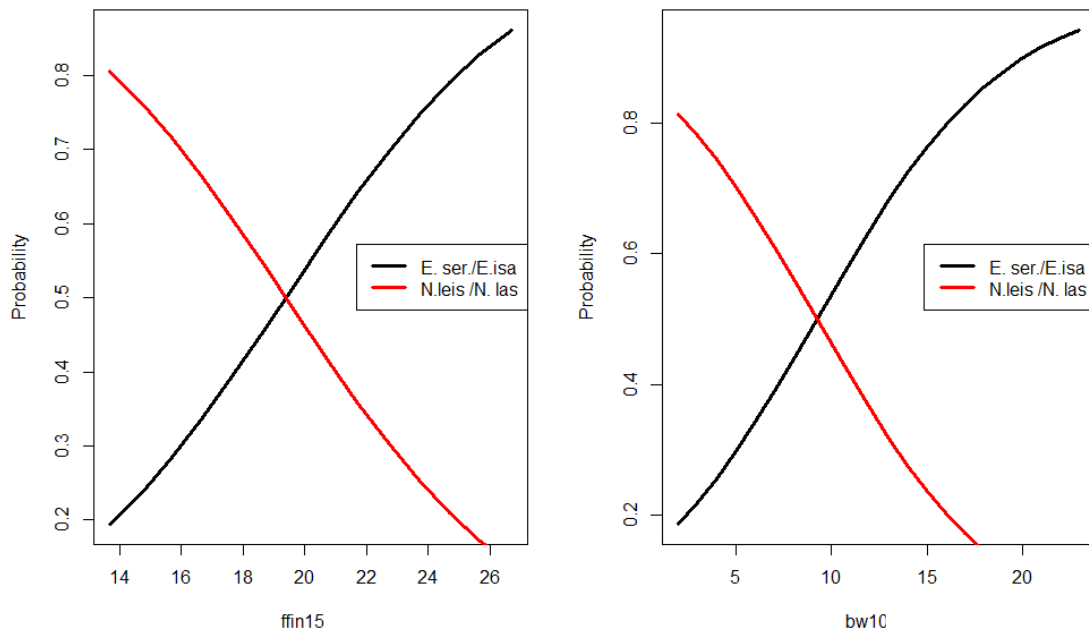


Figure 30 – Probabilities of a call belonging to *E. serotinus* / *E. isabellinus* or *N. leisleri* / *N. lasiopterus*. On the left bw10 is constant at 10 kHz and on the right ffin15 is fixed at 20 kHz

Next we present the multinomial ENN model for classifying *T. teniotis*, *E. serotinus* / *E. isabellinus* and *N. leisleri* / *N. lasiopterus*. As we can see in Figure 31, ensembles of networks with 20 neurons have the smallest SSE for the validation data. Nevertheless, and contrary to the ENN sets presented before the validation SSE remains almost constant regardless the number of neurons used, indicating that the generalization capabilities of these ENN are quite poor.

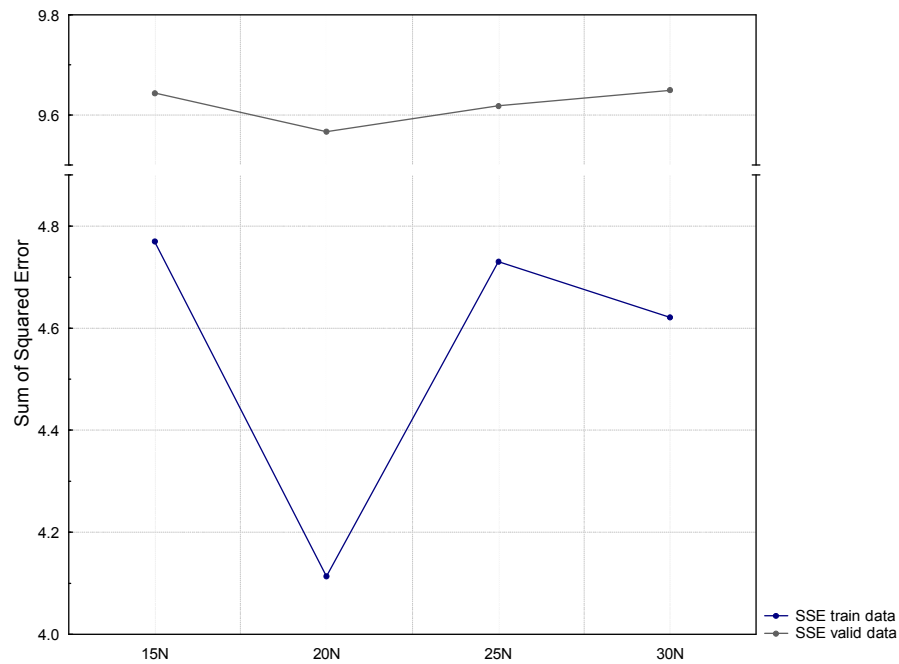


Figure 31 – Sum of squared errors of the training and validation data by the different ENN architectures tested for *Nyctaloid* species classification

In Table 19 are presented the results of classification of the 20 neurons ENN on the validation data. The overall classification rate of this ENN is 91.2% but the results vary a lot between the three categories. The ENN classified correctly 100% of *T. teniotis* calls, without any misclassification; for *E. serotinus* / *E. isabellinus* it classified correctly 95.6% but has almost 9% of wrong classifications and for *N. leisleri* / *N. lasiopterus* the results are very inaccurate, as it only classified correctly 50% of the calls with 33% of misclassifications.

By a close analysis of the misclassifications results we observe that the ENN tends to confound *E. serotinus* / *E. isabellinus* with *N. leisleri* / *N. lasiopterus* (and vice-versa) but never with *T. teniotis*, implying that when the ENN classifies a call as *T. teniotis* we can be 100% sure of the result, but if a call is classified as *E. serotinus* / *E. isabellinus* or *N. leisleri* / *N. lasiopterus* a more close inspection is required.

Table 19 – Classification results of the ENN for *Nyctaloid* species (Tten – *T. teniotis*; NN – *N. leisleri* / *N. lasiopterus*; EE – *E. serotinus* / *E. isabellinus*)

		Observed				Classification results	
		EE	NN	Tten	Total pred	Sens (%)	1-Spec (%)
Predicted	EE	43	4	0	47	95.6	8.5
	NN	2	4	0	6	50.0	33.3
	Tten	0	0	15	15	100.0	0.0
	Total obs	45	8	15	68	91.2	

We could realize that in this case, when just a small data sample is available the ENN was unable to obtain accurate classification results. This is probably the main drawback of ENN, and for certain very difficult to capture/record species, where the available samples will most often be small, the logistic regression models could present a valid alternative.

5 CONCLUSIONS & FUTURE WORK

We achieved in this work three main objectives: the foundations for the first database of bat calls recorded in Portugal; the development of a computer program to isolate and measure bat calls from a given recording; and statistical models to accurately classify bat calls from the measured parameters.

We started by obtaining recordings from 39 distinct locations all over the country after hand release of captured bats, outside known roosts or at foraging sites where the identification of the species by ultrasound was unambiguously. Free flight recordings were obtained for *B. barbastellus*, *P. pipistrellus*, *P. kuhlii* and *Plecotus spp.* (we didn't discriminate between *P. auritus* and *P. austriacus* in this study). The compiled database included 748 recordings (from approximately the same number of individual bats) with 2968 individual calls, of (at least) 20 different species.

We only used search phase calls for classification purposes. A program was implemented in R environment and used to filter the recordings, detect and isolate echolocation bat calls and measure parameters. There were some issues with the program though: it did not work for high-duty cycle bats (of which *Rhinolophus* genus is the example in Portugal) and it biased call duration for low-cycle bats. Nevertheless we expect this program to be a very useful tool for bat researchers since it can provide an almost immediate identification of a bat species from recorded echolocation calls in the field and can handle the amount of recordings produced by automatic recording stations, diminishing the costs associated with bat monitoring in environmental impact assessment studies and allowing for long term monitoring of bat populations for wildlife management purposes.

For the classification of echolocation bat calls we measured 19 variables, three temporal and 16 from the frequency domain, obtained around the point of maximum energy of the call, which is less prone to atmospheric attenuation, and used a hierarchical classification scheme. Two different statistical approaches were employed for classification of calls: ensembles of neural networks and logistic regression models. The statistical models were fitted with R.

We found considerable differences in initial frequencies between previously published studies and our sample for *P. pipistrellus*, *P. pygmaeus*, *P. kuhlii*, *E. serotinus* and *N. leisleri*, and even differences for this variable between previously published studies indicating that our option of not using this parameter for classification was correct.

We also found considerable *f_{maxe}* differences between previously published studies and our sample for *M. myotis*, *M. daubentonii*, *P. pygmaeus*, *E. serotinus*, *M. schreibersii* and *T. teniotis*. For *M. myotis*, *M. daubentonii*, *E. serotinus* and *T. teniotis* our sample sizes are quite small for definite conclusions but we think some of these differences could be related to different sampling methods, recording devices or even to geographic variation. For *P.*

pygmaeus our results suggest a distribution of fmaxe values with one peak around 52-53 kHz but also a large portion of values between 55-56 kHz, being this latter interval coincident with the mean values observed in other studies. This could indicate that this species has individuals with different echolocation characteristics and different proportions of these individuals could exist in different populations. For *M. schreibersii*, we can see that fmaxe values of our sample are effectively concentrated around 53 kHz. The differences from previous studies could in fact reflect geographic differences reinforcing the usage of a local recorded database for classification purposes.

Rhinolophus species could be discriminated from the remaining species by a simple fmaxe cutpoint, defined at 75 kHz. All calls above this cutpoint are considered to belong to *Rhinolophus* species and as fmaxe tends to be very constant within each *Rhinolophus* species this cutpoint can probably be generalized outside our sample without significant error. *Rhinolophus* species themselves could be discriminated using fmaxe cutpoints: *R. ferrumequinum* has values between 75.0 and 92.0 kHz, *R. mehelyi* / *R. euryale* between 92.0 and 106.3 and *R. hipposideros* above this latter value. Nevertheless a more robust sample size is needed before definite conclusions can be obtained.

For the discrimination of the remaining major groups of species the ENN correctly classified 98.3% of *Myotis* calls, 99.0% of *Pipistrelloid* calls, 97.1% of *Nyctaloid* calls, 96.6% of *Barbastella* calls and 92.9% of the *Plecotus* calls, having an overall correct classification rate of 98.3%.

For *Myotis* species the logistic regression model correctly classified 85.9% of the *M. myotis* / *M. blythii* calls and 77.5% of the *M. escalerae* / *M. daubentonii* / *M. emarginatus* calls with an overall correct classification rate of 80.7%. The ENN model correctly classified 88.9% of *M. myotis* / *M. blythii* calls and 96.3% of *M. escalerae* / *M. daubentonii* / *M. emarginatus* calls with an overall correct classification rate of 94.0%. The variables posFMAXE, ffin15, bwfini15 were identified by the LR model and the variables duration, ptFMAXE, bw15 and bwffin10 were identified by the ENN model as important for the discrimination of these two groups of species.

For *Pipistrelloid* species we observed that all *P. kuhlii* calls had fmaxe values below 43.0 kHz and lower than the remaining species. This cutpoint however cannot be trusted for generalization as the majority of *P. kuhlii* recordings were obtained from free flying bats and larger sample size for this species recorded under different circumstances has to be obtained before definite conclusions can be made. A multinomial logistic regression model was fitted for the remaining three species and correctly classified 96.9% of *P. pipistrellus* calls, 89.0% of *P. pygmaeus* calls and 89.3% of *M. schreibersii* calls with an overall correct classification rate of 90.7%. The ENN model was fitted to all the four species and correctly classified 93.3% of *P. pipistrellus* calls, 91.8% of *P. pygmaeus* calls, 100.0% of *P. kuhlii* calls and 96.2% of *M. schreibersii* calls, having an overall correct classification rate of 94.8%.

For *Nyctaloid* species we observed that all *T. teniotis* calls had f_{max} values below 15.2 kHz and lower than the remaining species. A logistic regression model was fitted for the remaining two species and correctly classified 97.6% of *E. serotinus* / *E. isabellinus* calls and 63.6% of *N. leisleri* / *N. lasiopterus* calls with an overall correct classification rate of 90.6%. The multinomial ENN model classified correctly 100% of *T. teniotis* calls, 95.6% of *E. serotinus* / *E. isabellinus* and 50% of *N. leisleri* / *N. lasiopterus* calls with an overall classification rate ENN of 91.2%.

We verified that ENN outperformed logistic regression in almost all classification tasks with the exception of *Nyctaloid* species. Here the small sample size prevented the ENN to obtain good generalization ability and logistic regression was able to fit better to the data. Also when a cutpoint was detected that allowed 100% of correct discrimination, true for *P. kuhlii* and *T. teniotis*, the corresponding ENN classified with 100% accuracy that same species.

The unbalanced sample sizes seemed to influence the training of the ENN models, probably because even though there are no priors defined, larger sample sizes will contribute more to the error function that regulates weight update and the fit tends to be better for these larger samples. Using bootstrap to balance the training datasets could help overcome this limitation.

For logistic regression models the greater difficulty found was the colinearity between the measured parameters that prevented the usage of much measured information in the models and made the modeling task extremely complex. Despite the lower classification rates observed in logistic regression models, the insights that these models bring to the relevant parameters and their influence in the classification, associated with a simple and quick method to calculate the probability of a given call belonging to a certain class, could be a great improvement in studies regarding bat calls classification.

The work presented in this thesis is part of an ongoing investigation project whose main objective is to develop an automated method for bat call analysis and classification for use in Portugal and to further investigate the echolocation characteristics of bat species, and in this sense the capture/recording of bat calls is still undergoing and a more robust computer program is being developed.

We also intend to study how gender and age influence echolocation characteristics and also how geographical variation influence echolocation. When a larger data sample is available, the improvement of the classification scheme to include more species and further development of artificial neural networks and multinomial regression models will also be the focus of future research.

BIBLIOGRAPHY

- Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed., p. 373). Hoboken, New Jersey: John Wiley & Sons.
- Ahlèn, I., & Baagoe, H. J. (1999). Use of ultrasound detectors for bat studies in Europe: experiences from field identification, surveys, and monitoring. *Acta Chiropterologica*, 1(2), 137–150.
- Airas, M. (2003). Echolocation in bats. *Proceedings of Spatial sound perception and reproduction. The postgrad seminar course of HUT Acoustics Laboratory*, 1–25.
- Allen, C., Romeling, S., & Robbins, L. (2011). Acoustic monitoring and sampling technology. *Protecting threatened bats at coal mines: a technical interactive forum*.
- Alttringham, J. D. (2011). *Bats: from evolution to conservation* (2nd ed., p. 324). Oxford: Oxford University Press.
- Anastasiadis, A. D., Magoulas, G. D., & Vrahatis, M. N. (2005). New globally convergent training scheme based on the resilient propagation algorithm. *Neurocomputing*, 64, 253–270. doi:10.1016/j.neucom.2004.11.016
- Barclay, R. M. (1999). Bats are not birds: a cautionary note on using echolocation calls to identify bats: a comment. *Journal of Mammalogy*, 80(1), 290–296. doi:10.2307/1383229
- Betts, B. J. (1998). Effects of interindividual variation in echolocation calls on identification of big brown and silver-haired bats. *Journal of Wildlife Management*, 62(3), 1003–1010.
- Boyles, J., Cryan, P., McCracken, G., & Kunz, T. (2011). Economic Importance of Bats in Agriculture. *Science*, 332, 11–12.
- Cabral, M. J., Almeida, J., Almeida, P. R., Dellinger, T., Ferrand de Almeida, N., Oliveira, M. E., Palmeirim, J. M., Queiroz, A. I., Rogado, L. & Santos-Reis, M. (Eds.). (2006). *Livro vermelho dos vertebrados de Portugal* (2nd ed., p. 659). Lisboa: Instituto de Conservação da Natureza/Assírio & Alvim.
- Chick, R., & Lumsden, L. (1999). Monitor mode: a new Anabat software feature that automatically saves bat calls to computer. *The Australasian Bat Society Newsletter*, 13, 16–19.
- Cochran, W. G. (1977). *Sampling techniques* (3rd ed., p. 452). New York, NY: John Wiley and Sons Inc.
- Collen, A. (2012). *The evolution of echolocation in bats: a comparative approach*. University College London.
- Denny, M. (2004). The physics of bat echolocation: signal processing techniques. *American Journal of Physics*, 72(12), 1465–1477. doi:10.1119/1.1778393
- Denzinger, A., Siemers, B., Schaub, A., & Schnitzler, H. (2001). Echolocation by the barbastelle bat, *Barbastella barbastellus*. *Journal of Comparative Physiology A*, 187(7), 521–528. doi:10.1007/s003590100223

- Fenton, M. B. (1994). Assessing signal variability and reliability - to thine ownself be true. *Animal Behaviour*, 47(4), 757–764. doi:10.1006/anbe.1994.1108
- Fenton, M. B. M., & Bell, G. P. (1981). Recognition of species of insectivorous bats by their echolocation calls. *Journal of Mammalogy*, 62(2), 233–243.
- Griffin, D. R., Webster, F. A., & Michael, C. R. (1958). The Echolocation of flying insects by bats. *Animal Behaviour*, 8(3-4), 141–154. doi:10.1016/0003-3472(60)90022-1
- Gu, L. (2002). *A new robust algorithm for isolated word endpoint detection*. Old Dominion University.
- Günther, F., & Fritsch, S. (2010). neuralnet: Training of Neural Networks. *The R Journal*, 2(1), 30–38.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference and prediction* (2nd ed., p. 745). New York, NY: Springer.
- Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*, 842.
- Heaton, J. (2005). *Introduction to neural networks with Java* (2nd ed., p. 439). Chesterfield: Heaton Research, Inc.
- Hogan, J. (1921). The Heterodyne Receiver. *Electric Journal*, 18(4), 116–119.
- Hosmer, D. W., & Lemeshow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics Theory and Methods*, 9(10), 1043–1069. doi:10.1080/03610928008827941
- Hosmer, & Lemeshow. (2000). *Applied logistic regression* (2nd ed., p. 392). New York, NY: John Wiley and Sons, Inc.
- Hozo, S., Djulbegovic, B., & Hozo, I. (2005). Estimating the mean and variance from the median, range, and the size of a sample. *BMC medical research methodology*, 5(13). doi:10.1186/1471-2288-5-13
- Ibáñez, C., Juste, J., García-Mudarra, J. L., & Agirre-Mendi, P. T. (2001). Bat predation on nocturnally migrating birds. *Proceedings of the National Academy of Sciences of the United States of America*, 98(17), 9700–9702. doi:10.1073/pnas.171140598
- Intrator, O., & Intrator, N. (2001). Interpreting neural-network results: a simulation study. *Computational Statistics & Data Analysis*, 37(13), 373–391.
- Iyer, M. S., & Rhinehart, R. R. (1999). A method to determine the required number of neural-network training repetitions. *IEEE Transactions on Neural Networks*, 10(2), 427–432. doi:10.1109/72.750573
- Jennings, N., Parsons, S., & Pocock, M. J. O. (2008). Human vs. machine: identification of bat species from their echolocation calls by humans and by artificial neural networks. *Canadian Journal of Zoology*, 86(5), 371–377. doi:10.1139/Z08-009
- Jones, G. (1999). Scaling of echolocation call parameters in bats. *The Journal of experimental biology*, 202(Pt 23), 3359–3367.

- Jones, G., Gordon, T., & Nightingale, J. (1992). Sex and age differences in the echolocation calls of the lesser horseshoe bat, *Rhinolophus hipposideros*. *Mammalia*, *56*(2), 189–193.
- Jones, K. (2002). Chiroptera (Bats). In *Encyclopedia of Life Sciences*. Chichester: John Wiley & Sons, Ltd.
- Kalko, E. K. V, & Schnitzler, H. U. (1993). Plasticity in echolocation signals of European pipistrelle bats in search flight: implications for habitat use and prey detection. *Behavioral Ecology and Sociobiology*, *33*(6), 415–428. doi:10.1007/BF00170257
- Killick, R., & Eckley, I. A. (2011). changepoint: An R package for changepoint analysis. *R package*.
- Kosmidis, I. (2007). brglm: Bias reduction in binary-response GLMs. *R package*.
- Kotsiantis, S. B. (2007). Supervised Machine Learning : A Review of Classification Techniques. *Informatica*, *31*, 249–268.
- Kunz, T., & Parsons, S. (2009). *Ecological and behavioral methods for the study of bats*. (T. Kunz & S. Parsons, Eds.) (2nd ed., p. 901). Baltimore, Maryland: The Johns Hopkins University Press.
- Larom, D., Garstang, M., Payne, K., Raspet, R., & Lindeque, M. (1997). The influence of surface atmospheric conditions on the range and area reached by animal vocalizations. *Journal of Experimental Biology*, *431*, 421–431.
- Mccullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models* (2nd ed., p. 511). London: Chapman and Hall.
- Mickleburgh, S. P., Hutson, A. M., & Racey, P. A. (2002). A review of the global conservation status of bats major threats. *Oryx*, *36*(1), 18–34. doi:10.1017/S0030605301000011
- Miller, R. L. (1939). Fractional-Frequency Generators Utilizing Regenerative Modulation. *Proceedings of the IRE*, *27*(7), 446–457. doi:10.1109/JRPROC.1939.228513
- Moss, C. F., & Sinha, S. R. (2003). Neurobiology of echolocation in bats. *Current Opinion in Neurobiology*, *13*(6), 751–758. doi:10.1016/j.conb.2003.10.016
- Nagelkerke. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*(3), 691–692. doi:10.1109/LPT.2009.2020494
- Nelder, J. A., & Wedderburn, R. W. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society*, *135*(3), 370–384.
- Norberg, U. M., & Rayner, J. M. V. (1987). Ecological morphology and flight in bats (Mammalia; Chiroptera): wing adaptations, flight performance, foraging strategy and echolocation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *316*(1179), 335–427. doi:10.1098/rstb.1987.0030
- Nowak, R. (1994). *Walker's bats of the world* (5th ed., p. 287). Baltimore, Maryland: The Johns Hopkins University Press.

- O'Shea, T., Ellison, L., & Stanley, T. (2004). Survival estimation in bats: historical overview, critical appraisal, and suggestions for new approaches. In W. L. Thompson (Ed.), *Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters* (Vol. 15, pp. 297–336). Washington, D.C.: Island Press.
- Obrist, M. K. (1995). Flexible bat echolocation: the influence of individual, habitat and conspecifics on sonar signal design. *Behavioral Ecology and Sociobiology*, *36*(3), 207–219. doi:10.1007/bf00177798
- Obrist, M. K., Boesch, R., & Flückiger, P. F. (2004). Variability in echolocation call design of 26 Swiss bat species: consequences, limits and options for automated field identification with a synergetic pattern recognition approach. *Mammalia*, *68*(4), 307–322.
- Palmeirim, J. M., & Rodrigues, L. (1992). Plano Nacional de Conservação de Morcegos Cavernícolas. *Estudos de biologia e conservação da natureza*, *8*.
- 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*, *75*(9), 1487–1494. doi:10.1139/z97-772
- 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*(Pt 17), 2641–2656.
- Pfalzer, G., & Kusch, J. (2003). Structure and variability of bat social calls: implications for specificity and individual recognition. *Journal of Zoology*, *261*(1), 21–33. doi:10.1017/S0952836903003935
- Pollack, G. S., & Imaizumi, K. (1999). Neural analysis of sound frequency in insects. *BioEssays*, *21*(4), 295–303. Retrieved from <http://doi.wiley.com/10.1002/%28SICI%291521-1878%28199904%2921%3A4%3C295%3A%3AAID-BIES5%3E3.3.CO%3B2-L>
- Racey, P. (2009). Bats: status, threats and conservation successes. Introduction. *Endangered Species Research*, *8*(July), 1–3. doi:10.3354/esr00213
- Rainho, A. (1995). *Inventariação das espécies e dos abrigos de morcegos nos Parques Naturais da Arrábida e da Serra de S. Mamede e determinação dos biótopos de alimentação de algumas espécies de morcegos* (p. 110). Lisboa: Instituto de Conservação da Natureza.
- Rainho, A. (1996). *Biótopos de alimentação dos morcegos presentes nos abrigos Marvão I e Moura I* (p. 131). Lisboa: Instituto de Conservação da Natureza.
- Redgwell, R. D., Szewczak, J. M., 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. doi:10.3390/a2030907
- Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster backpropagation learning: the RPROP algorithm. *IEEE International Conference on Neural Networks*, *1*, 586–591. doi:10.1109/ICNN.1993.298623
- Rodrigues, L. (1996). *Utilização de grades para protecção de abrigos de morcegos cavernícolas*. Faculdade de Ciências da Universidade de Lisboa.

- Rodrigues, L., & Palmeirim, J. M. (2008). Migratory behaviour of the Schreiber's bat: when, where and why do cave bats migrate in a Mediterranean region? *Journal of Zoology*, 274(2), 116–125.
- Rojas, R. (1996). *Neural Networks: A systematic introduction* (1st ed., p. 502). New-York, NY: Springer.
- Rumelhart, D., & McClelland, J. (1986). *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1. Foundations* (1st ed., p. 567). Cambridge, MA: MIT Press.
- Russ, J. (1999). *The Bats of Britain & Ireland: Echolocation Calls, Sound Analysis and Species Identification* (1st ed., p. 103). London: Alana.
- 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*, 258(1), 91–103. doi:10.1017/S0952836902001231
- Rydell, J. (1993). Variation in the Sonar of an Aerial-Hawking Bat (*Eptesicus-Nilssonii*). *Ethology*, 93(4), 275–284.
- Saerens, M., Latinne, P., & Decaestecker, C. (2002). Any reasonable cost function can be used for a posteriori probability approximation. *IEEE transactions on neural networks*, 13(5), 1204–1210. doi:10.1109/TNN.2002.1031952
- Schnitzler, H., & Kalko, E. (2001). Echolocation by insect-eating bats. *Bioscience*, 51(7), 557–569.
- Sherwin, H. A., Montgomery, W. I., & Lundy, M. G. (2012). The impact and implications of climate change for bats. *Mammal Review*, (Early view). doi:10.1111/j.1365-2907.2012.00214.x
- Siemers, B., Dietz, C., Nill, D., & Schnitzler, H. (2001). *Myotis daubentonii* is able to catch small fish. *Acta chiropterologica*, 3(1), 71–75.
- Skowronski, M. D., & Harris, J. G. (2006). Acoustic detection and classification of microchiroptera using machine learning: Lessons learned from automatic speech recognition. *The Journal of the Acoustical Society of America*, 119(3), 1817–1833. doi:10.1121/1.2166948
- Sueur, J., Aubin, T., & Simonis, C. (2008). Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics*, 18(2), 213–226.
- Thomas, D., Bell, G., & Fenton, M. B. (1987). Variation in echolocation call frequencies recorded from north american vespertilionid bats: a cautionary note.
- Vaughan, N., Jones, G., & Harris, S. (1997). Identification of British bat species by multivariate analysis of echolocation call parameters. *Bioacoustics The International Journal Of Animal Sound And Its Recording*, 7(3), 189–207. doi:10.1080/09524622.1997.9753331
- Walters, C. L., Freeman, R., Collen, A., Dietz, C., Brock, M., Jones, G., ... Jones, K. E. (2012). A continental-scale tool for acoustic identification of European bats. *Journal of Applied Ecology*, 49(5), 1064–1074. doi:10.1111/j.1365-2664.2012.02182.x

Waters, D., & Jones, G. (1995). Echolocation call structure and intensity in five species of insectivorous bats. *The Journal of experimental biology*, 198(Pt 2), 475–489.

Zingg, P. E. (1990). Acoustic species identification of bats (Mammalia: Chiroptera) in Switzerland. *Revue Suisse de Zoologie*, 97(2), 263–294.

APPENDIX 1 – IMPLEMENTED FUNCTIONS IN R

Function “imp_rec” to import, filter and obtain frequency of maximum energy over the entire recording

```
imp_rec <- function(tx, fileName) {  
  data <- readWave(fileName)  
  fs <- data@samp.rate  
  bt <- butter(10, 8000, type = "high")  
  data <- as.integer(filter(bt, data))  
  fmaxe <- floor(F.MAXE(data, fs, tx) * 1000)  
  blimit <- fmaxe * 0.33 / (fs * tx / 2)  
  bt <- butter(10, blimit, type="high")  
  data <- as.integer(filter(bt, data))  
  list(signal = data, fmaxe = fmaxe, fs = fs)  
}
```

Function “teager” to implement the Teager Energy Algorithm

```
teager <- function(signal) {  
  signal.new <- abs(signal[2:(length(signal) - 1)]) ^ 2 - diff(signal[1:(length(signal) - 1)]) *  
    iff(signal[2:length(signal)])  
  sign <- NaN  
  sign[signal < 0] <- -1  
  sign[signal >= 0] <- 1  
  sign <- sign[2:(length(sign) - 1)]  
  signal.new <- signal.new * sign  
  signal.new  
}
```

Functions “find_calls” and “sel_calls” to detect and isolate/extract bat calls from a recording

```

find_calls <- function(signal, pulsmax=10, fs, tx) {
  roll <- rollmean (abs(teager(signal)),100)
  signal.new <- approx (roll, y = NULL, method="linear", n = length(signal) / 200,
    rule = 1, f = 0, ties = mean)
  peaks.aux <- matrix(c(signal.new$x, signal.new$y), ncol = 2,
    nrow = length(signal.new$x))
  peaks <- fpeaks(peaks.aux, plot = FALSE, nmax = pulsmax)
  peaks <- c[,1]
  peaks.out <- NULL
  for (i in 2:pulsmax) {
    dif <- peaks[i] - peaks[i-1]
    if (dif < fs * tx * 25 / 1000) {
      peaks.out <- append(peaks.out,i)
    }
  }
  if (length(peaks.out) > 0) {
    peaks <- as.integer(peaks[-c(peaks.out)])
  } else {
    peaks <- as.integer(peaks)
  }
  peaks
}

sel_calls <- function(signal, fs, tx) {
  posfmaxe <- find_calls(signal, fs, tx)
  pos <- matrix(data = NA, nrow = length(posfmaxe), ncol = 2)
  maxpos <- NULL
  for (i in 1:length(posfmaxe)) {
    posi <- posfmaxe[i] - (fs * tx * 20 / 1000)
    posf <- posfmaxe[i] + (fs * tx * 20 / 1000)
    maxpos[i] <- which.max(signal[posi:posf]) + posi
    pos[i,1] <- posi
    pos[i,2] <- posf
  }
  np <- length(pos[,1])
  list(np = np, pos = pos, maxpos = maxpos)
}

```

Functions “zero_pad”, “fft_new” and “freq_maxe” to normalize a signal, compute FFT over the normalized signal and return the frequency with greater amplitude

```
zero_pad <- function (signal, n) {
  s <- signal
  np <- length(s)
  new.np <- n
  new.s <- 1:new.np
  new.s[1:new.np] <- 0
  new.s[1:np] <- s
  new.s
}

fft_new <- function(signal, fs, tx) {
  z <- zero_pad(signal, 32768)
  n <- length(z)
  f <- abs(fft(z))
  int <- f[1:(n / 2)]
  freq <- (0:(n / 2 - 1)) * (fs * tx) / n
  list(freq = freq, int = int)
}

freq_maxe <- function(signal, fs, tx) {
  aux <- fft_new(signal, fs, tx)
  pos <- which.max(aux$int)
  fmaxe <- round(aux$freq[pos] / 1000, 2)
  list (fmaxe = fmaxe, int = aux$int, freq = aux$freq)
}
```

Function “hosmerlem” to implement Hosmer and Lemeshow test

```
hosmerlem <- function(y, yhat, g=10) {  
  cutyhat <- cut(yhat,breaks = quantile(yhat, probs=seq(0, 1, 1 / g)),  
    include.lowest=TRUE)  
  obs <- xtabs(cbind(1 - y, y) ~ cutyhat)  
  expect <- xtabs(cbind(1 - yhat, yhat) ~ cutyhat)  
  chisq <- sum((obs - expect) ^ 2 / expect)  
  pval <- 1 - pchisq(chisq, g - 2)  
  list(chisq=chisq, p.value=pval)  
}
```

Function “splitbystr” randomize and split a dataset while maintaining the strata proportions of the original dataset

```
splitbystr <- function(data, strata, trainsize, seed) {  
  set.seed(seed)  
  n <- length(strata)  
  aux <- cbind(c(1:n))  
  pik <- rep(trainsize, times = n)  
  filter <- balancedstratification(aux, strata, pik, comment=F)  
  train <- data[filter == 1]  
  valid <- data[filter == 0]  
  ori <- cbind(filter, data)  
  list(train = train, valid = valid, ori = ori)  
}
```