Author: Jeff Laake at ~NMFS-NMML
Date: 1/22/1999 12:43 PM
Priority: Normal
TO: Tim Gerrodette at ~NMFS-L1
CC: kenb@lamar.colostate.edu at EXTERNAL, steve@mcs.st-and.ac.uk at EXTERNAL
Subject: Proration

Tim-

I thought some more about proration on the way back to Seattle. Prorating by the proportion of animals is correct. The concern I expressed about encountering groups rather than animals will only affect the manner of variance estimation.

Currently, you are prorating all of the unidentified. An alternative is to scale the estimates based on the proportion of animals identified near the line. If all unidentified animals were away from the line, they could be ignored entirely, because the definition of g(x) would simply be the probability of detecting and identifying the sighting. However, by using this definition if all groups were detected but some were not identified, g(0) < 1. An estimate of g(0) using animals (instead of sightings) and assuming all are detected on the line is simply the number of animals identified on the line (e.g., say within 200-300 m) divided by the total number of animals seen on the line (using same distance). The current estimates of abundance could simply be divided by this estimate of g(0) - thus multiplicatively scaling the estimates upward. This is somewhat ad-hoc and will be less precise than using all of the data but it avoids the problem of differing detectability and group sizes between species. However, like your current approach you are assuming that each species is equally likely to be identified near the line. It would be very easy to construct this to prorate unid spotted and spinners and unid dolphin. It is possible to extend this approach in a manner similar to the approach you proposed of fitting a curve to the unidentified. However, doing so would require simultaneous fitting of the detection curves for each species that are contained within the unidentified and would require an additional curve c(x) that described the probability of identifying (classifying) a species as a function of distance. What you did in your analysis of the unidentified implicitly assumed that c(x) was a uniform distribution. That is unlikely and is what caused the strange behavior in the tail of the distribution. It is likely that c(x) is constant for some distance from the line and then decreases with distance much like a detection curve. Thus the expected distribution (if all were detected) for the unidentified is 1-c(x). As an example consider two species which each have g(x)=1 for 0-W but for each c(x) = exp(-a-bx). The detection curves for the classified observations of each species would look like a negative exponential (g(x)=c(x))but for the unidentified it would be an increasing function 1-exp(-a-bx). If a=0, c(0)=1 and there is no reason to correct for the unidentified. Also, clearly if a>0, g(0)=c(0) and the abundance estimates from the classified detections can be corrected by dividing each by exp(-a). This is a rather simplistic and unrealistic example but it demonstrates both the problem in fitting the unidentified and my suggested solution for correcting the abundance estimates from the classified observations. I have written out the general formulae for fully analyzing the unidentified (unclassified) observations jointly with the classified observations but I need to check it and try it out. This could actually be useful in these multi-species surveys where it is not always possible to correctly classify each observation. When I get this written up more fully I'll pass it on.

I enjoyed yesterday's meeting. It's always nice to think about a different problem for awhile. It spurs the creative juices.

--jeff