Appendix S1: Optimization details
David L. Miller, Len Thomas
1 School of Mathematics and Statistics, and Centre for Research into Ecological and Environmental Modelling, University of St Andrews, St Andrews KY16 9LZ, Scotland
* E-mail: dave@ninepointeighthone.net

In practice, maximization is performed on the log-likelihood. However, as noted in the literature [1,2], mixture model likelihoods can be notoriously multimodal. This can cause problems when finding MLEs of the parameters. Here simulated annealing (SANN; [3, Chapter 10]) was used to explore the parameter space (for 500 iterations) then after that the Broyden-Fletcher-Goldfarb-Shanno method (BFGS; [3, Chapter 10]) was used to find the maxima (the implementations in the R function optim() were used). These two steps were run 5 times. This two step approach appears to be satisfactory in most cases.

The EM algorithm [4] was tested although there was no significant performance increase (in terms of computational time or parameter precision) over using BFGS with SANN. To aid the optimization, analytic derivatives were also used; these are given in Text S2.

Starting values
Beavers and Ramsay [5] give a method for estimating starting values for the scale parameter of a half-normal detection function. In the non-covariate case, the estimate is given as the intercept parameter from intercept only regression on log(y + \( \frac{w}{1000} \)) (where \( w \) denotes the truncation distance, as above). For covariate models, the equation used for the \( \sigma \) is used in the regression and the estimated parameters from the linear regression are used as the starting values for the \( \beta \)s.

A similar approach can be use in the mixture case by dividing the sorted distances into \( J \) equal parts. For each of these parts a Beavers and Ramsay-type estimate is used for the \( \beta \)s. The mixture weight \( \phi_j \) had starting values of \( 1/J \) since there is no reason a priori to believe anything else.

Parametrisation of the mixture proportions
When using 2-point mixtures, the constraint that the mixture proportions must sum to unity is enforced by definition (since \( \phi_2 = 1 - \phi_1 \)). However, in \( J \)-point mixtures when \( J > 2 \) ensuring that the proportions sum to 1 is not guaranteed. The obvious way to get around this would be to penalise the likelihood, should the optimisation procedure propose values for the \( \phi_j \)s that are not in accordance with this condition. This is, of course, inefficient and ugly. Instead, a parametrisation is used for the mixture proportions which yields \( \phi_j \)s that comply.

Rather than estimating the \( \phi_j \)s, estimate \( \alpha_p \)s \((p = 1, \ldots, J)\), where the relationship between the two is:

\[
\phi_j = F(\sum_{p=1}^{j} e^{\alpha_p}) - F(\sum_{p=1}^{j-1} e^{\alpha_p}) \quad \text{for } 1 \leq j \leq J - 1
\]

and

\[
\phi_J = 1 - \sum_{j=1}^{J-1} \phi_j
\]

where \( F \) is any continuous CDF on \((0, \infty]\). Exponentiation ensures that \( e^{\alpha_p} \geq 0 \), so \( \alpha_p \) may lie anywhere on the real line, allowing unconstrained optimisation. Summing these orders the \( \phi_j \)s, since only offsets are estimated. Finally, using the cumulative density function ensures that the \( \phi_j \)s sum to 1. In practise the Gamma(3, 2) CDF is (somewhat arbitrarily) used. Figure 1 illustrates the relationship.
To transform from the $\phi_j$s back to the $\alpha_p$s we simply re-arrange the above expression.

$$\alpha_p = \log\left(F^{-1}\left(\sum_{p=1}^{j-1} e^{\alpha_p}\right) - \sum_{p=1}^{j-1} e^{\alpha_p}\right).$$

Note that we only need as many $\alpha_p$s as we had $\phi_j$s, so we do not require any additional parameters.

References