

## Chapter 1

### Ancestral Reconstruction of Bat Echolocation Calls

J.P. Meagher\*

*Department of Statistics,  
University of Warwick,  
J.Meagher@Warwick.ac.uk<sup>†</sup>*

Something something bats something evolution something ancestral reconstruction something magic.

#### 1. Introduction

The emerging field of Data Science is driven by research which lies at the nexus of Statistics and Computer Science. Bioacoustics is one such area generating vast quantities of data, often through citizen science initiatives.<sup>12</sup> Bioacoustic techniques for biodiversity monitoring<sup>34</sup> have the potential to make real policy impacts, particularly with regard to sustainable economic development and nature conservation.

Bats (order *Chiroptera*) have been identified as ideal bioindicators for monitoring climate change and habitat quality,<sup>5</sup> and are of particular interest for monitoring biodiversity acoustically. Typically, a bat broadcasts information about itself in an ultrasonic echolocation call.<sup>6</sup> The development of automatic acoustic monitoring algorithms<sup>37</sup> means that large scale, non-invasive monitoring of bats is becoming possible.

Monitoring bat populations provides useful information, but an understanding the orders natural history is required to identify the cause and effect of any changes observed. The echolocation calls structure, which reflects a bats diet and habitat,<sup>8</sup> is a key aspect of this natural history. Reconstructing ancestral traits<sup>9</sup> relies on a statistical comparative analysis incorporating extant species and fossil records.<sup>10</sup> However, the fossil record

---

\* Author footnote.

<sup>†</sup> Affiliation footnote.

is of limited use in inferring ancestral bats echolocation calls. Therefore, statistical data science techniques may shed some light on this topic.

Previous studies of bat echolocation calls for both classification<sup>7</sup> and ancestral reconstruction<sup>11</sup> analysed features extracted from the call spectrogram. These call features relied domain knowledge to ensure they were sensibly selected and applied. More recently, general techniques for the classification of acoustic signals have been developed.<sup>124</sup> General techniques for the ancestral reconstruction of function-valued traits have also been proposed.<sup>13</sup> This piece of research applies some proposed techniques to bat echolocation calls.

A function-valued trait<sup>14</sup> is a characteristic of some organism measured along some continuous scale, usually time, and can be modelled as a continuous mathematical function by functional data analysis.<sup>15</sup> Jones & Moriarty<sup>16</sup> developed a method which extends Gaussian Process Regression<sup>17</sup> to model the evolution of function-valued traits over a phylogeny. A full demonstration of the phylogenetic Gaussian Process method for ancestral reconstruction on a synthetic dataset is presented by Hajipantelis et al.<sup>18</sup>

This general approach to evolutionary inference for function-valued traits is implemented here for a set of bat echolocation calls. Our goal in doing so is twofold. These techniques had previously been considered in the context of modelling the evolution of human speech sounds in language.<sup>13</sup> It is hoped that by applying these methods in a simpler context progress can be made towards resolving methodological problems. For example, how do we extend these methods to more realistic models of evolution?

We are also interested in what specifically these models tell us about bats and the evolutionary dynamics of echolocation. What impact might these results have on our understanding of ancestral bats and their behaviour?

This paper presents the early stages of our research and some preliminary results.

## 2. Echolocation Calls as Function-Valued Traits

A functional data object is generated when repeated measurements of some process are taken along a continuous scale, such as time.<sup>15</sup> These measurements can be thought of as representing points on a curve that varies gradually and continuously. In the context of phylogenetics, these functional data objects are function-valued traits.<sup>14</sup>

Denote the  $m^{th}$  call recording of the  $l^{th}$  individual bat of species  $S$  by  $\{\tilde{x}_{lm}^S(n) : n = 0, \dots, N_{lm}^S - 1\}$ . Thus,  $\{x_{lm}^S(\cdot)\}$  is a series of discrete

measurements of the function  $x_{lm}^S(\cdot)$ , the  $m^{th}$  call function for the  $l^{th}$  individual bat, observed at the time points given by  $\frac{n}{f_S}$ , where  $f_S$  is the sampling rate, in samples per second (Hz), of the recording. Assume then that  $x_{lm}^S(\cdot) = x_l^S(\cdot) + z_{lm}^S(\cdot)$ , where  $x_l^S(\cdot)$  is the representative call function for the  $l^{th}$  individual and  $z_{lm}^S(\cdot)$  is the noise process for the  $m^{th}$  call. Further to this, assume that  $x_l^S(\cdot) = x^S(\cdot) + z_l^S(\cdot)$  where  $x^S(\cdot)$  is the representative call function for the bat species  $S$  and  $z_l^S(\cdot)$  is the noise process for the  $l^{th}$  individual. It is the echolocation call functions at the species level that we are interested in.

The call recordings themselves are functional data objects, however modelling the phylogenetic relationships between  $\{x_{lm}^S(t)\}$  and  $\{x_{l'm'}^S(t)\}$  directly implies that the processes are comparable at time  $t$ . This is not the case for acoustic signals, a phenomenon which is often addressed by dynamic time warping.<sup>19</sup> Another approach to this issue is to consider an alternative functional representation of the signal.

The Fourier transform  $x_{lm}^S(\cdot)$  is given by

$$X_{lm}^S(f) = \int_{-\infty}^{\infty} x_{lm}^S(t) e^{-i2\pi ft} dt.$$

The energy spectral density of  $x_{lm}^S(\cdot)$  is the magnitude of the Fourier transform and the log energy spectral density is given by

$$\mathcal{E}_{lm}^S(\cdot) = 10 \log_{10} (|X_{lm}^S(\cdot)|^2). \quad (1)$$

Similarly to the call functions,  $\mathcal{E}_{lm}^S(\cdot)$  is the log energy spectral density of the  $m^{th}$  call of the  $l^{th}$  individual from species  $S$  where  $\mathcal{E}_{lm}^S(\cdot) = \mathcal{E}_l^S(\cdot) + \mathcal{Z}_{lm}^S(\cdot)$  and  $\mathcal{E}_l^S(\cdot) = \mathcal{E}^S(\cdot) + \mathcal{Z}_l^S(\cdot)$  where  $\mathcal{Z}_{lm}^S(\cdot)$  and  $\mathcal{Z}_l^S(\cdot)$  are noise processes. The log energy spectral density is a periodic function of frequency which describes the energy of a signal at each frequency on the interval  $F = [0, \frac{f_S}{2}]$ .<sup>20</sup> It is now assumed that  $\mathcal{E}_{lm}^S(f)$  and  $\mathcal{E}_{l'm'}^S(f)$  are comparable at frequency  $f$ .

The discrete Fourier Transform<sup>20</sup> of  $\{\tilde{x}_{lm}^S(n)\}$  provides an estimate for the log energy spectral density, the positive frequencies of which are denoted  $\{\mathcal{E}_{lm}^S(k) : k = 0, \dots, \frac{N_{lm}^S}{2} + 1\}$ . Smoothing splines<sup>21</sup> are applied to this series to obtain  $\hat{\mathcal{E}}_{lm}^S(\cdot)$ , a smooth function estimating  $\mathcal{E}_{lm}^S(\cdot)$ .

We now have a functional representation of each bats echolocation call where the pairs of observations  $\{f, \mathcal{E}_{lm}^S(f)\}$  and  $\{f, \mathcal{E}_{l'm'}^S(f)\}$  are directly comparable. These function-valued traits can now be modelled for evolutionary inference.

---

4 *J.P. Meagher et al.*

---

## References

1. P. E. Allen and C. B. Cooper. Citizen science as a tool for biodiversity monitoring. (2006).
2. N. Pettorelli, J. E. Baillie, and S. M. Durant, Indicator bats program: a system for the global acoustic monitoring of bats (2013).
3. V. Stathopoulos, V. Zamora-Gutierrez, K. E. Jones, and M. Girolami, Bat echolocation call identification for biodiversity monitoring: a probabilistic approach, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* (2017).
4. T. Damoulas, S. Henry, A. Farnsworth, M. Lanzone, and C. Gomes. Bayesian classification of flight calls with a novel dynamic time warping kernel. In *Machine Learning and Applications (ICMLA), 2010 Ninth International Conference on*, pp. 424–429 (2010).
5. G. Jones, D. S. Jacobs, T. H. Kunz, M. R. Willig, and P. A. Racey, Carpe noctem: the importance of bats as bioindicators, *Endangered species research*. **8**(1-2), 93–115 (2009).
6. D. R. Griffin, Echolocation by blind men, bats and radar, *Science*. **100**(2609), 589–590 (1944).
7. C. L. Walters, R. Freeman, A. Collen, C. Dietz, M. Brock Fenton, G. Jones, M. K. Obrist, S. J. Puechmille, T. Sattler, B. M. Siemers, et al., A continental-scale tool for acoustic identification of european bats, *Journal of Applied Ecology*. **49**(5), 1064–1074 (2012).
8. H. Aldridge and I. Rautenbach, Morphology, echolocation and resource partitioning in insectivorous bats, *The Journal of Animal Ecology*. pp. 763–778 (1987).
9. J. B. Joy, R. H. Liang, R. M. McCloskey, T. Nguyen, and A. F. Poon, Ancestral reconstruction, *PLoS Comput Biol*. **12**(7), e1004763 (2016).
10. J. Felsenstein and J. Felsenstein, *Inferring phylogenies*. vol. 2, Sinauer associates Sunderland (2004).
11. A. Collen. *The evolution of echolocation in bats: a comparative approach*. PhD thesis, UCL (University College London) (2012).
12. V. Stathopoulos, V. Zamora-Gutierrez, K. Jones, and M. Girolami. Bat call identification with gaussian process multinomial probit regression and a dynamic time warping kernel. In *Artificial Intelligence and Statistics*, pp. 913–921 (2014).
13. T. F. P. Group, Phylogenetic inference for function-valued traits: speech sound evolution, *Trends in ecology & evolution*. **27**(3), 160–166 (2012).
14. K. Meyer and M. Kirkpatrick, Up hill, down dale: quantitative genetics of curvaceous traits, *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. **360**(1459), 1443–1455 (2005).
15. J. O. Ramsay, *Functional data analysis*. Wiley Online Library (2006).
16. N. S. Jones and J. Moriarty, Evolutionary inference for function-valued traits: Gaussian process regression on phylogenies, *Journal of The Royal Society Interface*. **10**(78), 20120616 (2013).
17. C. E. Rasmussen, Gaussian processes for machine learning (2006).

18. P. Z. Hadjipantelis, N. S. Jones, J. Moriarty, D. A. Springate, and C. G. Knight, Function-valued traits in evolution, *Journal of The Royal Society Interface*. **10**(82), 20121032 (2013).
19. D. J. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, vol. 10, pp. 359–370 (1994).
20. A. Antoniou, *Digital signal processing*. McGraw-Hill Toronto, Canada: (2006).
21. J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. vol. 1, Springer series in statistics Springer, Berlin (2001).
22. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2017). URL <https://www.R-project.org/>.