Draft loggerhead sea turtle (Caretta caretta) 2009 status review under the U.S. Endangered Species Act

Comments on the draft review by:
Dr. Jeffrey Moore, Duke University Marine Laboratory, Beaufort, NC
jemoore@duke.edu
252-504-7678

Dear Barbara,

Thank you for giving me the opportunity to provide comments on this ESA status review for loggerheads.

The bulk of my comments are related to sections pertaining to extinction risk assessment, as this is my area of research expertise. I am currently conducting research on use of diffusion approximation models to northwestern loggerhead populations; I have nearly completed a report on this research (will be finished within days), and I will be preparing it for submission for peer-reviewed publication over the coming weeks as well. I do not know whether the review process from here forward would allow you to incorporate any of that work (if you would be interested).

In general, my comments are on methodological aspects of the risk assessments rather than on particular input parameters or results specific to a DPS. Any DPS-specific comments are limited the Northwest Atlantic DPS.

I have provided comments below and also directly to the draft report. Comments here are mostly focused on the diffusion-approximation work. Comments directly in the report are mostly on the matrix-model section.

All of my comments directly in the report can be found on pages 7, 11 – 12, 25, 30, and 33 – 48.

Quick minor comment: At the risk of self-promotion, I have highlighted a few areas in the report where I think information from Eckert et al. (2008), a paper on which I am a co-author, should be included.

Comments on diffusion approximation and SQE:

My first set of comments concern the use of Susceptibility to Quasi-extinction (SQE), the metric published by Snover and Heppell (2009). I have become quite familiar with the Snover/Heppell paper, and I think these authors have provided an important step forward for dealing with uncertainty associated with quasi-extinction estimation. Specifically, I am an advocate of interpreting distributions of extinction risk metrics through approaches other than means and confidence intervals. I am collaborating on a different project with both authors, and I admire their work. However, while I think SQE is useful, I think it's worth noting about few caveats about the method:

Most importantly, the low rates (~10%) of Type I and Type II error (falsely classifying a population as 'not at risk' or 'at risk', respectively) resulting from their analysis is likely to be unrealistically optimistic. This is because Snover and Heppell tested the SQE approach using simulated populations whose 'true' risk probabilities were for the most part very high (> 0.9) or very low (< 0.10). Where true risk is less certain (e.g., true Prob(QE) ~ 0.3 - 0.7), classification error rates would be expected to be much higher.

Secondly, SQE is in some ways a more complicated and arguably less transparent way of defining risk criteria that could be applied without this metric. In the example used in their paper (Snover and Heppell 2009), SQE is the proportion of bootstrapped quasi-extinction (QE) estimates that exceed 0.90, a value they defined as 'high risk'. As an aside, the same 0.90 value was used in the current status review as well, even though I would argue that 0.90 is too high under a precautionary mandate. I would argue that no value greater than 0.50 should be used, i.e., if the expected risk of decline to a threshold is more probable than improbable, then the population should be considered at risk (of reaching this threshold). Anyway, through simulation, Snover and Heppell determined that for a QE risk of 0.90, a critical SQE of 0.40 corresponded to balancing the minimization of Type I and Type II error rates. So, if > 0.40 of the bootstrapped quasi-extinction estimates are > 0.90, the population is defined as 'at risk'. But in addition to the concern I mentioned above (error rates likely to be higher than presented), I note that the SQE approach consists of merely doing two things:

(1) A level of unacceptable risk is defined. In their paper, and in this review, unacceptable risk is defined as Prob(quasi-extinction) > 0.90. But such a decision must be made in any quasi-extinction evaluation.

(2): A decision is made about the level of certainty required to conclude that the population is indeed at risk. Again, such a decision must be done in any risk assessment. Snover and Heppell have defined the certainty level as SQE, but one could just as easily summarize the bootstrap results of a PVA by saying, for example, "40% of the bootstrapped distribution was greater than the at-risk cutoff of 0.90" and then make a decision as to whether that's acceptably certain or not to conclude that the population risk is likely to the exceed the cutoff.

Now, the strength of Snover and Heppell's approach is that it attempts to help us quantitatively decide what level of certainty should be acceptable. For example, we might intuitively use a certainty level of 0.5 (the median), implying the population risk is more likely than not to be above the 0.90 cutoff (or whatever cutoff is used). Yet their simulations suggest that if the risk cutoff is 0.90, we should actually use a certainty level of 0.40 (not 0.50) if the goal is to minimize Type I and Type II errors. But, why should it be 0.40 instead of 0.50? I think it may be related to bias in quasi-extinction estimates under the diffusion approximation method.

Snover and Heppell (2009) suggested that SQE is robust to using different QE thresholds and different L (running sum lengths used in the QE estimation). But I fear this may again be because they tested the method on simulated populations that were unambiguously at risk or not at risk. I expect the critical value may actually be sensitive to what time horizons and running sum lengths are used when true risk is less certain than > 0.9 or < 0.1, and that even if Type I and Type II errors are optimally minimized at the critical value, these error rates may often be fairly high.

With respect to the use of diffusion approximation methods in general, there is a tradition in sea turtle applications of picking running sum lengths (L) so as to approximate the population

size of adult females (e.g., L = 2 or 3 in the case of loggerheads, corresponding roughly to inter-nesting intervals). But this needn't (and shouldn't) be the goal of picking an optimal L; there's no real theoretical basis for doing so under the state-space model paradigm that underlies the slopes estimation method of Holmes (2001, 2004, etc). Rather, L should be selected simply to minimize bias in the estimates of σ^2. Unfortunately, since we don't know truth, it's hard to know what that L should be, but simulation work I'm conducting suggests that L > 3 may be better than smaller L. Using L's that are too small lead to high-biased estimates of σ^2. The good news is that this bias is generally more precautionary.

Also, it is not necessary to estimate the number of adult females (from the number of nests) to conduct the diffusion approximation analysis. In fact this is philosophically inappropriate because the dynamics of adult females are not expected to behave like a diffusion process anyway (unpublished results). Nest counts can be used directly, realizing that the goal of the state-space model is to make inference on a latent variable (e.g., an age-class weighted estimate of whole population size). But as noted in the report, the choice of nests or adults should not affect the results since a constant conversion between nests and adults was used.

Finally, see Table 2. In my current analysis of Northwest Atlantic units (specifically PFRU and NRU), I do not get the same estimates (or confidence intervals) for σ^2 as those presented here. This makes me wonder about the methods that were used in the status review to obtain these estimates. Page 30 presents a somewhat vague explanation of how the parameters and confidence intervals were estimated, and it's not clear that appropriate methods were used (e.g., the 'slope method' for variance point estimates with appropriate distributions for bootstrapping as described on p. 1287 of Holmes 2004 and/or p. 2381 of Holmes and Fagan 2002).

Comments on Threat Matrix Analysis

Most of my comments on this section were made directly to the report draft. Here, I just wish to summarize some major points.

On the plus side, I appreciate all the work that went into this analysis, and there are certain aspects of it I like. I think work presented in Figures 11-17 is particularly useful (plausible combinations of life history parameters), and I thought the use of the negative binomial model to identify appropriate matrix dimensions for representing variation in AFR was especially creative and useful. I also think that conceptually, the use of expert opinion to explore plausible ranges of population growth rate due to anthropogenic effects has good potential.

My main concern, however, is the arbitrary nature by which different levels of threat ('very low' to 'high') were assigned actual mortality rate values, and the high values that were considered. More generally, I am concerned that using opinion provides an illusion of knowledge about vital rates for which there really is no good knowledge. The opinion-based threat 'estimates' resulted in population growth (λ) estimates for different DPSs that are so uncertain (e.g., spanning ranges from ~ 0.6 to 1.05) as to be arguably of little value, and I think the lower limits are unrealistically pessimistic in most if not all cases; this is partly because the mortality rates assigned to different

threat categories tend to span broad ranges and are probably, for the most part, too high. For example, "low" equals added mortality of 0.01 to 0.10 in an age class. But added mortality of 0.10 should not be considered 'low'. "Medium" mortality was considered 0.10 – 0.20, when in fact an additive mortality rate of 0.20 should be considered extremely high. This is surely higher than the harvest rate allowed for most fish or game populations!

That said, I think there may be value in comparing the maximum λ estimated for each DPS and in general for looking at the relative comparisons across DPS's. The fact that only NE Atl and Mediterranean DPS have max λ's less than 1 may be telling us something about the crisis state of these populations. These relative comparisons should perhaps be more strongly emphasized than the actual estimates themselves, and more discussion of caveats concerning the expert-opinion approach would be good to see.

Comments on SQE vs. threat matrix

Diffusion approximation models (used for SQE in this report) have their limitations, but they have the highly desirable property of minimal data requirements that in the case of sea turtles, reasonably match the data types available. Uncertainty and caveats notwithstanding, DA methods are simple and more easily tied to real data, and in the present example have therefore provided more straightforward and empirically grounded insight than the matrix analysis. I think it may be worth pointing this out.