Automated techniques for bat echolocation call analysis

Christopher David Scott

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Institute of Integrative and Comparative Biology

January, 2012

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

Thank you to my supervisor, John Altringham, for his support and guidance throughout the course of my PhD, and for inspiring me to study bats in the first place. This thesis would not have been possible without the help, advice and friendship from the Leeds bat lab – thank you all! I would also like to thank the many other people who assisted with the fieldwork in numerous ways, including Gail Armstrong, Jenny Blakemore, Joanna Hebbert, Chris Heward, Vicki Hodgetts, Aurelie Laurent, Luci Lay, John Martin, the Mason family, Kate Parker, Abbie Shaw, and Daniel Whitby. I am also deeply grateful to my parents and family who continue to support me in everything I do.

This project was funded by a University of Leeds Research Scholarship, and I'd like to thank the People's Trust for Endangered Species and the Countryside Council for Wales for added financial support.

Finally, I owe special thanks to Chloe Bellamy. Chloe introduced me to the Lake District and was the perfect guide, provided invaluable help during this project, and offered unfailing love and support throughout.

Abstract

Acoustic bat detectors are an extraordinarily valuable tool in bat research as they enable researchers to listen in on the otherwise secretive world of bats, providing the means to non-invasively survey and monitor bats in their natural habitats. Technological advances facilitate unprecedented data collection, considerably expanding the scope of field studies. However, the burden of manual analysis, and difficulty in identifying some species reliably from their calls, hampers the development of systematic survey and long-term monitoring methods. We developed a series of algorithms for the automated analysis of bat detector recordings, used to detect and extract calls from continuous recordings, and measure temporal and spectral call variables. By hand-labelling the location of calls in field recordings, we were able to evaluate the accuracy of the automated method at detecting calls. Comparison on the same dataset with two conventional bioacoustic signal detectors revealed our algorithm was more accurate and robust. Using machine learning (ML) classification algorithms that learn to identify calls following training using a reference library, we developed a fully automated species identification system. Evaluation of the system was carried out by cross-validation of our reference call library, containing recordings of >5000 calls from known British species, comparing classifier predictions to ground-truth labels. The ML approach outperformed conventional statistical analysis using discriminant function analysis (DFA). We applied our novel system to two field studies that highlight its utility. Firstly, monitoring multi-species bat activity at a remote cave system over a period of three months, analysing >20,000 audio files to investigate temporal patterns in activity. Secondly, separating acoustically cryptic *Myotis* species from data collected in the Lake District National Park, to generate presence data for species distribution modelling, facilitating the creation of species-specific habitat suitability maps projected over the entire Park (ca. 3,300 km^2).

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AUC – Area Under ROC Curve

ANN – Artificial Neural Network

CART – Classification and Regression Tree

CV – Cross-validation

DFA – Discriminant Function Analysis

EGV – Enviro-geographic Variable

FFT – Fast Fourier Transform

FLAC - Free Lossless Audio Codec

GAM – Generalised Additive Model

GIS – Geographical Information System

GPS – Global Positioning Satellite

HSI – Habitat Suitability Index

HSM – Habitat Suitability Model

LOESS – Locally Estimated Scatterplot Smoothing

LOOCV – Leave-one-out Cross-validation

MaxEnt – Maximum Entropy

ML – Machine Learning

PPP – Positive Predictive Power

PS – Power Spectrum

QDA – Quadratic Discriminant Analysis

RF – Random Forest

ROC – Receiver Operating Characteristic

SDM – Species Distribution Model

SNR – Signal to Noise Ratio

STFT – Short-time Fourier Transform

SVM – Support Vector Machine

VIF – Variance Inflation Factor

Chapter 1: General Introduction

1.1 Scope

The focus of this research is on advancing the tools and techniques available to study bat echolocation, specifically through developing automated analysis procedures, and the application of techniques developed from the field of machine learning (ML). The topics are technical in nature, and this introductory chapter aims to serve as a primer for ecologists. The nature of bat echolocation will be discussed, and the relevance to ecologists and conservationists. The latest techniques for recording ultrasound in the field will be discussed, as well as some fundamentals of digital signal processing, pertinent to the analysis of echolocation calls. ML classification and methodology will be introduced, illustrated with examples relevant to the study of bat echolocation. Lastly, the rationale behind the study will be presented, with the principle aims of the research.

1.2 Bats and echolocation

Bats emit very high frequency, short duration vocalisations and listen to the returning echoes to detect, localise and classify objects in their surroundings. This system of echolocation allows insectivorous bats to orientate and hunt at night by building up a sound picture of their environment. New World frugivorous bats make use of odour and vision, but also utilise echolocation in foraging and assessing fruit ripeness (e.g. Korine & Kalko, 2005). Bats are both the signallers and receivers of their own calls, and actively control the structure of their calls to influence the type, and quality of information encoded in the returning echoes (Schnitzler *et al.*, 2003; Surlykke *et al.*, 2009). The sensory demands that bats face in the environment are determined by the habitat they forage in, what they eat, and how they acquire it. These perceptual challenges have led to strong selection pressure on signal design, and bats have evolved signals that best suit their needs. Dawkins (1986) uses bat echolocation as an example of 'good design' through evolution by natural selection.

Bat species can be split broadly into guilds according to their habitat use, with associated adaptations in signal design for three major tasks in echolocation - *detection*, *localisation* and *classification* (Schnitzler & Kalko, 2001). Narrowband signals that concentrate their energy in a narrow range of frequencies are suited to *detection*. *Nyctalus* species hunt in open areas, and use relatively long, low frequency calls that return strong echoes from distant targets (Jones, 1995). *Myotis* and *Plecotus* species that take prey from close to vegetation use broadband calls that sweep through a range of frequencies, increasing *localisation* ability, and keep the duration of their calls short to prevent overlap between emitted calls and returning echoes. Frequency dependent atmospheric absorption rapidly attenuates high frequencies (Lawrence & Simmons, 1982), making broadband calls effective only over short distances. Moreover, there is a trade-off between *detection* and *localisation* ability, and a signal optimised for *detection* will be poorer at *localisation*. Bats such as *Pipistrellus* species are very flexible in their use of habitat, and alter their call types accordingly; in more open habitats their calls are narrowband and long, for increased *detection*, but become progressively shorter, and more broadband as they approach clutter to favour *localisation*. Horseshoe bats have a specialised call type that is long in duration, and of a high constant frequency. This sophisticated call facilitates the fine-tuned *detection*, and even *classification*, of insects from their fluttering wings (Schnitzler & Flieger 1983; Emde & Schnitzler, 1990).

1.3 Bat detectors and ultrasound recording

Due to the nocturnal habits of bats, visual identification in flight is rarely possible. Moreover, as small, fast flying animals with a highly evolved sensory system, capture using traps is a highly skilled, labour intensive process with typically low capture success rates (e.g. Berry *et al.*, 2004). However, the fact that bats call at a high repetition rate in flight, to continuously update their sound picture, offers enormous potential for researchers to eavesdrop on these calls. The majority of signals emitted by bats are ultrasonic (above the range of human hearing), necessitating some form of transformation to make them audible to humans. Bat detectors provide this function, and have been widely applied to gain insights into their ecology and behaviour, and are ubiquitous in practical conservation work for monitoring bat populations (Ahlen & Baagøe, 1999; Fenton, 2003).

There are three distinct methods typically employed for converting ultrasounds into audible sounds, although a single bat detector may employ more than one for flexibility: heterodyning, frequency-division, and time-expansion. The methods each have advantages and disadvantages (Parsons *et al.*, 2000), but time-expansion is the only method that retains the original structure of the recorded signal. Heterodyning and frequency-division both necessitate at least some loss of signal content in the process of transformation, making time-expansion most suited to subsequent detailed acoustic analysis. Time-expansion works on the principle of playing back a small sample of recorded ultrasound at a slower rate, thus reducing its pitch. A factor of ten is frequently employed, bringing ultrasonic frequencies up to ~200 kHz down into the audible range, at the expense of a tenfold increase in playback time. The output from a time-expansion detector can be recorded by a conventional audio recorder, for later acoustic analysis using a computer. The limitation with this technique is that current detectors will only make short recordings (a few seconds), and they are unable to playback and record concurrently, so during playback continued sampling of ultrasound cannot take place.

Ultrasound can be recorded in analog format using high speed tape recorders, but digital recording technology has largely replaced this practice. A digital recorder stores discrete samples of the signal at a very high rate. To avoid an undesirable effect known as aliasing, that causes different signals to become indistinguishable when sampled, the highest reproducible frequency, or Nyquist frequency, is half the sampling rate. For conventional audio material, since human hearing only extends to a maximum of ~20 kHz, digital audio is typically sampled at 44.1 kHz. This leads to a Nyquist frequency of 22.05 kHz. Accurate sampling of ultrasound requires even higher sampling rates, and the latest digital recorders are capable of sampling at up to 500 kHz, leading to faithful signal representation up to 250 kHz. These advances in digital recording technology in combination with increases in data storage capabilities, have led to a new generation of bat detectors that directly sample ultrasound received at the microphone for subsequent analysis. Recording length using direct sampling is limited only by the storage space available. This offers the advantage that a recorded 'bat pass' (a sequence of echolocation calls recorded from a bat passing the microphone), can be captured in its entirety, rather than sampling only a single call as is often the case with the limited sampling of time-expansion detectors. The limitation is that the recorded ultrasound remains inaudible at the time of recording, although they can later be time-expanded using a computer. Direct sampling is currently the most convenient way of recording high quality ultrasound for detailed acoustic analysis using computers.

1.4 Acoustic analysis

Recording ultrasound permits researchers to view and analyse the spectral content of a signal. A captured audio recording allows detailed and quantitative analysis of signals, capable of revealing temporal and spectral detail of signal structure. Recorded signals can be analysed using time or frequency domain techniques. Zero-crossing analysis is a time domain technique that analyses the recorded audio samples directly. A zero-crossing detector registers the transition of a signal waveform from positive and negative, which provides a means to track the harmonic with greatest amplitude (Parsons *et al.*, 2000). Zero-crossing analysis is very efficient due to its simplicity, but loses all harmonic information in the signal. The Fourier transform is a frequently employed frequency domain technique, which converts the time domain samples into the spectral composition of the signal. The Fourier transform permits a full spectrum analysis, revealing the harmonic content of the signal. The short time Fourier transform (STFT) slides a discrete analysis window through the signal, using the Fourier transform to reveal the changing frequency content of the signal over time. The Fourier transform is computationally intensive, and there is a trade-off between time and frequency resolution due to the Gabor limit (Gabor, 1946). To achieve a high frequency resolution, a large analysis window is required, which results in reduced temporal precision. In contrast, a shorter analysis window provides higher temporal resolution, but a coarser frequency resolution. There are alternative frequency domain techniques, for example wavelets (Graps, 1995), but they have not gained widespread use in analysing echolocation calls. This may be in part due to a very efficient algorithmic implementation of the Fourier transform, called the fast Fourier transform (FFT), which can be run in real time or faster on modern computers.

The output from the STFT can be used to plot the time-varying distribution of energy by frequency, called a spectrogram (Figure 1.1). In the spectrogram, time is displayed on the x-axis, and frequency is displayed on the y-axis. Energy, or amplitude, is typically represented by colour intensity or shade on the z-axis. The spectrogram is analogous to a musical score, with higher frequency sounds displayed in the upper regions of the plot. The spectrogram displays the time-varying structure or 'shape' of signals, and can be used to measure specific features of interest. A further representation is the power spectrum, which is a two-dimensional representation of power as a function of frequency (Figure 1.2). The power spectrum is time invariant, and can be used to summarise the frequency content over an entire call. Amplitude is usually expressed in decibels (dB), and is frequently relative to the loudest part of a call (e.g. 20 dB below the peak). In some cases amplitude may be expressed relative to digital full scale (dBFS), the maximum value that can be represented by a digital file. In this case dBFS reaches a maximum value of 0, and everything below takes negative values (e.g. -20 dBFS).

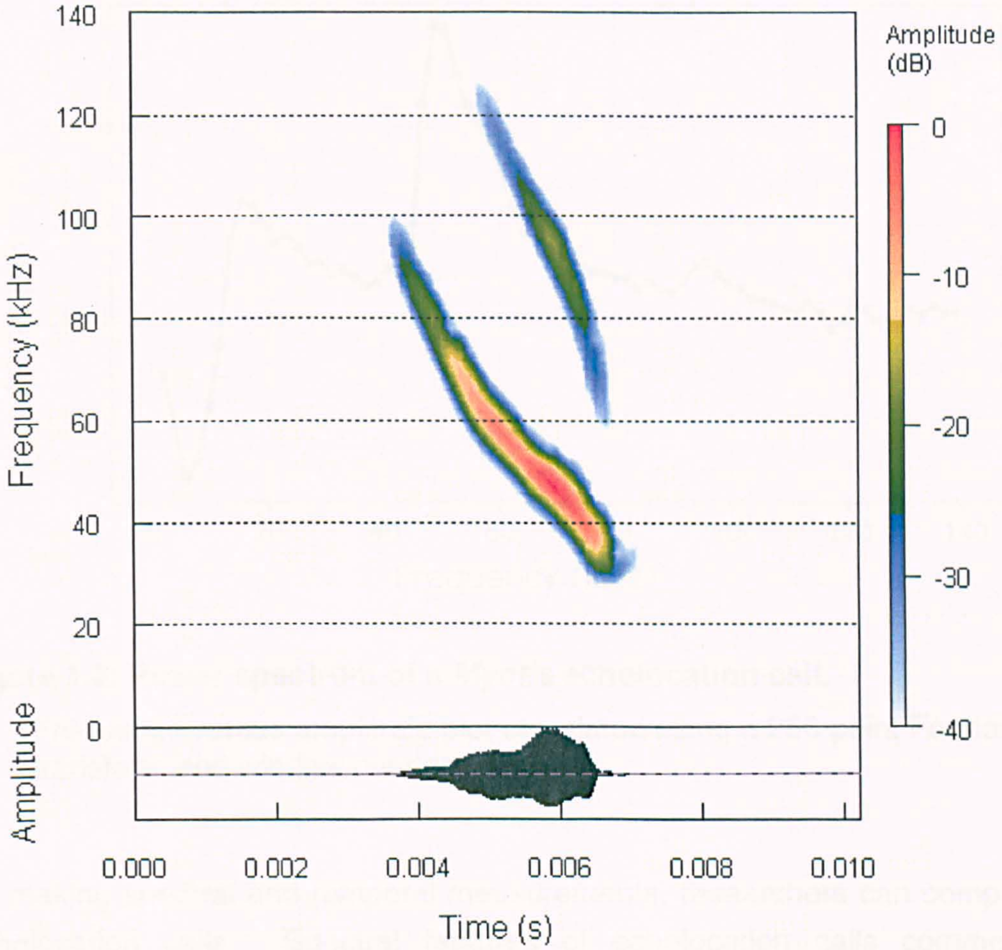


Figure 1.1: Spectrogram of a *Myotis* echolocation call.

Warmer colours indicate frequencies containing more energy. The time domain signal is plotted below as a waveform, showing the amplitude of the signal over time.

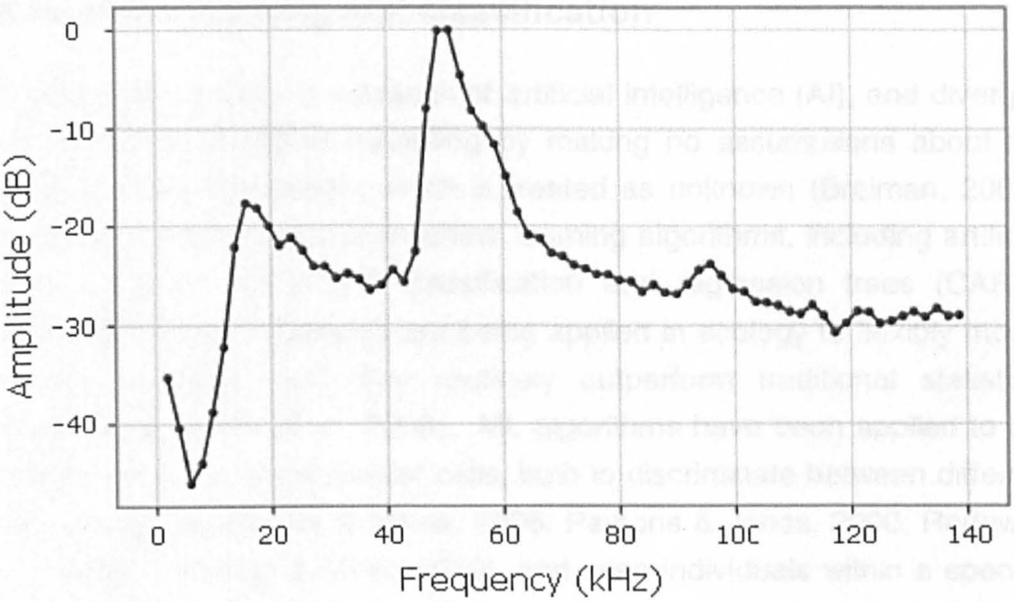


Figure 1.2: Power spectrum of a *Myotis* echolocation call.

Frequency versus amplitude plot calculated using a 256-point Fourier transform and window overlap of 75%.

By making spectral and temporal measurements, researchers can compare echolocation calls. Spectral features of echolocation calls commonly measured include the frequency at the start of the call, the frequency at the end of the call, and the frequency at the point of maximum energy in the call. A temporal feature, call duration, is most accurately measured from a plot of the time domain samples (waveform; Fig. 1.1), due to the higher temporal precision that can be achieved. Spectrogram displays and measured call features are frequently used to identify species on the basis of their calls. This can be achieved using subjective or quantitative statistical methods. For some species identification can be made quickly and reliably from call 'shape' and end frequency (Fenton & Bell, 1981). However, subjective separation of species in this way is not recommended, as it can vary significantly between researchers depending on their skills and experience, making results difficult to repeat (Parsons *et al.*, 2000). Quantitative statistical methods of separating species on the basis of measured call features, using techniques like discriminant function analysis (DFA; e.g. Obrist, 1995; Vaughan *et al.*, 1997; Russo & Jones, 2002), are therefore preferred.

1.5 Machine learning and classification

Machine learning (ML) is a branch of artificial intelligence (AI), and diverges from traditional statistical modelling by making no assumptions about the underlying data distribution, which is treated as unknown (Breiman, 2001). There exists a wide array of machine learning algorithms, including artificial neural networks (ANN) and classification and regression trees (CART). Increasingly these techniques are being applied in ecology to flexibly model complex problems, and they routinely outperform traditional statistical methods (e.g. Olden *et al.*, 2008). ML algorithms have been applied to the identification of bat echolocation calls, both to discriminate between different species (e.g. Skowronski & Harris, 2006; Parsons & Jones, 2000; Redgwell *et al.*, 2009; Armitage & Ober, 2010), and even individuals within a species (Burnett & Masters, 1999; Yovel *et al.*, 2009). However, whilst there is increasing use of ML algorithms in the literature, they are not widely used by ecologists despite the advantages they offer (Olden *et al.*, 2008). This may in part be because the ML algorithmic modelling approach diverges from the traditional statistical culture of data models, and remains poorly understood.

Classification or supervised learning is a form of predictive analysis. In classification a prediction is made directly on the data, to assign each case to one of a set of predefined classes. For example, classify an acoustic event as being an echolocation call or simply noise. This type of classification is considered "supervised", because the classes are determined by the researcher, rather than inferred from the data as in "unsupervised" clustering. There are many different algorithms that perform classification, but the process of building and using a classifier is shared, and can be broken into three main stages: training, testing and application. Training and testing necessitate ground truth data. To illustrate, in order to discriminate between *Pipistrellus pipistrellus* and *Pipistrellus pygmaeus* echolocation calls, example calls from each species are required. This necessitates building an echolocation call library, containing reference recordings from known species. To achieve this, recordings are typically made from bats that have been caught, identified and subsequently released; and from bats emerging from known roosts (e.g. Parsons & Jones, 2000; Obrist *et al.*, 2004). Once echolocation call features have been measured, a dataset is prepared containing multiple records for each class (in this case the two species), each record containing several call features and a class label (e.g. "*Pipistrellus pipistrellus*"). The data set is split into a training set and an independent test set. During a training phase, the classification algorithm attempts to iteratively "learn" to separate the classes based on the available features, hence the name machine learning. Once a classifier has fitted a function to the data, the training error is found by comparing the predicted class labels to the actual training data class labels. Results are frequently summarised in a confusion matrix, which displays the numbers of correctly and misclassified cases (Table 1 shows an example confusion matrix for a binary classification problem where 5 *Pipistrellus pipistrellus* calls are misidentified as *P. pygmaeus*, and 2 *P. pygmaeus calls* are misidentified as *P. pipistrellus*).

		Predicted	
		Pipistrellus pipistrellus	*Pipistrellus pygmaeus*
Actual	*Pipistrellus pipistrellus*	75	5
	Pipistrellus pygmaeus	2	78

Table 1.1: Example confusion matrix for a binary classification problem.

The training error estimate may be optimistic, and not a reliable measure of the future performance of the classifier on new data. To obtain an unbiased estimate of the generalisation error, the independent test set is used to estimate the error of the classifier. In this way, the performance of the classifier on data that were not used in training is established. To achieve good generalisation performance, i.e. make accurate predictions on unseen data, there is a balance to be made between under-fitting (high bias) and over-fitting (high variance). Under-fitting results in low accuracy on training and testing sets, whilst over-fitting results in high training set accuracy but lower test set accuracy. There is a bias-variance trade-off in fitting a function to a problem: under-fit and the function is too simplistic to accurately separate the classes, but over-fit and the function is overly-complex and models the training data too closely, which results in poor performance on previously unseen data. Figure 2 summarises these concepts.

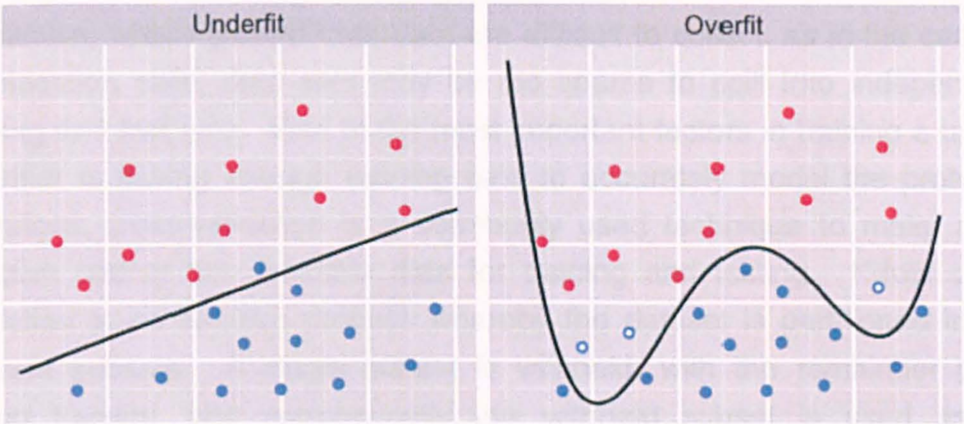


Figure 1.3: Examples of model underfitting and overfitting.

A linear decision boundary fails to fully separate the two classes (red and blue circles). In contrast, the example on the right shows a complex decision boundary that is overfit to the training data, and subsequently misclassifies independent test data from the blue class (shown as hollow blue circles).

Solving problems of bias or variance is complex and requires a deep understanding of the data and classifier being used. Some classifiers are high bias in design, and can only fit simple functions. Others have many user tuneable parameters, allowing them to fit complex and arbitrary functions (e.g. artificial neural networks). The classification problem itself may be trivial, with classes perfectly separated using a linear function, or may be more complex. Furthermore, the measured features have a large impact on the classification problem. A single feature may be enough to separate two classes, or many features may be required. For example, of the British species, only the two rhinolophid bats can be separated using the frequency of maximum energy alone (e.g. Parsons & Jones, 2000). The more features that are measured however, the greater the opportunity for a complex algorithm to overfit to the training data, highlighting the importance of the independent test set. Unfortunately, selecting features to use in a classification problem is domain specific, and remains empirical and heuristic. There is no independent and rigorous way of determining whether a dataset contains sufficient information to permit successfully classification (Janert, 2011). If a classifier fails, it may be unclear whether a different classifier would have been successful, or whether different/additional features may be useful.

In practice, where ground truth data are difficult to collect, as in the case of echolocation calls, data sets may be too sparse to split into independent training and test sets. One of the most important factors in training a useful classifier is having enough training data to accurately model the problem. Therefore, cross-validation is a commonly used technique to make most effective use of the available data for training and testing. K-fold cross validation is an iterative process whereby the dataset is partitioned into k discrete subsets. A single subset is withheld, with the remainder used during training, and subsequently the withheld subset is used as an independent test set. The process is repeated, each time withholding a different subset as the test set, until all k subsets have been used in this way. The estimated error is then averaged across subsets. Ten fold cross validation (k=10) is commonly adopted, as is leave-one-out cross validation (LOOCV), which splits the dataset into as many subsets as there are data points. Cross validation is a computationally expensive process, requiring k models to be trained and tested, but maximises the available training data and gives an almost unbiased estimate of the true error (Varma & Simon, 2006). It is important that any adjustments in the model building process are carried out prior to testing. It has been common for researchers to train and test a classifier, before adjusting model parameters and repeating the testing process until the highest accuracy has been attained (Simon *et al.*, 2003). This is overfitting to the dataset, and is highly likely to result in inflated estimates of classifier accuracy, as it allows the classifier to overfit to the total dataset. Where model parameter or architecture tuning is undertaken, a separate validation set should be employed. This allows different models to be trained and tested, with the aim of minimising the validation set error, and then once the final model has been selected, the independent test set is employed to estimate the unbiased generalisation error. In some cases the methodological details given are not sufficient to verify whether final models have been tested using independent test sets (e.g. Redgwell *et al.*, 2009).

Once a classifier has been optimised and tested, it can be used to classify unseen data, for which the correct class label is unknown (note that this is in contrast to the test set data, for which the class labels were known but withheld from the classifier). For this purpose, the classifier is first trained with all the available data. In making predictions, the classifications can be hard or soft. Hard classifications give the class output only, e.g. "*Pipistrellus*

pipistrellus", or "*Pipistrellus pygmaeus*", whereas soft classifications are probabilistic in nature, e.g. "*Pipistrellus pipistrellus*: 0.96", "*Pipistrellus pygmaeus*: 0.04". Not all classifiers provide probabilistic output, but where available they offer an insight into the classifier's confidence in the prediction.

1.6 Purpose of research

Acoustic methods are an extraordinarily valuable tool in bat research, facilitating rapid survey and non-invasive monitoring. With recent advances in bat detectors many hours of high quality digital recordings can be collected, extending the scope of current research. However, manual analysis is slow, laborious and subjective, and freely available methods to automate the procedure have not been forthcoming. The project evolved to address the following issues:

(i) To build an echolocation call library from British bat species, providing robust ground truth data for subsequent call classification experiments

(ii) The development of a robust method of automatically locating echolocation calls in audio recordings

(iii) The development of a robust method to automatically extract echolocation call parameters

(iv) To assess the ability of machine learning methods to classify calls to species from automatically extracted echolocation calls

(v) To assess the application of a fully automated acoustic identification system to two field studies

1.7 Thesis outline

Chapter two introduces a robust signal detection algorithm for locating bat echolocation calls in continuous recordings. The accuracy and computational cost of the algorithm is assessed on a dataset of field recordings, and compared to two conventional bioacoustic signal detection algorithms. Code is provided for the algorithm implementation. Chapter three builds on the signal detection algorithm, providing a means to automatically extract spectral and temporal features of echolocation calls suitable for quantitative description and classification. The methods used to collect a bat echolocation call library from sites across the UK are detailed. The call library is used to compare the ability of a machine learning classifier, random forest, to classify calls to genus and species level. The analysis procedure is compared to conventional methods of acoustic analysis. Chapter four explores the implications of novel signal types on automated acoustic classification, and details the use of outlier detection to mitigate against the effects. Chapter five applies the methods developed over the previous chapters to a field study, monitoring bats over a three month period visiting a remote cave system in northern England. Chapter 6 applies automated classification of acoustically cryptic *Myotis* bats to provide presence data for species distribution modelling in the Lake District National Park. Chapter seven discusses the implications of our developments and findings, and reviews the usefulness of automated acoustic methods as a bat research and conservation tool.

1.8 References

Armitage, D. W., & Ober, H. K. (2010). A comparison of supervised learning techniques in the classification of bat echolocation calls. *Ecological Informatics*, *5*(6), 465-473.

Ahlen, I., & Baagøe, 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.

Berry, N., O'Connor, W., Holderied, M. W., & Jones, G. (2004). Detection and avoidance of harp traps by echolocating bats. *Acta chiropterologica*, 6(2), 335–346.

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, *16*(3), 199–215. JSTOR.

Burnett, S.C., & Masters, W.M. (1999). The use of neural networks to classify echolocation calls of bats. *Journal of the Acoustical Society of America*, *106*, 2189.

Dawkins, R. (1986). The Blind Watchmaker. London: Penguin.

Emde, G., & Schnitzler, H.-U. (1990). Classification of insects by echolocating greater horseshoe bats. *Journal of Comparative Physiology, 167*(3), 423-430.

Fenton, M. B. (2003). Eavesdropping on the echolocation and social calls of bats. *Mammal Review*, *33*(3-4), 193-204.

Fenton, M. B., & Bell, G. P. (1981). Recognition of species of insectivorous bats by their echolocation calls. *Journal of Mammalogy*, *62*(2), 233–243. JSTOR.

Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Engineers-Part III: Radio and Communication.*

Graps, A. (1995). An introduction to wavelets. *IEEE Computational Science and Engineering*, *2*(2), 50-61.

Holderied, M. W., Baker, C. J., Vespe, M., & Jones, G. (2008). Understanding signal design during the pursuit of aerial insects by echolocating bats: tools and applications. *Integrative and comparative biology*, *48*(1), 74-84.

Janert, P. K. (2011). Data analysis using open source tools. O'Reilly.

Jones, G. (1995). Flight performance, echolocation and foraging behaviour in noctule bats Nyctalus noctula. *Journal of Zoology*, 303-312.

Jones, G. (2005). Echolocation. *Current Biology*, *15*(13), 484-488.

Jones, G., & Holderied, M. W. (2007). Bat echolocation calls: adaptation and convergent evolution. *Proceedings. Biological sciences / The Royal Society*, *274*(1612), 905-12.

Jones, G., Jacobs, D., Kunz, T., Willig, M., & Racey, P. (2009). Carpe noctem: the importance of bats as bioindicators. *Endangered Species Research*, *8*(July), 93-115.

Korine, C., & Kalko, E. K. V. (2005). Fruit detection and discrimination by small fruit-eating bats (Phyllostomidae): echolocation call design and olfaction. *Behavioral Ecology and Sociobiology*, *59*(1), 12-23.

Lawrence, B. D., & Simmons, J. a. (1982). Measurements of atmospheric attenuation at ultrasonic frequencies and the significance for echolocation by bats. *The Journal of the Acoustical Society of America*, *71*(3), 585-90.

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.

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.

Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine learning methods without tears: a primer for ecologists. *The Quarterly review of biology*, *83*(2), 171-93.

Parsons, S., & Jones, G. (2000). Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. *J. Exp. Biol*, *203*, 2641-56.

Parsons, S., Boonman, A. M., & Obrist, M. K. (2000). Advantages and Disadvantages of Techniques for Transforming and Analyzing Chiropteran Echolocation Calls. *American Society of Mammalogists*, *81*(4), 927-938.

Quinn, B. G. (1994). Estimating frequency by interpolation using Fourier coefficients. *Signal Processing, IEEE Transactions on*, *42*(5), 1264–1268. IEEE.

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.

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.

Schnitzler, H.-U., & Flieger, E. (1983). Detection of oscillating target movements by echolocation in the greater horseshoe bat. *Journal of Comparative Physiology*, *153*, 385-391.

Schnitzler, H.-U., & Kalko, E. K. V. (2001). Echolocation by Insect-Eating Bats. *BioScience*, *51*(7), 557.

Schnitzler, H. U., Moss, C. F., & Denzinger, A. (2003). From spatial orientation to food acquisition in echolocating bats. *Trends in Ecology & Evolution*, *18*(8), 386–394. Elsevier.

Siemers, B., & Schnitzler, H.-U. (2004). Echolocation signals reflect niche differentiation in five sympatric congeneric bat species. *Nature*, *429*(June), 657-661.

Simon, R., Radmacher, M. D., Dobbin, K., & McShane, L. M. (2003). Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *Journal of the National Cancer Institute*, *95*(1), 14-8.

Skowronski, M. D., & Harris, J. G. (2006). Acoustic detection and classification of microchiroptera using machine learning: lessons learned from automatic speech recognition. *J. Acoust. Soc. of Am*, *119*, 1817-1833.

Surlykke, A., Boel Pedersen, S., & Jakobsen, L. (2009). Echolocating bats emit a highly directional sonar sound beam in the field. *Proceedings. Biological sciences / The Royal Society*, *276*(1658), 853-60.

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*, *7*, 91.

Vaughan, N., Jones, G., & Harris, S. (1997). Identification of British bat species by multivariate analysis of echolocation call parameters. Bioacoustics, 7, 189-207.

Yovel, Y., Melcon, M. L., Franz, M. O., Denzinger, A., & Schnitzler, H.-U. (2009). The voice of bats: how greater mouse-eared bats recognize individuals based on their echolocation calls. *PLoS computational biology*, *5*(6), e1000400.

Chapter 2: Automated signal detection

2.1 Abstract

Acoustic methods are an invaluable tool in the study of bat ecology and in conservation work. Reliably detecting echolocation calls in bat detector recordings is a vital first stage in developing a fully automated species identification system, which would reduce the burden of manual analysis. We developed an algorithm designed to locate echolocation calls in continuous recordings. The algorithm automatically estimates and subtracts the background noise in the recording to increase the sensitivity of detection, and improve the robustness of the detection threshold to varying signal levels. The location of echolocation calls in field recordings were hand-labelled, providing ground truth data for evaluating the accuracy of automated call finding algorithms. By comparing the location of calls found by the automated algorithm to the hand-labelled ground truth data, we established the good detection rate (calls correctly detected) and error rates (missed calls and false detections caused by noise). We compared our algorithm to two conventional bioacoustic signal detection algorithms. Our algorithm achieved a high percentage of good detections (97.6%), with few false detections (1.8%), outperforming the conventional detectors. Our approach is more robust theoretically and practically.

2.2 Introduction

Since the introduction of portable bat detectors, enabling researchers to eavesdrop on the echolocation calls of bats, acoustic methods have become an invaluable tool in the study of bat ecology and in conservation work (Ahlen & Baagøe, 1999). As nocturnal fast flyers, bats are a difficult group to survey; visual identification in flight is rarely possible (Walsh & Harris, 1996), and capture requires skill, is labour-intensive, biased (e.g. MacSwiney *et al.*, 2008) and disturbs natural behaviour. As the echolocation calls of bats are readily detectable using portable ultrasonic detectors, they facilitate the non-invasive, rapid survey of bats in their natural habitats. Acoustic surveys have a wide range of applications, from compiling species inventories and assessing patterns of habitat use (e.g. Krusic *et al.*, 1996; Vaughan *et al.*, 1997; Russo & Jones, 2002, 2003; Davy *et al.*, 2007; Rebelo & Rainho, 2009; Webala *et al.*, 2011), to discriminating morphologically cryptic species (Jones & Parijs, 1993; Helversen *et al.*, 2001; Ramasindrazana *et al.*, 2011). While qualitative identification, based on expert opinion, remains a useful technique for identification of at least some species in the field, particularly in regions with few species, identification by quantitative analysis of recordings is repeatable, and not subject to researcher bias. It therefore has to be the method of choice for scientific research and survey, and monitoring for conservation.

There are two principle tasks involved in the bioacoustic analysis of recordings: signal detection and signal classification. Signal detection is the localisation of signals of interest in continuous recordings, e.g. bat echolocation calls. Signal classification labels signals into biologically relevant groups, for example to genus or species level. Quantitative acoustic identification has been applied to a wide range of animals, from birds (e.g. Peake & McGregor, 2001), to marine mammals (e.g. Mellinger & Clark, 2000; Yack *et al.* 2010), and insects (Mankin *et al.*, 2011). Classification of bat echolocation calls has been carried out using a variety of approaches, including multivariate statistical analysis (Vaughan *et al.*, 1997; Papadatou *et al.*, 2008), hidden Markov models (Skowronski & Harris, 2006), synergetic pattern recognition (Obrist *et al.*, 2004), artificial neural networks (Parsons & Jones, 2000), support vector machines (Redgwell *et al.*, 2009), and random forests (Armitage & Ober, 2010). However, despite the wealth of research on signal classification in the last two decades, the manual detection and extraction of calls from recordings, through the time-consuming visual inspection of spectrograms, remains commonplace in conservation work and research.

There is an increasing need to automate the process of signal detection, as developments in hardware and digital storage now facilitate unprecedented data collection. Remote loggers are able to operate for extended periods storing many hours of audio under field conditions (e.g. Pettersson D500X; http://www.batsound.com/). The use of automated loggers increases the scope of field studies, but the bottleneck of manual analysis effectively necessitates the use of automated methods of signal detection and extraction. As bat echolocation calls are pulses of energy, an intuitive way to detect calls is through measuring the changes in energy throughout the recorded audio file. Simple energy based methods have successfully been applied to extract and analyse a single call from a recorded 'bat pass' (e.g. Parsons & Jones, 2000; Redgwell *et al.*, 2009). However, it is desirable to develop methods suitable for detecting all calls in a recorded sequence, maximising the use of available data. In addition, the automated methods of Parsons and Jones (2000) had to be abandoned in a subsequent study, as the calls were of lower signal to noise ratio (SNR), and the methods did not prove robust (Jennings *et al.*, 2008). Recording bats under controlled conditions or where flight paths can be estimated (e.g. on emergence from roosts), can result in high quality calls with good SNR (strong recorded signal with quiet background noise), as bats may be <2m from the detector at the time of recording. In contrast, under field conditions bats are frequently recorded at much greater distances, resulting in lower SNR (weaker recorded signal with relatively higher levels of background noise). In addition, field recordings may be corrupted by varying levels of abiotic noise (for example, caused by wind or flowing water), and echoes may be recorded as a result of ground reflected calls received at the microphone.

Reliably determining the location of bat echolocation calls algorithmically in field recordings is not trivial, as calls may have low SNR and call echoes can easily be mistaken as calls. Sophisticated statistical model-based methods have been proposed (e.g. Skowronski & Fenton, 2008), but free and publicly available implementations remain elusive. Published mathematical descriptions of algorithms, if they are not supported by the code used to implement them, may not be easy to implement without significant guesswork and interpretation by others (Thimbleby, 2003). This makes it difficult to reproduce and build upon the work of others, creating a significant barrier to new researchers in the field, and restricting the widespread adoption of proposed methods.

More recently, commercial programs that automatically extract echolocation calls from bat detector recordings, and classify them to species have become available (e.g. ecoObs *bcAdmin,* http://www.ecoobs.com/; Elekon AG *BatExplorer,* http://www.elekon.ch/en/batlogger/home/; *SonoBat 3;* http://www.sonobat.com/SonoBat3.html). However, the high cost and current lack of scientific evaluation may restrict the adoption of commercial solutions by many. There remains a dearth of freely available automated software tools for bat researchers. This is in contrast to the field of marine bioacoustics, which has benefited from free and open source software tools specifically designed for detecting and classifying marine mammal vocalisations (e.g. PAMGUARD; Gillespie *et al.*, 2008). The open source nature of these tools allows other researchers to learn from, and adapt the code used to implement them. Furthermore, the openness and transparency of making code available encourages peer review, which may lead to improvements in methods (e.g. Barnes, 2010).

We developed a robust algorithm to locate bat echolocation calls in continuous recordings, using open source tools, as part of the ongoing development of methods for automated analysis. We describe the algorithm and its implementation, documenting the C++ source code. Our algorithm is evaluated on a dataset of field recordings with hand-labelled echolocation calls. Hand-labelling the location of calls allows us to compare the locations returned by the automated algorithm with the hand-labelled ground truth location of calls. The detection rate, error rates, and computational cost of our algorithm is compared to two conventional methods of bioacoustic signal detection described in the literature.

2.3 Methods

2.3.1 Signal detection algorithms

A recorded ‘bat pass’ contains broadband continuous background noise, and one or more discrete echolocation calls, which are pulses of acoustic energy (Figure 2.1). Here we define any discrete portion of a recorded signal above the level of background noise as an acoustic event. Under this definition, an acoustic event may include an echolocation call, a call echo and other noise sources, either biotic (e.g. stridulating insects) or abiotic (flowing water, rain, wind-induced vegetation noise).

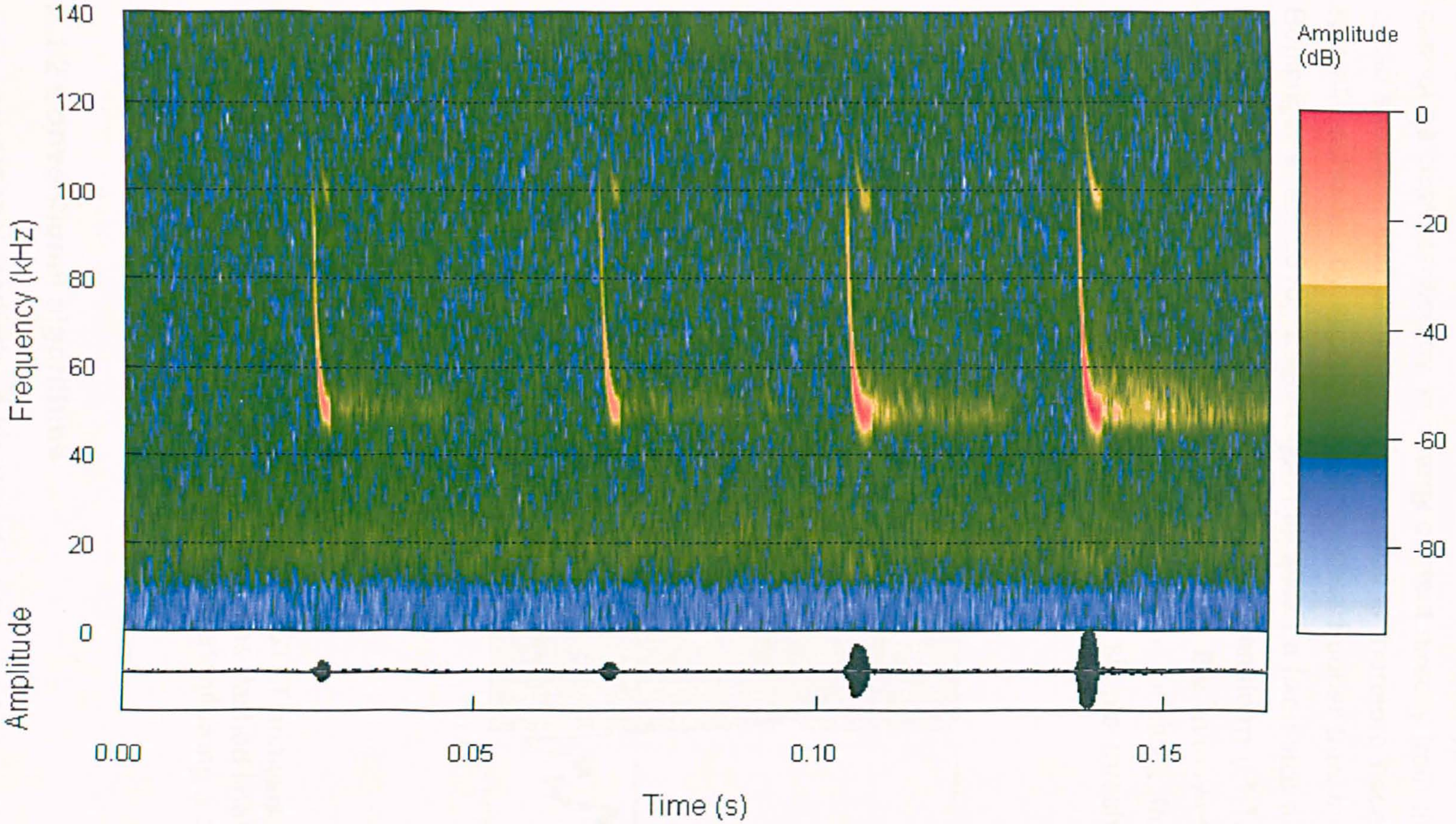


Figure 2.1: Example bat pass. Spectrogram (above) and waveform (below). Warmer colours represent areas of higher energy.

Acoustic events can be detected by first characterising a recording by its energy content, and defining a threshold rule for selecting events containing energy above the threshold. As audio is oscillatory, a recorded signal cannot be characterised by its energy content directly from the time domain audio samples, and an intermediate signal is therefore necessary. This is typically achieved by applying the short-time Fourier transform (STFT; e.g. Boulanger *et al.*, 2010), a technique that slides a fixed size analysis window through the signal, applying the fast Fourier transform (FFT) to reveal the spectral energy in the signal. The output from the windowed analysis is a discrete set of values forming the detection function, from which the locations of events can be identified through simple thresholding (Figure 2.2).

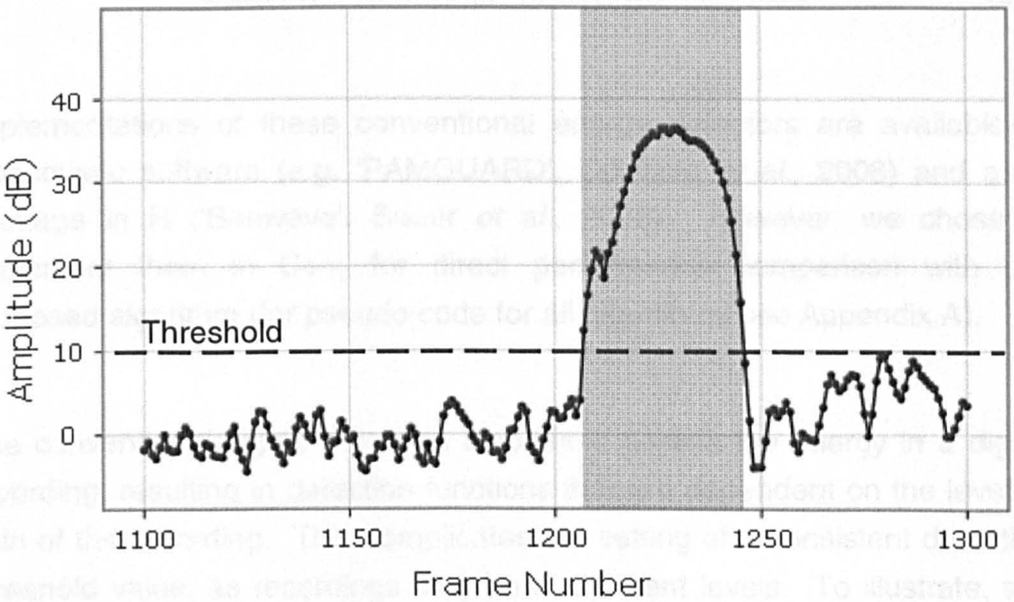


Figure 2.2: Example detection function

Plot of amplitude (dB) as a function of the STFT analysis frame number. A threshold set at 10 dB (shown as dashed line) results in a single detected event (grey highlighted area) following a simple threshold procedure.

2.3.2 Conventional algorithms

Two conventional methods of bioacoustic signal detection are the spectral sum, and spectral peak algorithms. For a signal x at time n, we define $X[n]$ as its STFT, where $|X_k[n]|$ is the spectral magnitude of the k^{th} FFT bin at n . The spectral sum algorithm calculates the sum of the STFT magnitudes (the total energy over the entire spectrum) at each consecutive window through the signal to create the detection function:

$$D_{sum}[n] = \sum_{k=0}^{N} |X_k[n]| \tag{2.1}$$

The spectral peak detector is similar in design, but uses the peak spectral magnitude from each analysis window, in contrast to the sum of all spectral magnitudes:

$$D_{peak}[n] = max\{|X_0[n]|, |X_1[n]|, \ldots, |X_N[n]|\} \tag{2.2}$$

Implementations of these conventional energy detectors are available in bioacoustic software (e.g. 'PAMGUARD', Gillespie *et al.*, 2008) and as a package in R ('Seewave'; Sueur *et al.*, 2008). However, we chose to implement them in C++, for direct performance comparison with our proposed algorithm (for pseudo code for all algorithms see Appendix A).

The conventional signal detection algorithms assess the energy in a digital recording, resulting in detection functions that are dependent on the level or gain of the recording. This complicates the setting of a consistent detection threshold value, as recordings may be at different levels. To illustrate, two copies of the same recording, one normalised to increase its level without changing the content, would require different threshold levels to detect the same acoustic events. In this study we normalised the detection function for each individual recording, by subtracting the median value over all analysis windows of a recording from each detection function data point (e.g. Skowronski & Fenton, 2009). The median value is an estimate of the noise floor of the recording, and the process of median offsetting allows the use of a fixed threshold parameter, that is then independent of the recording level. The process of normalisation requires that the entire signal must be acquired prior to processing, ruling out real time operation. A simple threshold algorithm selects candidate call locations from the detection functions normalised output. The threshold algorithm works by first marking the location at which the detection function crosses the trigger threshold level (in dB). From this point, the location of the maximum level of the detection function is tracked. On the detection function subsequently falling below the threshold, the location of the maximum level of the detection function whilst over the threshold is stored as a candidate acoustic event. Candidate events are subsequently filtered using the following rule: if the duration of the detected event is less than 1.5 milliseconds (ms) it is removed. This duration threshold is set to help remove spurious detections caused by transient noise.

2.3.3 Proposed noise subtraction algorithm

We developed a call detection algorithm to overcome some of the shortcomings of the conventional methods outlined above. Our algorithm is designed to estimate and remove the noise floor from recordings using only past values of the signal. By estimating the noise floor of the recording from a fixed window of previous analysis frames, *a priori* knowledge of the signal is not required for detection function normalisation, and real time operation remains a possibility. In addition, by estimating the noise floor locally, the algorithm can dynamically react to changes in the signal within a recording. We estimate and subtract the noise floor independently for each spectral bin of the FFT spectrum. Environmental noise and microphone self noise is rarely white in nature (equal power at all frequencies), and is typically weighted more heavily at the low end of the frequency spectrum (Figure 2.3). Frequency-specific noise subtraction can attenuate the noisier low frequency regions of the spectrum more heavily than the higher frequency and lower noise parts. This process increases the sensitivity of the call detection at higher frequency regions of the spectrum, as signals in those regions consequently have a higher SNR. We also employ a temporal mask to reduce the influence of call echoes on the detection function: an exponential decay curve is applied to the output of the detection function, which acts as an adaptive threshold. Echoes falling below the threshold do not contribute to the detection function as they are masked by the louder preceding call. The exponential decay curve is defined as:

$$F[n] = \max(\mathrm{D}[\mathrm{n}], \alpha \cdot \mathrm{F}[\mathrm{n}-1] + (1-\alpha) \cdot \mathrm{D}[\mathrm{n}]) \qquad (2.3)$$

where $F[n]$ is the threshold function, $D[n]$ is the detection function and α is the exponential decay factor (Dixon, 2006). This echo suppression is based on psychoacoustic principles (Moore, 1993), and aims to reduce the false alarms caused by echoes exceeding the energy threshold that triggers the detector. We experimented with various values of the α parameter, and found values >0.8 worked well in practice. When evaluating the algorithm in this study we fixed the parameter at a value of 0.9.

The detection function is generated by first summing all spectral magnitudes in frequency bands that are greater than both their local median values, and the temporal masking threshold. This value is considered to be the signal content. A noise estimate is then taken as the sum of all local median values. Finally, the detection function is expressed as a signal to noise ratio (SNR) in dB as:

$$SNR = 20 \cdot \log_{10}\left(\frac{signal}{noise}\right) \tag{2.4}$$

In this way, the proposed algorithm can be considered a modified spectral sum algorithm with background noise reduction and echo suppression.

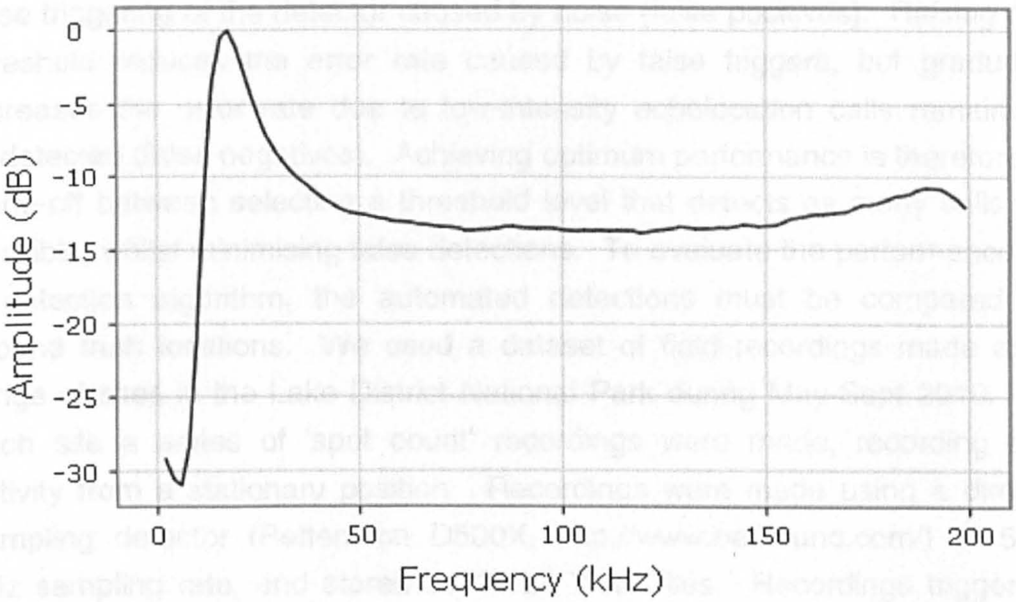


Figure 2.3: Ultrasound recording noise floor.

Power spectrum for Pettersson D500X detector noise floor, showing nonlinear frequency weighting. Calculated using a window size of 256-points and a window overlap of 75%. An inbuilt high pass filter on the D500X has strongly filtered frequencies below 15 kHz.

2.3.4 Implementation

Both conventional signal detection algorithms and our noise subtraction algorithm use a FFT window size of 256-points, with a Blackman Harris 4-term window to reduce spectral leakage (Harris, 1978). Larger FFT windows produce finer frequency resolution (an increased number of FFT bins, each with a narrower frequency span), but increase computation time and reduce temporal resolution due to Gabor's uncertainty principle (Gabor, 1946). All algorithms were implemented in C++ (for pseudo code see Appendix A), and used the FFTW (http://www.fftw.org/) and libsndfile (http://www.mega-nerd.com/libsndfile/) libraries, for Fourier transforms and sound file loading respectively.

2.3.5 Evaluation

An ideal call finding algorithm would detect all calls without falsely triggering due to noise. In practice, all calls can be detected simply by using a very low trigger threshold. However, this would result in a high error rate due to false triggering of the detector caused by noise (false positives). Raising the threshold reduces the error rate caused by false triggers, but gradually increases the error rate due to low-intensity echolocation calls remaining undetected (false negatives). Achieving optimum performance is therefore a trade-off between selecting a threshold level that detects as many calls as possible, whilst minimising false detections. To evaluate the performance of a detection algorithm, the automated detections must be compared to ground truth locations. We used a dataset of field recordings made at a range of sites in the Lake District National Park during May-Sept 2010. At each site a series of 'spot count' recordings were made, recording bat activity from a stationary position. Recordings were made using a direct-sampling detector (Pettersson D500X, http://www.batsound.com/) at 500 kHz sampling rate, and stored as 16 bit WAV files. Recordings triggered automatically on detecting ultrasound and recorded for a period of 1 sec, after which time the detector was ready to trigger again. We determined the start and end points of echolocation calls in recordings by visual inspection of waveforms and spectrograms using Adobe Audition (http://www.adobe.com/products/audition.html). The times of call start and end points were recorded as the number of samples through the recording in a CSV text file. Calls were classified to genus level using the overall shape of calls, and published descriptions of call frequencies (Vaughan, Jones & Harris, 1997; Parsons & Jones, 2000) to categorise recorded sequences by call type.

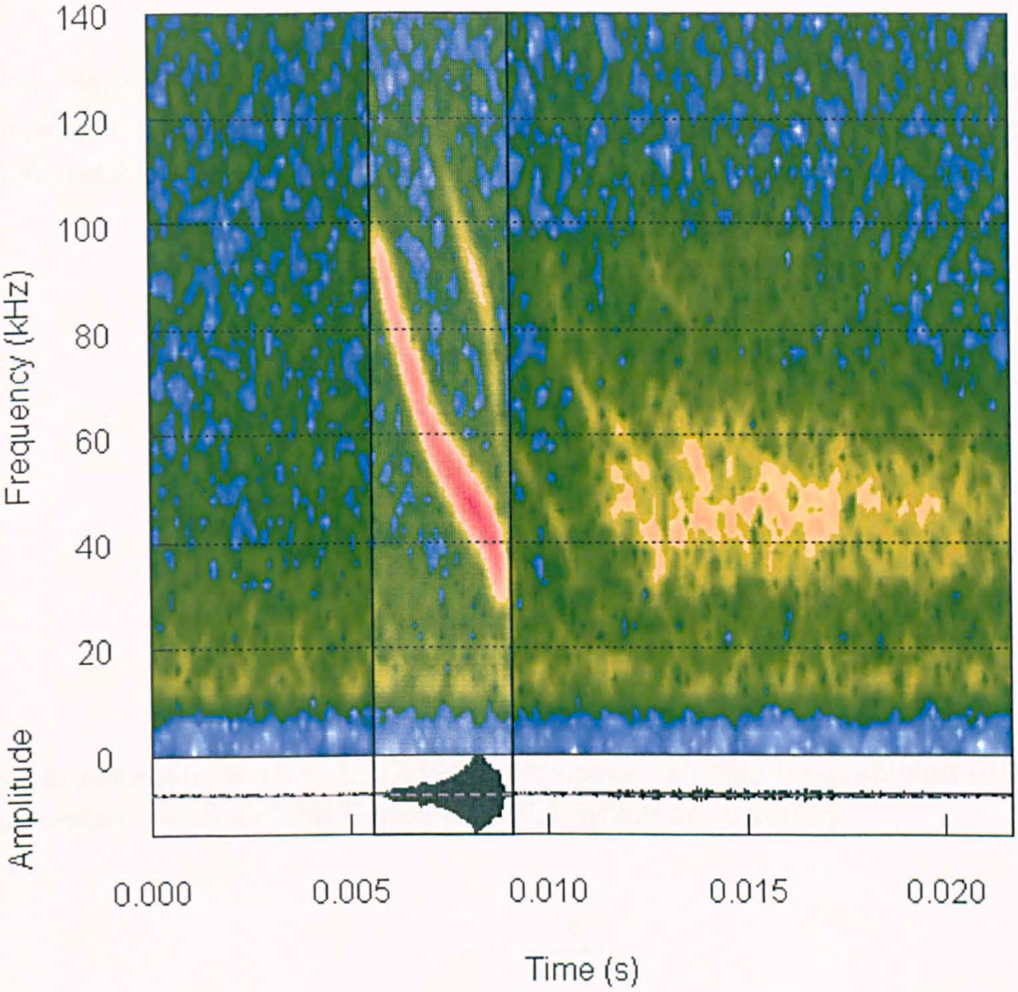


Figure 2.4: Spectrogram and waveform view illustrating a single hand-labelled call location.

The boxed area marks the duration of the echolocation call. The detection is only considered correct if it falls within the boxed area. Detections outside the boxed area, e.g. in the following echo, count as false positives. Failure to detect the call in the boxed region results in a false negative.

We considered a call to be correctly detected if the returned call location fell within the ground truth call start and end times. We did not penalise doubled detections (a single call recognised as two), since the purpose of automated detection was subsequent call extraction, and detections that overlap with previously extracted calls can simply be disregarded. An undetected call represents a false negative, and false detections caused by noise represent false positives.

The algorithms were evaluated using four statistics: good detections (correctly detected calls; GD), false positives (false alarms caused by noise; FP) and false negatives (missed calls; FN), which are given by:

$$GD = \frac{c}{c + fn} \tag{2.5}$$

$$FP = \frac{fp}{c + fn} \tag{2.6}$$

$$FN = \frac{fn}{c + fn} \tag{2.7}$$

A perfect score is $GD = 1$. Detection success can also be evaluated using precision, recall and the F-measure (F_N), which are given by:

$$P = \frac{c}{c + fp} \tag{2.8}$$

$$R = \frac{c}{c + fn} \tag{2.9}$$

$$F_N = \frac{(1 + N^2) \cdot P \cdot R}{N^2 \cdot P + R} \tag{2.10}$$

Note that the good detection rate (GD) is identical to the recall (R). Precision (P) compares the number of correctly detected calls to the number of automated detections, rather than the number of hand-labelled calls. Here we use the F_1-measure, and refer to it as the F-measure. The F-measure is the weighted harmonic mean of precision and recall, and represents the optimal point on the receiver operator characteristic (ROC) curve (van Rijsbergen, 1979). The F-measure is a useful single statistic to optimise algorithm parameters (e.g. Hoffmann *et al.*, 2001).

We evaluated the automated detection algorithms described above in two separate optimisation and testing experiments. Each algorithm was evaluated using 50 % of the input dataset stratified by genus/call type, using a range of trigger threshold values to find the optimum performance as measured by the F-measure statistic. Each detector was run with threshold values ranging from 0.0 to 60.0 dB in 0.25 dB steps. The trigger threshold producing the best performance was identified for each algorithm as the peak F-measure obtained over all threshold values. The remaining 50% of the dataset was then used as an independent test set, to obtain an unbiased estimate of algorithm performance using a single optimal threshold value.

The computational costs of the different automated algorithms in our implementation were estimated by calculating the time required to analyse 1 minute of 500 kHz audio. Low computational cost is important if real time operation is desirable, but also in offline processing applications where large datasets require analysing. Benchmarks were run on a 2.4 GHz laptop running Windows Vista.

2.4 Results

The database of field recordings was used to assess the signal detection algorithms' performance. It consists of 886 hand-labelled calls, in 102 recorded sequences (Table 2.1). The collection of recordings contains calls from *Myotis*, *Nyctalus* and *Pipistrellus spp.* (Figure 2.5).

Table 2.1: Summary of evaluation dataset.

Number of recorded sequences and individual hand-labelled calls, grouped by genus.

	Sequences	Calls
Myotis	32	362
Nyctalus	26	130
Pipistrellus	44	394
Total	102	886

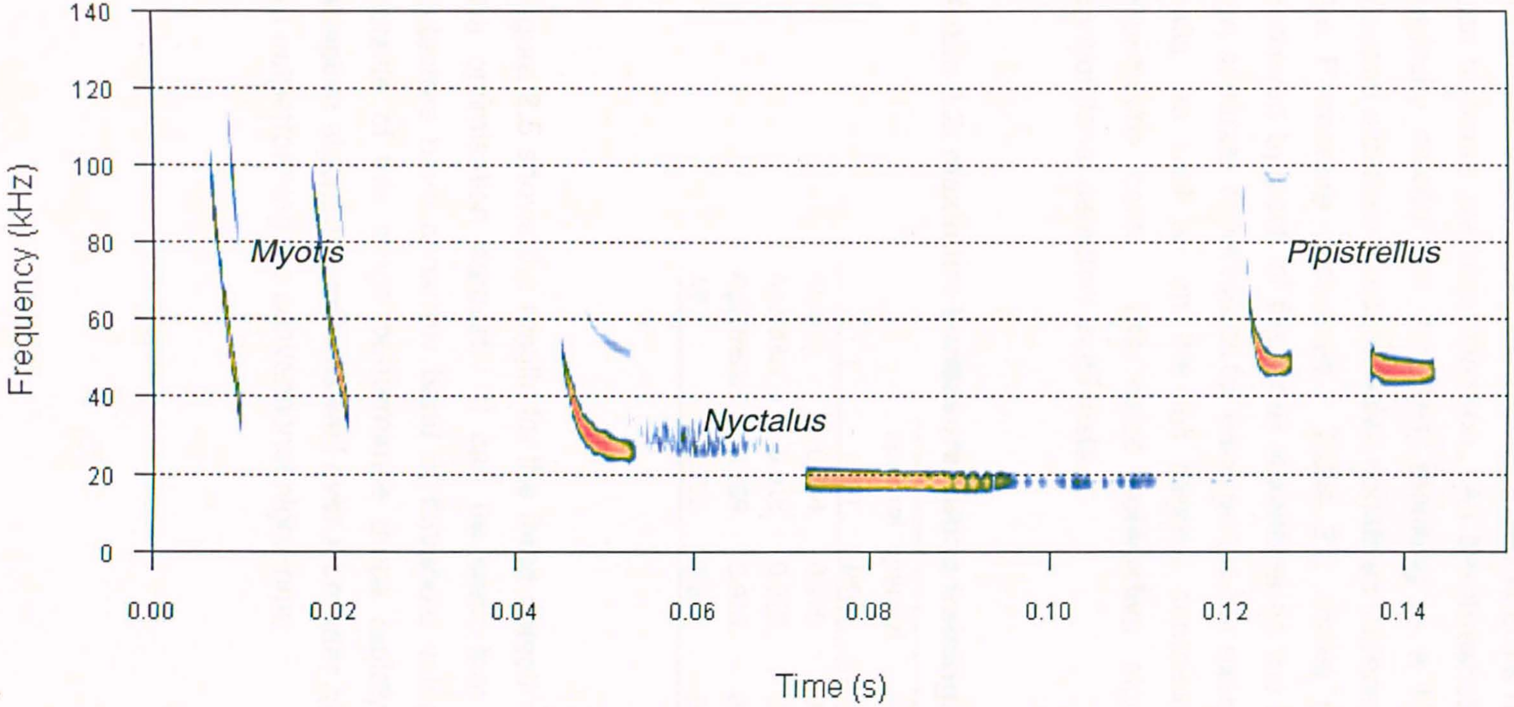


Figure 2.5: Spectrogram of call types.

Example call types for *Myotis*, *Nyctalus*, and *Pipistrellus spp*. within the evaluation dataset.

2.4.1 Threshold optimisation

The threshold setting influences the numbers of calls detected and the numbers of false positives and negatives, reflected in the F-measure. At low thresholds the number of calls detected is at its highest, but false positives due to noise are also common. As the threshold is increased, fewer low intensity noises are detected, resulting in a higher overall F-measure. Beyond a certain threshold false negatives increase as calls are missed, and the F-measure decreases. Table 2.2 shows the maximum F-measure achieved by each of the three algorithms on the optimisation dataset. We ran separate experiments for each genus, for example detecting only *Myotis* calls, as well as on the full dataset containing *Myotis*, *Nyctalus* and *Pipistrellus* calls. The noise subtractive algorithm outperforms both conventional detectors in all tests.

Table 2.2: maximum F-measure during training.

	spectral sum	spectral peak	noise subtractive
Myotis	0.734	0.796	0.977
Nyctalus	0.918	0.929	0.962
Pipistrellus	0.969	0.974	0.995
All	0.873	0.851	0.989

Figure 2.5 shows the results for the three detection algorithms evaluated on the optimisation dataset. It can be seen that the conventional energy detectors have a narrow band of threshold values that perform well, and outside of this range performance drops rapidly. In contrast, the noise adaptive algorithm performs well over a broader range of thresholds, as well as outperforming the conventional algorithms.

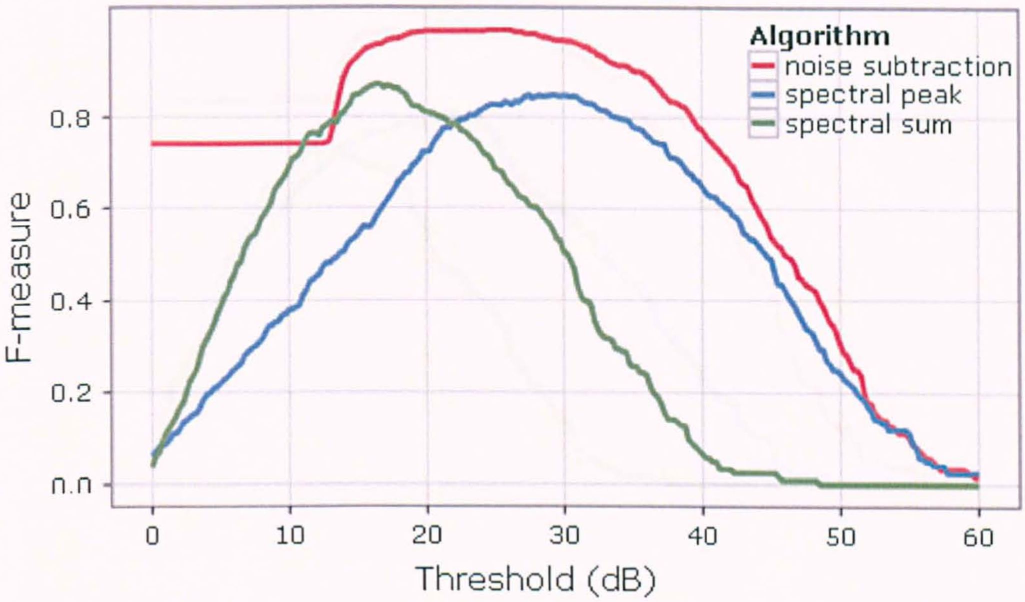


Figure 2.5: Effect of the threshold on the F-measure.

Results obtained on the mixed species optimisation dataset for each algorithm.

The individual results for *Myotis*, *Nyctalus* and *Pipistrellus spp.* as plotted in Figures 2.6 to 2.8 reveal that for the conventional detectors, the optimum threshold is a narrow band, with the peak located at different thresholds for each group. In contrast, the noise subtraction algorithm is optimal over a broad range of threshold values, and these regions are largely similar for each call type.

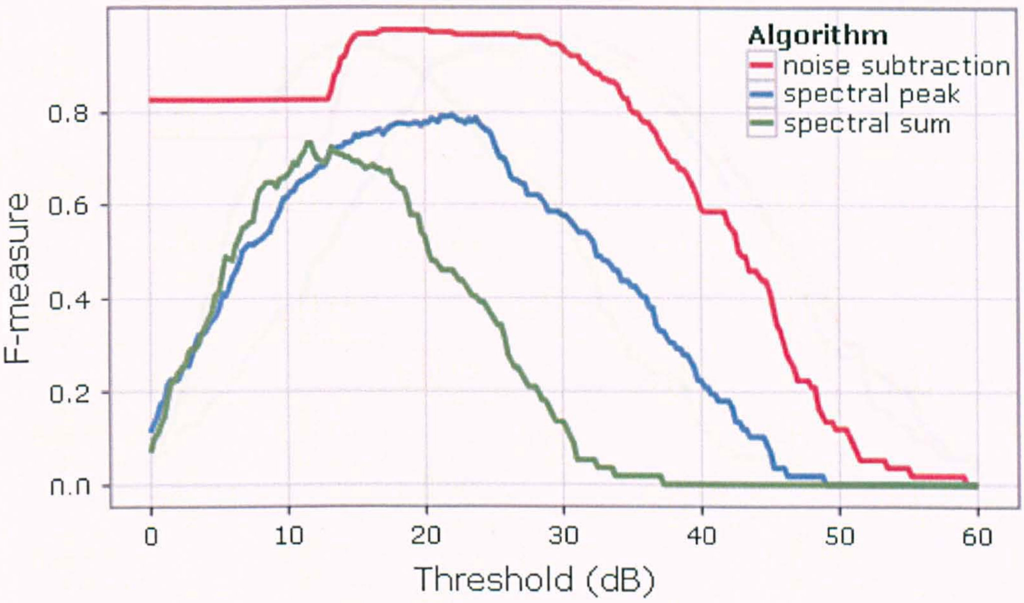


Figure 2.5: Effect of the threshold on the F-measure for *Myotis*.

Results obtained for the *Myotis spp.* in the optimisation dataset for each algorithm.

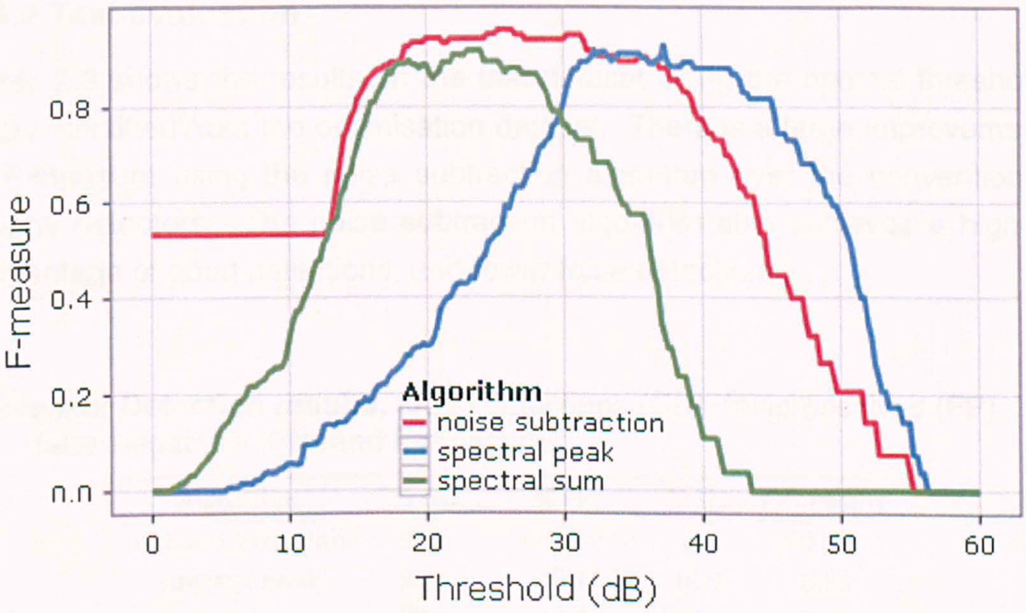


Figure 2.6: Effect of the threshold on the F-measure for *Nyctalus*.

Results obtained for the *Nyctalus spp.* in the optimisation dataset for each algorithm.

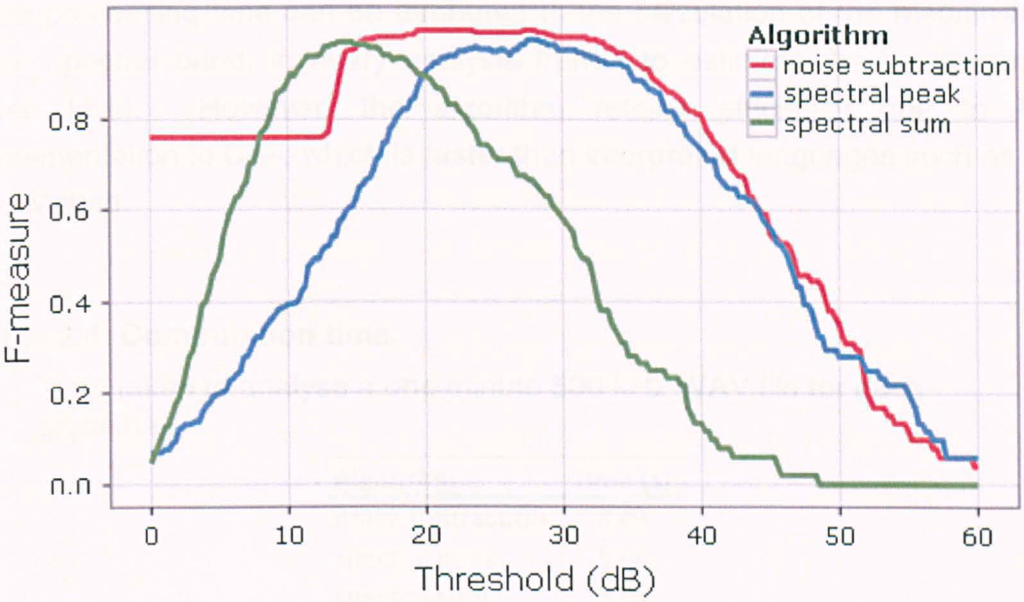


Figure 2.7: Effect of the threshold on the F-measure for *Pipistrellus*.

Results obtained for the *Pipistrellus spp.* in the optimisation dataset for each algorithm.

2.4.2 Test evaluation

Table 2.3 shows the results on the test dataset using the optimal threshold value identified from the optimisation dataset. There is a large improvement in F-measure using the noise subtraction algorithm over the conventional energy detectors. The noise subtraction algorithm also achieves a higher percentage of good detections, and fewer false detections.

Table 2.3: Detection results. Good detections (GD), false positives (FP), false negatives (FN) and F-measure.

Algorithm	% GD	% FP	% FN	F-measure
noise subtraction	97.6	1.8	2.4	0.98
spectral peak	80.1	10.4	19.9	0.85
spectral sum	83.1	14.8	16.9	0.83

Computation time is dependent on the hardware used to run the algorithms, as well as algorithm efficiency. All algorithms performed significantly faster than real time (Table 2.4). Our noise subtraction algorithm was the slowest of those evaluated, yet still more than 7 times faster than real time. The extra processing time can be attributed to the calculation of the median for each spectral band, in every analysis frame, to estimate the background noise level. However, the algorithm retains efficiency due to its implementation in C++, which is faster than interpreted languages such as R and Matlab.

Table 2.4: Computation time.

Time taken to analyse a one minute 500 kHz WAV file for each algorithm.

Algorithm	Time (s)
noise subtraction	8.49
spectral peak	3.68
spectral sum	3.68

2.5 Discussion

We developed a bioacoustic signal detection algorithm for the automated location of bat echolocation calls. Tests on a real-world dataset of field recordings confirmed that the new method outperformed two conventional approaches in terms of accuracy. The good detection rate was higher than both conventional approaches, indicating that subtracting the local noise floor estimate increased call detection sensitivity. In addition, the lower false positive rate can be attributed to the effectiveness of the temporal mask in reducing false triggers caused by call echoes. Our algorithm was the slowest evaluated, although implemented in C++, an efficient low-level language, it performed significantly faster than real time. Under field conditions, where levels of background noise may change, the robustness of our algorithm to the threshold setting represents a clear improvement on conventional methods. The fact that no prior knowledge of the signal is required for processing, offers the potential for the algorithm to be implemented in real time. This opens up the possibility of integrating the algorithm into suitable hardware, and developing a field portable identification system.

Comparison with previous call detection studies (e.g. Skowronski & Fenton, 2009) is not possible in the absence of a shared evaluation dataset. The difficulty of the detection task is affected by the quality of the recordings, the SNR of the recorded calls, and the type of noise sources present. As a result, fair comparisons can only be made between algorithms on the same dataset. The approach taken by Skowronski and Fenton (2009) was to generate synthetic bat calls and embed them in Gaussian noise. This gave them control over the SNR of calls, providing a precise evaluation of detector performance, and enabling the generation of a large dataset of synthesised recordings with known call locations for evaluation. However, they acknowledge that this approach fails to provide a truly real-world test for the detection algorithms, as the synthetic calls do not accurately match the variety and quality of signal types encountered under field conditions.

A balanced evaluation dataset should be representative of the real-world data that the detection algorithms are intended for, which is application specific. For example, an automated detection algorithm may be desired to process recordings of captive bats under controlled conditions, in which case a very simple algorithm may perform well, as recordings are likely to be high quality. In contrast, the same algorithm may perform poorly when applied to field recordings where the SNR of calls are lower. For evaluation, we used field recordings made during stationary spot counts, as they closely match our intended application of an automated system. Recordings made during walked transects frequently contain extraneous noises, as the detector records researcher-induced noise, for example the movement of vegetation. As a result, a higher false positive rate would be expected in applying automated detection. However, our dataset provided a reliable means to compare algorithms. Moreover, when call detection is used as a front-end for subsequent call extraction and classification, it is false negatives that should be minimised, as calls missed at the detection stage never reach the classifier. If the classifier is robust and can reliably discriminate echolocation calls from noise sources, then the cost of a high false positive rate is only one of efficiency, with more processing time spent extracting and classifying noise sources for rejection.

The main problem with the conventional signal detection algorithms is that they performed best at different threshold settings for different signal types. This complicates their effective use, as achieving good performance may require considerable manual analysis to determine a suitable threshold setting, undermining the objective of automated analysis. In addition, they require normalisation to account for possible differences in recording level, which means they cannot be implemented in real time. Furthermore, the approach does not scale well with recording length, as the entire detection function must be stored in memory for median offsetting, restricting their use to shorter recordings. In contrast, our noise subtraction algorithm performed well with all signal types across a broad range of thresholds, making it simpler to apply in practice where the recording content is not known *a priori*. As normalisation is applied in real time through local background noise subtraction, the length of the recording to be analysed has no effect on algorithm efficiency.

Implemented our algorithm in C++ ensured efficient performance, which is an important factor where large numbers of files need processing in long term studies. Using open source libraries for calculating the Fourier transform (FFTW, http://www.fftw.org/) and reading WAV files (libsndfile, http://www.mega-nerd.com/libsndfile/) increased the development time, freeing us to concentrate on other aspects of the development. The detection algorithm forms the front-end for a fully automated species identification system we are developing. As an open source project, we hope other researchers will benefit from our developments. Although beyond the scope of this study, the detection algorithm may prove useful for the bioacoustic signal detection of other taxa.

2.6 References

Ahlen, I., & Baagøe, 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.

Armitage, D. W., & Ober, H. K. (2010). A comparison of supervised learning techniques in the classification of bat echolocation calls. *Ecological Informatics*, *5*(6), 465-473.

Barnes, N. (2010). Publish your computer code: it is good enough. *Nature*, *467*(7317), 753.

Boulanger, R., Lazzarini, V., Mathews, M. (2010). The Audio Programming Book. MIT Press

Davy, C. M., Russo, D., & Fenton, M. B. (2007). Use of native woodlands and traditional olive groves by foraging bats on a Mediterranean island: consequences for conservation. *Journal of Zoology*, *273*(4), 397-405.

Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Engineers-Part III: Radio and Communication.*

Harris, F. J. (1978). On then Use of Windows with the Discrete for Harmonic Analysis Fourier Transform. *Proceedings of the IEEE*, *66*(1), 51-83.

Helversen, O. V., Heller, K.-G., Mayer, F., Nemeth, A., Volleth, M., & Gombkötö, P. (2001). Cryptic mammalian species: a new species of whiskered bat (Myotis alcathoe n. sp.) in Europe. *Naturwissenschaften*, *88*(5), 217-223.

Hoffmann, F. (2001). *Advances in Intelligent Data Analysis.* ISBN 3-540-42581-0

Hothorn, T., Hornik, K., Van De Wiel, M. A., & Zeileis, A. (2006). A lego system for conditional inference. *The American Statistician*, *60*(3), 257–263.

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.

Jones, G., & van Parijs, S. M. (1993). Bimodal echolocation in pipistrelle bats: are cryptic species present? *Proceedings. Biological sciences / The Royal Society*, *251*(1331), 119-25.

Krusic, R. A., Yamasaki, M., Neefus, C. D., & Pekins, P. J. (1996). Bat habitat use in white mountain national forest. *The Journal of wildlife management*, *60*(3), 625–631.

MacSwiney G., M. C., Clarke, F. M., & Racey, P. A. (2008). What you see is not what you get: the role of ultrasonic detectors in increasing inventory completeness in Neotropical bat assemblages. *Journal of Applied Ecology*, *45*(5), 1364-1371.

Mankin, R. W., Hagstrum, D. W., Smith, M. T., Roda, A. L., & Kairo, M. T. K. (2011). Perspective and Promise□: a Century of Insect Acoustic Detection and Monitoring. *American Entomologist*, 30-44.

McCauley, R., Jenner, C., & Bannister, J. (2000). Blue whale calling in the Rottnest trench, Western Australia, and low frequency sea noise. *Australian Acoustical Society Conference* (Vol. 1993, pp. 1-6).

Mellinger, D. K., & Clark, C. W. (2000). Recognizing transient low-frequency whale sounds by spectrogram correlation. *The Journal of the Acoustical Society of America*, *107*(6), 3518-29.

Menke, J., & Martinez, T. R. (2004). Using permutations instead of student's t distribution for p-values in paired-difference algorithm comparisons. *IEEE International Joint Conference on Neural Networks*, *2*, 1331-1335.

Moore, B. (1993). Characterization of Simultaneous, Forward and Backward Masking. *Perception of Reproduced Sound. Paper number 12-003.* Audio Engineering Society, New York.

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.

Papadatou, E., Butlin, R. K., & Altringham, J. D. (2008). Identification of bat species in Greece from their echolocation calls. *Acta Chiropterologica*, *10*(1), 127-143.

Parsons, S., & Jones, G. (2000). Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. *J. Exp. Biol*, *203*, 2641-56.

Peake, T., & McGregor, P. K. (2001). Corncrake Crex crex census estimates: a conservation application of vocal individuality. *Animal Biodiversity and Conservation*, *1*, 81-90.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Ramasindrazana, B., Goodman, S. M., Schoeman, M. C., & Appleton, B. (2011). Identification of cryptic species of Miniopterus bats (Chiroptera: Miniopteridae) from Madagascar and the Comoros using bioacoustics overlaid on molecular genetic and morphological characters. *Biological Journal of the Linnean Society, 104*, 284-302.

Rebelo, H., & Rainho, A. (2009). Bat conservation and large dams: spatial changes in habitat use caused by Europe's largest reservoir. *Endangered Species Research, 8*, 61-68.

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.

Roads, C. (1996). The Computer Music Tutorial. MIT Press.

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.

Russo, D., & Jones, G. (2003). Use of foraging habitats by bats in a Mediterranean area determined by acoustic surveys: conservation implications. *Ecography, 26*(2), 197-209.

Skowronski, M. D., & Fenton, M. B. (2008). Model-based automated detection of echolocation calls using the link detector. *The Journal of the Acoustical Society of America, 124*(1), 328-36.

Skowronski, M. D., & Fenton, M. B (2009). Detecting Bat Calls: An Analysis of Automated Methods. *Acta Chiropterologica, 11*(1), 191-203.

Skowronski, M. D., & Harris, J. G. (2006). Acoustic detection and classification of microchiroptera using machine learning: Lessons learned from automatic speech recognition. *Acoustical Society of America*, (March), 1817-1833.

Sueur J., Aubin T., Simonis C. (2008). Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics, 18*: 213-226

Swiston, K. a., & Mennill, D. J. (2009). Comparison of manual and automated methods for identifying target sounds in audio recordings of Pileated, Pale-billed, and putative Ivory-billed woodpeckers. *Journal of Field Ornithology, 80*(1), 42-50.

Thimbleby, H. (2003). Explaining code for publication. *Software: Practice and Experience, 33*(10), 975-1001.

Vaughan, N., Jones, G., & Harris, S. (1997). Habitat use by bats (Chiroptera) assessed by means of a broad-band acoustic method. *Journal of Applied Ecology*, *34*(3), 716-730.

Walsh, A. L., & Harris, S. (1996). Factors determining the abundance of vespertilionid bats in Britain: geographical, land class and local habitat relationships. *Journal of Applied Ecology*, *33*(3), 519–529.

Webala, P. W., Craig, M. D., Law, B. S., Armstrong, K. N., Wayne, A. F., & Bradley, J. S. (2011). Bat habitat use in logged jarrah eucalypt forests of south-western Australia. *Journal of Applied Ecology*, *48*(2), 398-406.

Yack, T. M., Barlow, J., Roch, M. A., Klinck, H., Martin, S., Mellinger, D. K., & Gillespie, D. (2010). Comparison of beaked whale detection algorithms. *Applied Acoustics*, *71*, 1043-1049.

Chapter 3: Acoustic identification of British bats from their echolocation calls

3.1 Abstract

This study describes a robust method for automatically extracting and measuring echolocation call features from calls embedded in continuous recordings, and represents the first use of random forest for the identification of bats from directly sampled recordings. Directly sampled echolocation calls from all genera of British bats were recorded and analysed. Temporal and spectral variables were automatically extracted from calls, and used to train a random forest classifier, a machine learning (ML) algorithm. In addition to four commonly measured 'base' call variables (call start frequency, end frequency, frequency at the point of maximum call energy, and total call duration), we included power spectra of calls as features for classification. A process of cross-validation was used to test the ability of classifiers to classify previously unseen data. Training and testing were carried out using sequences of recorded calls, and we report accuracies at both the call, and sequence level. An overall accuracy of 97.6% and 98.4% was achieved for classification to genus at the call and sequence level respectively, and 93.1% and 95.9% for classification of *Myotis* bats to species. We statistically compared the overall accuracy achieved by random forest classification of power spectra, to classification using the 'base' call features, and to quadratic discriminant function analysis (QDA), a traditional statistical classifier. Random forest had significantly higher accuracy to both genus and species level than QDA using the commonly measured temporal and spectral call features ($p<0.001$).

3.2 Introduction

The bioacoustic analysis of vocalisations has been applied to the identification of species in a wide range of taxa, and is of great conservation importance (for a review see Laiolo, 2010). The identification of bats from their echolocation calls is a vital tool in the study of their ecology and in conservation work. Phylogeny, ecology and morphology influence echolocation call signal structure, offering significant potential for discriminating between species on the basis of their calls. Some species are readily identified, because their calls occupy a niche in the frequency spectrum, e.g. *Rhinolphus hipposideros*, or because they emit a characteristic alternating call type, e.g. *Barbastella barbastellus*. However, for many species the situation is more complicated, and identification cannot necessarily be made with certainty. Whilst echolocation is influenced to some degree by phylogenetic constraints, the importance of environmental factors in shaping signal design has led to bats that feed in similar habitats evolving similar designs of echolocation calls, despite being distantly related (Jones & Holderied, 2007). Echolocation is functional, and bats change the structure of their calls in relation to situation. For example, vespertilionid bats flying in open areas use relatively longer duration calls with a narrower bandwidth, compared to the shorter and more broadband calls they emit in increasingly cluttered habitats (e.g. Schnitzler & Kalko, 2001). Echolocation calls may also vary due to the presence of conspecifics (Obrist, 1995), age and gender (Russo *et al.*, 2001).

In addition to the variation in echolocation calls caused by ecology and morphology, variability is introduced between the emission of a call and the signal that is received by the bat detector. The recorded signal is heavily influenced by both the distance and relative position of the bat to the bat detector. At increasing distances, high frequency attenuation caused by atmospheric absorption (Lawrence & Simmons, 1982) low pass filters the signal, resulting in a recorded call that has high frequency loss. Additionally, the frequency content of the received signal is affected by the relative angle of the bat to the bat detector. Echolocation calls are projected in front of the bat in a relatively narrow beam, and the bat detector is most sensitive to sounds arriving within a narrow arc in front of the microphone. Recorded calls from bats off-axis from the detector will therefore suffer some form of frequency-specific attenuation. Finally, echoes caused by ground reflection may add to the degradation of the recorded signal, and the overall result is that the recorded signal may not closely match the call actually emitted by the bat. For all but the most acoustically characteristic species, quantitative analysis that gives an objective measure of confidence in the identification is therefore preferred to the subjective judgement of human experts (Jones *et al.*, 2000).

The extraction of calls from recordings, and the subsequent measurement of call features, should be automated to fully remove any subjectivity from species identifications (Jones *et al.*, 2000). Automated methods remove the burden of hand measurement, a time-consuming process, and therefore costly in studies that generate significant amounts of data. They are also objective and repeatable, and not subject to researcher bias. Moreover, detailed call information can be extracted precisely, which may not be readily measurable by hand. However, the practical application of quantitative analysis of bat echolocation calls in field studies remains focused on the use of small numbers of hand measured call features, typically followed by traditional statistical classification techniques such as discriminant function analysis (DFA; e.g. Russo & Jones, 2003; Davy *et al.*, 2007; MacSwiney *et al.*, 2008; Georgiakakis *et al.*, 2010). This is presumably because there is a lack of viable alternatives. Parsons and Jones (2000) developed an automated method of call extraction and measurement, and used it to successfully classify twelve species of British bat from high quality echolocation call recordings. However, Jennings and colleagues (2008) found the proposed automated call measurement system was not robust to the low signal to noise ratios (SNRs) typical of calls encountered during field studies. There are significant technical challenges in tracking the harmonic with most energy in echolocation calls recorded under field conditions, as there may be significant low frequency noise, calls may be notched (heavily attenuated at specific frequencies), or they may suffer interference from echoes. Without the ability to track the harmonic with most energy, some of the call features that researchers have typically measured, e.g. the frequency at the start and end points of the call, cannot be accurately and robustly measured. However, the features that are most intuitive to researchers may not be the most important for discriminating between species. Modern non-linear machine learning classification algorithms such as artificial neural networks (ANN) and support vector machines (SVM) can handle large numbers of call features, whilst making few assumptions about the underlying data distributions, making them powerful tools relative to traditional statistical techniques like DFA. Parsons (2001) applied an ANN to the classification of two species of New Zealand bats, and found they could be unambiguously identified using traditional measured parameters, or using power spectra from time-expanded calls. A power spectrum describes the power of a call as a function of frequency, independent of temporal information. Similarly, Yovel and colleagues (2009) applied power spectra and SVM to the task of identifying individual greater mouse-eared bats (*Myotis myotis*), and achieved high levels of accuracy (81-90%).

In this study we consider direct sampling recordings, a method that retains more detail of recorded calls than other methods (Parsons *et al.*, 2000), and apply an automated call analysis algorithm to extract full spectrum call data. We apply random forest, a machine learning classifier, and statistically compare its ability to classify calls from British bat species.

3.3 Methods

3.3.1 Recording methods and call analysis

We recorded search-phase echolocation calls of bats (the calls emitted in free flight rather than immediately prior to insect capture; Griffin *et al.*, 1960). The majority of calls were recorded as bats left or returned to roosts, but at a distance of 10-20 m from the roost to ensure that normal search-phase calls were recorded. Where it was not possible to obtain recordings outside known roosts, calls were recorded from hand-released bats following capture under licence using harp traps (*M. bechsteinii*). Recordings were made at foraging sites where bats could be identified unambiguously (*B. barbastellus* and *Nyctalus noctula,* identified from their distinctive alternating call types). Recordings were made during 2009-2010 between May and October at a range of sites across the UK. Sites included the Lake District National Park (*M. brandtii/M. mystacinus, M. daubentonii, M. nattereri, N. noctula, Pipistrellus spp.*), the North York Moors (all *Myotis* species excluding *M. bechsteinii*), the Peak District National Park (*M. daubentonii*), the Yorkshire Dales National Park (*M. brandtii/M. mystacinus, M. daubentonii, M. nattereri, N. noctula, Pipistrellus spp.*), Dorset (*Eptesicus serotinus*), Herefordshire (*B. barbastellus*), Monmouthshire (*R. ferrumequinum* and *R. hipposideros*), Norfolk (*B. barbastellus, M. nattereri*), Sussex (*B. barbastellus, M. alcathoe, M. bechsteinii*). We obtained recordings from the seven genera of British bat species, *Barbastella, Eptesicus, Myotis, Nyctalus, Pipistrellus, Plecotus* and *Rhinolphus.* Of the bats known to breed in Britain, *P. nathusii, P. austriacus,* and *N. leisleri* were not recorded.

Recordings were made using a D500x direct sampling detector (Pettersson, www.batsound.com), and stored as 16 bit 500 kHz WAV files. Recordings of several seconds were made to capture a sequence of calls from each individual bat as it approached the detector. This procedure resulted in a large dataset of calls, capturing multiple calls per individual bat that varied in their characteristics. This variability was caused by behavioural differences: bats changing their call structure in relation to their environment, and acoustical differences: the distance and position of the bat relative to the bat detector microphone differed between emitted calls in flight, resulting in calls that varied in their signal to noise ratio (SNR), and degree of high frequency filtering due to atmospheric attenuation (Lawrence & Simmons, 1982). We pooled data from two morphologically cryptic species, *M. brandtii* (Brandt's bat) and *M. mystacinus* (whiskered bat), due to the difficulty in separating them on morphology alone (e.g. Berge, 2007). Small numbers of hand-released bats were subsequently confirmed genetically, but sample sizes are currently too small for comparison.

As features for classification we used the power spectrum of calls. The power spectrum is the square of the FFT magnitude, and describes the power of a signal as a function of frequency (see Figure 3.1). Power spectrums were calculated through the short time Fourier transform (STFT), sliding a 256 point window through recordings with an overlap between consecutive windows of 75%, resulting in a frequency resolution of 2 kHz and a time resolution of 0.13 ms. A 4-term Blackman Harris smoothing window was applied prior to the FFT (Harris, 1978). Background noise was estimated as the median spectrum of the previous 55 FFTs (6.91 ms), and subtracted from each analysis window. Individual calls were segmented as follows: the start of a call was estimated as the point at which a signal 12 dB above the background noise estimate was encountered, as measured by the total spectral magnitude above the background noise estimate. Subsequent analysis windows were summed for the generation of power spectra, and the end of the call was signalled by a drop in energy of more than 40 dB from the peak energy encountered during the call. Power spectra were taken from the first 0.75 ms of the call, and from the total call duration. Power spectra were normalised to have a maximum of 1 by dividing by their maximum value. Call duration was calculated as the time in ms between the point of triggering and the call end. Frequency at the point of maximum energy was calculated as the frequency containing the maximum power spectral value. Start and end frequencies were estimated from the initial and final power spectra respectively, as the frequency at the maximum power spectral value. We measured the call signal to noise ratio (SNR) taking the estimate of the signal as the average spectral energy of the call after background subtraction, and calculating the ratio to the background noise estimate. We excluded from our analysis triggered regions that had durations of less than 1.5 ms, or SNRs of less than 20 dB, to exclude small fragments of calls or feeding buzzes. The background subtraction protocol led to power spectra that were independent of the acoustic transfer function of the recording equipment, and the environmental noise present in the recording.

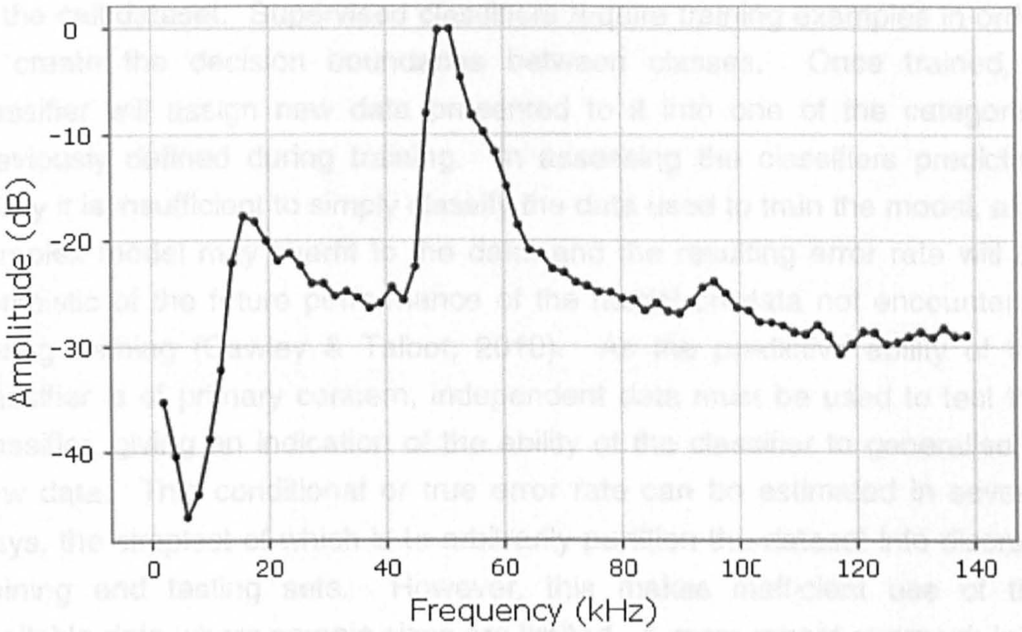


Figure 3.1: Example power spectrum for an echolocation call.

3.3.2 Classifiers

Quadratic discriminant function analysis (QDA) is a traditional statistical classification technique that has been widely applied to the classification of bat echolocation calls (e.g. Parsons & Jones, 2000; Russo & Jones, 2002; Preatoni *et al.*, 2005; Papadatou *et al.*, 2008; Armitage & Ober, 2010; Hughes *et al.*, 2010; Redgwell *et al.*, 2009). We used the implementation of QDA in the MASS package for R (Venables & Ripley, 2002).

Random forest (RF) is an ensemble learning method that combines multiple classification and regression trees (CARTs) using a process termed bagging (bootstrap aggregating). Bagging aims to increase the diversity of each tree by ensuring they grow from different subsets of the training data. In the prediction of new data, each tree contributes a single vote and the majority vote determines the final classification of the data (Breiman, 2001b). In this study, calls were classified using the randomForest R package (Liaw & Wiener, 2002). All RF models were built using the default parameters.

3.3.3 Model evaluation

The predictive ability of each classifier was evaluated using cross-validation of the call dataset. Supervised classifiers require training examples in order to create the decision boundaries between classes. Once trained, a classifier will assign new data presented to it into one of the categories previously defined during training. In assessing the classifiers predictive ability it is insufficient to simply classify the data used to train the model, as a complex model may overfit to the data, and the resulting error rate will be optimistic of the future performance of the model on data not encountered during training (Cawley & Talbot, 2010). As the predictive ability of the classifier is of primary concern, independent data must be used to test the classifier, giving an indication of the ability of the classifier to generalise to new data. This conditional or true error rate can be estimated in several ways, the simplest of which is to arbitrarily partition the dataset into discrete training and testing sets. However, this makes inefficient use of the available data where sample sizes are limited. A more recent approach is to use cross-validation. In this approach a portion of the dataset is removed or held out, the remainder is used to train the model, and the held out portion of data is then used as an independent test set. This process of training and testing models is repeated, each time holding out a different portion of the dataset until all the data have been used for testing. The results of all iterations are then averaged to provide an estimate of the true error rate. Whilst computationally intensive, cross-validation makes efficient use of the available data, and has been demonstrated to provide an almost unbiased estimate of the true error rate (Varma & Simon, 2006).

An important property of the test set in evaluating classifiers is that it is independent of the training data. In this study we measured call parameters from echolocation call sequences, containing multiple calls from each individual. As calls from the same individual are temporal pseudoreplicates (Hurlbert, 1984), cross-validation is performed on the sequence level. Cross-validation on the call level would result in *twinning*, training and test sets with calls from the same individual. In this study we use leave one out cross validation (LOOCV), holding out all the calls from an individual sequence at each iteration for testing. The result of the cross-validation procedure is a prediction for every call in the dataset.

In addition to total classifier accuracy, class specific metrics were calculated. We used positive predictive power (PPP). PPP is the proportion of calls predicted to be a class that are actually of it. It is the conditional probability that a case is truly positive given it is predicted to be positive.

To compare pairs of classifiers tested on the same dataset we used McNemar's test (McNemar, 1947) (Salzberg, 1997), a non-parametric test based on the standardised normal test statistic:

$$z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} - f_{21}}} \tag{3.1}$$

where f_{12} represents the total number of cases correctly classified by classifier 1, but misclassified by classifier 2; and f_{21}represents the total number of cases correctly classified by classifier 2, but misclassified by classifier 1 (see table 3.1). We used the exact McNemar test implemented in R by the extract2x2 package (Fay, 2010).

Table 3.1: Illustration of the 2x2 table required for the McNemar test.

	Classifier 2	
Classifier 1	Correctly classified	Misclassified
Correctly classified	f_{11}	f_{12}
Misclassified	f_{21}	f_{22}

3.4 Results

We recorded a total of 609 bat passes of known identity, from which we extracted 5370 individual calls for analysis (Table 3.2). Within the *Myotis* we obtained 296 recordings from four species and one species group, comprising 3208 individual calls (Table 3.3).

Table 3.2: Summary of echolocation call library.

	Seqs	Calls
Barbastella	62	186
Eptesicus	13	330
Myotis	300	3207
Nyctalus	27	502
Pipistrellus	67	782
Plecotus	80	270
Rhinolophus	16	93

Table 3.3: Summary of *Myotis* echolocation call library.

Abb.	Species	Seqs	Calls
M. alc	*Myotis alcathoe*	23	309
M. bec	*Myotis bechsteinii*	16	191
M. bra./mys.	*Myotis brandtii/mystacinus*	88	1221
M. dau	*Myotis daubentonii*	96	775
M. nat	*Myotis nattereri*	73	712

3.4.1 Random Forest classification

A confusion matrix summarises the results of the model classification of test data. The relation between the actual identity of calls and the predicted identity of calls is displayed in a matrix which has one row and column for each class. Diagonal elements represent correct classifications, whilst all elements off the diagonal represent misclassifications. Random forest using power spectrum features classified calls to genus with an overall accuracy of 97.6% and 98.4%, at the call and sequence level respectively (confusion matrices in Tables 3.4 and 3.5). At the sequence level, *Barbastella*, *Eptesicus*, *Nyctalus*, *Pipistrellus* and *Rhinolophus* were all identified with 100% PPP. A single *Pipistrellus* sequence was misclassified as *Myotis* and three *Barbastella* sequences were misclassified as *Plecotus*. At the call level, all *Rhinolophus* calls were unambiguously identified. *Barbastella*, *Myotis*, *Nyctalus*, *Pipistrellus* and *Plecotus* were all identified with a PPP of more than 90%. *Eptesicus* was identified with the lowest PPP, due to confusion with *Nyctalus*.

Table 3.4: Genus sequence classification.

Confusion matrix showing the classification of sequences to genus level using power spectra features and a random forest classifier. Positive predictive power (PPP) indicates the percentage of calls predicted to be from a group that were actually from that group. Overall accuracy was 98.4%, with an average PPP over all classes of 99.2%.

		Predicted class							
		Barbastella	*Eptesicus*	*Myotis*	*Nyctalus*	*Pipistrellus*	*Plecotus*	*Rhinolophus*	unclassed
Actual class	*Barbastella*	**35**	0	0	0	0	3	0	2
	Eptesicus	0	**12**	0	0	0	0	0	1
	Myotis	0	0	**296**	0	0	0	0	0
	Nyctalus	0	0	0	**26**	0	0	0	0
	Pipistrellus	0	0	1	0	**66**	0	0	0
	Plecotus	0	0	0	0	0	**54**	0	1
	Rhinolophus	0	0	0	0	0	0	**16**	0
	PPP	100.0	100.0	99.7	100.0	100.0	94.7	100.0	

Table 3.6: Genus call classification.

Confusion matrix showing the classification of individual calls to genus level using power spectra features and a random forest classifier. Positive predictive power (PPP) indicates the percentage of calls predicted to be from a group that were actually of that group. Overall accuracy was 97.6%, with an average PPP over all classes of 95.6%.

		Predicted class						
		Barbastella	*Eptesicus*	*Myotis*	*Nyctalus*	*Pipistrellus*	*Plecotus*	*Rhinolophus*
Actual class	*Barbastella*	**174**	3	4	0	2	3	0
	Eptesicus	1	**293**	0	36	0	0	0
	Myotis	2	0	**3196**	0	9	0	0
	Nyctalus	0	38	0	**455**	0	9	0
	Pipistrellus	2	0	10	0	**770**	0	0
	Plecotus	2	3	0	3	0	**262**	0
	Rhinolophus	0	0	0	0	0	0	**93**
	PPP	96.1	86.9	99.6	92.1	98.6	95.6	100.0

Random forest classified *Myotis* species with an overall accuracy of 93.1% and 95.9%, at the call and sequence level respectively. At the sequence level, *M. bechsteinii*, *M. daubentonii* and *M. nattereri* were all identified with 100% PPP. *M. alcathoe* and *M. brandtii/mystacinus* were identified with more than 90% PPP. *M. bechsteinii* was the most frequently misclassified, with sequences being confused with *M. alcathoe* and *M. brandtii/mystacinus*. At the call level *M. alcathoe*, *M. daubentonii* and *M. nattereri* were classified with more than 90% PPP. Few *M. brandtii/mystacinus* calls were misclassified, but all other species were misclassified to some degree as *M. brandtii/mystacinus*, resulting in a PPP of 88.8%. *M. bechsteinii* was frequently misclassified as *M. brandtii/mystacinus*. *M. alcathoe* and *M. brandtii/mystacinus* were misclassified as *M. bechsteinii* resulting in a PPP of 87.4%.

Table 3.5: *Myotis* sequence classification.

Confusion matrix showing the classification of *Myotis* sequences using power spectra features and a random forest classifier. Positive predictive power (PPP) indicates the percentage of calls predicted to be from a group that were actually from that group. Overall accuracy was 95.9%, with an average PPP over all classes of 97.1%.

		Predicted class					
		M.alc	*M. bra./mys.*	*M.bec*	*M.dau*	*M.nat*	unclassed
Actual class	*M. alc*	**23**	0	0	0	0	0
	M. bra./mys.	1	**86**	0	0	0	1
	M.bec	1	6	**8**	0	0	1
	M.dau	0	0	0	**94**	0	2
	M.nat	0	0	0	0	**73**	0
	PPP	92.0	93.5	100.0	100.0	100.0	

Table 3.7: *Myotis* call classification.

Confusion matrix showing the classification of individual *Myotis* calls using power spectra features and a random forest classifier. Positive predictive power (PPP) indicates the percentage of calls predicted to be from a group that were actually from that group. Overall accuracy was 93.1%, with an average PPP over all classes of 92.5%.

		Predicted class				
		M.alc	*M.bra./mys.*	*M.bec*	*M.dau*	*M.nat*
Actual class	*M.alc*	**285**	19	4	1	0
	M.bra./mys.	16	**1169**	9	23	4
	M.bec	14	84	**90**	3	0
	M.dau	0	39	0	**735**	0
	M.nat	0	5	0	0	**707**
	PPP	90.5	88.8	87.4	96.5	99.4

3.4.2 Classifier comparisons

For classification at the genus level, the largest increase in overall accuracy was achieved as a result of the change in classifier, from QDA to RF (5.6% increase, p<0.001). In classifying calls from the genus *Myotis*, there was a significant improvement in overall accuracy as a result of the switch from QDA to RF (3.2% increase, p<0.001). However, the greatest increase in classification accuracy was a result of additional features, in the form of the power spectrum (PS) values. The addition of the overall call PS features improved accuracy by 6.4% (p<0.001), with a further 4.7% (p<0.001) increase with the addition of the PS features measured from the start of the call.

Table 3.8: Classifier comparisons at the genus level.

Overall accuracies (Acc) and statistical comparisons of classifiers in the classification of calls to genus. Classifiers are quadratic discriminant function analysis (QDA) and random forest (RF). Features are the conventional 'base' call features, power spectrum (PS) and power spectra (2 x PS). Classifier 1 accuracy is statistically compared to classifier 2 accuracy, e.g. there is a statistical difference in the accuracy of QDA using 'base' features compared with RF using 'base' features. 95% CI = confidence intervals for the McNemar test.

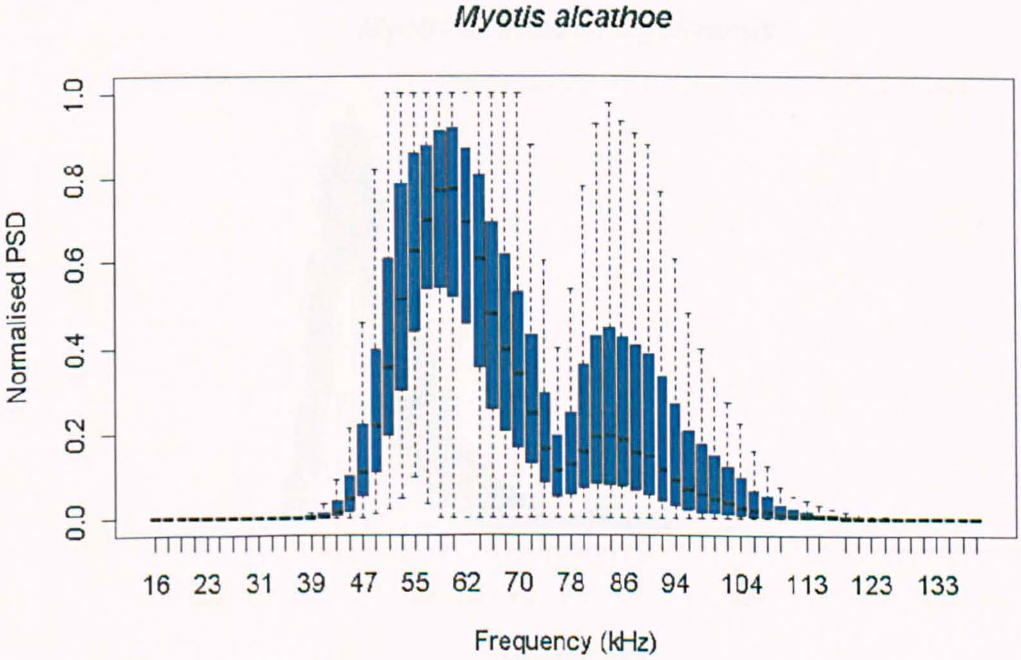
Classifier 1	Acc (%)	Classifier 2	Acc (%)	McNemar test	95% CI	Odds ratio
QDA 'base'	90.6	RF 'base'	96.2	$p < 0.001$	3.29 to 5.18	4.11
RF 'base'	96.2	RF PS	95.7	$p = 0.15$	0.67 to 1.06	0.84
RF 'base'	96.2	RF 2 x PS	97.6	$p < 0.001$	1.44 to 2.50	1.89

Table 3.9: Classifier comparisons for the *Myotis*.

Overall accuracies (Acc) and statistical comparisons of classifiers in the classification of calls to genus. Classifiers are quadratic discriminant function analysis (QDA) and random forest (RF). Features are the conventional 'base' call features, power spectrum (PS) and power spectra (2 x PS). Classifier 1 accuracy is statistically compared to classifier 2 accuracy, e.g. there is a statistical difference in the accuracy of QDA using 'base' features compared with RF using 'base' features. 95% CI = confidence intervals for the McNemar test.

Classifier 1	Acc (%)	Classifier 2	Acc (%)	McNemar test	95% CI	Odds ratio
QDA 'base'	78.8	RF 'base'	82.0	$p < 0.001$	1.37 to 2.07	1.68
RF 'base'	82.0	RF PS	88.4	$p < 0.001$	0.43 to 0.61	0.51
RF PS	88.4	RF 2 x PS	93.1	$p < 0.001$	0.13 to 0.28	0.2

Myotis spp. calls were very variable. Figures 3.2 – 3.6 show Boxplots of average power spectra for each species.



Myotis alcathoe

Figure 3.2: *M. alcathoe* average power spectrum.

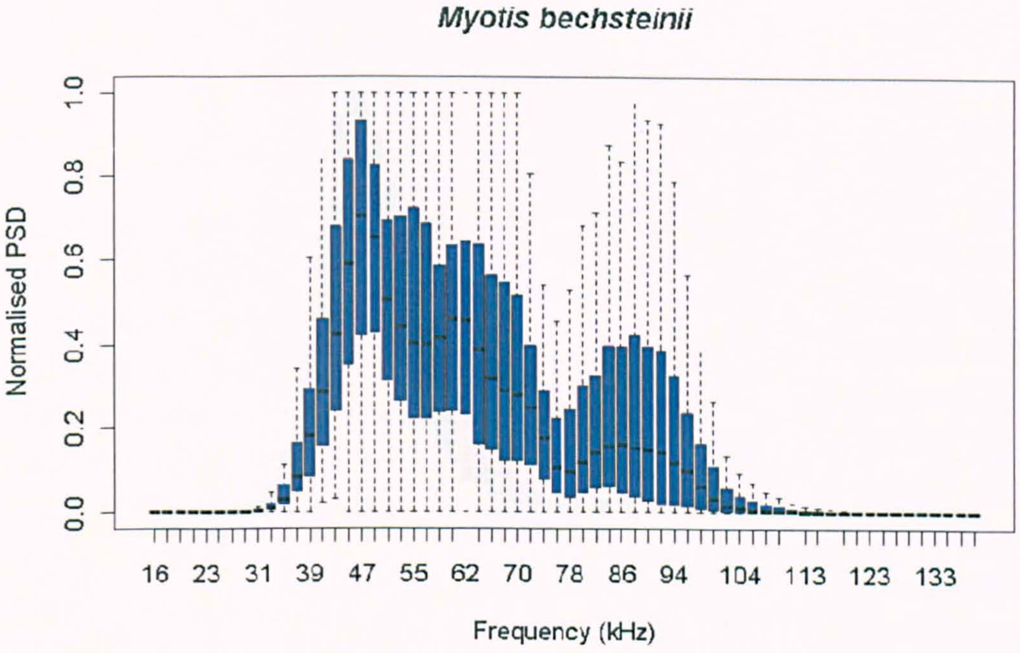


Figure 3.3: ***M. bechsteinii*** **average power spectrum.**

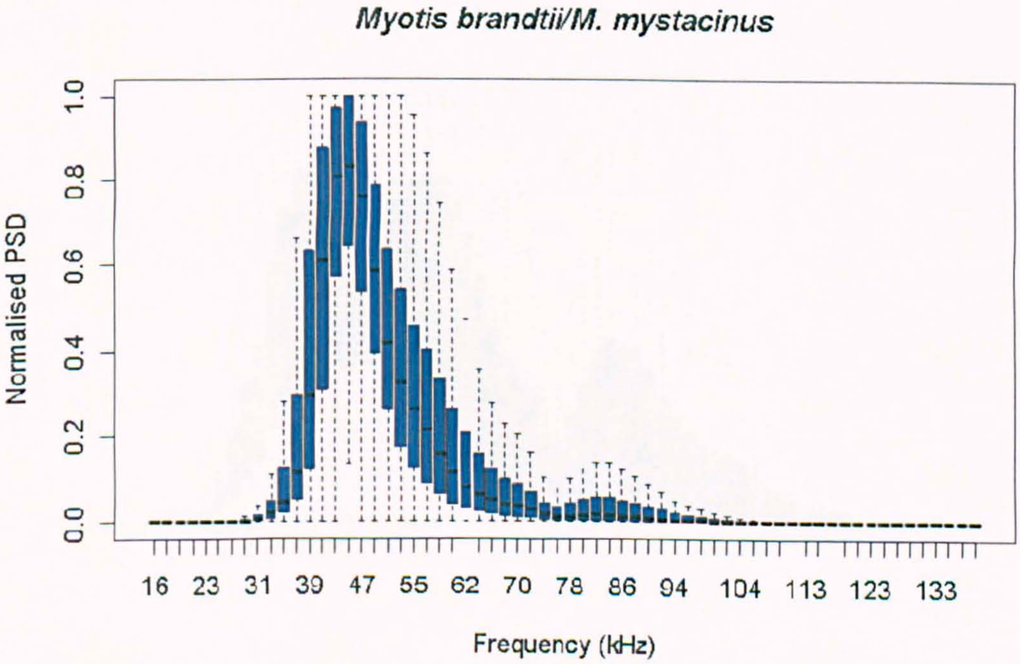


Figure 3.4: ***M. brandtii/M. mystacinus*** **average power spectrum.**

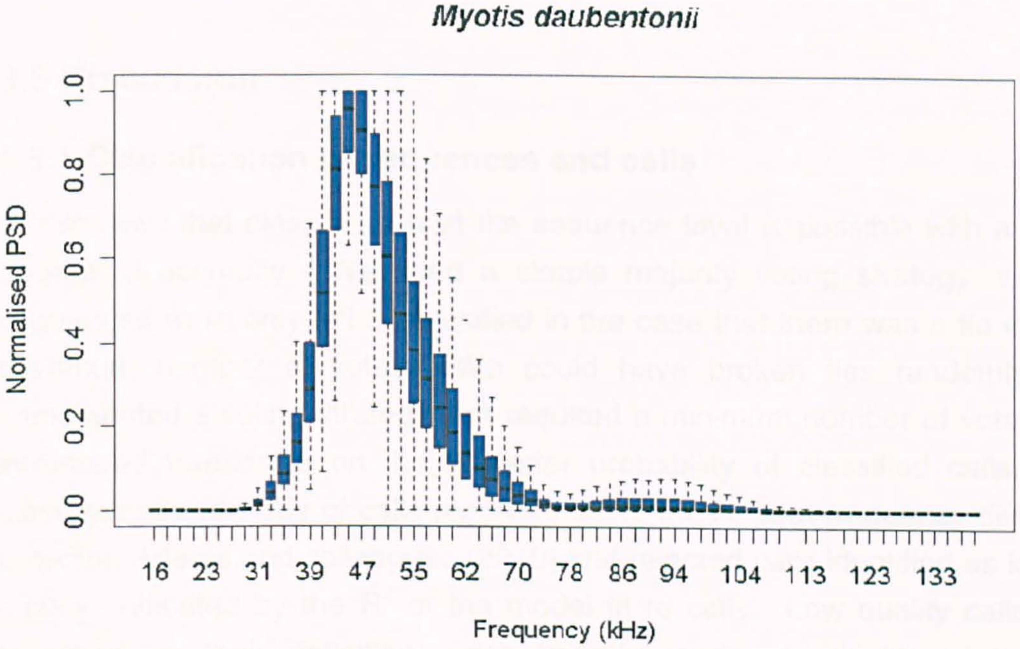

Myotis daubentonii

Figure 3.5: ***M. daubentonii.***

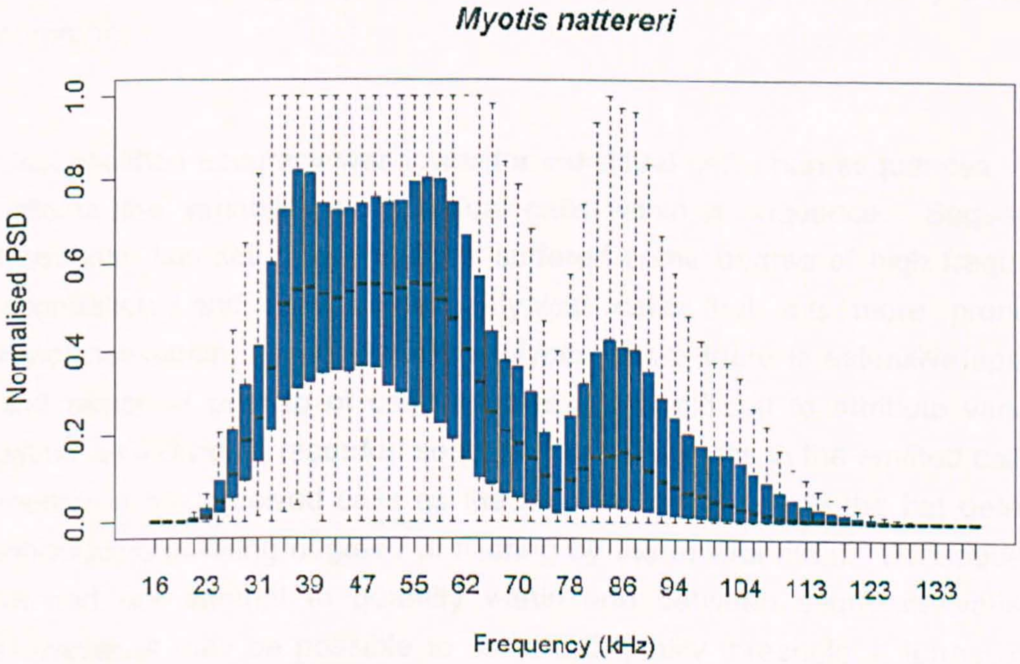

Myotis nattereri

Figure 3.6: ***M. nattereri*** **average power spectrum.**

3.5 Discussion

3.5.1 Classification of sequences and calls

We showed that classification at the sequence level is possible with a high degree of accuracy. We used a simple majority voting strategy, where sequences were only left unclassified in the case that there was a tie in the maximum number of votes. We could have broken ties randomly, or implemented a voting strategy that required a minimum number of votes, or introduced thresholds on the posterior probability of classified calls. In classifying sequences of calls recorded using the Anabat frequency division detector, Adams and colleagues (2010) first rejected calls identified as lower quality, indicated by the R^2 of the model fit to calls. Low quality calls, as identified by their definition, were rejected and not considered in the subsequent voting scheme. Our procedure produced few misclassified and unclassified call sequences, but a more elaborate procedure that rejects low quality, or low confidence classifications, may prove effective in situations where encountering sequences containing exclusively low quality calls are common.

Classification accuracy was lower for individual calls than sequences. This reflects the variability of individual calls within a sequence. Sequences frequently contain calls that have suffered some degree of high frequency attenuation, and may contain atypical calls that are more prone to misclassification, especially in the *Myotis* where there is extensive spectral and temporal overlap of call variables. It is difficult to attribute variation within an individual recorded sequence to differences in the emitted calls, or merely in the received calls as the bat moves relative to the bat detector, introducing differing degrees of filtering by the environment. Consequently, we did not attempt to quantify within and between sequence variation. However, it may be possible to set a call quality threshold in terms of call SNR, to facilitate a quantitative comparison of variation among the higher quality recorded calls. *M. bechsteinii* was often misclassified as *M. brandtii/M. mystacinus*. However, this may be a consequence of the relatively low sample size for this species. Whilst a large error rate can be attributed to the difficulty of the classification problem, insufficient training data may also degrade classifier performance (Raudys & Jain, 1991). We will continue to add to our echolocation call library, to investigate the effects of sample size on classification accuracy. *M. bechsteinii* was frequently misclassified as *M. brandtii/M. mystacinus* which lowers the probability of detecting this rare species using acoustic methods. However, all *M. bechsteinii* sequences classified as *M. bechsteinii* were of *M. bechsteinii* (100% positive predictive power). Whilst we must remain cautious given the small sample size (n=8), this suggests acoustic methods may still prove a reliable way of generating presence data for *M. bechsteinii.* Presence data can then be used to build species distribution models, and produce habitat suitability maps, valuable conservation tools (species distribution modelling is the focus of Chapter 6, and more fully discussed there).

Ideally an identification system should be transferable between areas, but geographic variation and differing species assemblages may complicate the transferability of trained systems. A concern is that any technique that has been trained from calls from a species from one part of its geographic range may have learned features that are specific to that area, and may not transfer to other areas with the same accuracy in classifications (e.g. Barclay & Brigham, 2002). However, a statistical finding of call variation across a species' geographic range may not affect the ability to identify the species by call structure (e.g. O'Farrell, 2000). Also, in practice, a local call library may simply not be available, and references calls may have to be taken from locations as close to the study area as possible that share similar habitats (Barclay & Brigham, 2002). Davy and colleagues (2007) tested two DFAs trained to identify calls from Italian bats, using a small sample of calls from known species of Greek bats, and found the functions proved effective.

3.5.2 Comparison with other studies

High classification accuracies were reported by Redgwell *et al.* (2009) for British species, with correct identification rates that varied from 91 – 100% for five species of *Myotis*, using an ensemble of 21 neural networks. A previous study using the same dataset achieved classification rates that varied from 75-90%, using a single neural network (Parsons & Jones, 2000). Direct comparison with this study is not possible as we use a different dataset, which is not biased towards high quality calls, a caveat acknowledged by the previous studies' authors (Parsons & Jones, 2000; Redgwell *et al.*, 2009). In addition, the exact methodology that led to the improvements in accuracy attained by Redgwell *et al.* (2009) is not clear. There is no mention of a validation set in determining the best performing neural networks prior to their aggregation into ensembles, and final testing on an independent test set. It appears that the best performing neural networks may have been selected on the basis of their performance on the final test set, although it was not possible to determine for certain because insufficient information was provided. If this was the case, it is an example of overfitting, i.e. the models are given the opportunity to take advantage of statistical peculiarities of the dataset, and it is expected that many of the results obtained will therefore be overly optimistic (Cawley & Talbot, 2010). To ensure unbiased estimates of accuracy, any tuning of model parameters or changes to model architecture must be performed prior to seeing the final test set (Salzberg, 1997).

3.5.3 Concluding remarks

The results of this study show that the random forest classifier and echolocation call power spectra can be used to reliably identify British bats from their calls. This approach outperformed conventional statistical analysis. However, echolocation calls are variable within a species, and some calls may be confused with other species. Classifying sequences of calls, and using a voting scheme to assign a final identification, produces more accurate results. We achieved high positive predictive power at the sequence level even for the *Myotis* bats, indicating that acoustic methods can reliably be used to collect accurate presence data for species distribution modelling and assessing habitat use (Chapter 6). A trained machine learning classifier can operate in almost real-time. As our approach classifies calls in a single pass through a recording, this opens up the possibility that the identification system could be built into equipment that could be used in the field to acoustically identify bats *in situ*.

3.6 References

Adams, M. D., Law, B. S., & Gibson, M. S. (2010). Reliable Automation of Bat Call Identification for Eastern New South Wales, Australia, Using Classification Trees and AnaScheme Software. *Acta Chiropterologica*, *12*(1), 231-245.

Armitage, D. W., & Ober, H. K. (2010). A comparison of supervised learning techniques in the classification of bat echolocation calls. *Ecological Informatics*, *5*(6), 465-473.

Berge, L. 2007. The effects of phylogenetic differences on resource partitioning between the cryptic species whiskered bat (*Myotis mystacinus*) and Brandt's bat (*M. brandtii*). Unpublished Thesis, University of Bristol, School of Biological Sciences, Bristol, UK, 162 pp.

Breiman, L. (2001a). Statistical modeling: The two cultures. *Statistical Science*, *16*(3), 199–215. JSTOR.

Breiman, L. (2001b). Random forests. *Machine learning*, *45*(1), 5–32.

Britzke, E. R., Duchamp, J. E., Murray, K. L., Swihart, R. K., & Robbins, L. W. (2011). Acoustic identification of bats in the eastern United States: A comparison of parametric and nonparametric methods. *The Journal of Wildlife Management*, *75*(3), 660-667.

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd international conference on Machine learning,* 161–168.

Cawley, G., & Talbot, N. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning*, *11*, 2079-2107.

Chapman, D. S., Bonn, A., Kunin, W. E., & Cornell, S. J. (2010). Random Forest characterization of upland vegetation and management burning from aerial imagery. *Journal of Biogeography*, *37*(1), 37–46.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783-92.

Davy, C. M., Russo, D., & Fenton, M. B. (2007). Use of native woodlands and traditional olive groves by foraging bats on a Mediterranean island: consequences for conservation. *Journal of Zoology*, *273*(4), 397-405.

Fay, M. P. (2010). Two-sided Exact Tests and Matching Confidence Intervals for Discrete Data. *R Journal*, *2*(1), 53-58.

Georgiakakis, P., Vasilakopoulos, P., Mylonas, M., & Russo, D. (2010). Bat species richness and activity over an elevation gradient in Mediterranean shrublands of Crete. *Hystrix-the Italian Journal of Mammalogy, 21*(1), 43-56.

Hughes, A. C., Satasook, C., Bates, P. J. J., Soisook, P., Sritongchuay, T., Jones, G., & Bumrungsri, S. (2010). Echolocation Call Analysis and Presence-Only Modelling as Conservation Monitoring Tools for Rhinolophoid Bats in Thailand. *Acta Chiropterologica, 12*(2), 311-327.

Hurlbert, S. H. (1984). Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs, 54*(2), 187.

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.

Jones, G., Vaughan, N., & Parsons, S. (2000). Acoustic identification of bats from directly sampled and time expanded recordings of vocalizations. *Acta Chiropterologica, 2*(2), 155–170.

Krusic, R. (1995). Habitat associations of bat species in the White Mountain National Forest. *Bats and forests symposium, October.*

Laiolo, P. (2010). The emerging significance of bioacoustics in animal species conservation. *Biological Conservation, 143*(7), 1635-1645.

Lance, R., Bollich, B., Callahan, C., & Leberg, P. (1995). Surveying forest-bat communities with Anabat detectors. *Bat Research News, 36*, 28.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R news, 2*(3), 18–22. Retrieved from http://www.webchem.science.ru.nl/PRiNS/rF.pdf

MacSwiney G., M. C., Clarke, F. M., & Racey, P. a. (2008). What you see is not what you get: the role of ultrasonic detectors in increasing inventory completeness in Neotropical bat assemblages. *Journal of Applied Ecology, 45*(5), 1364-1371.

Marckmann, U., & Runkel, V. (2010). Automatic bat call analysis with the batcorder-system, http://www.ecoobs.com/.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*(2), 153-157.

D. Meyer. (2001) Support vector machines, R News 1(3) 23–26. http://CRAN.R-project.org/ doc/Rnews/.

O'Farrell, M. J., Corben, C., & Gannon, W. L. (2000). Geographic variation in the echolocation calls of the hoary bat (Lasiurus cinereus). *Acta Chiropterologica*, *2*(2), 185–196.

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.

Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine learning methods without tears: a primer for ecologists. *The Quarterly review of biology*, *83*(2), 171-93.

Parsons, S. (2001). Identification of New Zealand bats (*Chalinolobus tuberculatus* and *Mystacina tuberculata*) in flight from analysis of echolocation calls by artificial neural networks. *Journal of Zoology*, *253*(4), 447-456.

Parsons, S., Boonman, A. M., & Obrist, M. K. (2000). Advantages and Disadvantages of Techniques for Transforming and Analyzing Chiropteran Echolocation Calls. *American Society of Mammalogists*, *81*(4), 927-938.

Parsons, S., & Jones, G. (2000). Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. *J. Exp. Biol*, *203*, 2641-56.

Preatoni, D. G., Nodari, M., Tosi, G., & Wauters, L. U. C. A. (2005). IDENTIFYING BATS FROM TIME-EXPANDED RECORDINGS OF SEARCH CALLS: COMPARING CLASSIFICATION METHODS. *Journal of Wildlife Management*, *69*, 1601-1614.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Raudys, S. J., & Jain, A. K. (1991). Small sample size effects in statistical pattern recognition: recommendations for practitioners and open problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(3), 252-264.

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.

Rodriguez-Galiano, V., Abarca-Hernández, F., Ghimire, B., Chica-Olmo, M., Atkinson, P., & Jeganathan, C. (2011). Incorporating Spatial Variability Measures in Land-cover Classification using Random Forest. *Procedia Environmental Sciences*, *3*, 44–49.

Russo, D., Jones, G., & Mucedda, M. (2001). Influence of age, sex and body size on echolocation calls of Mediterranean and Mehely's horseshoe bats, Rhinolophus euryale and R. mehelyi (Chiroptera: Rhinolophidae). *Mammalia*, *65*(4), 429-436.

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.

Russo, D., & Jones, G. (2003). Use of foraging habitats by bats in a Mediterranean area determined by acoustic surveys: conservation implications. *Ecography*, *26*(2), 197-209.

Salzberg, S. (1997). On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Mining and knowledge discovery*, *327*, 317-327.

Schnitzler, H.-U., & Kalko, E. K. V. (2001). Echolocation by Insect-Eating Bats. *BioScience*, *51*(7), 557.

Vapnik, V. The Nature of Statistical Learning Theory. Springer-Verlag, New York, 1995.

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*, *7*, 91.

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Yovel, Y., Melcon, M. L., Franz, M. O., Denzinger, A., & Schnitzler, H.-U. (2009). The voice of bats: how greater mouse-eared bats recognize individuals based on their echolocation calls. *PLoS computational biology*, *5*(6), e1000400.

Chapter 4: Removing errors due to unexpected species and noise

4.1 Abstract

Supervised classification assumes that all classes to be encountered in practical use were included in training the classifier. For ecological applications such as classifying bat echolocation calls, this is rarely likely to be the case, and the number of classes may be unknown or unbounded. The failure to satisfy the assumption of an exhaustively defined set of training classes leads to the misclassification of outliers, signals very different from the training examples. We used two techniques to try and remove outliers, in an attempt to reduce misclassifications. Firstly, we used the soft classification output, in the form of posterior probability estimates, to try and identify and remove untrained classes using a simple threshold procedure. Post-classification thresholding proved ineffective in reducing the misclassifications caused by untrained classes. Secondly, a one-class support vector machine (SVM) was applied to each species group used in training, to protect the classifier from outliers not representative of the training data. One-class SVM reduced the misclassifications caused by untrained species at the cost of rejecting some known data. We suggest that in a fully automated identification system, methods of outlier detection must be employed for reliable classification.

4.2 Introduction

Statistical classification methods are widely used in ecology, with applications ranging from species distribution modelling (e.g. Garzon *et al.*, 2006; Olden *et al.*, 2008), to the characterisation of vegetation types from aerial imagery (e.g. Chapman *et al.*, 2010; Bradter *et al.*, 2011). A diverse range of classification techniques has been investigated for the identification of bat echolocation calls, from traditional statistical methods such as disciminant function analysis (DFA; Parsons & Jones, 2000; Russo & Jones, 2002; Preatoni *et al.*, 2005; Papadatou *et al.*, 2008; Armitage & Ober, 2010; Hughes *et al.*, 2010; Redgwell *et al.*, 2009), to an array of more modern machine learning approaches, including artificial neural networks (ANNs; Parsons & Jones, 2000; Preatoni *et al.*, 2005; Armitage & Ober, 2010; Redgwell *et al.*, 2010), holographic neural networks (Broders *et al.*, 2004), Gaussian mixture models (GMMs; Skowronski & Harris, 2005), hidden Markov models (HMMs; Skowronski & Harris, 2005), synergetic pattern recognition (Obrist *et al.*, 2004), classification and regression trees (Adams *et al.*, 2010), cluster analysis (Preatoni *et al.*, 2005), support vector machines (SVM; Armitage & Ober, 2010; Redgwell *et al.*, 2009) and random forests (RF; Armitage & Ober, 2010). Machine learning approaches are suited to solving non-linear and high-dimensional problems, and have generally been found to outperform traditional statistical methods such as DFA (e.g. Armitage & Ober, 2010; Redgwell *et al.*, 2009). High classification accuracies have been achieved, even where the spectral and temporal characteristics of calls exhibit extensive overlap (e.g., *Myotis* bats – Redgwell *et al.*, 2009). However, models have frequently been trained using only high quality calls, recorded as bats were within 2 m of the bat detector (e.g. Parsons & Jones, 2000; Redgwell *et al.*, 2009), and it has been demonstrated that the performance of such models degrades when faced with classifying lower quality calls, more typical of those encountered during field studies (Jennings *et al.*, 2008).

Where classifiers have been applied to field studies, it has been typical for the researchers to hand select the calls for classification, using only those with the highest signal to noise ratio (SNR) from each recorded bat pass (e.g. Davy *et al.*, 2007). This procedure ensures that calls of low SNR or that have suffered high-frequency attenuation, can be excluded from analyses to reduce the risk of misclassification. The manual measurement of call parameters remains a time-consuming and error-prone task, and it is desirable to fully-automate the process of call detection, extraction and classification. However, there are numerous challenges associated with moving from semi-automated, where there is still significant human interaction in the process, to fully-automated classification. The most significant challenge is perhaps the problem of how to deal with "unknowns" (Gaston & O'Neill, 2004). The basic underlying assumption of supervised classification is that the number of classes is exhaustively defined (Foody, 2001). When making a prediction, a classifier assigns the incoming data to one of the classes defined during training. If the incoming data is novel, for example an echolocation call from a species not encountered during training, it will be misclassified. This represents a significant problem where classifiers are to be applied to field data, where additional species or novel signals may be encountered. Compiling echolocation call libraries, containing representative calls of known species, is a time-consuming and difficult task (Obrist *et al.*, 2004). Due to the potential for intraspecific geographic variation (e.g. Barclay *et al.*, 1999; O'Farrell *et al.*, 2000; Papadatou *et al.*, 2008), libraries local to the area to which they are to be applied are desirable. Obtaining reference calls of all species in an area may be very difficult, especially in areas of high species richness. Additionally, species assemblages should not be assumed to be static. Regional species composition is dynamic and likely to change over time due to factors such as climate change driven range expansion (e.g. Lundy *et al.* 2010), or the identification of morphologically cryptic species (Barratt *et al.*, 1997; Helversen *et al.*, 2001; Spitzenberger, 2006). Even where comprehensive call libraries can be obtained, known species may emit novel call types that are not represented in the training data, e.g. social calls (Fenton, 2003). Moreover, novel signals may be encountered due to noise. Noise sources may be biotic (other calling animals or stridulating insects) or abiotic (flowing water, rain, wind-induced vegetation noise). During manual analysis, researchers may filter the data that is presented to a classifier, ensuring that calls significantly different from the training data are not passed on to the classification stage. However, in a fully automated system there must be some provision for identifying and rejecting novel data, so that the classifier can return "unknown" rather than force a misclassification. This issue has been overlooked in the past (but see Adams *et al.*, 2010), perhaps because fully automated systems were not yet developed.

Previous research on classifying bat echolocation calls has been largely concerned with hard (crisp) classification methods. In this approach, a call is a member of a class or not, and the decision is binary. This may suffice where classes are discrete, mutually exclusive and exhaustively defined, but this is often not the case (Foody, 2001). An alternative to hard classification is soft (fuzzy) classification, which provides posterior probability estimates for class membership. Soft classification is therefore able to represent ambiguity between classes, and the input data can be predicted to belong to more than one class. A 'winning' class can still be determined by selecting the class with the highest probability estimate. A single class probability estimate dominating all others suggests confidence in the classification. In contrast, if estimated probabilities are almost equal in value between classes, the prediction can be considered for rejection, and labelled as having an unknown class membership. Alternatively, a prediction may be assigned to multiple classes, for example a species group. This procedure has been applied to improve classification rates in a range of problem domains, for example handwritten digit recognition (Le Cun *et al.*, 1989), diagnosing sleeping disorders (Gudmundsson *et al.*, 2005), and the acoustic identification of insects (e.g. Chesmore, 2004). However, soft classification has largely been ignored for the task of bat echolocation call classification. Post-classification thresholds on probability estimates offer the potential to reject novel signals that would otherwise be misclassified (Morris *et al.*, 2001). A further possibility is the use of outlier detection methods, to reject data that are significantly different from the training data (e.g. Tax & Duin, 1999). One-class support vector machines have been applied to the problem of outlier detection, and have proved particularly successful on high dimensional data (Tax & Duin, 2004).

The aim of this study was to explore the effects of novel data, i.e. untrained classes, on the accuracy of hard and soft classifications produced by a support vector machine classifier. Post-classification probability thresholds and one-class support vector machines were investigated to reduce the misclassifications caused by untrained classes.

4.3 Methods

4.3.1 Database of recordings

During 2009-2011 we recorded bats at a range of sites across the UK to build an echolocation call library for British bat species (Chapter 3). In addition to our echolocation call library, we collected a database of noise recordings, to investigate the effects of noise sources being presented to the classifier. Noise recordings were obtained by carrying out walked transects, where noise was frequently encountered of sufficient amplitude to trigger the detector.

4.3.2 Call analysis

Calls were automatically extracted from recordings using the methods detailed in Chapter 2.

For the purposes of machine classification, discriminatory features must be extracted from calls. Features, in classification terms, are quantifiable attributes that provide useful information for the discrimination of different classes. Features can be continuous (e.g. frequency at the start of the call), or binary (presence or absence of a characteristic). Four call features were automatically extracted from each call: call duration (ms), frequency at the start of the call (kHz), frequency at the end of the call (kHz), and frequency at the location of maximum energy of the call (kHz). These call variables are commonly employed in echolocation call analysis (e.g. Vaughan *et al.*, 1997; Parsons & Jones, 2000; Russo & Jones, 2001), and are described in Redgwell *et al.* (2009) as 'base' parameters, a convention we adopt here.

4.3.3 Classification algorithms

Support vector machines (SVM) are a supervised learning technique derived from statistical learning theory (Vapnik, 1995). Training data features are mapped into a higher dimensional space using the 'kernel trick', to find the optimal separating hyperplane that maximises the distance between two groups of data. The hyperplane is a non-linear decision boundary, and during classification of novel data, the assigned class is determined from the side of the hyperplane that the data point falls on. In constructing the optimal hyperplane, the algorithm only considers the support vectors, those data points that lie on the boundary between classes. Mapping into a higher dimensional space can be achieved using non-linear kernels where classes are not linearly separable, e.g. radial basis function (RBF) or sigmoidal. For a more detailed mathematical derivation of SVMs refer to (Burges, 1998). We implemented SVMs using LIBSVM, a library for support vector machines developed by Chang and Lin (2001). The LIBSVM package estimates posterior probabilities by fitting a sigmoid function that maps the SVM outputs (Gudmundsson *et al.*, 2005).

SVMs are essentially a binary (two-class) classification technique, and to achieve multiclass classification, the outputs of multiple models must be combined. Multiclass classification in this study was achieved using the 'one vs. one' approach implemented as default in LIBSVM (Chang and Lin, 2001). An SVM classifier is constructed for each pair of classes, resulting in $N(N-1)/2$ models. For example, a classifier trained to discriminate *Pipistrellus*, *Myotis* and *Nyctalus* species requires three separate SVM classifiers. Classifier 1 is trained to discriminate *Pipistrellus* and *Myotis*, classifier 2 *Pipistrellus* and *Nyctalus*, and classifier 3 *Myotis* and *Nyctalus*. In classifying novel data, each model predicts the class in turn, and the class with the majority of votes over all models is selected as the predicted class. A common alternative is the 'one vs. all' approach, which consists of building one SVM per class, trained to discriminate that class from all others, but this has not been shown to improve predictive performance over the 'one vs. one' procedure (Schwenker, 2000).

The one-class SVM algorithm is a variation on the standard SVM technique, adapted to use examples from only one class in training (Tax & Duin, 1999). Input data are mapped into a high-dimensional feature space using a kernel function as in SVM. However, the origin is treated as the only example from other classes. The algorithm iteratively finds the maximum margin separating hyper-plane between the training data and the origin. In this way a hyper-sphere is fitted around the training data. Adjusting the kernel parameter *nu* changes the shape of the boundary, and provides a method of fitting a specified proportion of the data. During prediction of new data, points falling outside the boundary are classed as outliers to the class. For example, a one-class SVM fitted to training data from *Myotis* species should reject all non-*Myotis* calls as outliers. However, there is a trade-off between rejecting all outliers, and accepting all true data. In practice, some true data must be rejected to ensure the majority of outliers are rejected correctly. In this study we used a fixed *nu* value of 0.05, effectively fitting the boundary around 95% of the training data of a class. Consequently, a one-class SVM fitted to *Myotis* species would reject ~5% of true *Myotis* calls in order to reliably reject non-*Myotis* calls. The literature provides little guidance on parameter settings for one-class SVM, but a *nu* value of 0.05 is default in LIBSVM.

4.3.4 Model evaluation

A support vector machine was first trained using data from all species, to illustrate the results obtained when all classes are known. That is, during testing the classifier was not presented with data from species that it had not been trained with. To evaluate the effects of novel signal types on the classifications, the noise dataset was used to test the predictions of the classifier. In addition, we created a series of classifiers form which a single species group had been excluded during training. Each classifier was then tested with the data from the excluded group, for example *Barbastella* calls withheld during training, but subsequently presented to the classifier during testing. Following classification, a series of thresholds on the posterior probability estimates were applied to reject low-confidence classifications. Additionally, a one-class support vector machine was fitted to each class, and used to reject outliers following prediction by the standard SVM classifier.

To obtain an unbiased estimate of the true error rate of different classification algorithms, models must be tested using a set of data independent from that used to train the model. Cross-validation procedures use the available data more efficiently than a simple division into a separate training and test set. A small amount of data is withheld, with the remaining data being used to train the classifier. The withheld data are then used to test the classifier. This process is repeated, withholding a new set of data each time, until all data have been classified. A nearly unbiased estimate of classifier accuracy is then obtained by averaging the results from all iterations (Varma & Simon, 2006). We performed validation of models using the leave-one-out cross-validation (LOOCV) method, taking call sequences as the sampling unit. A single sequence of calls (i.e. an individual bat) is held out and used to assess the accuracy of a model trained on the remaining sequences of calls, by classifying each individual call from the withheld sequence. This process is repeated until all sequences have been used as test sets. This approach maximises the information available to train models, whilst avoiding training and testing models using calls from the same individual, which would introduce problems of pseudo-replication (Hurlbert, 1984).

Models were evaluated using positive predictive power (PPP; percentages of predictions that were actually calls of that species), and sensitivity (percentage of calls correctly classified). Accuracy metrics were calculated for individual classes, as overall measures of model accuracy can be biased by unbalanced class sizes.

4.4 Results

We used an evaluation database of 5518 echolocation calls, from 565 recorded sequences (Table 4.1). In addition, to evaluate the effects of noise being presented to the classifier we collected a database comprising 414 noise samples.

Table 4.1: Summary of call library used for evaluation.

	Seqs	Calls
Barbastella	62	222
Eptesicus	13	306
Myotis	300	3185
Nyctalus	27	512
Pipistrellus	67	829
Plecotus	80	354
Rhinolophus	16	110

When all classes were included in classifier training, the accuracy with which the test set was classified using leave-one-out cross-validation (LOOCV) was high, with a minimum positive predictive power (PPP) of 88.6% for *Eptesicus*, with all other groups above 90% (Table 4.2).

Table 4.2: Confusion matrix for an exhaustively defined set of classes.

Support vector machine classification, evaluated using leave-one-out cross-validation (LOOCV). Positive predictive power (PPP) indicates the percentage of calls that were predicted to be from a genus that were actually of that genus.

	Predicted						
Actual	*Barbastella*	*Eptesicus*	*Myotis*	*Nyctalus*	*Pipistrellus*	*Plecotus*	*Rhinolophus*
Barbastella	210	0	0	0	0	12	0
Eptesicus	0	279	0	27	0	0	0
Myotis	0	0	3183	0	2	0	0
Nyctalus	0	31	0	476	0	5	0
Pipistrellus	0	0	1	0	828	0	0
Plecotus	5	5	0	0	0	344	0
Rhinolophus	0	0	0	0	0	0	110
N	215	315	3184	503	830	361	110
PPP	**97.7**	**88.6**	**100.0**	**94.6**	**99.8**	**95.3**	**100.0**

Table 4.3 shows the results of withholding all *Barbastella* calls during classifier training, but including them during testing. Omitting a class in this way violates the classifier's assumption of an exhaustively defined set of classes. It can be seen that calls from *Barbastella* are misclassified as four other genera, with the majority of calls misclassified as *Plecotus*. As a result, the PPP for Plecotus is reduced from 95.3% to 66.5%.

Table 4.3: Confusion matrix with *Barbastella* excluded during training.

Support vector machine classification, evaluated using leave-one-out cross-validation (LOOCV). *Barbastella* calls are withheld during training, but presented to the classifier during testing. Positive predictive power (PPP) indicates the percentage of calls that were predicted to be from a genus that were actually of that genus.

	Predicted					
Actual	*Eptesicus*	*Myotis*	*Nyctalus*	*Pipistrellus*	*Plecotus*	*Rhinolophus*
Barbastella	42	0	4	8	168	0
Eptesicus	279	0	27	0	0	0
Myotis	0	3183	0	2	0	0
Nyctalus	31	0	476	0	5	0
Pipistrellus	0	1	0	828	0	0
Plecotus	5	0	0	0	344	0
Rhinolophus	0	0	0	0	0	110
Σ	357	3184	507	838	517	110
N	315	3184	503	830	361	110
PPP	**78.2**	**100.0**	**93.9**	**98.8**	**66.5**	**100.0**

Table 4.4 shows the distributions of misclassifications as a result of excluding a single class in turn from the training set, but including it during testing. The average classifier confidence is presented, calculated over all the calls misclassified as a particular genus. It can be seen that when *Barbastella* calls were excluded, they were classified with a high confidence as *Plecotus* (median confidence 0.93). *Barbastella* calls were also misidentified as *Eptesicus*, *Nyctalus* and *Pipistrellus*, although the classifier confidence is lower (median confidence 0.63, 0.44, and 0.53 respectively). Excluding *Eptesicus* calls during training results in their misclassification as four other genera, with the majority and highest confidence being *Nyctalus* (median confidence 0.94). *Myotis* calls were also misclassified as four other genera, with the majority misclassified as *Pipistrellus* (median confidence

0.92). *Nyctalus* calls were misclassified as three other genera, the majority as *Eptesicus* (median confidence 0.8). *Pipistrellus* calls were misclassified as two other genera, the majority as *Myotis* (median confidence 0.99). *Plecotus* calls were misclassified as three other genera, the majority as *Barbastella* (median confidence 0.84). *Rhinolophus* calls were exclusively misclassified as *Myotis* (median confidence 0.71). In general, excluding a class during training led to misclassifications which had falsely high levels of confidence, with median confidence frequently higher than 0.9.

Table 4.4: Confusion matrix showing the distribution of misclassifications on testing a classifier with calls from a class excluded during training.

The median and inter-quartile ranges are given for the probability estimates of each misclassified class; e.g. a classifier not trained with calls from the *Barbastella* class misclassifies them as *Eptesicus*, *Nyctalus*, *Pipistrellus*, and *Plecotus*; 42 calls are misclassified as *Eptesicus*, and the median probability estimate for those calls is 0.63, while the inter-quartile range is 0.23.

	Misclassifications							
excluded class	*Barbastella*	*Eptesicus*	*Myotis*	*Nyctalus*	*Pipistrellus*	*Plecotus*	*Rhinolophus*	Σ
Barbastella	x	42	0	4	8	168	0	222
Median (IQR)		0.63 (0.23)		0.44 (0.06)	0.53 (0.19)	0.93 (0.18)		
Eptesicus	35	x	1	209	0	61	0	306
Median (IQR)	0.57 (0.18)		0.47 (0)	0.94 (0.20)		0.55 (0.19)		
Myotis	0	63	x	0	3102	1	19	3185
Median (IQR)		0.63 (0.34)			0.92 (0.14)	0.18 (0)	0.37 (0.19)	
Nyctalus	0	481	4	x	0	27	0	512
Median (IQR)		0.80 (0.34)	0.49 (0.05)			0.70 (0.22)		
Pipistrellus	114	0	715	0	x	0	0	829
Median (IQR)	0.73 (0.45)		0.99 (0.04)					
Plecotus	246	42	0	66	0	x	0	354
Median (IQR)	0.84 (0.24)	0.53 (0.22)		0.80 (0.22)				
Rhinolophus	0	0	110	0	0	0	x	110
Median (IQR)			0.71 (0)					

Using the soft classifier output and a post-classification threshold, calls can be rejected that have been classified with low confidence. Table 4.5 shows the results of varying the post-classification thresholds. It is evident that increasing the threshold results in more calls being rejected. *Myotis* and *Rhinolophus* are identified with 100% PPP over all thresholds, but all other genera show increasing PPP with increasing threshold. This indicates that the low confidence classifications were those more likely to be misclassified, and were rejected by the threshold procedure. However, sensitivity decreases with increasing threshold, with the rejection of more calls.

Table 4.5: Classifier accuracy in response to a post-classification threshold.

Classifications below the threshold are rejected. Accuracy is assessed in terms of positive predictive power (PPP; percentages of predictions that were actually calls of that species), and sensitivity (percentage of calls correctly classified). The minimum classification rate (i.e. species with the lowest correct identification rate) is highlighted in bold for each threshold. The number of rejected calls left unclassified for each threshold level is shown.

		Genus						
Threshold		*Barbastella*	*Eptesicus*	*Myotis*	*Nyctalus*	*Pipistrellus*	*Plecotus*	*Rhinolophus*
none	**PPP (sensitivity)**	97.7 (94.6)	**88.6 (91.2)**	100 (100)	94.6 (93.0)	99.8 (99.9)	95.3 (97.2)	100 (100)
0.5	**rejected**	0	1	0	8	0	2	0
	PPP (sensitivity)	97.7 (94.6)	**90.3 (91.2)**	100 (100)	94.8 (92.8)	99.8 (99.9)	95.8 (96.9)	100 (100)
0.6	**rejected**	6	13	1	25	0	3	0
	PPP (sensitivity)	97.7 (93.7)	**92.5 (88.2)**	100 (100)	95.3 (91.2)	99.9 (99.9)	97.1 (96.6)	100 (100)
0.7	**rejected**	12	30	1	42	0	9	0
	PPP (sensitivity)	99 (92.8)	**94.9 (85.0)**	100 (100)	96.6 (89.8)	99.9 (99.9)	98.9 (95.8)	100 (100)
0.8	**rejected**	16	42	3	60	2	15	0
	PPP (sensitivity)	99 (91.0)	**95.8 (82.0)**	100 (99.9)	97.2 (87.0)	99.9 (99.6)	98.8 (94.1)	100 (100)
0.9	**rejected**	33	93	7	87	9	38	0
	PPP (sensitivity)	99.5 (84.2)	**98.1 (68.0)**	100 (100)	98.8 (82.8)	99.9 (98.9)	99.4 (88.1)	100 (100)
	N	222	306	3185	512	829	354	110

To investigate the ability of a post-classification threshold to reject noise data, a classifier trained on all genera was tested with noise, at a range of threshold values (Table 4.6). At a threshold of zero all noise files are misclassified as echolocation calls, distributed between the *Myotis*, *Nyctalus* and the majority misclassified as *Plecotus*. 82% of noise files remain misclassified even at a high threshold level of 0.9. Noise files are totally rejected at a threshold level of 0.99, a level that would reject the majority of echolocation calls.

Table 4.6: Confusion matrix showing classification of noise recordings.

The classification of noise files are shown for varying post-classification thresholds. At a threshold of zero, all 414 noise files are misclassified. A threshold of 0.99 is required to correctly reject all noise files.

Threshold		*Barbastella*	*Eptesicus*	*Myotis*	*Nyctalus*	*Pipistrellus*	*Plecotus*	*Rhinolophus*	Σ
0	Noise	0	0	3	27	0	384	0	414
0.5		0	0	3	27	0	375	0	405
0.6		0	0	3	27	0	370	0	400
0.7		0	0	0	24	0	368	0	392
0.8		0	0	0	22	0	358	0	380
0.9		0	0	0	17	0	323	0	340
0.95		0	0	0	7	0	154	0	161
0.97		0	0	0	4	0	29	0	33
0.98		0	0	0	1	0	9	0	10
0.99		0	0	0	0	0	0	0	0

In contrast to probability thresholding, outlier detection proved more effective. The use of outlier detection, in the form of a one-class SVM, protected the classifier from untrained classes. Table 4.7 shows the results of excluding the *Barbastella* class during training. Ten calls were misclassified as *Plecotus*, and all others were rejected by the classifier. When *Barbastella* calls were included during training of a standard classifier without outlier detection, twelve *Barbastella* calls were misclassified as *Plecotus*. This indicates that whilst there is some confusion between *Barbastella* and *Plecotus* due to their similar call types, the outlier detection effectively protects the classifier from untrained classes. The classifier was also presented with noise, and all noise files were correctly rejected by the classifier.

Table 4.8: Confusion matrix for hard classification of calls, followed by outlier rejection by a one-class SVM.

The Barbastella class has been excluded during training, to illustrate the effects of a novel class during testing. In addition, the classifier is tested with noise files. Classifier accuracy for each class is assessed in terms of positive predictive power (PPP; percentages of predictions that were actually calls of that species), and sensitivity (percentage of calls correctly classified). The number of rejected calls is also shown.

	Predicted						
Actual	*Eptesicus*	*Myotis*	*Nyctalus*	*Pipistrellus*	*Plecotus*	*Rhinolophus*	Rejected
Barbastella	0	0	0	0	10	0	212
Eptesicus	279	0	25	0	0	0	2
Myotis	0	3021	0	1	0	0	163
Nyctalus	31	0	458	0	5	0	18
Pipistrellus	0	0	0	787	0	0	42
Plecotus	5	0	0	0	327	0	22
Rhinolophus	0	0	0	0	0	99	11
Noise	0	0	0	0	0	0	414
Σ	315	3021	483	788	342	99	884
N	315	3184	503	830	361	110	
PPP	**88.6**	**100.0**	**94.8**	**99.9**	**95.6**	**100.0**	
sensitivity	**88.6**	**94.9**	**91.1**	**94.8**	**90.6**	**90.0**	

4.5 Discussion

Classifiers implicitly assume that the data used in training is representative of the real world data they are expected to make predictions on. This presents a problem for ecological applications classifying species, where during training some species may be excluded because suitable data was unavailable, or because they were unknown to the researcher. In the absence of publically available echolocation call libraries, researchers must develop their own. Collecting ground truth data for classifying bat echolocation calls is labour intensive and time consuming. Bats must be captured so that they can be identified and recorded on release, or species-specific roosts must be found, enabling recordings of bats of known identity to be recorded emerging from or returning to their roost. Collecting sufficient data for training a classifier from rare species may not be possible without considerable effort, and the species may not even be of interest for the particular applications of the researcher. However, this study has clearly shown that without a comprehensive library of calls, misclassifications will occur in the event of novel calls being presented to the classifier, unless measures are taken to prevent it. Previous echolocation classification studies have rarely considered the implications of untrained classes, as their focus has largely been comparing methods of classification (e.g. Parsons & Jones, 2000; Redgwell *et al.*, 2009). Jones and colleagues (2000) acknowledge that classifiers are limited by the data they are trained with, and that misclassification will result from a species a trained system has not seen before. However, they make no suggestion as to how this can be resolved. Where classifiers have been applied to practical ecological research, calls have been hand-selected for classification, enabling the researcher to filter the calls reaching the classifier, protecting it from novel signals, for example untrained social calls and feeding buzzes (e.g. Davy *et al.*, 2007).

In a fully automated identification system without a human screening process, a range of novel signal types may reach the classifier. By forcing data into the classes predefined during training, species not present during training will be misclassified. In addition, other signals such as social calls or environmental noise will be misclassified as echolocation calls, unless measures are taken to prevent it. Training data for an exhaustive set of classes may not be available, particularly in ecological applications, and some form of outlier detection is necessary. The posterior probability estimates from soft classifiers provide a relative measure of class membership, and can therefore be used to identify calls that are not clearly identifiable as belonging to a single species or species group. Soft classifier outputs are increasingly used in land cover classification from remotely sensed data, where pixels may not comprise homogeneous cover of a single class (Foody, 2001). Despite the extensive overlap in the spectral and temporal parameters in the echolocation calls of some species, soft classifier outputs have rarely been made use of. We used a simple post-classification processing of the soft output to reject classifications falling below a confidence threshold. This procedure effectively rejected calls falling close to the boundary of known classes that had been classified with low certainty. This improved the positive predictive power for many species. However, we found probability estimates to be routinely over confident in the face of outliers, data significantly different from which the classifier had been trained with. Novel data are classified arbitrarily according to where they fall in relation to the decision boundaries created during training. It appears that where data fall far from a decision boundary, they are classified with falsely high confidence. This may be in part due to probability estimates being normalised to sum to unity over all classes for any one prediction. This means data are never classified as having a low probability over all classes, and renders probability thresholding an ineffective method for rejecting novel data.

In contrast to post-classification thresholding, one-class support vector machines, a dedicated form of outlier detection, robustly rejected novel data. Noise and species not represented in the training data were successfully rejected as outliers, preventing misclassification. With a conventional classifier, considerable effort must be directed on obtaining training data for an exhaustive set of classes, for example all species in a geographic area, to ensure misclassifications are not made. However, only a subset of those species may be of interest to the researcher. In addition, in some cases, for example rare species, obtaining representative training data may not be feasible. Here, it is suggested that a one-class-classification approach could provide a solution. The one-class approach could be particularly useful when interest focuses on a specific species or species group, and the goal of classification is not to produce full inventories. In this case, researchers could rapidly obtain training data for the target group, and use outlier detection to filter calls from other species. One-class SVMs are simple to apply, but their use comes at the cost of falsely rejecting some known data. However, in applications where the number of classes cannot be known, the alternative is misclassification of novel data, or significant human intervention to manually filter the input to the classifier. The use of outlier detection methods enables the reliable application of fully automated species identification systems, removing the bottleneck of manual analysis, increasing the scope of field studies.

Once in practical use in the field, the classifications rejected as outliers from an automated species identification system may prove informative. For example, clustering techniques such as k-means (MacKay, 2003) may be applied to the outlying data, to group similar call types and to try and identify the number of unknown classes that are being encountered. This technique could prove particularly insightful applying acoustic methods in species rich areas, where only a proportion of the species in bat assemblages may be known. Acoustic methods could be applied to survey and monitor those species for which training data is available, with the accumulating outlying data used to identify morphospecies not present in the training data. Morphospecies richness as identified by acoustic methods may be a valuable first step in understanding the bat biodiversity of a region.

4.6 References

Adams, M. D., Law, B. S., & Gibson, M. S. (2010). Reliable Automation of Bat Call Identification for Eastern New South Wales, Australia, Using Classification Trees and AnaScheme Software. *Acta Chiropterologica, 12*(1), 231-245.

Armitage, D. W., & Ober, H. K. (2010). A comparison of supervised learning techniques in the classification of bat echolocation calls. *Ecological Informatics, 5*(6), 465-473.

Barclay, R. M. R., Fullard, J. H., & Jacobs, D. S. (1999). Variation in the echolocation calls of the hoary bat (*Lasiurus cinereus*): influence of body size, habitat structure, and geographic location. *Canadian Journal of Zoology, 77*(4), 530-534.

Barratt, E. M., Deaville, R., Burland, T. M., Bruford, M. W., Jones, G., Racey, P. A., & Wayne, R. K. (1997). DNA answers the call of pipistrelle bat species. *Nature, 387*, 138-139.

Bradter, U., Thom, T. J., Altringham, J. D., Kunin, W. E., & Benton, T. G. (2011). Prediction of National Vegetation Classification communities in the British uplands using environmental data at multiple spatial scales, aerial images and the classifier random forest. *Journal of Applied Ecology, 48*(4), 1057-1065.

Broders, H., & Findlay, C. (2004). Effects of clutter on echolocation call structure of Myotis septentrionalis and M. lucifugus. *Journal of Mammalogy, 85*(2), 273-281.

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery, 2*(2), 121–167.

Chang, C. C., & Lin, C. J. (2001). LIBSVM: a library for support vector machines. Retrieved from http://en.scientificcommons.org/43278956

Chesmore, D. (2004). Automated bioacoustic identification of species. *Anais da Academia Brasileira de Ciências, 76*(2), 436–440.

Davy, C. M., Russo, D., & Fenton, M. B. (2007). Use of native woodlands and traditional olive groves by foraging bats on a Mediterranean island: consequences for conservation. *Journal of Zoology, 273*(4), 397-405.

Erik de Castro Lopo. Libsndfile, a library for reading and writing files containing sampled sound. http://www.mega-nerd.com/libsndfile/, 2011. Last visited: July 2 2011.

Fenton, M. B. (2003). Eavesdropping on the echolocation and social calls of bats. *Mammal Review, 33*(3-4), 193-204.

Foody, G. M. (2001). The effect of a non-exhaustively defined set of classes on neural network classifications. *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium.*

Garzón, M. B., Blazek, R., Neteler, M., Dios, R. S. D., Ollero, H. S., & Furlanello, C. (2006). Predicting habitat suitability with machine learning models: The potential area of Pinus sylvestris L. in the Iberian Peninsula. *Ecological Modelling, 197*(3-4), 383-393.

Gaston, K. J., & O'Neill, M. a. (2004). Automated species identification: why not? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 359*(1444), 655-67.

Gudmundsson, S., Runarsson, T. P., & Sigurdsson, S. (2005). Automatic Sleep Staging using Support Vector Machines with Posterior Probability Estimates. *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06), 2*, 366-372.

Helversen, O. V., K.-G., H., F., M., a., N., M., V., & P., G. (2001). Cryptic mammalian species: a new species of whiskered bat (Myotis alcathoe n. sp.) in Europe. *Naturwissenschaften, 88*(5), 217-223.

Hughes, A. C., Satasook, C., Bates, P. J. J., Soisook, P., Sritongchuay, T., Jones, G., & Bumrungsri, S. (2010). Echolocation Call Analysis and Presence-Only Modelling as Conservation Monitoring Tools for Rhinolophoid Bats in Thailand. *Acta Chiropterologica, 12*(2), 311-327.

Hurlbert, S. H. (1984). Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs, 54*(2), 187.

Jones, G., Vaughan, N., & Parsons, S. (2000). Acoustic identification of bats from directly sampled and time expanded recordings of vocalizations. *Acta Chiropterologica, 2*(2), 155–170.

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.

MacKay, D. (2003). Chapter 20. An Example Inference Task: Clustering. *Information Theory, Inference and Learning Algorithms.* Cambridge University Press. pp. 284–292.

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.

Le Cun, Y., Jackel, L., Boser, B., Denker, J., Graf, H., Guyon, I., Henderson, D., *et al.* (1989). Handwritten digit recognition: Applications of neural network chips and automatic learning. *Communications Magazine, IEEE*, *27*(11), 41–46.

Lundy, M., Montgomery, I., & Russ, J. (2010). Climate change-linked range expansion of Nathusius' pipistrelle bat, Pipistrellus nathusii (Keyserling & Blasius, 1839). *Journal of Biogeography*, *37*(12), 2232-2242.

Morris, C. W., Autret, A., & Boddy, L. (2001). Support vector machines for identifying organisms — a comparison with strongly partitioned radial basis function networks. *Ecological Modelling*, *146*(1-3), 57-67.

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.

O'Farrell, M. J., Corben, C., & Gannon, W. L. (2000). Geographic variation in the echolocation calls of the hoary bat (Lasiurus cinereus). *Acta Chiropterologica*, *2*(2), 185–196.

Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine learning methods without tears: a primer for ecologists. *The Quarterly review of biology*, *83*(2), 171-93.

Papadatou, E., Butlin, R. K., & Altringham, J. D. (2008). Identification of bat species in Greece from their echolocation calls. *Acta Chiropterologica*, *10*(1), 127-143.

Parsons, S., & Jones, G. (2000). Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. *J. Exp. Biol*, *203*, 2641-56.

Preatoni, D. G., Nodari, M., Tosi, G., & Wauters, L. U. C. A. (2005). IDENTIFYING BATS FROM TIME-EXPANDED RECORDINGS OF SEARCH CALLS: COMPARING CLASSIFICATION METHODS. *Journal of Wildlife Management*, *69*, 1601-1614.

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.

Schwenker, F. (2000). Hierarchical support vector machines for multi-class pattern recognition. *KES'2000. Fourth International Conference on*

Knowledge-Based Intelligent Engineering Systems and Allied Technologies.

Skowronski, M. D., & Harris, J. G. (2006). Acoustic detection and classification of microchiroptera using machine learning: lessons learned from automatic speech recognition. *J. Acoust. Soc. of Am*, *119*, 1817-1833.

Spitzenberger, F., Strelkov, P. P., Winkler, H., & Haring, E. (2006). A preliminary revision of the genus Plecotus (Chiroptera, Vespertilionidae) based on genetic and morphological results. *Zoologica Scripta*, *35*(3), 187-230.

Tax, D. M. J., & Duin, R. P. W. (1999). Support vector domain description. *Pattern Recognition Letters*, *20*(11-13), 1191-1199.

Tax, D., & Duin, R. (2004). Support vector data description. *Machine Learning*, *54*(4), 45-66.

Vapnik, V. The Nature of Statistical Learning Theory. Springer-Verlag, New York, 1995.

Chapter 5: Species-specific swarming activity using automated echolocation call loggers

5.1 Abstract

We used acoustic loggers to monitor species-specific bat activity at Link Pot, part of a remote cave system in the north of England, over a period of three months. In total, >20,000 audio files were analysed to investigate the temporal patterns in activity. The use of automated methods allowed the rapid analysis of the acquired data and the separation of species with high accuracy in an objective, repeatable way. We showed that for monitoring purposes, acoustic logging is a viable alternative to catching, which is a labour intensive and intrusive technique that need only be used when necessary. We also demonstrate the use of generalised additive models (GAM), a flexible non-linear statistical technique, to produce a quantitative phenological model of late autumn swarming, relating temporal bat activity and environmental variables.

5.2 Introduction

During the autumn many species of bats gather at underground sites such as caves and mines, in an activity known as 'autumn swarming'. Many hundreds of bats may visit a single site each night, arriving from a large catchment area, and staying for only a few hours before returning to their day roosts (e.g. Fenton, 1969; Parsons & Jones, 2003; Rivers *et al.*, 2006). Within this time the bats fly in and around the entrances to underground sites, often chasing other individuals in what can be a spectacular social display. The most widely accepted function of swarming is that it is a mating event (Thomas *et al.*, 1979; Rivers *et al.*, 2005; Furmankiewicz & Altringham, 2007), although it may also play an important role for the bats in assessing hibernacula before the onset of winter (e.g. Fenton, 1969; Veith *et al.*, 2004). The species that take part in this activity belong to the genera *Myotis*, *Plecotus* and *Barbastella* and include some of Britain's rarest mammal species, including Bechstein's bat (*Myotis bechsteinii*) and the newly discovered Alcathoe bat (*M. alcathoe*). Swarming individuals are typically faithful to a single location (Glover and Altringham, 2008), and given the large number of bats and the rarity of some of the species involved, these sites have a vital role to play in conservation (Parsons *et al.*, 2003a). With the bats flying in to sites from such large catchment areas, swarming also offers an opportunity to monitor bat populations at the landscape scale.

The identification and monitoring of swarming sites is challenging. Answering many of the ecological questions surrounding swarming has necessitated the capture and recapture of ringed bats, helping provide estimates as to the numbers of individuals visiting a site, and the catchment areas involved (e.g. Parsons & Jones, 2003; Rivers *et al.*, 2006). Whilst these studies have provided an invaluable insight into the use of swarming sites by bats, capture is an intrusive and extremely labour intensive survey method that can only be carried out by highly trained individuals. Small, portable devices that log the echolocation calls of bats have been used both to discover and monitor sites without disturbing the natural activity of the bats (Parsons *et al.*, 2003b; Glover & Altringham, 2008). In addition, light barrier systems that log the movement of bats as they break carefully placed beams, have also been applied to swarming studies (e.g. Berková & Zukal, 2010). Both types of activity logger can be left for extended time periods to collect large quantities of data. Whilst they give only an index of activity (a single bat may trigger a logger multiple times), logged activity at a site has been shown to be strongly correlated with the number of bats caught (Parsons *et al.*, 2003b). Loggers remain the most practical method of monitoring long-term population trends at large numbers of underground sites (Glover & Altringham, 2008). However, these devices have been limited to monitoring the overall bat activity, with no ability to discriminate between species. With advances in bat detector technology it is now feasible to collect long term monitoring data that stores high quality digital recordings of the bats' calls, from which species information can subsequently be extracted.

In this study we applied automated acoustic analysis techniques to extract species-specific data on swarming activity at a known swarming site. Ecological data are frequently complex and non-linear in nature, making it challenging to identify meaningful patterns using traditional statistical techniques such as linear regression. The use of simple linear statistical techniques is still widespread in ecology, but increasingly ecologists are applying flexible non-linear models such as Generalised additive models (GAM; e.g. Baker, 2008; Maloney *et al.*, 2011). GAMs offer a flexible modelling approach that retains the interpretability of standard linear models, fitting smooth functions to identify nonlinearities in the relationship between predictor and response variables (Wood, 2006). We demonstrate the use of GAM to produce a quantitative phenological model of late autumn swarming, relating temporal bat activity and environmental variables.

5.3 Methods

5.3.1 Study site

The study was carried out at the entrance to Link Pot, part of the Ease Gill cave system on the Cumbria/Lancashire border (OS grid reference SD668803; Figure 5.1).

Figure 5.1: A view over Casterton Fell with the Ease Gill cave system below. The entrance to Link Pot is situated approximately in the centre of the picture.

5.3.2 Acoustic data collection

A direct sampling detector (Pettersson D500X, www.batsound.com) was positioned within 2 m of the narrow, vertical entrance to the cave, and left to collect data for a total of 48 nights during August-October 2010. The detector was secured to a tripod ~1m above ground level. The detector was set to trigger on detecting ultrasound, and subsequently record for a period of 1 s. Recordings were stored internally at 16bits/500kHz on removable compact flash cards. The effective frequency response of the system was

10-250 kHz which enabled all British species of bat to be recorded. The detector trigger threshold and gain were kept constant throughout the study period to maintain consistency in sensitivity. Sampling was not carried out in periods where heavy rain was forecast, as the detector had limited weather proofing.

5.3.3 Acoustic analysis

Echolocation calls were automatically extracted and classified to species level using the methodology outlined in Chapters 2-4. At the start of each WAV file the D500X detector embedded the date and time as the number of seconds elapsed since midnight Coordinated Universal Time (UTC) of Thursday, January 1st, 1970 (a convention known as Unix time). Dates and times were subsequently extracted during analysis using a custom written C++ function (Appendix C.1). For each recorded file, all individual calls detected were classified to genus, with a further stage of classification for *Myotis* and *Pipistrellus* calls to species level. The results from a single file were converted into a list of presences and absences, with a record for a presence requiring at least 3 classified calls with posterior probabilities obtained from the classification algorithm of >0.5. This procedure enabled a recorded sequence to be attributed to more than one group (multiple species may fly in the same airspace and be recorded concurrently). Such an approach also prevents a single, or pair of misclassified calls, from being incorrectly attributed as a presence. A log of absences was made to help ensure that in sampling multiple nights it was possible to separate (using an automated script), an 'absence of presence' or true absence, from the absence of data due to a period without data collection. The automated analysis procedure stored the presence/absence log, date and time for all recordings in a simple text file format (comma separated file; CSV). Text files were then imported into the statistical program R (R Development Core Team, 2011), where scripts were used for aggregating and plotting data, and for statistical analysis.

Two main R scripts were used to summarise bat activity from the raw presence absence data extracted from each sound file over the study period. We summarised activity by night, producing a nightly activity index for each species as the sum of passes for that species. This was achieved using the period.apply function from the package xts (Ryan & Ulrich, 2011; Appendix C.2). Additionally, we summarised activity measured as hours after sunset, summed over the swarming season. This script automatically calculated the time of sunset for the study site, for each day of sampling, using the function sunriset from the package maptools (Lewin-Koh & Bivand, 2011; Appendix C.3).

5.3.4 Environmental variables

Weather data were obtained from the closest publically available records to the study site, at Storth (Weather Underground, http://www.wunderground.com/), situated ~20 km Link Pot. Percentage of moon face illuminated was obtained from the Solar and Moon Calculator (Dexter, 2011; http://ftdevelop.pcriot.com/index.php). Data are shown in Appendix C.4.

5.3.5 Statistical analysis

Previous studies have indicated that activity at swarming sites is highest throughout August-October (the swarming period), with a peak of activity in September (Parsons *et al.*, 2003a; Glover & Altringham, 2008). We used penalized regression splines in generalised additive models (GAM), to model the non-linear development in nightly bat activity over the swarming period. Variance inflation factors (VIF) were used to identify and drop correlated variables prior to analysis (Zuur *et al.*, 2009). VIF values were calculated using the corvif function from the AED package (Zuur, 2010). All GAM models were fitted in R using the package mgcv (Wood, 2011) assuming a negative binomial distribution. For count data, Poisson or negative binomial models should be fitted in preference to log-transforming the data to satisfy parametric test assumptions (O'Hara & Kotze, 2010). Predictors were fitted using smooth terms and 4 degrees of freedom.

5.4 Results

5.4.1 Total activity and species composition

7699 bat passes (recorded sequences containing >2 calls) from 47 nights of sampling were classified (Tables 5.1-5.3).

Table 5.1: Bat passes classified to genus.

The number of passes classified (N) and the percentage of the total passes (%) for each genera classified.

	N	%
Myotis	6250	81.2
Nyctalus	363	4.7
Pipistrellus	1031	13.4
Plecotus	55	0.7
TOTAL	7699	100.0

Table 5.2: *Pipistrellus* passes.

The number of passes classified (N) and the percentage of the total passes (%) for *Pipistrellus* bats classified.

	N	%
P. pipistrellus	864	83.8
P. pygmaeus	167	16.2
TOTAL	1031	100.0

Table 5.3: *Myotis* passes.

The number of passes classified (N) and the percentage of the total passes (%) for *Myotis* bats classified.

	N	%
M. alcathoe	57	0.9
M. brandtii/ M. mystacinus	3090	49.4
M. daubentonii	695	11.1
M. nattereri	2408	38.5
TOTAL	6250	100.0

Activity was dominated by *Myotis* bats, comprising 81.2% of total bat passes. *Pipistrellus* species accounted for of the bulk of the remaining activity, with >10% of the total passes. Small amounts of *Nyctalus* and *Plecotus* activity were also recorded, at 4.7% and 0.7% of the total activity respectively. The grey long-eared bat (*Plecotus austriacus*) has a very restricted distribution in Britain (Harris *et al.*, 1995), and the *Plecotus* activity recorded in this study can confidently be attributed to the brown long-eared bat (*P. auritus*). Calls attributed to the genus *Nyctalus* were long duration and low frequency, characteristic of *N. noctula*.

Activity within the *Myotis* was dominated by the species group *M. brandtii/M. mystacinus* and *M. nattereri*, with 49.4% and 38.5% of the total activity respectively. A previous study carried out catching at swarming sites in the wider study area, and found both *M. brandtii* and *M. mystacinus* (n=103 & n=76 respectively; Glover & Altringham, 2008), so the activity found in this study attributed to the species group is likely to be split fairly evenly between the two species, making *M. nattereri* the likely dominant swarming species. *M. daubentonii* accounted for >10% of the total *Myotis* activity, and a very small number of passes were attributed to *M. alcathoe* (0.9% of the total). The majority of *Pipistrellus* activity was *P. pipistrellus*, with *P. pygmaeus* occurring in lower proportions (83.8% and 16.2% respectively).

Species composition changed throughout the three months of the swarming season, with a transition of dominance from *M. brandtii/M. mystacinus* to *M. nattereri* (Figure 5.2). The proportion of *M. daubentonii* decreased steadily throughout the season, and the proportion of *M. alcathoe* was consistently low.

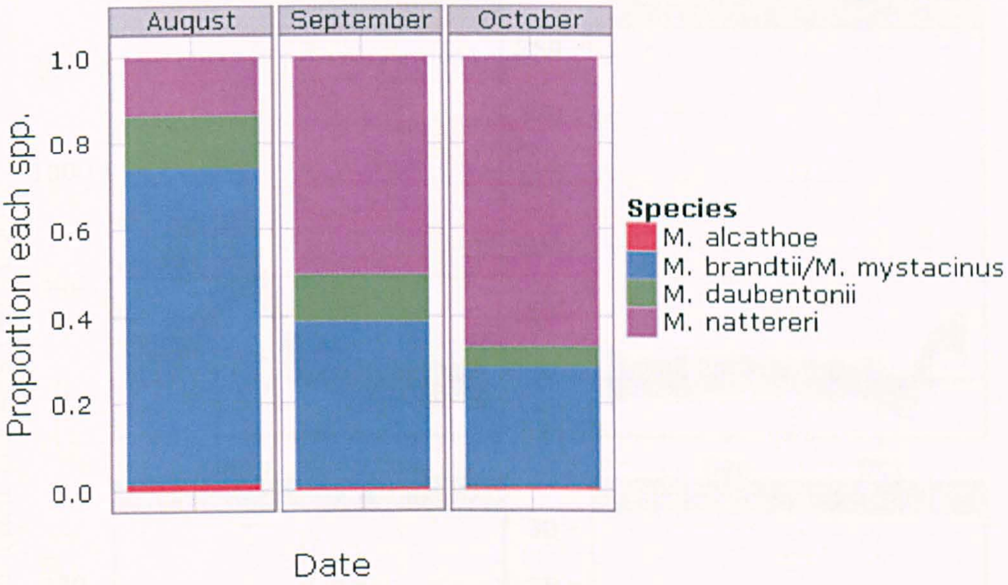


Figure 5.2: ***Myotis*** **species composition by month of acoustic sampling.**

5.4.2 Changes in activity through the swarming season

Myotis, *Pipistrellus* and *Plecotus* activity peaked in early September, whilst *Nyctalus* activity was consistently low until a late peak in October, dominated by a single night (14th October; Figure 5.3). Visual inspection of sound files revealed *Nyctalus* feeding buzzes (highly increased call repetition rate), consistent with foraging activity. *Myotis* and *Plecotus* activity remained low throughout October, but relative *Pipistrellus* activity showed signs of continuation into the second half of October. There was considerable night to night variation in activity in all groups.

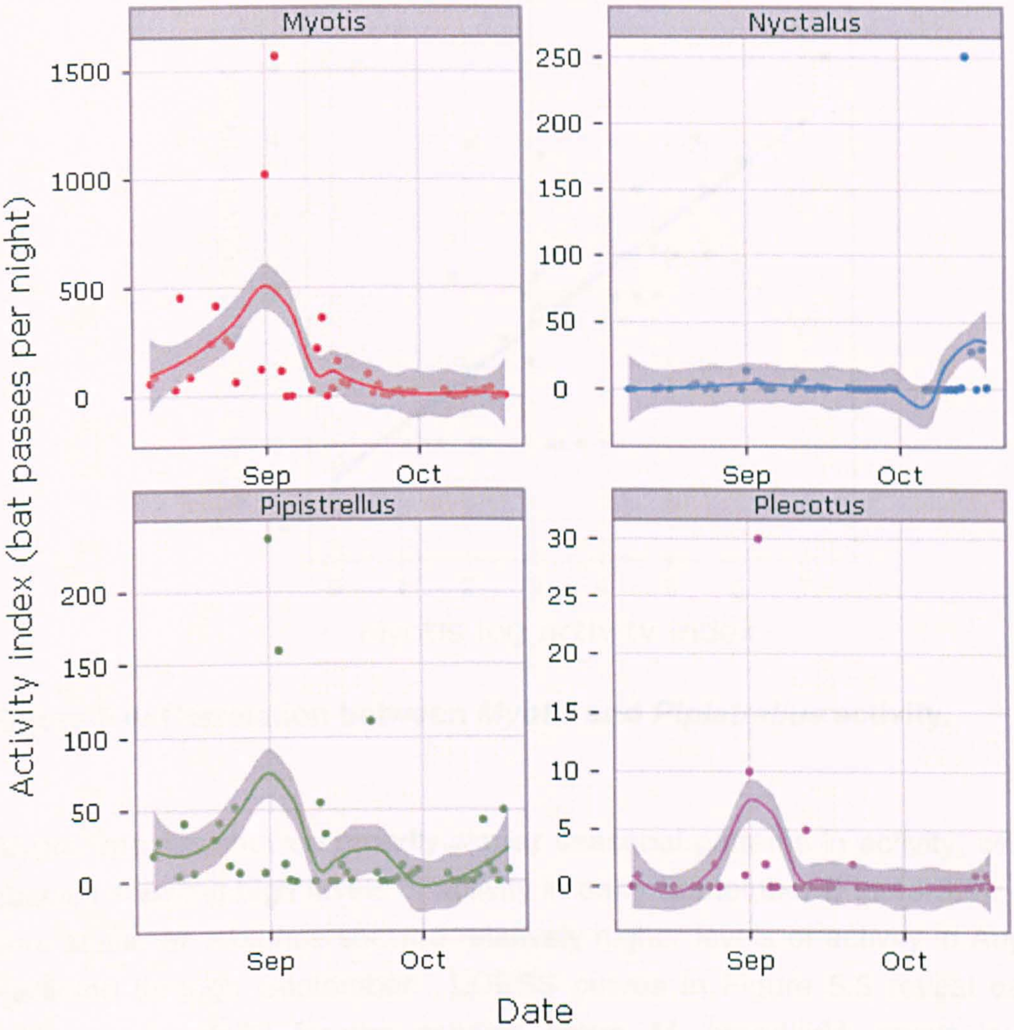


Figure 5.3: Seasonal activity.

Myotis, *Nyctalus*, *Pipistrellus* and *Plecotus* nightly activity over the swarming period (August-October).

The similarity between seasonal *Myotis* and *Pipistrellus* activity was unexpected, as catching records indicate *Pipistrellus spp.* do not engage in autumn swarming activity at the site (Glover & Altringham, 2008). There was a significant moderate to strong positive correlation between *Myotis* and *Pipistrellus* activity (Pearson's correlation coefficient = 0.742, d.f. = 45, $p<0.001$; Figure 5.4).

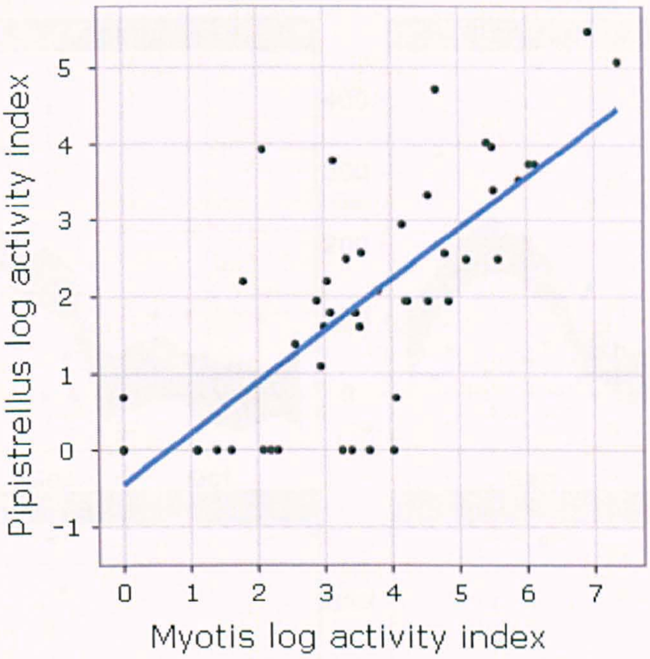


Figure 5.4: Correlation between *Myotis* and *Pipistrellus* activity.

Myotis species showed broadly similar seasonal patterns in activity, with all species showing high levels of activity in early September (Figure 5.5). At a finer scale, *M. alcathoe* showed relatively higher levels of activity in August, declining through September. LOESS curves in Figure 5.5 reveal earlier patterns of activity for the species group *M. brandtii/M. mystacinus* in comparison to *M. daubentonii*, and for *M. daubentonii* compared to *M. nattereri*.

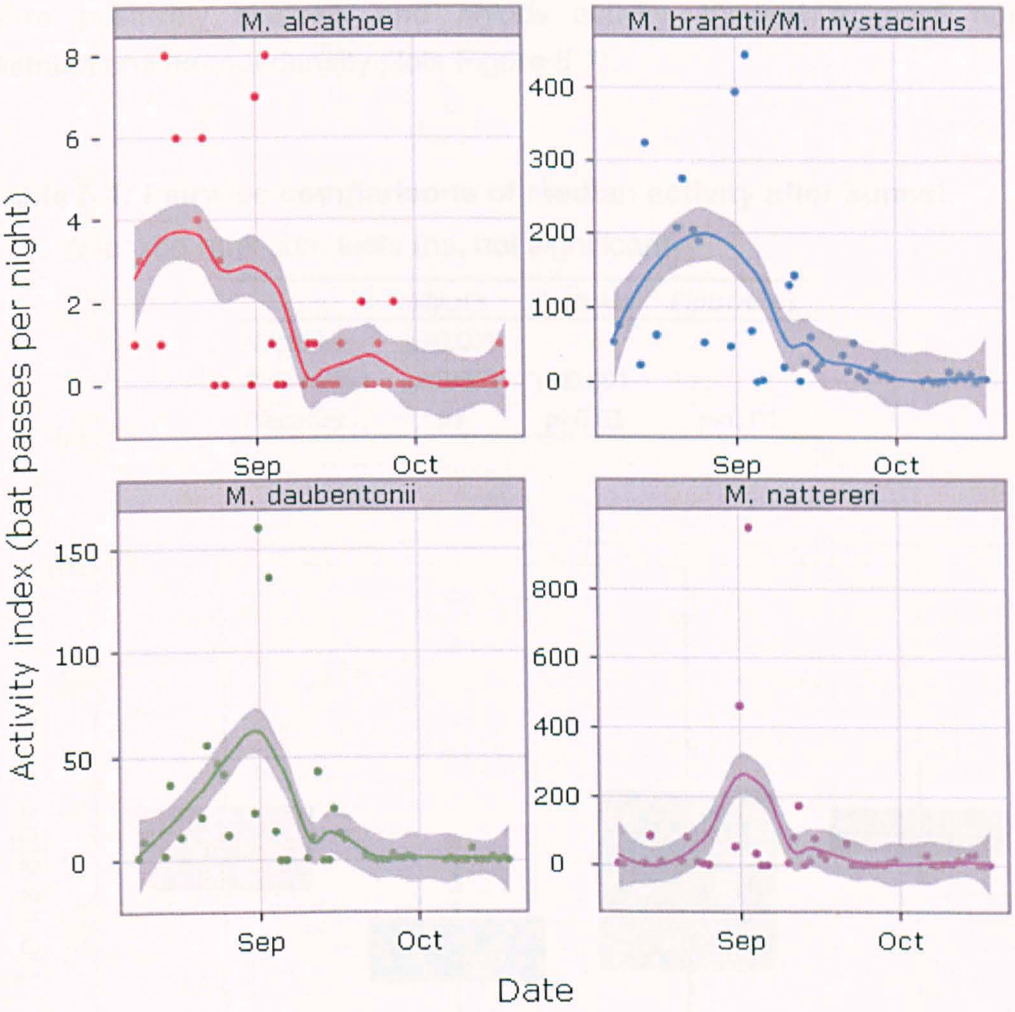


Figure 5.5: Seasonal *Myotis* activity.

Myotis spp. nightly activity over the swarming period (August-October).

5.4.3 Changes in activity through the night

The median time of activity after sunset was significantly different between genera (Kruskal-Wallis H= 575.5, df = 3, $p < 0.001$), with all pairwise tests significant ($p < 0.01$) except between *Myotis* and *Plecotus* (Wilcoxon rank sum tests, using Bonferroni adjustment for multiple tests; Table 5.4). Median *Nyctalus* and *Pipistrellus* activity were approximately 2 and 3 hours after sunset respectively, with median *Myotis* and *Plecotus* activity both approximately 4 hours after sunset (Figure 5.6). The overall nightly distribution of activity was significantly different at the genus level between all groups (pair wise two-sample Kolmogorov–Smirnov tests with Bonferroni correction for multiple tests, Table 5.5). *Pipistrellus* and *Nyctalus* activity were positively skewed, and *Myotis* activity showed a more normal distributions (kernel density plots Figure 5.7).

Table 5.4: Pairwise comparisons of median activity after sunset. Wilcoxon rank sum tests (ns, not significant).

	Myotis	*Nyctalus*	*Pipistrellus*
Nyctalus	p<0.001		
Pipistrellus	p<0.001	p<0.001	
Plecotus	ns	p<0.01	p<0.01

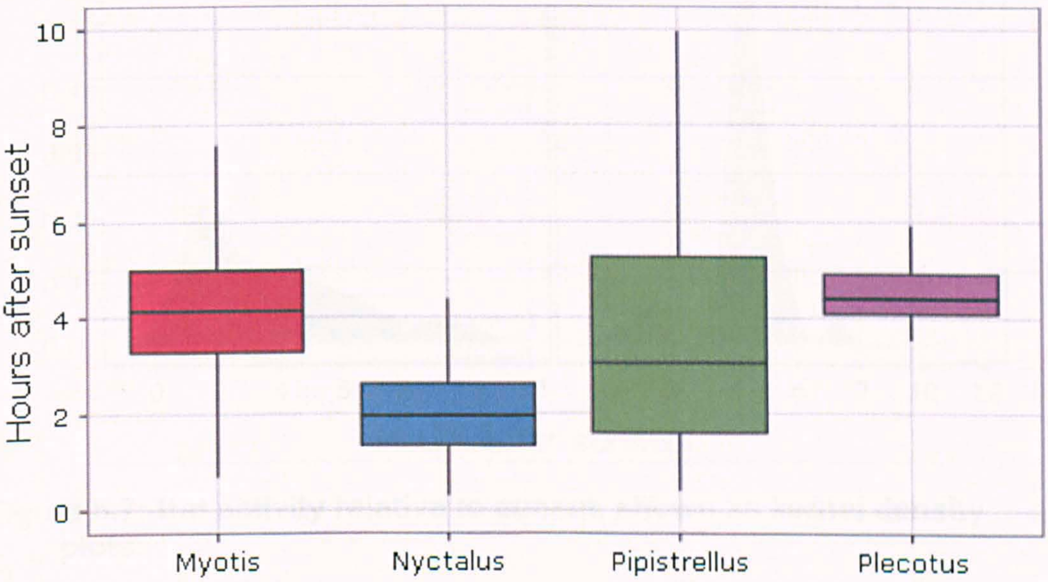


Figure 5.6: Boxplot of time of activity (bat passes) relative to sunset.

Table 5.5: Comparison of nightly distribution of activity. Pairwise two-sample Kolmogorov-Smirnov tests.

	Myotis		*Nyctalus*		*Pipistrellus*	
	D	p-value	D	p-value	D	p-value
Nyctalus	0.65	p<0.01				
Pipistrellus	0.32	p<0.01	0.35	p<0.01		
Plecotus	0.26	p<0.01	0.81	p<0.01	0.47	p<0.01

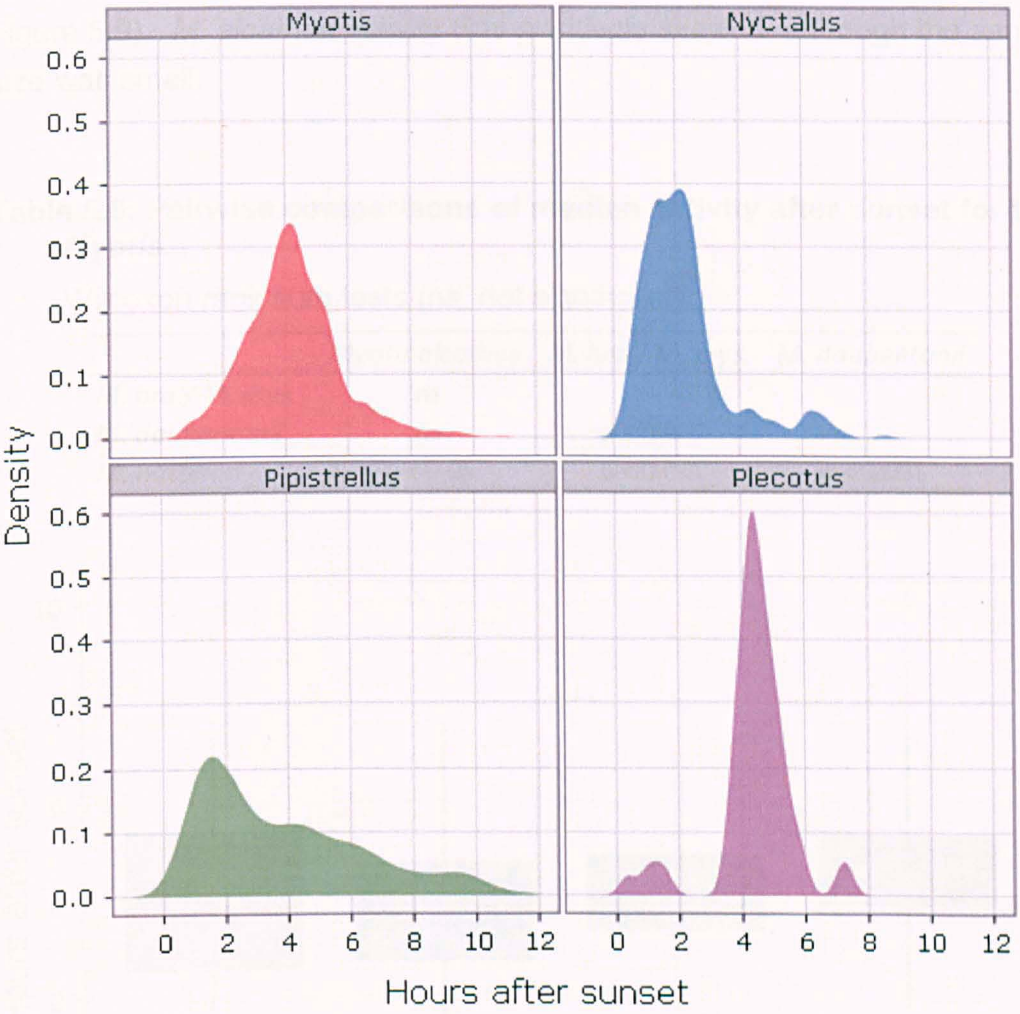


Figure 5.7: Bat activity relative to sunset, shown as kernel density plots.

The median time of activity after sunset was significantly different between the *Myotis* species (Kruskal-Wallis H= 459.3, d.f. = 3, p < 0.001), with all pairwise tests significant except between *M. alcathoe* and *M. brandtii/M. mystacinus* and between *M. alcathoe* and *M. daubentonii* (Wilcoxon rank sum tests, using Bonferroni adjustment for multiple tests; Table 5.6, Box plots shown in Figure 5.8). Between the *Myotis* species the nightly distribution of activity was significantly different between all pairs of species except for between *M. alcathoe* and *M. daubentonii*, and between *M. alcathoe* and the species group *M. brandtii/M. mystacinus* (pair wise two-sample Kolmogorov–Smirnov tests with Bonferroni correction for multiple tests, Table 5.7). The data for *M. brandtii/M. mystacinus* were more positively skewed than *M. daubentonii* and *M. nattereri* (kernel density plots

Figure 5.9). *M. alcathoe* activity was positively skewed, although the sample size was small.

Table 5.6: Pairwise comparisons of median activity after sunset for the *Myotis*.

Wilcoxon rank sum tests (ns, not significant).

	Myotis alcathoe	*M. bra./M. mys.*	*M. daubentonii*
M. bra./M. mys.	ns		
M. daubentonii	ns	$p < 0.001$	
M. nattereri	$p < 0.05$	$p < 0.001$	$p < 0.001$

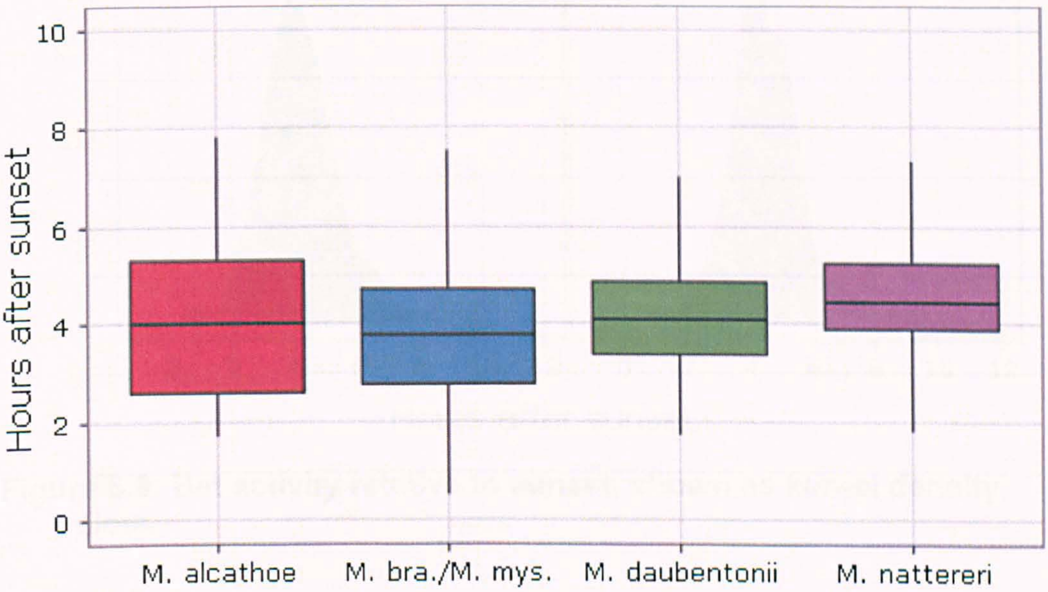


Figure 5.8: Boxplot of time of activity (bat passes) relative to sunset.

Table 5.7: Comparison of nightly distribution of activity for the *Myotis*.

Pairwise two-sample Kolmogorov-Smirnov tests.

	Myotis alcathoe		*M. bra./M. mys.*		*M. daubentonii*	
	D	p-value	D	p-value	D	p-value
M. bra./M. mys.	0.12	$p<0.01$				
M. daubentonii	0.21	$p<0.01$	0.16	$p<0.01$		
M. nattereri	0.30	$p<0.01$	0.29	$p<0.01$	0.18	$p<0.01$

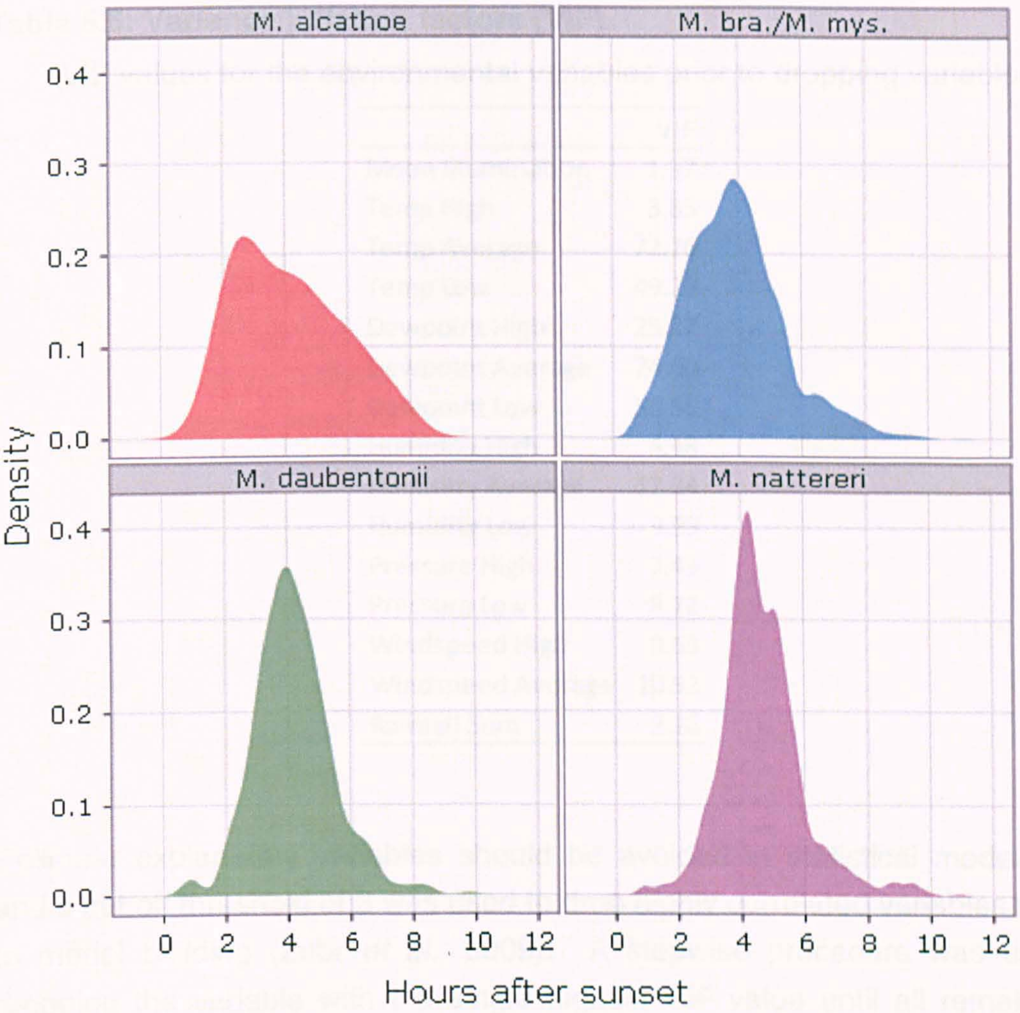


Figure 5.9: Bat activity relative to sunset, shown as kernel density plots.

5.4.4 Generalised additive modelling (GAM)

An initial inspection of the environmental variables was undertaken to identify outliers. Average daily wind speed and rainfall sum were subsequently square root transformed because of large values. Variance inflation factors (VIF) revealed that many environmental variables were highly correlated (Table 5.6).

Table 5.6: Variance inflation factors (VIF).

VIF values for the environmental variables prior to dropping variables.

	VIF
Moon illumination	1.37
Temp High	3.85
Temp Average	72.76
Temp Low	49.28
Dewpoint High	25.27
Dewpoint Average	74.60
Dewpoint Low	50.55
Humidity High	4.48
Humidity Average	17.24
Humidity Low	9.95
Pressure High	7.43
Pressure Low	8.72
Windspeed High	9.63
Windspeed Average	10.92
Rainfall Sum	2.28

Collinear explanatory variables should be avoided in statistical modelling, and a cut off threshold of 3 was used to drop highly correlated variables prior to model building (Zuur *et al.*, 2009). A stepwise procedure was used, dropping the variable with the single highest VIF value until all remaining variables had VIFs less than the threshold of 3. This process removed eight of the environmental variables leaving a final set of seven variables for subsequent GAM analysis (Table 5.7). Temp high showed a seasonal trend, which led us to create a further variable, residual temperature, taken as the residuals from a GAM fitted to temp high.

Table 5.7: Variance inflation factors (VIF).

VIF values for the final environmental variables.

	VIF
Moon illumination	1.17
Temp High	1.29
Humidity High	1.32
Humidity Low	1.52
Pressure High	1.30
Windspeed Average	1.40
Rainfall Sum	1.62

GAM explained 60.6% of the deviance of *M. nattereri* nightly activity over the season. There was a significant non-linear effect of day on activity levels ($p < 0.001$), reflecting a seasonal change in swarming behaviour. In addition, percentage moon illumination ($p < 0.001$), average wind speed ($p < 0.01$) and daily humidity high ($p < 0.05$) had significant effects on activity (Figure 5.10). No significant relationships were found between activity and daily humidity low, daily precipitation sum, daily pressure high, daily temperature high, or residual temperature.

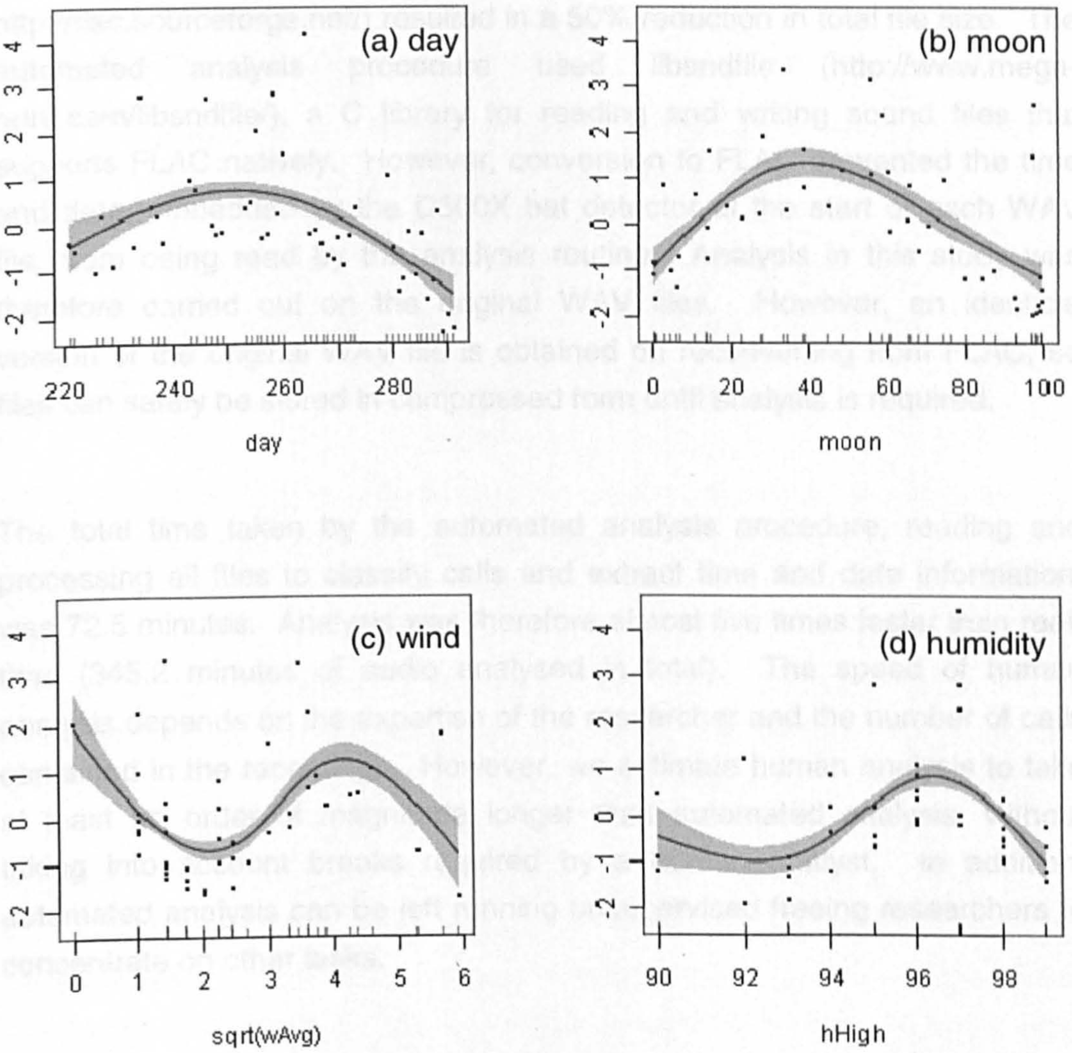


Figure 5.10: GAM plot for Myotis nattereri nightly activity.

Plots show the effects of: (a) day of the year: day, (b) moon illumination, moon (%), (c) daily average wind speed, sqrt(wAvg) (km/h) and (d) daily high humidity, hHigh (%). Rug plots at the base of each scatter plot show the observed values of each explanatory variable. The solid line in each plot shows the estimate of the smooth function, with 95% confidence limits represented by the shaded area. Note that average wind speed was square root transformed to reduce variance heterogeneity.

5.4.5 Data compression and computational cost of automated analysis

This study generated a total of 20,713 one second 16bit/500kHz WAV files, requiring 19.3 GB in total storage space. Lossless compression of the audio files using the Free Lossless Audio Codec (FLAC;

http://flac.sourceforge.net/) resulted in a 50% reduction in total file size. The automated analysis procedure used libsndfile (http://www.mega-nerd.com/libsndfile/), a C library for reading and writing sound files that supports FLAC natively. However, conversion to FLAC prevented the time and date, embedded by the D500X bat detector at the start of each WAV file, from being read by the analysis routine. Analysis in this study was therefore carried out on the original WAV files. However, an identical version of the original WAV file is obtained on reconverting from FLAC, so files can safely be stored in compressed form until analysis is required.

The total time taken by the automated analysis procedure, reading and processing all files to classify calls and extract time and date information, was 72.5 minutes. Analysis was therefore almost five times faster than real-time (345.2 minutes of audio analysed in total). The speed of human analysis depends on the expertise of the researcher and the number of calls contained in the recordings. However, we estimate human analysis to take at least an order of magnitude longer than automated analysis, without taking into account breaks required by a human analyst. In addition, automated analysis can be left running unsupervised freeing researchers to concentrate on other tasks.

5.5 Discussion

In addition to activity from swarming bats of the genera *Myotis* and *Plecotus*, notable activity was recorded from *Pipistrellus* and *Nyctalus* bats. These species are very rarely caught at swarming sites (*e.g.* Glover and Altringham, 2008), and have a different mating strategy (McCracken & Wilkinson, 2000), indicating that they were foraging on site and their calls were of sufficient intensity to trigger the acoustic logger.

57 passes (0.9% of total *Myotis* activity) were attributed to *M. alcathoe*, only recently identified in Britain (Jan *et al.*, 2010). *M. alcathoe* has been caught at a swarming site in the North York Moors, ca. 90 km to the east, but there are currently no catching records for the present site. *M. alcathoe* has a distinctive call among the *Myotis* (with a call end frequency typically > 40 kHz), and the automated classifier used in this study had a high positive predictive power for this species. Some caution must be used in attributing a species presence from acoustic records, as bats are capable of great plasticity in call design, and swarming sites where individuals are engaging in chasing behaviour may result in atypical calls being recorded. Furthermore, in a study where thousands of calls are recorded, misclassification rates as low as 1% will result in hundreds of misclassified calls. This complicates the extraction of reliable data from species that may be present in very low numbers. However, manual inspection of sequences classified as *M. alcathoe* revealed search phase echolocation calls typical of reference calls recorded from the species. In light of this, it seems likely that *M. alcathoe* is indeed a rare swarming visitor to Link Pot and the Ease Gill system of caves. The lack of catching records may be a reflection of the fact that it has only recently been discovery in Britain, and may have been overlooked due to its similarity to other *Myotis* species, and that the probability of capture is extremely low given its scarcity.

Catching results from the sites in the area (Glover & Altringham, 2008), support our acoustic data, with abundant *M. nattereri* and *M. daubentonii*. We recorded levels of *M. brandtii/mystacinus* higher than might have been expected from catching alone. This may represent behavioural differences, with *M. brandtii/mystacinus* spending more time swarming outside the cave, triggering the detector multiple times, although we have no data to support this. Further years of monitoring, with additional catching data, would help determine how closely catching and acoustic data represent the relative species proportions.

5.5.1 Changes in activity through the swarming season

Consistent with previous studies we found large variations in temporal bat activity, both for foraging (Hayes, 1997) and swarming species (Parsons *et al.*, 2003b). Two nights in particular (September 1st and 3rd) dominated the average activity for many species. Species-specific automated logging revealed trends supporting catching data, with early peaks of *M. daubentonii* and *M. brandtii/mystacinus* as found at sites in the same area (Glover & Altringham, 2008), in the south of England (Parsons *et al.*, 2003), and in Poland (Furmankiewicz & Gorniak, 2002).

5.5.2 Changes in activity through the night

Myotis and *Plecotus* bats showed patterns of low activity at dusk and dawn, with peak activity ~4 hours after sunset, consistent with late emergence relative to other species, and of bats travelling from some distance away from the site to swarm. In contrast, *Pipistrellus* and *Nyctalus* activity began around dusk, peaking 1-2 hours later. This suggests early evening foraging activity of local bats, rather than swarming behaviour, further supported by the low incidence of captures reported by a previous study of swarming sites in the area (Glover & Altringham, 2008). Patterns of activity within the *Myotis* showed fine grain differences, but an overall trend for a peak in activity ~4 hours after sunset, consistent with catching data from the same site (Glover & Altringham, 2008). Kernel density plots showed evidence that *M. alcathoe* may swarm earlier in the evening than the other *Myotis* species in this study, although due to its scarcity, further data are required to draw robust conclusions.

5.5.3 Environmental effects on swarming: the generalised additive model (GAM)

Day of the year had a clear non-linear trend on the activity of *M. nattereri*, consistent with the seasonal development of swarming behaviour. The combined effect of day, moon illumination, average daily wind speed and daily humidity high explained 60% of the variation in activity. However, considerable unexplained variation remains, and further nonlinear effects are likely. We found no significant effect of rainfall on activity, although sampling was not undertaken in periods where heavy rainfall was forecast. Parsons and colleagues (2003b) found heavy rainfall suppressed swarming activity. In contrast, Navo and colleagues (2002) noted that activity at a swarming site remained high during a thunderstorm. As previously noted by Parsons and colleagues (2003b), it seems plausible that heavy rainfall at the time of emergence suppresses swarming activity, whereas if conditions at emergence are favourable and bats commit to swarming, later rainfall has a diminished effect on activity levels. We found evidence for a nonlinear effect of moon illumination, which has not previously been reported to significantly influence swarming activity (Parsons et al., 2003b; Karlsson et al., 2006). Cloud cover was not included in our analysis, which may have confounded the relationship with moon illumination. In addition, we had no data on other factors that affect the brightness of the moon, such as its height in the sky. The influence of humidity on bat activity has been little explored. However, little brown bats (Myotis lucifugus) were found to reduce their activity at lower humidity levels (Lacki, 1984). A possible hypothesis for this is that the bats reduce their activity to avoid excess water loss because of extremes in vapour pressure deficits during flight (Lacki, 1984; Adam et al., 1994). In north of England it is unlikely that humidity is an important factor in swarming activity, and the effect may be a result of the relationship between humidity and rainfall. We found a significant non-linear effect of wind. Strong wind may negatively affect swarming activity as is likely to significantly increase the energy demands for bats flying long distances to sites from their roosts. The non-linear relationship we found for wind, with an increase in activity with increasing wind speed, before a strong negative influence, is difficult to interpret. It is possible that an interaction with another environmental variable is confounding the GAM analysis (Zuur *et al.*, 2009).

The effects of environmental variables on activity are difficult to interpret, and are complicated by the high correlation among many of the variables. However, Pipistrellus and Myotis activity was positively correlated, supporting the hypothesis that favourable nights for swarming are also favourable for foraging (Parsons et al., 2003b). Revealing associations between activity and environmental variables is complicated by the large catchment areas that swarming sites support. Bats have been recorded travelling maximum distances of between 20 to 60 km between diurnal roosts or hibernacula to swarming sites (Parsons & Jones 2003, Rivers et al. 2006). In addition to site-specific conditions, local environmental factors at the time of emergence may be critical in bats' decisions whether to swarm or not. As such, models may benefit from averaged weather data at larger spatial scales around the site. Furthermore, daily averages of weather data may be too general, as daytime weather may not directly affect activity; rainfall around emergence time is likely to be more critical. Preliminary results here are based on a single season of data, and the addition of data from additional years would permit a more rigorous evaluation of the environmental factors influencing swarming activity. Due to the remoteness of the study site, the nearest available weather records were from a site almost 20 km away, which may not always have been representative of the weather on site. Small, portable weather loggers may be the only option to obtain reliable data regarding on site temperature and humidity, but this still does not account for weather conditions at the roost sites that bats travel from, which may be the more critical factor. However, attempting to describe the intensity of autumn swarming in relation to time, and identifying environmental factors that influence activity levels, highlights the utility of GAM as a flexible nonlinear modelling tool.

5.5.4 Data compression and computational cost of automated analysis

The high sampling rates required to directly record ultrasound results in larger file sizes than conventional audio applications. Many detectors utilise a simple triggering system so that they are not continuously sampling, but only activate on detecting ultrasound to make short recordings. This reduces the storage demands on the detector, which typically store digital audio on removable media. The duration researchers can leave detectors to collect data, before returning to download information collected, is limited by the digital storage capabilities of the particular device. However, technological advances are occurring rapidly, providing increasingly large and inexpensive digital storage. In practice, power requirements may be the limiting factor in the duration detectors can be left remotely monitoring. In this study the detector ran for less than five nights before requiring a new power source, and the removable media never filled in this time (note – the detector was run on internal batteries, and subsequently connections to larger external sources of power have become available). An important aspect of quantitative acoustic studies is data archival. Stored recordings are available for reanalysis as new methods become available, or may act as acoustic voucher specimens (O'Farrell *et al.*, 1999). It is therefore desirable to apply data compression to archived recordings to reduce space requirements. Lossy compression, where reduction of file size is achieved by discarding some data, is widely used for audio (e.g. MP3) where an effective trade-off between file size and sound quality can be achieved. However, for detailed bioacoustic analysis it is desirable to preserve the original audio data, as it is not predictable whether important information may be lost in conversion. Lossy compression is not reversible and discarded data are lost permanently. FLAC (Free Lossless Audio Codec; http://flac.sourceforge.net/) is a relatively recent development facilitating lossless compression, meaning that there is no loss in audio quality as a result of data conversion. Moreover, FLAC files can be decoded back to their original WAV file format. We applied FLAC conversion to the files generated in this study and effectively halved the storage requirements, helping us achieve reliable and relatively inexpensive data archival.

Automated loggers can produce voluminous data, presenting challenges to researchers in terms of analysing and archiving sound files. Traditional bioacoustic software for echolocation call analysis, e.g. BatSound Pro (www.batsound.com), are designed for analysing single files and performing manual analysis. To address this we applied custom written programs designed to batch process large numbers of files. Automated analysis of sound files was faster than real-time, and we estimate several orders of magnitude faster than possible by human analysis. Moreover, the automated method is repeatable and not subject to researcher bias. Automated methods can be applied to data collected over a variety of time frames (hours, days, months, years), unconstrained by the high costs of manual analysis. In addition, by archiving recordings, as new analysis techniques become available, the data can be re-evaluated. We aim to make the software we have developed available as free and open source software, enabling other researchers to apply our methods to large data sets.

5.5.5 Conservation implications

Studying autumn swarming presents significant challenges to researchers. Sites may be remote (Link Pot is only accessible by foot, over the open moorland of Casterton Fell) and open, making effective capture of bats using mist nets or harp traps difficult. Automated acoustic loggers offer intensive and expansive sampling capabilities, combined with non-invasive data collection. The interpretation of conventional acoustic logger data is complicated by the assemblage of species, each of which may display different patterns of activity throughout the night and over the season, clouding the overall data obtained. In addition, our results suggest that at least at some sites, the activity registered from non-swarming species is not negligible. Acoustic loggers are also sensitive to rain, causing them to falsely trigger, which in the case of conventional loggers, can then not be separated from bat activity. The use of direct-sampling loggers and subsequent automated acoustic analysis can address all these problems, extending current capabilities, and helping researchers better understand species-specific temporal variation in swarming activity. Conventional acoustic loggers are likely to remain a powerful tool in discovering swarming sites however due to their low cost, which enables multiple units to be deployed, facilitating rapid acoustic survey.

Acoustic monitoring of bats is a powerful tool for researchers and conservation workers. However, the effective long-term monitoring of swarming and other sites is complicated by a number of factors. Equipment must be suitable for deployment in a field environment, with the associated hazards from environmental damage and theft. Conventional heterodyne bat loggers (e.g. Glover & Altringham, 2008) are relatively inexpensive and can be made robust to environmental conditions through placement in waterproof cases with small holes cut for the microphone to receive data. As conventional loggers only need to detect the presence of ultrasound, rather than produce high quality recordings for analysis, the quality of the signal is less important, considerably simplifying their deployment. In contrast, direct sampling detectors are expensive in comparison, and high quality recordings are necessary for successful automated analysis. Ground reflection causes interference in recorded signals degrading their quality, and so careful placement of detectors 1-2 m from ground level and other reflective surfaces is required to achieve the best recordings. Providing effective weather protection without compromising the quality of the recorded signal is also challenging. In addition, running electrical equipment for long periods under field situations is often difficult. In this study the bat detector was run on internal batteries which lasted 3-5 days before a visit to the site was required to renew the power source. This was a time-consuming and labour-intensive task. Although the detector never exceeded its storage capabilities in this time, it provided an opportunity to download data, and check on the security of the equipment. The use of external battery supplies could feasibly extent the useful deployment of equipment to a period of weeks or even months. However, issues of equipment security remain for many sites, and the long-term installation of monitoring equipment remains problematic.

The direct sampling detectors used in this study allowed detailed acoustic analysis and classification of calls to genus and in many cases species level. This represents a substantial advantage over conventional loggers, both acoustic and beam-break systems, that do not permit the extraction of species information. This benefit must be weighed against the increased cost of direct sampling detectors, and the higher level of expertise required in their effective setup and post data collection analysis. A remaining limitation is that acoustic loggers are only able to determine presence at a swarming site, and not movement of the bats. Double-beam light barriers provide data on the direction of flight, which with careful setup can be used to distinguish between flights into or out from caves or mines. Previous studies have also used video techniques to show a net influx into a cave in late autumn, to identify the onset of hibernation (Rivers *et al.* 2006). However, automated logging remains the most practical method of identifying and monitoring long-term trends at multiple sites (Glover & Altringham, 2008), and the ability to discriminate species using automated methods extends current capabilities.

5.6 References

Adam, M. D., Lacki, M. J., & Shoemaker, L. G. (1994). Influence of environmental conditions on flight activity of *Plecotus townsendii virginianus* (Chiroptera: Vespertilionidae). *Brimleyana, 21*, 77-85.

Baker, J. (2008). Variation in the relationship between offspring size and survival provides insight into causes of mortality in Hawaiian monk seals. *Endangered Species Research*, *5*(September), 55-64.

Berková, H., & Zukal, J. (2010). Cave visitation by temperate zone bats: effects of climatic factors. *Journal of Zoology*, *280*(4), 387-395.

Fenton, M.B. (1969). Summer activity of *Myotis lucifugus* (Chiroptera: Vespertilionidae) at hibernacula in Ontario and Quebec. *Canadian Journal of Zoology, 47*, 597-602.

Furmankiewicz, J., & Altringham, J. (2007). Genetic structure in a swarming brown long-eared bat (*Plecotus auritus*) population: evidence for mating at swarming sites. *Conservation Genetics*, *8*(4), 913-923.

Furmankiewicz, J., & Gorniak, J. (2002). Seasonal changes in number and diversity of bat species (Chiroptera) in the Stolec mine (SW Poland). Przyroda Sudetow Zachodnich Suppl. 2, 49-70.

Glover, A, and Altringham, J. (2008). Cave selection and use by swarming bat species. *Biological Conservation*, *141*(6), 1493-1504.

Harris, S., Morris, P., Wray, S., Yalden, D., & others. (1995). *A review of British mammals: population estimates and conservation status of British mammals other than cetaceans*. JNCC.

Hayes, J. (1997). Temporal variation in activity of bats and the design of echolocation-monitoring studies. *Journal of Mammalogy*, *78*(2), 514-524.

Jan, C. M. I., Frith, K., Glover, A. M., Butlin, R. K., Scott, C. D., Greenaway, F., Ruedi, M., *et al.* (2010). *Myotis alcathoe* Confirmed in the UK from Mitochondrial and Microsatellite DNA. *Acta Chiropterologica*, *12*(2), 471-483.

Karlsson, B.-L., Eklöf, J., & Rydell, J. (2006). No lunar phobia in swarming insectivorous bats (family Vespertilionidae). *Journal of Zoology*, *256*(4), 473-477.

Komsta, L., and Novomestky, F. (2011). moments: Moments, cumulants, skewness, kurtosis and related tests. R package version 0.12. http://CRAN.R-project.org/package=moments

Lacki, M. J. (1984). Temperature and humidity-induced shifts in the flight activity of little brown bats. *The Ohio Journal of Science*, *84*(5), 264–266.

Lewin-Koh, N. J. and Bivand, R. (2011). maptools: Tools for reading and handling spatial objects. *R package version 0.8-10.* http://CRAN.R-project.org/package=maptools

Maloney, K. O., Schmid, M., & Weller, D. E. (2011). Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages. *Methods in Ecology and Evolution.*

McCracken, G.F., and Wilkinson, G.S. (2000). Reproductive Biology of Bats. In: *Reproductive Biology of Bats.* (E.G. Crichton and P.H. Krutzsch, eds.). pp. 321-362. Academic Press, New York

Navo, K. W., Henry, S. G., & Ingersoll, T. E. (2002). Observations of swarming by bats and band recoveries in Colorado. *Western North American Naturalist*, *62*(1), 124–126.

O'Farrell, M. J., Miller, B. W., & Gannon, W. . (1999). Qualitative identification of free-flying bats using the Anabat detector. *Journal of Mammalogy*, *80*(1), 11-23.

O'Hara, R. B., and Kotze, D. J. (2010). Do not log-transform count data. *Methods in Ecology and Evolution*, *1*(2), 118-122.

Parsons, K. N., & Jones, G. (2003). Dispersion and habitat use by *Myotis daubentonii* and *Myotis nattereri* during the swarming season: implications for conservation. *Animal Conservation*, *6*(4), 283-290.

Parsons, K., Jones, G., Davidson-Watts, I., & Greenaway, F. (2003a). Swarming of bats at underground sites in Britain—implications for conservation. *Biological Conservation*, *111*(1), 63-70.

Parsons, K. N., Jones, G., & Greenaway, F. (2003b). Swarming activity of temperate zone microchiropteran bats: effects of season, time of night and weather conditions. *Journal of Zoology*, *261*(3), 257-264.

Piksa, K., Bogdanowicz, W., & Tereba, A. (2011). Swarming of Bats at Different Elevations in the Carpathian Mountains. *Acta Chiropterologica*, *13*(1), 113-122.

Rivers, N. M., Butlin, R. K., & Altringham, J. D. (2005). Genetic population structure of Natterer's bats explained by mating at swarming sites and philopatry. *Molecular ecology*, *14*(14), 4299-312.

Rivers, N., Butlin, R., & Altringham, J. (2006). Autumn swarming behaviour of Natterer's bats in the UK: Population size, catchment area and dispersal. *Biological Conservation, 127*(2), 215-226.

Ryan, J. A., & Ulrich, J. M. (2011). xts: eXtensible Time Series. *R package version 0.8-2.* http://CRAN.R-project.org/package=xts

Šuba, J., Vintulis, V., & Pētersons, G. (2008). Late summer and autumn swarming of bats at Sikspārņu caves in Gauja National Park. *Acta Universitatis Latviensis, 745,* 43-52.

Thomas, D. W., Brock Fenton, M., & Barclay, R. M. R. (1979). Social behavior of the little brown bat, Myotis lucifugus. *Behavioral Ecology and Sociobiology, 6*(2), 129–136.

Veith, M., Beer, N., Kiefer, A., Johannesen, J., & Seitz, A. (2004). The role of swarming sites for maintaining gene flow in the brown long-eared bat (*Plecotus auritus*). *Heredity, 93*(4), 342-9.

Wickham, H. (2009). ggplot2: elegant graphics for data analysis. Springer New York.

Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B, 73*(1), 3-36.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). Mixed Effects Models and Extensions in Ecology with R. Springer.

Zuur, A. F. (2010). AED: Data files used in Mixed effects models and extensions in ecology with R. (2009). Zuur *et al.* (2009). R package version 1.0.

Chapter 6: Acoustic monitoring and species distribution modelling as a non-invasive conservation tool for bats[1] [2]

6.1 Abstract

Species distribution models (SDMs) can be used to predict areas of potential distribution, and investigate the habitat requirements of species. Consequently, SDMs are becoming a valuable tool in landscape scale conservation efforts. In this study, we used acoustic surveys and catching techniques to collect data on the spatial distribution of foraging bats, from 30 field sites in the Lake District National Park. Echolocation calls were identified to species/species groups, using machine learning (ML) classification to separate acoustically cryptic *Myotis* species. GPS technology provided precise locations for all foraging bats, and a geographic information system (GIS) was used to generate fine scale habitat data. We employed the presence-only modelling software MaxEnt to investigate the patterns of geographic distribution, and produce species-specific habitat suitability maps for the entire Lake District National Park (ca. 3,330 km^2). The resulting maps were used to generate a species richness map, highlighting hotspot areas of potential conservation priority within the Park. The robustness of models to geographic transferral was tested using independent data collected from eight field sites outside the range of the training data area.

---

[1] A paper based on this work is being prepared for submission to Journal of Applied Ecology.

[2] Chloe Bellamy planned fieldwork and carried out acoustic transects, GIS and MaxEnt modelling work (Bellamy, 2011). Catching work was carried out jointly by Chloe Bellamy and Chris Scott. Chris Scott carried out the classification of *Myotis* calls and assisted with acoustic transects.

6.2 Introduction

For effective conservation planning and management, knowledge of the geographic distribution of species and their habitat requirements are prerequisites. For many species, large scale survey is impractical, time-consuming and costly. Bats are a prime example, as their nocturnal behaviour makes them difficult to detect and survey. Increasingly, practical information on the spatial patterns of distribution of species is being provided by species distribution models (SDMs). SDMs use occurrence data, and associated ecogeographic variables (EGVs; e.g. altitude, slope, vegetation), to predict areas of potential distribution between and beyond the known data which fulfil the species' niche requirements (Elith & Leathwick, 2009). Using geographical information systems (GIS), model output can be visualised as habitat suitability maps. SDMs are now widely used to gain insight into species ecological requirements and to predict distributions across landscapes (Elith & Leathwick, 2009), and are becoming valuable tools in conservation planning and management, assessing potential impacts from human activities and climate change (e.g. Lundy *et al.*, 2010).

For some species, suitable data for building SDMs may already be held by natural history museums and herbaria (e.g. Elith & Leathwick, 2007), and atlas data (e.g. Niamir *et al.*, 2011). For bats, survey is complicated and labour-intensive, and consequently existing records may be deficient. Increasingly sophisticated modelling techniques are maximising the information that can be extracted from small datasets. For example, the freely available modelling software Maximum Entropy Species Distribution Modelling (MaxEnt; http://www.cs.princeton.edu/~schapire/maxent/; Phillips *et al.*, 2006) can generate accurate models from presence-only data, and with small sample sizes (Wisz *et al.*, 2008). However, SDMs perform better if the presence data is unbiased and free of error (Graham *et al.*, 2007), and where presence data is sparse and its reliability is questionable, improving the quality of the occurrence data is suggested (Lobo, 2008). The ability to use presence-only data considerably simplifies survey work, as it may be difficult to obtain accurate absence data for mobile species with large home ranges (Brotons *et al.*, 2004).

In this study we used a combination of acoustic and catching techniques to collect data on foraging bats in the south of the Lake District National Park. Acoustic methods facilitated the rapid and systematic survey of sites, recording the echolocation calls of foraging bats for subsequent identification to species/species groups. Automated methods of call analysis were employed to aid in the separation of acoustically cryptic *Myotis* species, which are not reliably identified using conventional analysis, due to extensive overlap of call parameters (e.g. Parsons & Jones, 2000; Chapter 3). In addition to acoustic surveys, we caught foraging bats to increase the available presence data for *Plecotus auritus*, a low-intensity 'whispering' species (Waters & Jones, 1995), rarely recorded during acoustic surveys (e.g. Bellamy, 2011). The robustness of models to geographic transferral was tested using independent data collected from eight field sites outside the range of the training data. This study represents the first use of machine learning techniques to classify acoustically cryptic *Myotis* calls to provide presence data for species distribution modelling.

The aim of the present study was to demonstrate the use of acoustic methods and machine learning techniques to provide species-specific presence data, in producing species distribution models (SMDs) and habitat suitability maps.

6.3 Methods

6.3.1 Study area

The study was carried out in the Lake District National Park, in Cumbria, north-west England. This is a diverse and complex landscape with a high density of deciduous and ancient woodland, given National Park status in 1951.

6.3.2 Acoustic transects

Bat presence data for training models were collected by recording echolocation calls during 2-3 km walked transects, in thirty 1 km^2 field sites in the south of the park during 2008 and 2009 (Figure 7.1). Sites were selected using a stratified sampling design to ensure a range of habitats and elevations were represented (for further details see Bellamy (2011)). Transects were walked twice each year during the period May-Sept, to cover the main foraging activity of UK bat species. During the second visit transect routes were reversed to reduce possible bias due to the effect of time of night on activity. A single transect was walked per night, starting one hour after sunset to avoid bats commuting from roosts, and to cover all species' peak foraging activity (Barlow & Jones 1997). Transects lasted ~90 minutes, walking at a slow, steady pace. Surveys were not carried out on nights when weather conditions were adverse for bat activity, i.e. low temperatures (<10 °C) or strong winds (>20 km/h).

Bats were recorded using the time-expansion output of an ultrasound detector (D240x; www.batsound.com) and an Edirol R-09 digital recorder (www.edirol.com). The detector triggered automatically on detecting ultrasound, sampling 100 ms of audio at 307 kHz, which was then time-expanded (10x) and recorded at 16 bit, 44.1 kHz by the Edirol R-09. Transects were walked with a GPS (Garmin GPSmap 60Cx; www.garmin.com/uk/) which logged a breadcrumb trail storing accurate position, speed, direction and altitude every 20 s. This enabled each recorded bat passes to be linked to a specific geographic position.

To provide an unbiased test of model accuracy, an independent test set of presence data was collected from eight field sites, four to the north and four to the west of the training sites (Figure 6.1). Test data from outside the region of training data were used to assess the geographic transferability of models. Data were collected using the same field methods as the training data.

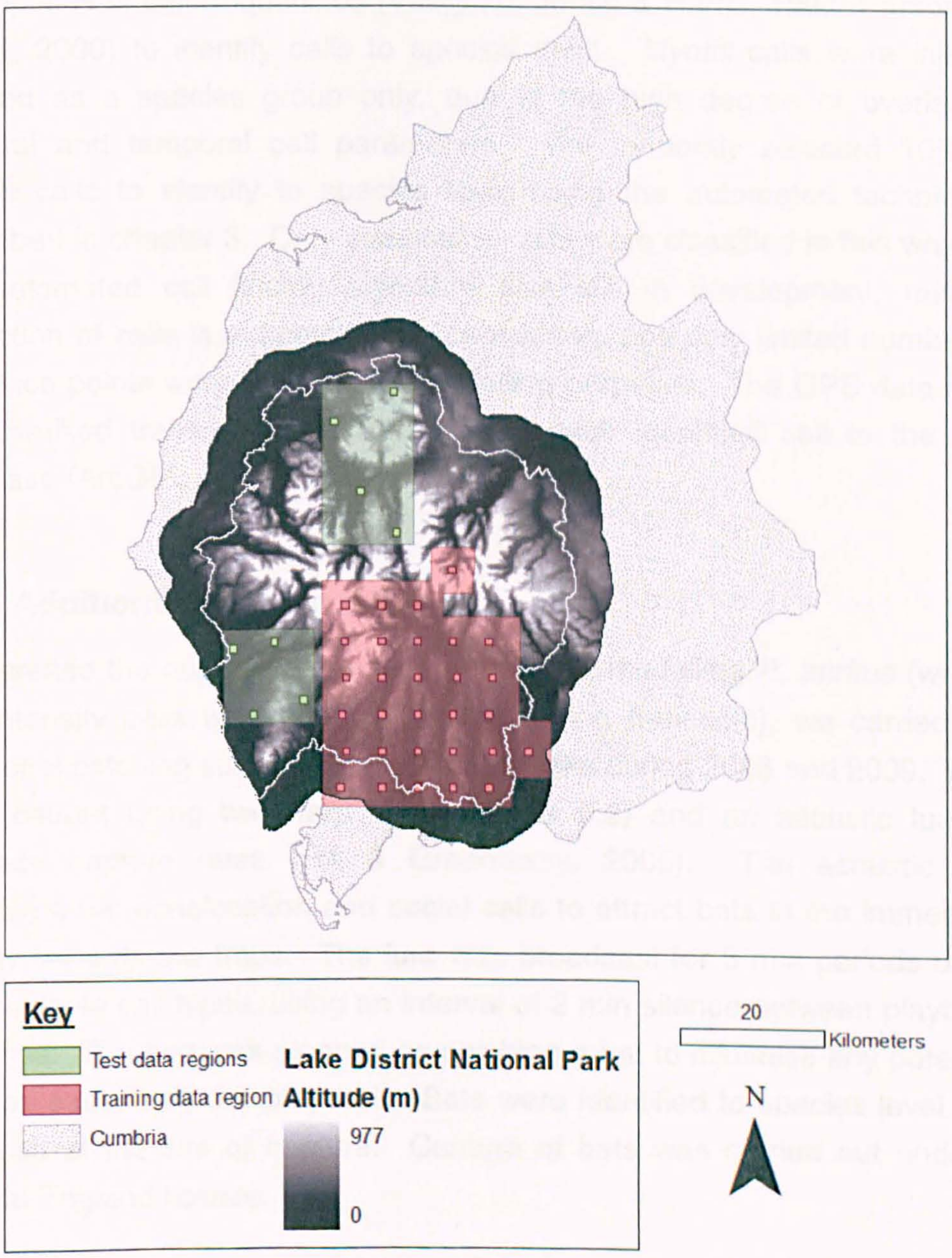


Figure 6.1: Map of study area showing field sites.

Field sites (illustrated with small squares) for training and testing regions used to build and validate models. From Bellamy (2011).

6.3.4 Sound analysis

Bat detector recordings were visualised as spectrograms using BatSound Pro (www.batsound.com), using the overall shape of calls and published descriptions of call frequencies (Vaughan, Jones & Harris, 1997; Parsons & Jones, 2000) to identify calls to species level. *Myotis* calls were initially classed as a species group only, due to the high degree of overlap in spectral and temporal call parameters. We randomly selected 10% of *Myotis* calls to identify to species level using the automated techniques described in chapter 3. Only a subset of calls were classified in this way, as the automated call finding algorithm was still in development, manual extraction of calls is extremely time consuming, and only limited number of presence points were required for modelling purposes. The GPS data from each walked transect were used to add each identified call to the GIS database (ArcGIS; version 9.3, www.esri.com).

6.3.5 Additional catching data

To increase the number of presence points for modelling *P. auritus* (whose low intensity calls were rarely recorded during transects), we carried out additional catching surveys within the field sites during 2008 and 2009. Bats were caught using two harp traps (Figure 6.2) and an acoustic lure to increase capture rates (Hill & Greenaway, 2005). The acoustic lure broadcast bat echolocation and social calls to attract bats in the immediate vicinity towards the traps. The lure was broadcast for 5 min periods using the available call types, using an interval of 2 min silence between playback sessions. The lure was stopped on catching a bat to minimise any potential distress caused by the playback. Bats were identified to species level and released at the site of capture. Capture of bats was carried out under a Natural England license.

Figure 6.2: Harp trap. A single harp trap set up at a catching site.

6.3.6 Distribution modelling

Presence-only modelling was used to predict the distribution of species using MaxEnt. MaxEnt is a machine learning presence-only method (Phillips *et al.*, 2006), that uses spatially distributed presence data and associated ecogeographic variables (EGVs) to predict areas of potential distribution. EVGs consisted of fifteen habitat variables, including distance to water, distance to woodland edge, mean altitude and slope, and the percentage cover of five habitat classes (buildings, water, coniferous woodland, deciduous woodland, and manmade surface and road). These were measured over a range of spatial extents (100 – 6,000 m) using a moving window analysis in GIS (ArcGIS 9.3, www.esri.com; Store & Jokimaki, 2003) and were represented at a fine resolution (100 x 100 m for *Plecotus auritus* and 50 x 50 m for all other species). The variables were tested for their predictive accuracy over a range of spatial scales using 5-fold cross validation. Each variable was selected at its best performing scale to enter into a species' SDM. These models were then pruned using a backwards stepwise reduction procedure to produce minimum adequate models (Parolo *et al.*, 2008). Full methodological details are given in Bellamy (2011). All models were built using Maxent v3.3.2, using mainly default settings (http://www.cs.princeton.edu/~schapire/maxent/; Phillips *et al.*, 2006).

Models were built using training data, and tested using the independent test data to assess model accuracy. The area under the ROC curve (AUC) statistic was used to measure model accuracy (e.g. Fielding & Bell, 1997). MaxEnt models presence-only data by first adding randomly generated data points or “pseudo-absences” to the data. Pseudo-absences were extracted from the region of the independent test sites in assessing the AUC score. AUC was calculated in R using the package ROCR (R Development Core Team, 2011; Sing *et al.*, 2009). 2,000 bootstrap iterations were used to generate 95% confidence intervals.

A species richness map was produced using a simple aggregation method in ArcGIS, summing the logistic habitat suitability values for each species in each cell, producing a continuous measure of potential species richness (Aranda & Lobo, 2011).

6.4 Results

Walked transects covered 334 km, recording approximately 180 hours of acoustic survey data. 15,466 echolocation calls were manually categorised into five species groups (Bellamy, 2011), with 266 *Myotis* calls subsequently classified using automated methods (Table 6.1).

Species/species group	Presence points collected				
	(i) Call ID	(ii) Machine learning	(iii) Caught	(iv) Incidental/ museum	Total
Pipistrellus pipistrellus	5,210	N/A	17	0	5,227
P. pygmaues	6,489	N/A	43	0	6,532
Plecotus auritus	22	N/A	18	34	74
Nyctalus spp.	675	N/A	1	0	676
Myotis spp.	3,051	N/A	71	0	2,477
M. daubentonii	0	101	13	0	114
M.bra./mys.	0	139	45	0	184
M. nattereri	19	26	13	0	58

Table 6.1: Total number of presence records per species for modelling collected from field sites in the southern Lake District by (i) examination of call spectrograms, (ii) classification using machine learning algorithm, (iii) capture using harp traps, and (iv) Tullie House Museum (Cumbria County Council) records.

Models performed well with AUC scores on the independent test data all above 0.7 (Table 6.2). Habitat suitability maps were created for each species model (two examples are shown in Figures 6.3 & 6.4). The predictive species richness map is shown in Figure 6.5.

Spp.	N	Test AUC	CI
P. pip	113	0.704	0.656 - 0.754
P. pyg	117	0.751	0.719 - 0.784
N. noc	16	0.803	0.731 - 0.875
M. bra./mys.	26	0.733	0.656 - 0.809
M. dau	25	0.760	0.692 - 0.827
M. nat	7	0.842	0.622 - 1.062

Table 6.2: Performance of models on independent test data.

AUC CI = 2000 bootstrap confidence intervals.

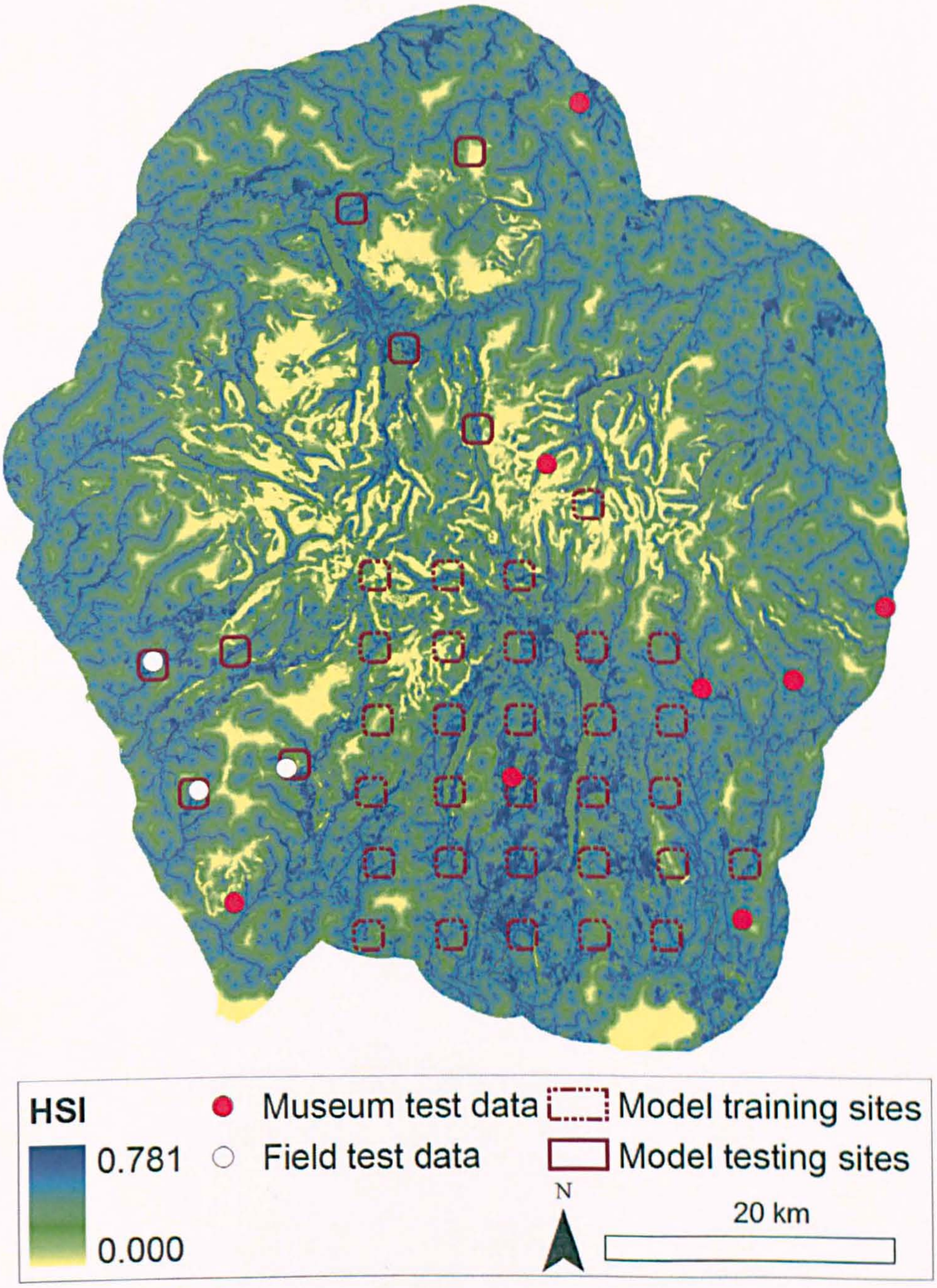


Figure 6.3: Habitat suitability map for foraging *M. nattereri*.

Colour-mapped Habitat Suitability Index (HSI) projected over the Lake District National Park, showing areas of high suitability in deep blue. From Bellamy (2011).

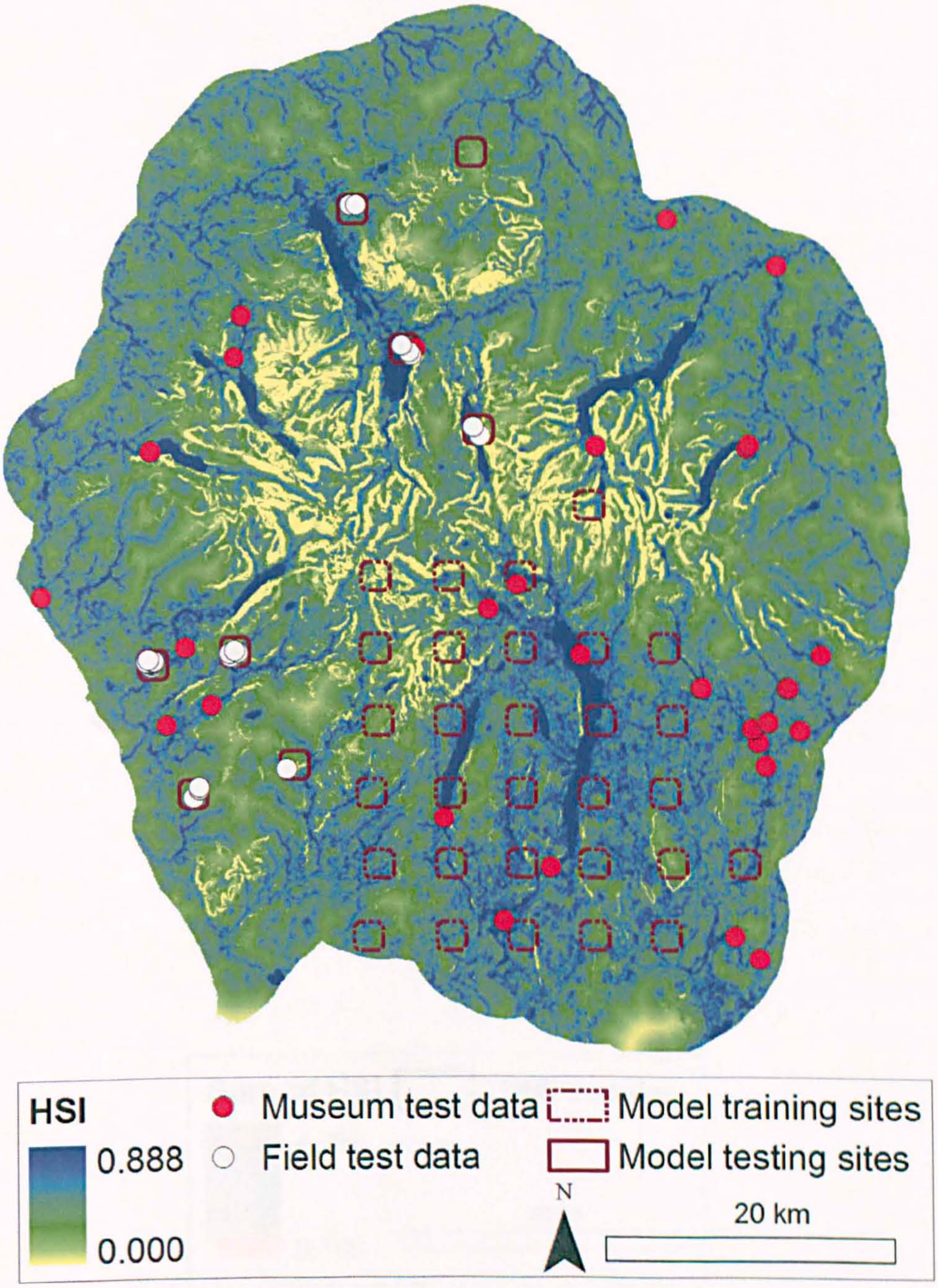


Figure 6.4: Habitat suitability map for foraging *M. daubentonii.*

Colour-mapped Habitat Suitability Index (HSI) projected over the Lake District National Park, showing areas of high suitability in deep blue. From Bellamy (2011).

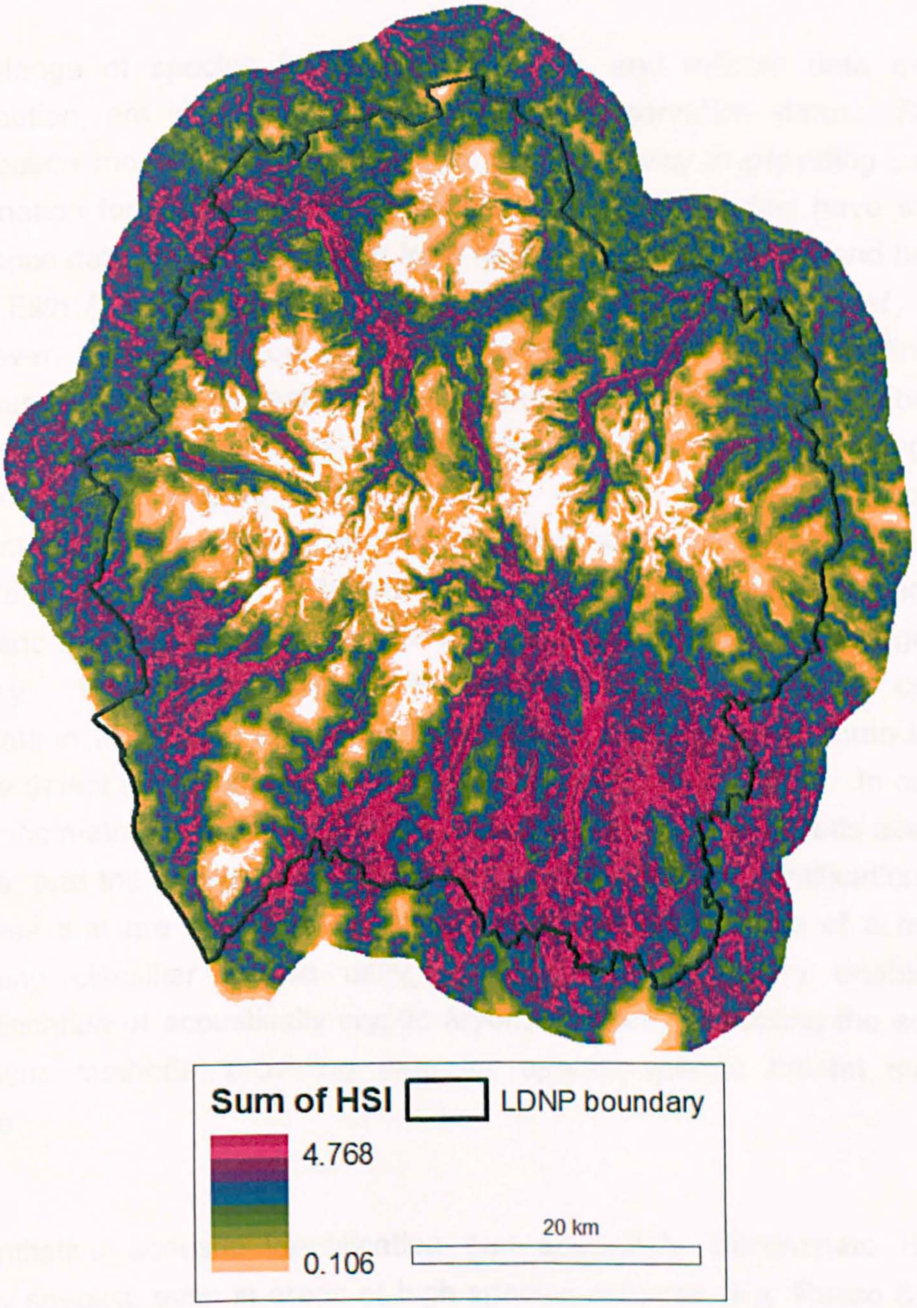


Figure 6.5: Species richness map.

Predictive species richness map, produced by summing the predicted HSI values for each individual species. Higher scores indicate hotspot areas for bat foraging activity. LDNP = Lake District National Park. From Bellamy (2011).

6.5 Discussion

Knowledge of species habitat requirements, and reliable data on their distribution, are vital to adequately assess conservation status. Species distribution modelling increasingly has a role to play in providing practical information for difficult to survey species. Previous studies have sourced presence data for building SDMs from natural history museums and herbaria (e.g. Elith & Leathwick, 2007), and atlas data (e.g. Niamir *et al.*, 2011). However, records may be biased due to non-representative sampling, and inaccurate due to misidentifications (Graham *et al.*, 2004). In addition, not all records may be suitable for modelling due to their coarse geographic resolution (e.g. Rebelo & Jones, 2010). Where data remain sparse or absent due to difficulties in survey, new methodologies may be required to facilitate data collection. Capture is intrusive to bats, requires considerable skill and training, and is labour-intensive, making it inefficient for large scale survey. Furthermore, capture is restricted to the structurally complex habitats in which nets and traps are effective, and biased as some species easily detect and avoid nets and traps (e.g. Larsen *et al.*, 2007). In contrast, acoustic methods facilitates the rapid sampling of diverse habitats over large areas, and the analysis of echolocation calls allows the identification of bat species that are difficult to capture. In this study, the use of a machine learning classifier trained using a reference call library enabled the identification of acoustically cryptic *Myotis* species, increasing the scope of acoustic methods, providing data for species-specific habitat suitability maps.

Quantitative acoustic identification can accurately discriminate between many species, even in areas of high species richness (e.g. Russo & Jones, 2002; Obrist *et al.*, 2004; Papadatou *et al.*, 2008). However, not all species are equally acoustically apparent, as some low-intensity 'whispering' species are only detected at close range (e.g. *Plecotus auritus;* Waters & Jones, 2005). These differences in the probability of detection mean acoustic methods used in isolation may under-record some species (e.g. O'Farrell & Gannon, 1999). In addition, separating some species on the basis of their echolocation calls remains challenging. For *Myotis brandtii* and *M. mystacinus,* separation morphologically is possible but complicated (Berge, 2007), and in such cases catching combined with genetic analysis may be the only reliable way of obtaining accurate species-specific presence data (Mayer *et al.*, 2007). However, acoustic methods may still be employed to collect accurate data for a species group. In this study we used a combination of acoustic and catching techniques to reliably survey a large number of sites in the most efficient way possible. The use of an acoustic lure (Hill & Greenaway, 2005) aided the capture of 'whispering' species under-recorded during acoustic surveys (e.g. *Plecotus auritus*). In future research we will apply an acoustic lure to attract bats, to bring 'whispering' species closer to the bat detectors. By doing so, we expect to obtain recordings suitable for automated classification from 'whispering' species, without necessitating capture. This novel use of a lure would further increase the efficiency of collecting presence-only bat data, leaving labour-intensive capture techniques for situations where genetic separation is desired.

MaxEnt produced useful models that performed reasonably well (AUC > 0.7) using presence-only data, making it a potentially powerful tool for researchers where reliable absence data cannot be acquired. The habitat suitability maps produced in this study provide reliable baseline data for the Park, and the combined species richness map can be used to identify 'hotspot' areas of high conservation priority. Moreover, model output can be used to determine the species-specific environmental factors that drive patterns of distribution. A strong positive relationship between the presence of woodland and fresh water habitats was apparent for all species. However, there were also species- and scale-specific effects (Bellamy 2011).

By projecting beyond the region they were fitted, species distribution models can be used to predict whether a species is likely to occur outside of its known range (Randin *et al.*, 2006). As ongoing work, we are building models for the North York Moors and Yorkshire Dales National Parks, where predictive maps are being generated, and ground-truth data collected to investigate the geographic transferability of the models developed in this study. The long term goal is accurate bat habitat suitability maps for the whole of the UK. A priority is the collection of data from rare species. *Myotis alcathoe* has only recently been discovered in the UK (Jan *et al.*, 2010), and acoustic methods offer the opportunity to rapidly survey and model suitable habitat, which could be used to predict its distribution and target further surveys and research.

6.6 References

Aranda, S. C., & Lobo, J. M. (2011). How well does presence-only-based species distribution modelling predict assemblage diversity? A case study of the Tenerife flora. Ecography, 34(1), 31-38.

Bellamy, C. C. (2011). Predictive modelling of bat-habitat relationships on different spatial scales. PhD Thesis, University of Leeds, Institute of Integrative and Comparative Biology, Leeds, UK.

Berge, L. (2007). The effects of phylogenetic differences on resource partitioning between the cryptic species whiskered bat (*Myotis mystacinus*) and Brandt's bat (*M. brandtii*). Unpublished Thesis, University of Bristol, School of Biological Sciences, Bristol, UK.

Brotons, L., Thuiller, W., Araujo, M., & Hirzel, A. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography, 27*, 437-448.

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R. J., *et al.* (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography, 29*(2), 129–151.

Elith, J., & Leathwick, J. (2007). Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions, 13*(3), 265-275.

Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics, 40*(1), 677-697.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), 38-49.

Graham, C. H., Ferrier, S., Huettman, F., Moritz, C., & Peterson, a T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in ecology & evolution, 19*(9), 497-503.

Graham, C. H., Elith, J., Hijmans, R. J., Guisan, A., Townsend Peterson, a, & Loiselle, B. a. (2007). The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology, 45*(1), 239-247.

Hill, D. A., & Greenaway, F. (2005). Effectiveness of an acoustic lure for surveying bats in British woodlands. *Mammal Review, 35*(1), 116–122.

Jan, C. M. I., Frith, K., Glover, A. M., Butlin, R. K., Scott, C. D., Greenaway, F., Ruedi, M., *et al.* (2010). Myotis alcathoe Confirmed in the UK from Mitochondrial and Microsatellite DNA. *Acta Chiropterologica*, *12*(2), 471-483.

Joint Nature Conservation Committee. (2007). *Second Report by the United Kingdom under Article 17 on the implementation of the Directive from January 2001 to December 2006.* Peterborough.

Larsen, R. J., Boegler, K. a., Genoways, H. H., Masefield, W. P., Kirsch, R. a., & Pedersen, S. C. (2007). Mist netting bias, species accumulation curves, and the rediscovery of two bats on Montserrat (Lesser Antilles). *Acta Chiropterologica*, *9*(2), 423-435.

Lobo, J. M. (2008). More complex distribution models or more representative data? *Biodiversity Informatics*, *5*, 14-19.

Lundy, M., Montgomery, I., & Russ, J. (2010). Climate change-linked range expansion of Nathusius' pipistrelle bat, *Pipistrellus nathusii* (Keyserling & Blasius, 1839). *Journal of Biogeography*, *37*(12), 2232-2242.

Mayer, F., Dietz, C., & Kiefer, A. (2007). Molecular species identification boosts bat diversity. *Frontiers in zoology*, *4*, 4.

Niamir, A., Skidmore, A. K., Toxopeus, A. G., Muñoz, A. R., & Real, R. (2011). Finessing atlas data for species distribution models. *Diversity and Distributions*, *17*, 1173-1185.

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.

O'Farrell, M. J., & Gannon, W. L. (1999). A Comparison of Acoustic versus Capture Techniques for the Inventory of Bats. *Journal of Mammalogy*, *80*(1), 24.

Papadatou, E., Butlin, R. K., & Altringham, J. D. (2008). Identification of bat species in Greece from their echolocation calls. *Acta Chiropterologica*, *10*(1), 127-143.

Parolo, G., Rossi, G., & Ferrarini, A. (2008). Toward improved species niche modelling: *Arnica montana* in the Alps as a case study. *Journal of Applied Ecology*, *45*(5), 1410–1418.

Parsons, S., & Jones, G. (2000). Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. *J. Exp. Biol*, *203*, 2641-56.

Phillips, S., Anderson, R., & Schapire, R. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling, 190*(3-4), 231-259.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Randin, C. F., Dirnböck, T., Dullinger, S., Zimmermann, N. E., Zappa, M., & Guisan, A. (2006). Are niche-based species distribution models transferable in space? *Journal of Biogeography, 33*(10), 1689-1703.

Rebelo, H., & Jones, G. (2010). Ground validation of presence-only modelling with rare species: a case study on barbastelles Barbastella barbastellus (Chiroptera: Vespertilionidae). *Journal of Applied Ecology, 47*(2), 410-420.

Rebelo, H., Tarroso, P., & Jones, G. (2010). Predicted impact of climate change on European bats in relation to their biogeographic patterns. *Global Change Biology, 16*(2), 561-576.

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.

Russo, D., & Jones, G. (2003). Use of foraging habitats by bats in a Mediterranean area determined by acoustic surveys: conservation implications. *Ecography, 26*(2), 197-209.

Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T. (2009). ROCR: Visualizing the performance of scoring classifiers. R package version 1.0-4. http://CRAN.R-project.org/package=ROCR

Store, R., & Jokimaki, J. (2003). A GIS-based multi-scale approach to habitat suitability modeling. *Ecological Modelling, 169*(1), 1-15.

Waters, D. & Jones, G. (1995). Echolocation call structure and intensity in five species of insectivorous bats. *The Journal of experimental biology, 198*(Pt 2), 475-89.

Wickramasinghe, L. P., Harris, S., Jones, G., Vaughan, N., Harris, S., Jones, G., & Vaughan, N. (2003). Bat activity and species richness on organic and farms: impact of agricultural intensification conventional. *Journal of Applied Ecology, 40*(6), 984-993.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, a. T., Graham, C. H., & Guisan, a. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions, 14*(5), 763-773.

Chapter 7: General Discussion

Acoustic methods are central to bat research as they provide the means to non-invasively survey and monitor bats in their natural habitats, revealing their presence and facilitating the study of their habitat preferences and temporal patterns of activity. Technological advances are permitting researchers to monitor acoustically at greater spatial and temporal resolution, opening new avenues of research. The major problem with expanding the scope of current research is that acoustic analysis requires skill, and is repetitive and time consuming. Automation of analysis removes the burden of routine identifications, and provides a replicable methodology (Gaston & O'Neill, 2004). Additionally, automated methods may also enable faster and more accurate identifications than possible by human experts. However, the adoption of automated methods of analysis by researchers and conservation workers has been hampered by a lack of reliable and freely available tools.

We developed and described an algorithm to locate, extract and measure objective spectral and temporal call features from continuous bat detector recordings. In a comparison with two conventional bioacoustic energy detectors, our algorithm proved more accurate and robust at locating calls in field recordings, minimising false detections caused by echoes. Through efficient implementation in an open source programming language, C++, the algorithm runs faster than real time and an order of magnitude faster than possible by human analysis. These developments allow the location of signals of interest in a fast and reliable manner, facilitating intensive acoustic sampling over extended time periods, without the bottleneck of manual analysis. The automated measurement of temporal and spectral call parameters provides data suitable for the quantitative bioacoustic description of signals, and for statistical analysis or machine learning (ML) classification.

Using ML algorithms, we implemented a fully automated acoustic identification system for bat echolocation calls. To evaluate the system we built a reference echolocation call library, recording known species of British bats at a range of sites across the UK. The combination of full spectrum call parameters and a non-parametric ML classifier achieved a high level of accuracy and outperformed conventional statistical analysis using discriminant function analysis (DFA). Automated classification also provided levels of confidence in the identifications, which allowed us to reduce misclassification rates by leaving classifications below a confidence threshold as ‘unknown’. We found that in a fully automated unsupervised system, steps must be taken to mitigate for the effects of novel signals presented to the classifier. Classifiers implicitly assume that training data are representative of the real world data to be encountered in practical use, and force all new data into the categories defined during training. However, in ecological applications this presents a significant problem, as it may be impossible to collect representative data from all species in a study area, or to know the full species assemblage in advance, leading to the misclassification of call types that are novel to the classifier. We showed that ML methods of outlier detection effectively protect a classifier from novel signal types, not representative of those encountered during training. This is a vital step in ensuring unsupervised classifiers produce sensible output, and allows researchers to focus their efforts on collecting training data for target species, as classification methods can then be applied without an exhaustive call library for the area of study. This facilitates the early adoption of automated acoustic survey and monitoring methods in new areas, where development of comprehensive call libraries may take several years.

To assess the automated identification system we applied it to two real world field studies. Firstly, we investigated the multi-species bat activity at a remote cave in the north of England over a three month period. The use of automated methods allowed the rapid analysis of the acquired data (>20,000 audio files) and the separation of species with high accuracy in an objective, repeatable way. Through the use of the freely available statistical program R (R Development Core Team, 2011; http://www.r-project.org/) and its related packages, we have written scripts to visualise the results of our automated acoustic analysis, revealing temporal trends at scales from minutes to months. We showed that for monitoring purposes, acoustic logging is a viable alternative to catching, which is a labour intensive and intrusive technique that need only be used when necessary (Parsons *et al.*, 2003; Rivers *et al.*, 2006; Glover & Altringham, 2008). These automated techniques allow researchers to take hundreds of hours of acoustic data and produce multi-species plots of activity in a period of hours. Secondly, we separated acoustically cryptic *Myotis* species from data collected in the south of the Lake District National Park. The identified calls had precise GPS locations, provided spatially explicit presence data to generate species distribution models (SDMs) using MaxEnt. Models were used in combination with geographic information systems (GIS) to create species-specific predictive habitat suitability maps covering the entire Park (ca. 3,300 km^2). These results can be used to aid the development of management plans and identify areas of conservation priority. This work is currently being extended to the North York Moors and Yorkshire Dales National Parks, where further predictive maps are being generated, and ground-truth data being collected to validate models with independent test data. The long term goal is accurate bat habitat suitability maps for the whole of the UK.

In developing tools for the automated acoustic analysis of bat detector recordings, this work has extended the scope of acoustic monitoring studies. Automated tools provide efficient data analysis and fill a gap in current capabilities, providing replicable methods and making long term monitoring feasible. Future research aims to integrate the automated methods we developed into a woodland bat survey protocol. The UK's rarest bat species are all woodland species, their declines in recent centuries a reflection of the degradation, fragmentation and loss of habitat as a result of human activity (Altringham, 2011). However, the application of acoustic surveys to the monitoring of many of our woodland species has been limited by two problems. Firstly, there has been the difficulty in identifying the six species of *Myotis* bats from their echolocation calls using conventional methods. Secondly, there is the risk of under-recording low-intensity 'whispering' species, that are only recorded when they fly very close to the detectors. We will apply our automated methods of acoustic identification, and overcome the issue of under-recording 'whispering' species through the novel use of an acoustic lure. Broadcasting ultrasonic bat calls has increased capture rates, by attracting bats into nets and traps (Hill & Greenaway, 2005). We aim to attract bats into close range of a bat detector, to ensure high quality recordings of all species suitable for classification. This offers the potential for a new national woodland bat monitoring scheme, collecting data that is vital for effective conservation in a manner that is scientifically rigorous, efficient, and capable of being carried out by trained volunteers. Through the combined use of GPS and GIS technology, survey data can contribute to ongoing predictive habitat mapping. As part of this work we will provide a user friendly front end to our algorithms, to maximise the ease with which the research community and conservation workers can benefit from these tools.

We restricted this study to the identification of bat species' from their echolocation calls. Despite their high variability, some species may encode individual-specific information in their echolocation calls sufficient for recognition (e.g. Yovel *et al.*, 2009). If vocal individuality can be extracted reliably, it offers the potential to estimate numbers of individuals acoustically, thus extending the current scope of acoustic methods to studying the population ecology of bats. In addition to echolocation calls, bats also emit social calls for the purpose of communication, carrying information to conspecifics (Fenton, 2003). Social calls may be species- and individual-specific (Pfalzer & Kusch, 2003), providing further opportunity for researchers to extract information from the vocalisations of bats.

References

Altringham, J. D. (2011). *Bats: From Evolution to Conservation.* OUP, Oxford.

Fenton, M. B. (2003). Eavesdropping on the echolocation and social calls of bats. *Mammal Review*, *33*(3-4), 193-204.

Gaston, K. J., & O'Neill, M. A. (2004). Automated species identification: why not? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *359*(1444), 655-67.

Glover, A., & Altringham, J. (2008). Cave selection and use by swarming bat species. *Biological Conservation*, *141*(6), 1493-1504.

Hill, D. A., & Greenaway, F. (2005). Effectiveness of an acoustic lure for surveying bats in British woodlands. *Mammal Review*, *35*(1), 116–122.

Parsons, K., Jones, G., Davidson-Watts, I., & Greenaway, F. (2003). Swarming of bats at underground sites in Britain—implications for conservation. *Biological Conservation*, *111*(1), 63-70.

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.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rivers, N., Butlin, R., & Altringham, J. (2006). Autumn swarming behaviour of Natterer's bats in the UK: Population size, catchment area and dispersal. *Biological Conservation*, *127*(2), 215-226.

Yovel, Y., Melcon, M. L., Franz, M. O., Denzinger, A., & Schnitzler, H.-U. (2009). The voice of bats: how greater mouse-eared bats recognize individuals based on their echolocation calls. *PLoS computational biology*, *5*(6), e1000400.

Appendix A
Signal detection code

C++ pseudo code for the signal detection algorithms described in Chapter 2.

A.1 Spectral peak algorithm

```
// Input WAV samples stored in a std::vector
// Output detection function as a std::vector
vector<double> spectralPeak(vector<float> const& audioSamples)
{
    int fftSize = 256;
    int binN = fftSize/2.0;
    step = 0.25*binN;//75% overlap

    //calculate how many iterations
    int frames = 1 + floor((audioSamples.size()-fftSize)/(double)step);
    vector<double> detectionFunc(frames);

    int startN = 0;
    double magnitude, peak;
    for(int i=0; i<frames; ++i) {
        Calculate FFT from audio samples at startN
        peak = 0.0;
        for  (int j=1; j<binN; j++) {
            magnitude = FFT[j];
            peak = max(peak, magnitude);
        }
        //convert peak value to dB and store
```

```
        detectionFunc.at(i) = 20. * log10(max(peak, 0.000001));

        //advance frame position by step

        startN += step;

    }

    return detectionFunc;

}
```

A.2 Spectral sum algorithm

```
// Input WAV samples stored in a std::vector
// Output detection function as std::vector
vector<double> spectralPeak(vector<float> const& audioSamples)
{
    int fftSize = 256;
    int binN = fftSize/2.0;
    step = 0.25*binN;//75% overlap

    //calculate how many iterations
    int frames = 1 + floor((audioSamples.size()-fftSize)/(double)step);
    vector<double> detectionFunc (frames);

    int startN = 0;
    double magnitude, sum;
    for(int i=0; i<frames; ++i)
    {
        Calculate FFT from audio samples at startN
        sum = 0.0;
        for  (int j=1; j<binN; j++) {
            magnitude = FFT[j];
            sum += magnitude;
        }
        //convert peak value to dB and store
        detectionFunc.at(i) = 20. * log10(max(sum, 0.000001));
        //advance frame position by step
        startN += step;
    }
```

```
    return detectionFunc;

}
```

A.3 Noise subtraction algorithm

```
// Input WAV samples stored in a std::vector

// Output detection function as std::vector

vector<double> noiseSubtraction(vector<float> const& audioSamples)

{

   int frames = 1 + floor((audioSamples.size()-fftSize)/(double)step);

   int binN = fftSize/2.0;

   step = 0.25*binN;//75% overlap

   vector<double> detectionFunc;


   int startN = 0;

   double signal, noise;

   signal = noise = 0;

   for(int i=0; i<frames; ++i) {

        Calculate FFT from audio samples at startN

        for(int j=1; j<binN; ++j) {

          double magnitude = FFT[j];//bin magnitude at j

          calculate median from circular buffer of previous magnitudes

          //subtract median

          double tmp = magnitude - median;

          //half-wave rectify

          tmp = (tmp + fabs(tmp)) * 0.5;


          tmp = max(tmp, 0.000001);

          median = max(median, 0.000001);

          //skip any bin less than background

          if(tmp < tmpMed) continue;

          signal += tmp;
```

```
            noise += median;
         }
            startN += step;//advance frame position by step
            //calculate signal to noise ratio in dB
            double dB = 20. * log10( signal / noise );
            detectionFunc.push_back(dB);
}
return detectionFunc;
}
```

Appendix B: Chapter 6 source code

C.1 - C++ function to read date and time from D500X file

```
// this C++ function returns the date and time as Unix time (seconds elapsed
// since 00:00 hours, Jan 1, 1970 UTC), read from the start of a D500X
// recording. Input is the path to the filename as a std::string, and
// Unix time output is as a long unsigned integer.


#include <iostream>
#include <cstdlib>
#include <cstring>
#include <string>
#include <fstream>
#include <iomanip>
#include <sys/time.h>
#include <sstream>


long unsigned int readTime(std::string filePath){
    long unsigned int epoch = -1;
    const char * astr = filePath.c_str();
    fstream callTime( astr, std::ios::in | std::ios::binary);
    if (callTime.is_open() ){
        callTime.seekg( 240, std::ios::beg );
        std::string waveChunk ("D500X");
        char chunk1[6] = {0,0,0,0,0};
        callTime.get(chunk1, 6);
        // extract date and time
```

```
if(!waveChunk.compare(chunk1)) {
   //this is a D500X recording
   callTime.seekg( 224, std::ios::beg );
   char chunk[16];
   callTime.get(chunk, 16); // read date and time from D500X
   int dateTime[16];
   int d;
   for(int i=0; i<16; ++i){
      std::stringstream ss;
      ss << std::hex << chunk[i];
      ss >> d;
      dateTime[i] = d;
   }
   // create a time struct and fill
   // with data from D500X recording
   // seconds elapsed since 00:00 hours, Jan 1, 1970 UTC
   // http://www.epochconverter.com/
   int dec[2] = {1,10};
   int tmp = 2000 + dateTime[0]*dec[1] + dateTime[1]*dec[0];
   dt.tm_year = tmp - 1900; // year - 1900 (years since 1900)
   tmp = dateTime[2]*dec[1] + dateTime[3]*dec[0];
   dt.tm_mon = tmp - 1; // month - 1 (months since January 0-11)
   tmp = dateTime[4]*dec[1] + dateTime[5]*dec[0];
   dt.tm_mday = tmp;
   dt.tm_hour = dateTime[7]*dec[1] + dateTime[8]*dec[0];
   dt.tm_min = dateTime[10]*dec[1] + dateTime[11]*dec[0];
   dt.tm_sec = dateTime[13]*dec[1] + dateTime[14]*dec[0];
   dt.tm_isdst = -1;//less than zero if not known
```

```
            dt.tm_wday = 0;
            dt.tm_yday = 0;
            epoch = mktime(&dt);
            callTime.close();
        }
    }
    return epoch;
}
```

C.2 - R code to aggregate bat activity by night and plot as a time-series by species group.

```
#load necessary R packages
library(chron)
library(zoo)
library(xts)

# load in data from csv file
#columns – “time”, “N”, “species”
# time = Unix time (an integer)
# N = binary presence absence (1 / 0)
# species = name of genus or species, e.g. “Pipistrellus”
setwd("C:/Users/...")#location of file
dataset<-read.csv(file = "filename.csv", header=TRUE)# filename
#convert Unix time (epoch) in column named ‘time’ to POSIX format
dataset$time<-dataset$time - 21600#minus 6 hours as bat activity runs overnight
dataset$time<-as.POSIXct(dataset$time, origin=ISOdatetime(1970,1,1,0,0,0), tz="GMT")

#convert to xts format
species<-mat.or.vec(0,1)
dates<-mat.or.vec(0,1)
values<-mat.or.vec(1,0)
#make time/date a factor
dataset $time<-factor(dataset $time)
#aggregate nightly activity
```

```
for(i in levels(dataset $species))
{
        #subset data by species
        subset<- dataset [which(dataset $species==i),]
        #subset$N is a column with binary presence absence (1 / 0)
        xts.ts <- as.xts(subset$N, order.by=subset$time)
        #sum the activity
        nightly<-period.apply(xts.ts, endpoints(xts.ts,"days"), sum)
        dates<-c(dates,index(nightly))
        values<-c(values,as.vector(nightly))
        species<-c(species, rep(i,length(as.vector(nightly))))
}
dat<-data.frame(species, dates, values)
dat$dates<-as.POSIXct(dat$dates, origin=ISOdatetime(1970,1,1,0,0,0),
tz="GMT")
dat$dates<-as.Date(dat$dates)


#code to plot nightly activity as time-series
library(ggplot2)
ggplot(dat, aes(dates, values, color = species))+
geom_point()+
facet_wrap(~ species, scales="free") +
#facet_grid(species ~ ., scales="free") +
scale_colour_brewer(palette="Set1") +
stat_smooth(alpha = 0.6, span = 0.4) +
ylab("Activity index (bat passes per night)") +
xlab("Date") +
opts(legend.position = "none")
```

C.3 - R code to summarise bat activity after sunset and plot the kernel density by species group.

```
#load necessary R packages
library(chron)
library(zoo)
library(xts)
library(maptools)

# load in data from csv file
#columns – “time”, “N”, “species”
# time = Unix time (an integer)
# N = binary presence absence (1 / 0)
# species = name of genus or species, e.g. “Pipistrellus”
setwd("C:/Users/...")#location of file
dataset<-read.csv(file = "filename.csv", header=TRUE)# filename
#convert Unix time (epoch) in column named ‘time’ to POSIX format
times<-as.POSIXct(dataset$time, origin=ISOdatetime(1970,1,1,0,0,0), tz="GMT")
times<-as.POSIXlt(times)
dataset$time<-dataset$time - 21600# minus 6 hours as bat activity runs overnight
dataset$time<-as.POSIXct(dataset$time, origin=ISOdatetime(1970,1,1,0,0,0), tz="GMT")
dataset<-data.frame(dataset, times)
#keep only presence data
dataset<-dataset[ which(dataset$N > 0),]
subset<-dataset
```

```
#convert to xts
subset$species<-factor(subset$species)
species<-mat.or.vec(0,1)
dates<-mat.or.vec(0,1)
hours<-mat.or.vec(1,0)
#day only
subset$time<-as.Date(subset$time)
#sunset times http://www.earthtools.org/
link <- matrix(c(-2.51857,54.22762), nrow=1)#Bullpot farm lat long
for(i in 1:length(subset$N))
{
        down <- sunriset(link, as.POSIXct(subset$time[i]), direction="sunset",
POSIXct.out=TRUE)
        sunset<-down$time#time of sunset
        hours<-c(hours, difftime(subset$times[i], sunset,
units="hours"))#hours after sunset
        dates<-c(dates,subset$time[i])
        species<-c(species, as.character(subset$species[i]))
}
dat<-data.frame(dates, species, hours)
dat$dates<-as.Date(dat$dates)
dat$dates<-factor(dat$dates)


#plot kernel density
library(ggplot2)
ggplot(dat, aes(hours, fill = species)) +
facet_wrap(~ species) +
scale_fill_brewer(palette = "Set1") +
xlim(-1, 12) +
```

```
geom_density(color = 'transparent', alpha = 0.7) +
ylab("Density") +
xlab("Hours after sunset") +
opts(legend.position = "none")
```

C. 4 – Environmental variables

Table C1: Environmental variables by date used for generalised additive modelling (GAM) of swarming data. Date, moon illumination in % (Moon), daily temperature high (Temp High), daily humidity high (Hum High) and low (Hum Low), daily pressure high (Press High), daily average windspeed (Wind Avg), and daily rainfall sum (Rain Sum).

Date	Moon (%)	Temp High (°C)	Hum High (%)	Hum Low (%)	Press High (hPa)	Wind Avg (km/h)	Rain Sum (cm)
09/08/2010	0.7	19	96	79	1017	6	0.03
10/08/2010	0.3	16	99	79	1009	14	0.15
14/08/2010	27.6	21	96	62	1023	4	0
15/08/2010	38.6	22	96	64	1025	5	0
17/08/2010	60.7	19	98	72	1012	15	0.33
21/08/2010	93.1	19	95	78	1014	19	0
22/08/2010	97.2	18	95	71	1013	12	0.05
24/08/2010	99.4	16	93	72	1011	21	0.48
25/08/2010	99.8	18	96	53	1013	5	0
26/08/2010	98.3	20	96	52	1009	2	0
31/08/2010	65.7	17	98	64	1027	2	0
01/09/2010	55.4	20	94	93	1023	0	0
03/09/2010	33.4	22	97	61	1022	1	0
04/09/2010	22.9	20	97	68	1021	2	0.03
05/09/2010	13.6	21	92	61	1020	4	0.05
06/09/2010	6.2	19	93	55	1014	6	0.48
10/09/2010	7	17	99	92	1015	9	2.18
11/09/2010	13.6	17	98	82	1014	13	0.46
12/09/2010	14.4	17	94	65	1024	13	0.05
13/09/2010	34	17	98	85	1023	28	2.34
14/09/2010	44.8	17	97	70	1012	35	0.56
15/09/2010	55.4	15	92	72	1009	32	0.1
16/09/2010	65.4	15	90	69	1012	11	0.03
17/09/2010	74.6	14	90	56	1018	5	0
21/09/2010	98	19	97	75	1017	2	0.03
22/09/2010	98.85	18	98	76	1016	2	0.51
23/09/2010	99.7	18	98	83	1007	3	0.89
24/09/2010	99.5	13	90	65	1017	11	0
25/09/2010	97.3	13	95	63	1020	3	0
26/09/2010	93.3	15	95	70	1020	2	0
27/09/2010	87.4	15	93	86	1016	1	0
28/09/2010	79.8	16	96	85	1016	1	0.13
29/09/2010	70.6	15	98	87	1014	6	1.32

Date	Moon (%)	Temp High (°C)	Hum High (%)	Hum Low (%)	Press High (hPa)	Wind Avg (km/h)	Rain Sum (cm)
30/09/2010	60.3	16	99	75	1013	2	0.03
06/10/2010	2.9	14	97	76	1010	9	0.94
07/10/2010	0.3	19	98	58	1017	2	0
11/10/2010	19	18	93	62	1023	1	0
12/10/2010	28.4	13	99	76	1024	1	0.03
13/10/2010	38.4	9	98	87	1024	1	0
14/10/2010	48.6	12	95	80	1023	1	0
15/10/2010	58.6	14	97	73	1022	5	0
16/10/2010	68.1	13	96	70	1027	1	0
17/10/2010	76.7	12	98	81	1027	7	0.05
18/10/2010	84.4	13	94	84	1019	18	0.13