MSC response to comments from Janet Coetzee, Chair, Small Pelagics Working Group, Republic of South Africa

Note that the MSC responses are given in bold italics.

Comments from 1 August 2011

In summary, and without detracting from the valuable contribution which the Smith *et al.* paper makes to the topic in question, including strengthening the strategic basis for higher target levels for such fisheries in comparison to higher trophic level fisheries, in the context of the proposal for a 75%B_0 target figure for LTL fisheries as a threshold for certification, consideration of the material in the paper now available to us **exacerbates our concerns and strengthens our view that these specifics of this MSC proposal are premature,** for the reasons following.

1) We note that on the third page of the Smith *et al*. paper its authors state: "For this reason we do not consider that these models should be used to provide tactical management decisions." Yet it seems to us that in advocating a choice of a 75%B_0 target based on the results provided in this paper, the MSC is in direct contradiction of this by putting forward a specific tactical decision concerning threshold choice.

There is a distinction to be made between tactical use of a particular model and the general results found in Smith et al. (2011) that have been found to be robust across a range of ecosystems, model types, and low trophic level species. The MSC is keen to remove ambiguity in application of the FAM and this necessarily involves providing quantitative guidance where possible. The results found in Smith et al. (2011) are the best information currently available for low trophic level species in this respect.

2) We originally asked if the parameters of the models considered "*have been estimated through fitting to time series of abundances of predators and prey in the studies conducted".*

a) The paper comments: "*Each of the models has been validated against time series data from well studied systems*". Although these models may indeed all have had some of their parameters estimated by fitting to such data, the reliability of any inferences to be drawn from the models are dependent on the quality of these fits - the extent to which they at least approach the customary standards required for diagnostic tests for single species model fits that are used in providing fisheries management advice. This we have not been placed in a position to be able to judge adequately. In the short time available we have been able to access publications referenced for only two of the examples considered by Smith *et al*. The one shows no fits of the model to such data, and the other shows only some, a number of which could hardly be considered acceptable under the norms above. In asking consultees to comment on these findings, the MSC should minimally provide all such fits for all the models considered, to allow an overall appraisal of the acceptability of these model fits for the purpose for which MSC wishes to use their results.

The "norms" mentioned are a matter of some debate (e.g. Plaganyi 2007, FAO 2008). All the models used in the study have been tuned to data for each of the ecosystems involved. In most cases this has involved fitting to time series data over extended periods but in some cases (e.g. the Humboldt) the fits were to mean conditions over a recent period of time. The papers that outline and illustrate the model fits are listed in Smith et al. (2011) with additional references listed in Appendix 1 below. The fits to means and trends are generally good but most models do not fit the variability particularly well. This is a feature common to many ecosystem models particularly in variable ecosystems (see comments in Gaichas et al. in prep.). Clearly the state of the art in fitting ecosystem models to data is still a long way from that for stock assessment models, which is one of the reasons that their use for tactical decision making is not yet widely advocated. However models broadly and adequately fitted to trends in time series data are suitable for answering the strategic questions being addressed in this case. While it is to some extent a matter of judgement, models that fit not only trends but variability could be used for more tactical purposes though we are not aware of such applications to date.

b) It is important in framing an assessment of the reliability of the models used to have some idea of the level of rigorous evaluation to which they have been subjected by local scientists responsible for providing fisheries management advice in the regions in question. To indicate towards that end, can we be advised of the extent to which any are taken into account directly in the provision of tactical scientific advice for fisheries controls (catch, effort limitations etc.) in these regions? The sense of this question is not whether any of the models themselves are used to provide such advice directly (we suspect not), but rather whether any have been used in making the choice for overall controls such as medium term biomass target levels.

To our direct knowledge none of the models has been used to provide tactical advice so none has been subject to rigorous evaluation equivalent to that provided for stock assessment models by those responsible for providing fishery management advice. However all the models have been developed by scientists who are very familiar with the local ecosystems and who are part of their local scientific communities. Indeed most of the model authors work for local scientific agencies with direct responsibilities for providing fishery management advice (Kaplan, Field, Tam, Mackinson, Fulton, Bulman, Johnson, Smith). Apart from normal peer review processes involved in journal publication, most of the models used have been subject to review and scrutiny by local scientists familiar with the ecosystems and fisheries involved. In the case of management agency scientists, the formal reports and publications have been subject to internal agency review and approval, and in several cases to external review as well.

c) A similar recent model to those considered in the paper - an EwE-like based analysis of the Gulf of Alaska ecosystem: Sarah K Gaichas, Kerim Y Aydin and Robert C Francis: What drives dynamics in the Gulf of Alaska? Integrating hypotheses of species, fishing and climate relationships using ecosystem modeling (CJFAS, accepted) - has similar difficulties in finding acceptable fits to all the time series of data to which it is fitted. When questioned about this and its implications for the reliability of more tactical inferences from the model following a recent presentation, Aydin responded that it was unrealistic

to expect acceptable fits to all series, but that where inferences about specific species were required, it was necessary to fit the model giving much higher weight to series associated with that and closely related species (in the food web) at the expense of other species and the quality of fits to time series for them. This suggests that single overall fits of ecosystem models should not be relied upon for the exercises attempted by studies such as Smith *et al.*, but rather that a series of such fits should be examined, each focussing in turn on fitting more closely to the data for the species (and ones closely related) that are the focus of the study. If this has not yet been done for the Smith *et al.* study, we suggest that it needs to be considered as part of further robustness testing.

The Smith et al. study team is also aware of the Gaichas et al study and has been in touch with the authors. There is broad agreement between the two teams on issues surrounding fitting such models to data, though as noted previously the state of the art in this respect is still developing. Several of the models in Smith et al. have explored uncertainties in a similar fashion to that proposed by Gaichas et al. For example the Atlantis model for south eastern Australia investigated many alternative parameterizations and reported on these (see the table in the supplementary material). For major commercial species such as sardines, herring and anchovies, good time series data are generally available and selection of all models used in the study involved parameterizations that fit these data well. For some non-targeted groups such as mesopelagics and krill, the time series were less extensive.

d) The Smith *et al.* study most appropriately considers the robustness of results to applications of structurally different ecosystem models to the same ecosystem. In the broadest of terms, results are indeed similar, but at the next level of scrutiny some major differences are apparent. For example for the Southern Benguela, the estimated impact of the depletion of anchovy is substantially different for the EwE and OSMOSE models (Fig. 2 of Smith *et al.*). This must raise serious questions if it is the results of such modelling, as at present, that were to provide, even if in somewhat more of a strategic than a tactical sense, the basis of the MSC's threshold target requirement for certifying the South African fishery on anchovy, were this to be requested for consideration.

The revisions to the FAM for LTL species envisions and allows the use of "credible" ecosystem models for more tactical purposes (such as supporting system and species specific reference points) provided certain criteria are met (see TAB directive D-036 paragraphs 9b and 14a). The directive states that ""Credible" should be interpreted here to mean 1) publicly available and well documented, 2) fitted to time series data and 3) comprehensive (dealing with the whole ecosystem including all trophic levels)". In cases where such models are used in this way, MSC would also envisage that such models would be subject to the kind of scrutiny and review proposed in point 2b above.

3) Empirical evidence of extent of impact [Note our original request that "*empirical evidence as well as broad modelling studies needs to be presented demonstrating impacts at this level on predators for such a (75%B_0) target*"]

a) Convincing evidence to support the extent of impact on predators suggested by these ecosystem models would come most powerfully from the models' fits to the time series data for the prey - prey fishery - predator triplets under particular consideration. Disappointingly however, the "Empirical support" section of the on-line supporting material for the Smith *et al*. paper appears not to offer such but rather primarily qualitative correlations, implying by omission that there is not actual evidence to back the **quantitative** extent of the impacts predicted by these models - is that correct?

The quantitative support for the models resides primarily with the fits to data and particularly time series, already discussed above. MSC is also aware of a number of purely empirical analyses that examine particular associations (e.g. seabirds in relation to their prey) that are in review and that generally support the findings in Smith et al. However the advantage of using models is that impacts throughout the food chain can be considered, and "experiments" run involving different levels of depletion of selected LTL species. Data sets do not exist for any ecosystem that would allow this range of experiments to be evaluated across large parts of the ecosystem and for large numbers of LTL species.

b) Examples of support offered can be open to alternative explanation, such as predator reductions reflecting a temporary distributional shift to another area offering improved feeding conditions at the time, rather than that a reduction of predator abundance caused reduction in biomass of a LTL species as the ecosystem models used to project effects are implying. For example, the supporting material for the North Sea models cites broad agreement of EwE model predictions and empirical data that minke whales have declined (would decline?) by more than 40% in response to a 60% reduction in sand eels. However, Norwegian cetacean sightings survey results (Nils Oien, *pers. commn*) do not indicate any decline in the North-east Atlantic minke whale population in question, though the proportion of these whales in the North Sea does vary over time.

We agree that multiple interpretations of particular empirical associations are possible and that data do not "speak for themselves". This is why we prefer to rely on the use of reasonably well-validated models where the relationships and assumptions are at least clearly stated, particularly if results are fairly robust across model types and assumptions.

4) As pointed out in our original submission, the choice of 40% as the threshold for impact on a predator in the evaluations of impact that lead to the $75\%B_0$ target recommendation is essentially arbitrary. It is unfortunate that the Smith *et al*. contribution shows results for this choice only, and gives no basis for the choice other than referring to it as reflecting a "severe" impact. Other choices could be defended, and one would then want to see what corresponding LTLF depletion target would be implied.

A justification of the choice of 40% impact is that it is a level that would tend to avoid listing of any species impacted by LTL depletion as vulnerable under IUCN criteria. The IUCN criteria for vulnerable listing identify two thresholds – a 50% impact where the causes are well understood and the cause has ceased, and a 30% impact where the cause is generally understood and the impact has not ceased. A sense of where the 40% impact

level sits relative to the broader distribution of impact levels can be seen in Figures S1 and S2 in the supplementary material to Smith et al. Figure S1 shows that most (but not all) of the impacts occur within the 40% level at 75% B_0, but that much larger impacts are more common at B_{MSY} levels.

5) Surplus production function (Fig. 4 of Smith *et al*. Implies a loss of only 20% of yield in targeting 75%B_0 instead of B_{msy})

a) It is the squarish shape of this surplus production function in Fig. 4 that leads to this conclusion of a lesser loss than might normally be expected. No details are given of the calculations that led to this curve, and the question is begged of what particular aspect (input, assumption, ...) of the models examined is leading to this - how robust is this result? Note also that the argument of lesser fishing costs at a higher abundance would not apply in many LTLFs where shoaling behaviour leads to highly non-linear catch-rate vs abundance relationships.

The shape of the surplus production function for each LTL species is a direct output from each ecosystem model and is a complex function of the particular assumptions in each model. The values of yield and biomass at each depletion level were determined by running the model to "near" equilibrium (generally 50 years) under fixed Fs for each focal LTL species with other species fished at status quo levels. Figure 4 shows the overall results across all models, ecosystems and LTL species. The fact that the shape of the surplus production function differs from that derived from single species models is interesting but not necessarily surprising. Both are the product of particular model assumptions.

b) For most of the models considered in the Smith *et al*. paper, the impacts on predators of fishing the LTL species were calculated under deterministic assumptions. This might be expected to over-estimate those impacts, as LTL species frequently fluctuate appreciably in abundance at a time scale too short to allow the larger predators with their typically slower dynamics to take full advantage of upward fluctuations in the abundance of their prey. Thus these predators are unable to grow to the level that deterministic analyses would suggest in the absence of fishing, and hence would likely not be impacted as greatly by fishing on their prey.

OSMOSE is not a deterministic model and the Atlantis models were driven by fluctuations in oceanographic conditions. Most of the Ecosim models were fitted using observed variations in primary production, though the future projections were deterministic. Diets in OSMOSE are not pre-determined, but are size based and emerge from patterns of local abundance. While we agree that it would be interesting to explore further the role of fluctuations in abundance of prey on predator dynamics, we do not envisage that this would substantially change the overall results from the study.

6) We suspect that there may be problems with the manner in which recruitment fluctuations in LTF species are being generated in some of these models, which could lead to misrepresentative results. Because of the shortness of time afforded for these comments, we have not been able to check this sufficiently. We would not normally comment before

completing such checks, but since we would wish to do so to you if these suspicions are confirmed in due course, we mention this now so that our motivation in such possible further submission later is not misunderstood.

 This issue is addressed further below.

In conclusion we consider the Smith *et al*. paper a valuable first step, but certainly not sufficient as a basis for the decisions on target thresholds for certification which the MSC is proposing. The next step should be the organising of a vigorous and wide-ranging review of this work, leading to suggestions to take it further towards the stage where it might provide such sufficiency. The level of rigour for which we are asking may seem high, but the MSC is suggesting this work to be the key basis for novel criteria for LTL fisheries management. If applied and followed, these criteria would quite likely have very heavy negative socio-economic impacts, while the associated predictions of subsequent gains have limited reliability given the current state of development of the field. Decisions should not be taken before the underlying science has been subject to much more careful scrutiny than might more customarily apply, i.e. they need to be taken very carefully and not rushed.

The issue of low trophic level species and how they should be assessed by MSC has been raised by many stakeholders over several years. MSC set up a consultative process that sought wide stakeholder input and ran two public workshops on this issue. MSC also commissioned the research presented in Smith et al., which was presented in detail at the second of these public workshops. While the results in Smith et al. are unlikely to be the last word on the matter, and the authors of that study agree that many of the uncertainties raised by the South African Small Pelagic Scientific Working Group deserve further study, MSC considers that the results in Smith et al. do encapsulate current best understanding of these issues and form a reasonable basis on which to introduce changes to the MSC Fishery Assessment Methodology. As with all such issues, further developments in the science will be kept under close review, and future changes will be made as new information comes to light.

Further comments 12 August 2011

In our earlier comments to you on the Smith et al. paper, under point 6) we remarked:

"6) We suspect that there may be problems with the manner in which recruitment fluctuations in LTF species are being generated in some of these models, which could lead to misrepresentative results. Because of the shortness of time afforded for these comments, we have not been able to check this sufficiently."

We have now had the opportunity to look a little further into this point, which concerns the EwE modelling approach used for five of the nine models considered by Smith et al. We understand (though have yet to confirm for all cases) that in these EwE models, recruitment fluctuations have been modelled through putting a (potentially annually estimable) multiplier ("forcing function") on primary production, and/or by temporally varying one or

more of the parameters of the foraging arena interaction model (for the interaction of the species concerned with either or both of its prey and its predators).

However, particularly with LTL small shoaling pelagic species (e.g. sardine, anchovy) in mind, the general understanding is that year-class strength is primarily determined at the egg-larval stage, where key events can happen on spatial and temporal scales much smaller than those customarily used for fisheries (including EwE) population modelling. Thus, for example, it is having feeding conditions just right for a few pockets of eggs-larvae shortly following spawning that can lead to the occasional very strong year-class.

The question that then arises is whether the EwE forcing function approach can mimic this process adequately. Clearly all the methods listed above can produce variations in year-class strength for a LTL species under consideration, but the important associated issue is whether at the same time they perhaps have unrealistic side effects. For example, varying the primary production will increase phytoplankton abundance, thus creating better conditions for the species under consideration, but also for other species that are reliant directly or indirectly on phytoplankton in a way that therefore correlates closely with the recruitment fluctuations for the species under consideration - is that realistic/appropriate for all ecosystems?

There are similar questionable side effects from introducing temporal variation in the foraging arena model interaction terms between the species of interest and its predators. That method would produce good recruitment by reducing predation mortality. But that means that feeding conditions are worse for the predators normally responsible for such predation, so that their abundances are adversely affected, when that would not be happening if the recruitment variability is instead being driven by the egg-larval mechanism described above.

If the forcing functions are instead acting through the foraging arena interaction terms linking the species of interest with its prey, this means that good recruitment is produced in EwE by the species feeding more successfully. That has the problem under the egg-larval mechanism that the amount of extra food consumed by those further surviving eggs-larvae would be negligible compared to other predation on that prey species, but EwE will elevate that impact to the detriment of the prey's abundance and hence to other species feeding from it. Those further surviving larvae will, as they grow, require more food, but is that not already being taken care of through the standard relationships - indeed unless life stages in the species of interest are very highly resolved in setting up the EwE model, older fish of that species will be benefitting as well (and inappropriately) from the adjustment to the parameters of the interaction term.

Thus given the reality of generally large recruitment fluctuations in LTLFs, and concerns that the side-effects produced by the way EwE mimics these effects could impact overall model behaviour strongly but unrealistically, is it not premature to draw quantitative conclusions

from the EwE models considered by Smith et al. before such aspects are investigated (unless naturally they already have been)? We note from Aydin (pers. commn) that the EwE-like ecosystem models he develops deliberately model recruitment fluctuations more directly (see, for example, Appendix 1 of the Gaichas, Aydin and Francis paper referenced in our earlier comments) to avoid these same side effects of the more standard forcing functions of EwE approaches which concern us.

Given that the majority of the models upon which the Smith et al. conclusions are based are EwE, this again leads us to the conclusion that the MSC would be premature in using the Smith et al. analysis as the basis for certification criteria, at least until the above (together with the other issues we raised earlier) have been more thoroughly investigated.

These are useful insights and we thank the SPSWG for taking the trouble to consider these issues in depth and to communicate them in such detail. In fact, although most of the Ecosim models were fitted to historical time series data by incorporating variability in primary production forcing, the projections on which the results were based were deterministic. This would seem to imply that the issues raised above by the Working Group do not arise in this instance. We do however note that they may have influenced some aspects of the parameterization of the Ecosim models, and this issue should be explored further. For OSMOSE and Atlantis, the variability was reflected in both the historical period and the projections. While agreeing that the SPSWG has raised some legitimate issues, we do not see that this should prevent decisions being made in the mean time on the basis of the results presented.

References cited

FAO (2008)
Gaichas et al. (in review)
Plaganyi (2007)
Smith et al. (2011)