Center for Independent Experts (CIE) Independent Peer Review of the Gulf of Alaska Walleye Pollock Assessment

Carmen Fernández

August 2012

Executive Summary:

This report provides a review of the Gulf of Alaska walleye pollock assessment. The 6 ToRs for the review are presented in the "Description of review activities" section and center around the input data, the assessment model and methodology, reporting of outputs, and SPR percentage-based reference points in non-stationary environments. A Review meeting took place at the Alaska Fisheries Science Center (AFSC), in Seattle, WA, during July 17-20 2012, in which presentations on the different aspects of the assessment were given and discussions held.

The Gulf of Alaska sustains an important commercial fishery on walleye pollock. The stock is assessed annually using a flexible purpose-built statistical catch-at-age model, fitted by maximum likelihood. Inputs to the model are commercial fishery data (catch in weight and age or length compositions) and survey data (abundance indices and age or length compositions). Biological parameters (natural mortality, maturity, weight at age) are treated as known in the assessment. The stock has been exploited lightly (fishing mortality is estimated to have been less than $F_{40\%}$ in most years) and predation mortality (not modelled in the assessment) is very high, believed to be well above fishing mortality. The Gulf of Alaska ecosystem has been experiencing changes in the last decades, with a regime shift in the late 1970s. The abundance of main pollock predators show diverse trends, several of which are increasing, highlighting the relevance of taking environmental and ecosystem processes into account in the assessment and management. Doing this in a formal, quantitative fashion is, however, difficult. Although the current pollock assessment does not incorporate these processes, the work conducted so far on Management Strategy Evaluation suggests that the current management strategy is precautionary in the face of them.

From my perspective, the current assessment is sound. The model has been constructed sensibly and makes appropriate use of the available data. A number of questions were raised during the Review process, and several aspects that merit further investigation identified, together with suggestions for how the issues may be taken forward. A detailed discussion is provided later in this report and a complete bullet point list of suggestions and recommendations is presented at the end of the main body of this report. I recommend that they are all considered and explored. Here I provide a concise summary of only the points I identify as most relevant.

In terms of the input data, much of the discussion versed around the difficulties in obtaining representative length frequency samples by trawling in acoustic-trawl surveys, and the implications this can have for the surveys' ability to provide an abundance index for the whole population. It is recommended that this aspect is further explored, and suggestions in this respect were provided during the Review meeting.

Regarding the assessment itself, my two main recommendations are that: (1) serious consideration is given to estimating the catchability of the bottom trawl survey in the assessment, instead of fixing it at q=1; (2) further exploration is conducted on ways of constraining the selectivities at the older ages (considering also the selectivity of the commercial fishery), examining the impact this has on the assessment's results. Some exploratory assessment runs were conducted during the Review meeting, and I recommend that further analysis of model fit diagnostics takes place before reaching definite conclusions. All exploratory runs conducted during the meeting excluded the historical (pre-1984) bottom trawl survey indices, because of the uncertainty surrounding their properties, given the difficulties in deriving those indices. The exploratory runs illustrated the sensitivity of the assessment results for those early years (1960s to early 1980s) to the data used and assumptions made. Having seen the sensitivity of the results, I (tentatively) suggest keeping the historical indices in the assessment (assuming some reliability can be attached to them, which I understand is the case, despite the difficulties associated with their derivation).

Concerning the presentation of inputs and reporting of outputs, my main recommendation is that a more comprehensive exploratory analysis of input data be conducted and presented, as well as providing a more comprehensive presentation of model fit diagnostics.

The ToR about SPR percentage-based reference points in non-stationary environments was not discussed during the Review meeting. Having read again through documents after the meeting, I believe that AFSC scientists have the issues of relevance for pollock clearly identified. I suggest that their Management Strategy Evaluation work could be expanded to test further the performance of different ways of defining F_{MSY} and B_{MSY} proxies in the context of the changing GOA ecosystem. I also encourage them to continue their work on developing models and simulation-testing for situations involving environmental and ecosystem processes (mainly predation on pollock). An assessment model incorporating these effects could be developed and tried in phases and, if/when considered appropriate, might eventually become the main assessment model for pollock. I expect this would be a longer term (rather than a short term) goal.

Background:

The Gulf of Alaska (GOA) sustains an important commercial fishery on walleye pollock (hereafter denoted as pollock). It started as a foreign fishery in the early 1970s, then developed into a mixture of foreign, joint venture and domestic fisheries in the late 1970s, and it has been fully domestic since the late 1980s. Catches were just a few thousand tonnes (t) in the 1960s, but subsequently increased strongly, reaching a maximum of around 300,000 t in 1984. After this peak, catches declined to around 90,000 t in 1986 and have fluctuated around this level since then. The fishery is closely monitored and regulated. Annual catches have for over 2 decades been restricted by, and generally close to, the set TACs. Pollock is a semi-pelagic schooling fish caught mainly with pelagic trawl gear. Incidental catch of other species in the GOA directed pollock fishery is low. A much bigger pollock fishery exists in the Eastern Bering Sea (EBS), but pollock in the GOA and EBS are considered to be different stocks and assessed and managed separately. Studies of pollock stock structure within the GOA are, however, not conclusive. Peak spawning at the two major spawning areas in the GOA occurs at different times (around the second half of February in the Shumagin Islands area, and around the second half of March in the Shelikof Straight), but it is not clear what causes the difference.

The state of the GOA pollock stock has been assessed annually since the late 1970s, using information from the commercial fishery and several research surveys (bottom trawl, acoustic-trawl and egg surveys). Not all surveys are conducted every year and the egg survey stopped in 1992; the assessment incorporates the years available for each of the surveys. The data used in the assessment are commercial catch, survey biomass indices, age composition in the commercial fishery and in the survey samples, and length composition data (only for the years/surveys when age composition data are lacking). Natural mortality, proportion mature and weight at age are also inputs to the stock assessment (treated as fixed values, not as parameters to be estimated within the assessment model). Several statistical catch-at-age assessment models have been used in the past: CAGEAN during the 1980s, Stock Synthesis during the 1990s and, since 1999, a purpose-built model for the pollock stock, coded in AD Model Builder and fitted by maximum likelihood. The same fundamental model structure and assumptions have been used since first implemented in AD Model Builder in 1999, although minor changes have been implemented to deal with novel situations as they arose. The assessment model is fairly standard, following the usual exponential equation for decay in abundance within cohorts, with catches-at-age modelled via the Baranov catch equation and with observation equations (likelihoods) assumed to be log-Normal (for catch and survey biomass indices) and Multinomial (for compositional data). Some model features deal with specific aspects of the pollock assessment, but they are still well within the realm of standard modelling tools (*e.g.* a random walk is used to model changes in fishery selectivity over time). Dorn *et al*. (2011) describe the pollock assessment, including the model, in detail. Discussion of some aspects of the input data, assessment model and output reporting is provided later in this report, under ToRs 1-3.

The current GOA pollock assessment model includes ages 2-10+ and assumes a constant natural mortality rate M=0.3 across ages and years. There is no evidence of a stock-recruit relationship holding for this stock, and annual recruitment (at age 2) is estimated in the assessment with a separate parameter for every year. The assessment starts in 1961 and results indicate a very strong stock increase from the 1960s to the early 1980s, followed by a long and strong decrease. SSB is estimated to have stabilised since the late 1990s at low levels. The stock has been exploited lightly, with $F \leq F_{40\%}$ throughout the entire time period.

Pollock is a mid trophic level species and a key component of the GOA ecosystem, with many predators and preys. Dorn *et al.* (2011) and other documents presented for the Review indicate that predation mortality on pollock is likely to be very high (well above fishing mortalit*y*), age-dependent (generally higher for younger ages of pollock) and time-varying (as a consequence of changes in the GOA ecosystem). A regime shift occurred in the GOA in the late 1970s and the abundance of main species in the ecosystem show diverse trends. Whereas pollock biomass has decreased strongly since the early 1980s, biomass of arrowtooth flounder has been continuously increasing through the same period and is the biggest source of pollock mortality. Pacific halibut, Pacific cod and Steller sea lions are other main predators of pollock, and have shown diverse trends through time. Diet analyses indicate that pollock constitute a very high proportion of the diets of Pacific halibut (48% in weight) and Steller sea lions (40% in weight), whereas it has lower importance in the diets of Pacific cod and arrowtooth flounder. This suggests that a low pollock stock could strongly impact the dynamics of Pacific halibut and Steller sea lions, but is less likely to have an effect on Pacific cod and arrowtooth flounder.

The issues mentioned in the previous paragraph suggest that pollock assessment and management could be improved by including ecosystem processes in the assessment. AFSC scientists have invested considerable effort working in this direction and this work was presented for the Review. For example, the 4 papers by A'mar *et al.*, the 3 papers by Gaichas *et al.*, and the papers by Hollowed *et al.* (2000) and Dorn (2004), all explore aspects connected to ecological (mainly predation) and environmental processes, and potential ways of incorporating them in the pollock assessment or, alternatively, the effect that ignoring them in the assessment and management may have on the resource. The issues are obviously complex and difficult to model, and no alternative pollock assessment model was presented in the Review as a potential replacement for the current stock assessment.

The harvest control rule used to derive the advised catch (ABC proposal), has F depend on the estimated value of current SSB in relation to reference points (with lower SSB implying lower F). In particular:

- If $SSB \geq SSB_{47\%}$, then $F=F_{40\%}$;
- If $SSB < SSB_{47\%}$, then F decreases linearly from $F=F_{40\%}$ when $SSB=SSB_{47\%}$ to F=0 when $SSB=0.05*SSB_{47\%}$;
- If $SSB \leq SSB_{47\%}$, then F=0.

On top of this, no directed pollock fishery is allowed if $SSB < SSB_{20\%}$, with the aim of protecting the endangered Steller sea lions, for which pollock is the main prey species.

$F_{x\%}$ is defined as the value of F that results in SPR = x% of SPR_0 (where SPR_0 is the SPR value corresponding to F=0) and $SSB_{x\%}$ is here defined as x% of SPR_0 times average recruitment. In a non-stationary context, as is the case here, questions arise concerning the appropriate ranges of years on which inputs to $F_{x\%}$ calculations (fishery selectivity, natural mortality, weight at age, maturity), and recruitment for $SSB_{x\%}$, should be based. This will be discussed under ToR 4.

Description of review activities:

The Review was organised around a meeting held at the Alaska Fisheries Science Center (AFSC), in Seattle, during July 17-20, 2012. The documents marked with (*) in the Bibliography section of this report were provided to the reviewers about 2 weeks in advance of the meeting and constitute the central material for the review. Additional documents were made available during the meeting and are also all listed in the Bibliography, although given the extensiveness of the material, the reviewing effort concentrated on the material provided in advance of the meeting.

The meeting followed quite closely the planned agenda of presentations, developing as follows:

Tuesday, July 17, 2012

Anne Hollowed	- Welcome and Introductions, Adoption of Agenda
Martin Dorn	- Overview of biology, surveys, fishery, management system

Michael Martin	- Gulf of Alaska bottom trawl survey
Chris Wilson	- Acoustic surveys in the Gulf of Alaska
Kresimir Williams	- Evaluation of net selectivity in acoustic surveys
Lisa Thompson	- Fishery monitoring of the GOA pollock fishery
Kerim Aydin	- Role of pollock in the GOA ecosystem

Wednesday, July 18, 2012

Martin Dorn	- Pollock stock assessment model
Teresa A'mar	- Management Strategy Evaluation of GOA pollock assessment
Martin Dorn	- Discussion of proposed assessment model changes

Thursday, July 19, 2012

Martin Dorn	- Evaluation of alternative model configurations
Martin Dorn	- Continued evaluation of alternative model configurations

Friday, July 20, 2012
Continued evaluation of alternative model configurations and informal discussions with AFSC scientists.

The following ToRs were given for the review process:

1. Evaluate and provide recommendations on data collection procedures and analytical methods used to develop assessment model input.
2. Evaluate and provide recommendations on model structure, assumptions, and estimation procedures.
3. Evaluate and provide recommendations for the reporting of assessment results and characterization of uncertainty.
4. Evaluate and provide recommendations on F35% spawning biomass per recruit as an appropriate proxy for FMSY under non-stationarity in vital rates. Also evaluate and provide recommendations on the B35% biomass reference point as a proxy for BMSY.
5. Recommendations for further improvements.
6. Brief description on panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

There was no specific division of tasks among the three reviewers, each of us fully participating in all aspects of the review. The procedure I followed to provide this review was to read carefully in advance the documents provided, then to exchange views and clarify questions with AFSC scientists and the other reviewers during the meeting and, finally, to review the documents once again (benefiting from the insights gained during the meeting) and go through additional literature as a follow up to some of the discussions held during the meeting.

The AFSC scientists were very helpful clarifying doubts and questions during the discussions held at the meeting. I was impressed by the team of people and the obvious quality of their work, at the cutting edge of marine science and research, as their many publications in top international journals make clear. The breadth and depth of material presented made it a challenging task for the reviewers, but also a very interesting and rewarding experience, from my perspective.

Summary of findings for each ToR:

This section presents the main points that arose during the review, according to my own perspective and understanding of the issues discussed. Thoughts from following up (after the meeting) on some aspects of the work presented and discussed there, are also included. This section is organised following the 6 ToRs.

ToR 1. Evaluate and provide recommendations on data collection procedures and analytical methods used to develop assessment model input.

The datasets used in the assessment are:

- estimates of annual catch in tonnes and fishery length or age composition (depending on year),
- NMFS summer bottom trawl survey (starting 1984) biomass indices and age composition,
- winter (March) acoustic-trawl survey (Shelikof Straight, starting 1981) biomass indices and age composition,
- spawning biomass indices from egg survey in the Shelikof Straight (1981-1992),
- ADF&G summer bottom trawl survey (starting 1989) biomass indices and length or age composition (depending on year),
- biomass indices from historical surveys (conducted in some years prior to 1984), and length or age composition depending on year.

All elements above were described in different presentations during the Review meeting and discussed in detail following the presentations. Here I highlight main points of discussion.

NMFS summer bottom trawl survey:

This survey starts in 1984, and was triennial from 1984 to 1999 and biennial thereafter. It uses chartered commercial vessels, fitted with standardised NMFS bottom trawl gear, and is conducted according to NMFS bottom trawl survey protocols. Usually three vessels take part in a survey. The survey is aimed at a range of groundfish species, not just pollock. It follows a stratified random design, with 59 strata based on regulatory areas, bathymetry and major gullies.

The main points raised related to how homogeneous the survey time series could be considered (the assessment model assumes that the survey catchability q and selectivity-at-age is constant through the range of survey years). It was explained that in 1996 tow duration changed from 30 minutes to 15 minutes and the way of measuring tow duration also changed in the same year (from brake set – haulback before 1996, to on bottom - off bottom since 1996). It was also explained that in 1984 and 1987 the survey was conducted cooperatively with Japanese vessels, with some issues concerning the gear used by those vessels and their objectives. Years 1984 and 1987 stood out as different in time series graphs shown for net spread and net height. Additionally, in 1990 and 1993 the survey was conducted by a different organisation east of 144ºW, and their primary objective was rockfish. A historical graph of survey timing (day of year versus longitude) indicated high consistency in the timing of this survey since 1996, but much less before then (even though it was always conducted in the May-September months). All this raises the question of whether the survey before 1996, and particularly in 1984 and 1987, is consistent enough with later years (to be used in the assessment as a single homogeneous abundance index for the whole period). I note that the two other pollock abundance indices for the mid 1980s (from acoustic-trawl and egg surveys) both indicate a decline in biomass in that period (particularly the acoustic-trawl survey index), not consistent with the stability indicated by the NFMS bottom trawl survey. These conflicting signals cause a misfit of the assessment model to the survey biomass indices in the mid 1980s.

AFSC scientists are well aware of all this. They indicated that the NMFS survey series is used from 1984 in the assessments of all groundfish species, and that treating it differently for pollock would require a clear rationale as to why this was required for this species. Exploratory assessment runs for pollock, removing the 1984 and 1987 surveys, were conducted during the Review meeting, with results presented in this report under ToR 2.

Another point raised, which is important from my perspective, is the fact that the assessment assumes catchability q=1 for this survey. Because the catchability-at-age is modelled in the assessment as q*s(a), where the maximum value of s(a) over the ages is 1, q in reality represents the highest survey catchability among all ages included in the assessment. Pollock are distributed in different parts of the water column, and insights from survey experts as well as results from assessment model runs where q is estimated instead of fixed to 1, all point to q being less than 1, likely somewhere between 0.5 and 1. I find it difficult to see a justification for assuming q=1 in the assessment, and recommend that the possibility of removing this assumption is seriously considered (as happens for the vast majority of stock assessments with which I am familiar). Exploratory assessment runs estimating q were conducted during the Review meeting and results reported below under ToR 2.

Acoustic-trawl winter (March) survey in Shelikof Straight:

Two main points were highlighted. One of them (length selectivity of mid-water trawls) was raised by AFSC scientists, whereas the other one (whether the length frequency sampling design in the survey was appropriate for the purpose of deriving a biomass index of the whole population) was raised by a reviewer. I discuss the two points in sequence.

Length selectivity of mid-water trawls

Acoustic-trawl surveys operate by receiving an acoustic backscatter signal, from which aggregations of biomass in the water column can be identified. Trawl gears are then used to sample some of the aggregations in order to identify the fish species composition and length frequencies of the fish in the aggregation. Ideally, the trawl gear has the same selectivity and catchability for all species and lengths in the aggregation, so that the correct composition is observed. If that is the case, the observed frequencies of different species and lengths in the aggregation can be combined with the backscatter signal, taking into account the target strength (TS) of each species by length, in order to obtain an estimate of numbers-at-length per species in the aggregation. Normally, combining the observed hauls with backscatter signal is based on strata rather than individual aggregations, but this does not change the basic problem raised at the meeting, which is that if the trawl selectivity and catchability vary with species or with length within a given species, the numbers-at-length calculated by the method just described will be biased towards those species and lengths with higher trawl selectivity and catchability. This will, in turn, result in wrong estimates of survey biomass (obtained multiplying numbers-at-length by weight-at-length and then summing over the lengths) and proportions-at-age in the survey. Since larger fish often have higher trawl selectivity than smaller fish, it may be expected that total numbers will often be underestimated (as the aggregation will be wrongly believed to contain a bigger proportion of large fish for the same amount of acoustic backscatter signal), whereas total biomass may be over or underestimated (resulting from a combination of underestimating the number of fish and overestimating the proportion of large fish, which are heavier than small fish). These errors depend on the population length structure, which varies between years depending, among other things, on cohort strength. If the problem is severe enough, it can compromise the use of the survey as an index of stock abundance (in other words, the assumptions of constant catchability and selectivity-at-age through time can be severely compromised).

This problem arises for acoustic-trawl and not for standard bottom trawl CPUE surveys, as it is related to the fact that acoustic-trawl surveys use the trawls only to estimate proportions-at-length and the scaling up to numbers (later converted to biomass) is done by combining the trawl sampling information with the acoustic backscatter signal.

AFSC scientists have examined this problem in the context of mid-water trawls for the Shelikof Straight acoustic-trawl survey (Williams *et al.*, 2011, and Williams, 2012). Pollock of a wide range of lengths can appear in the hauls, so trawl gear selectivity is potentially an important source of error in survey abundance estimates. Williams *et al.* (2011) focuses on estimating length-selectivity (*i.e.* length-dependent retention of fish that have come into contact with the gear) based on three experiments, two conducted in the GOA and one in the EBS. Each experiment was analysed separately and results presented during the Review meeting. The experiments and analysis performed are described in detail in Williams *et al.* (2011), so I will not repeat that here. I expand on a few technical remarks I made during the meeting, in case they may help for future work by the authors.

A Bayesian analysis is conducted for analysing the results of the experiments, using improper priors (*i.e.* prior distributions that do not integrate to a finite value, such as the uniform prior used for μ_i, where μ_i is the number of fish of length i entering the mouth of trawl). Although improper priors are very often used in Bayesian analyses, they do not guarantee the existence of a proper posterior distribution (*i.e.* a distribution that integrates to 1), which is a requirement for Bayesian results to make sense. This often seems to be overlooked when conducting Bayesian inference and, even though improper priors often result in proper posterior distributions, there is no guarantee that it will happen for a particular case and there is no guarantee either that, if the posterior distribution turns out to be improper, this will be detected when calculating it using computational algorithms (MCMC). There have been instances in the published literature where this problem has arisen because it had gone undetected when performing the MCMC computations. Clearly, if a density function is finite everywhere and is restricted to a bounded domain, then it will integrate to a finite constant (and, hence, correspond to a proper distribution). So if there is a good basis for restricting the unknown parameters to a particular bounded domain and the density function is finite everywhere in that domain, then no problems will occur. However, the choice of the bounded domain to which the parameters are restricted will not always be obvious. For example, if a uniform prior is set on a positive parameter μ and one wanted to turn it into a proper prior distribution by restricting μ to some bounded domain, one might consider limiting the upper range for μ to 100, or perhaps to 1,000 or 1,000,000. All these values may, in a particular instance, be considered as potential upper limits for μ. However, if one was interested in *e.g.* the probability that μ is less than 1, this probability would be 0.01=1/100, 0.001=1/1000 or 0.000001=1/1000000, depending on the upper limit chosen for μ. Hence, these seemingly uninformative priors (which may all be considered feasible and "realistic" in a particular situation) are saying very different things about the probability that μ is less than 1. Depending on the shape and curvature of the likelihood, this may have stronger or weaker effects on the posterior distribution.

I am not implying that these problems necessarily arise in the length-selectivity analysis presented in Williams *et al.* (2011), as they are very case-specific and generally difficult to check, but I make the general comment that using improper prior distributions in Bayesian analyses without checking (somehow, it is not an easy matter in general) that the posterior distribution is proper can cause problems. Even if a computing program internally selected some finite domains (ranges) for the parameters in question, hence formally solving the technical matter, inference on some quantities of interest may potentially be substantially impacted by the choice of bounded domain.

In the particular context of the Williams *et al.* (2011) model, an alternative to the uniform prior on μ_i (μ_i is the number of fish of length i entering the mouth of trawl, intervening in equation (6) of the paper) that could be worth trying is a prior density proportional to $1/\mu_i$, which is the "usual non-informative" prior for positive quantities. Even though this is also an improper prior distribution, in this case it is easy to see that it does not cause the posterior distribution to become improper. Using this prior, the parameter μ_i can be marginalized out of the Poisson likelihood, very similarly to how it has been done for the uniform prior in Williams *et al.* (2011), and the resulting marginal likelihood is seen to correspond to the one obtained if the number of fish of length i in each of the 13 locations (12 pocket nets and codend), $\{x_{im}, m=1,...,13\}$, had been modelled conditionally on the total number in the 13 locations, $\sum_m \{x_{im}\}$, as Multinomial with probability $F_{im}/\sum_m \{F_{im}\}$ of being encountered in location m (F_{im} defined as in the paper). The fact that marginalising out the parameter μ_i leads to a result that corresponds to a well-defined likelihood (the Multinomial just described) proves that using the improper prior distribution with density proportional to $1/\mu_i$ does not create problems in this instance.

My recommendation when conducting Bayesian analyses is to use proper priors (unless a way is found to check that the posterior distribution under a certain improper prior is proper), so that the problems mentioned above do not arise, and then to conduct a sensitivity analysis (in other words, vary the prior and see what impact this has on posterior results) to explore whether the posterior quantities of interest are unduly impacted by the prior choices.

Along the same lines, my recommendation when conducting Bayesian analyses is to report results based on posterior quantiles (median, 5 and 95 percentiles, etc.) instead of posterior moments (mean, variance, CV, etc.) given that all quantiles are well-defined for any proper distribution, but the same can not be said about moments, whose existence is strongly dependent on the thickness of the tails of the distribution. Checking existence of moments of the posterior distribution is generally also difficult and, hence, my preference for using quantiles, which cause no such problems.

Following after the work of Williams *et al.* (2011), Williams (2012) explores the impact of correcting the acoustic-trawl survey indices using the trawl length-selectivity estimates. The document shows that the degree to which survey indices are corrected is year-specific, depending on the population length-structure that year, generally requiring bigger corrections when higher proportions of small individuals are present. The work by Williams (2012) also indicates that stock assessment estimates of F and SSB appear to be reasonably robust to the selectivity corrections, whereas bigger changes occurred for estimated Recruitment. The estimated trawl selectivity in the 2 GOA experiments is quite different, with a considerably larger L50 value in the 2007 experiment (26 cm) than in 2008 (15 cm). Within each experiment, the posterior-predictive distribution of the length selectivity curve, upon which survey corrections to indices rely, is highly uncertain. This suggests that trawl selectivity may be highly variable and that further analyses and development would be required before attempting to correct the acoustic-trawl survey series currently used in the assessment.

As Williams (2012) says, given the various other sources of error that can happen in this type of survey (such as sampling error, target strength, performance of acoustic instrumentation, etc.) the question of their relative importance, so as to be able to optimize survey results within practical limitations, is pertinent.

Is the trawl sampling design and processing of hauls' data able to provide a biomass index of the whole population?

The acoustic-trawl survey provides an index of stock biomass and proportions-at-age, which are used in the assessment model in the same way as the bottom trawl survey data. In other words, a selectivity-at-age and catchability are assumed to relate the survey data to the fish population, with the same selectivity-at-age values applying both to the survey biomass index and the proportions-at-age from the survey.

During the Review meeting, doubts were raised concerning the appropriateness of the trawl sampling design leading to the survey biomass index and proportions-at-age. Only a rather limited number of hauls is carried out during a survey (11 hauls in the 2012 survey). From discussion with AFSC scientists I was left with the impression that there is no very clear sampling design for where/when the trawls are conducted, they seem to take place essentially according to expert judgment, trying to target higher fish density areas and following experts' expectations of what type of fish aggregation a given acoustic signal is likely to represent. The length frequency distributions (LFD) from the sampled hauls are post-stratified, aiming to have similar LFDs within each stratum. In 2012, 4 strata were defined. All acoustic signals that are not sampled by the trawl are assigned to one of the strata. I did not fully understand how the non-sampled signals are allocated to strata but, as far as I could see from the results presented, each stratum is contiguous in space and I presume the allocation procedure is, to a large extent, based on expert judgment. The average LFD of the hauls sampled in each stratum is used to calculate the numbers-at-length for the entire stratum (taking target strengths into account).

The main question raised was whether the sampling scheme is able to produce reliable LFDs for each stratum, given the small number of hauls carried out, as well as the difficulties in sampling at the selected trawling locations appropriately (*e.g.* are the dense aggregations sampled well enough, with the trawl traversing through them, or are the trawls only capable of going through their margins? If structure exists within aggregations, with fish of different lengths or ages distributed in different parts of the aggregation, is the trawl able to get a representative sample of the aggregation? In some cases where the acoustic signal shows several aggregations at different depths in the water column, but only one of the aggregations is sampled by the trawl, may this lead to biased results (*e.g.* because larger fish tend to be close to the sea bottom and juveniles higher up in the water column)?). If pollock were mixed with other fish species, this would significantly increase the difficulties in obtaining a representative length frequency sample and raising it to the stratum, since the other species would also need to be taken into account. The main species that appears mixed with pollock in the Shelikof Straight is eulachon, which has very low target strength and is therefore ignored. The question was raised, however, that even if its target strength is very low, eulachon could still have an impact on the acoustic signal if it was present in much larger numbers than pollock. Clearly, the length-selectivity of mid-water trawls, discussed in detail in the previous subsection, is one more element that can introduce error in the process.

The above issues were discussed at great length during the Review meeting. Not being an expert in acoustic-trawl survey methodology, I feel there is little I can contribute to this discussion, except to say that the questions raised seem relevant and to encourage the very capable AFSC acoustic survey team to investigate them, which I suspect they have already done to some extent.

It was suggested during the meeting that focusing the acoustic survey index only on the spawning individuals, as opposed to the entire population, would likely result in fewer problems. Fish aggregate to spawn, so it may be possible to identify the aggregations specifically corresponding to spawning fish and to obtain a representative sample of them by trawl. A spawning biomass index could be constructed from the acoustic-trawl survey. The assessment model could then link the survey spawning biomass index to the pollock stock using as selectivity the maturity-at-age values. Exploring the properties of this procedure would require going through the historical survey data and reanalyse them, selecting the aggregations and hauls deemed to correspond to spawning fish. For exploration purposes during the Review meeting, a spawning biomass index was constructed from the current survey numbers-at-length distributions, splitting them at 43 cm (length at 50% maturity) and the assessment was ran with this index. Results from the experiment are presented under ToR 2.

Whereas I consider these issues relevant and certainly worth investigating, I also note that plots of the current LFD and numbers-at-age from the survey do not seem too bad at tracking cohorts through time (see the two bubble plots below, under ToR 3), specially up to about age 7, which to me suggests that the problems with the way the survey is currently used are possibly not too severe.

ADF&G summer bottom trawl survey:

This survey has been conducted annually since 1989, in nearshore areas of the Gulf of Alaska. It is designed to monitor population trends of crab species, but also used in the pollock assessment. The survey does not cover the entire shelf area and catches mostly large pollock, with the LFD mode typically above 45 cm. The LFDs are quite similar in most years, no cohort tracking is apparent from visually inspecting them, but this is not surprising given the range of lengths the survey catches (with several ages most likely contributing to the mode of the LFD). There seemed to be a slight lack of in-depth information concerning this survey, with no presentation made about it and no document describing it available to the reviewers in advance either (a document was made available during the meeting).

The questions raised in connection with this survey are whether its spatial coverage is sufficient to provide an abundance index for pollock and whether much is gained by using it in the assessment.

Pre-1984 bottom trawl survey:

The current assessment runs from 1961, but no survey information is available before 1981, when the acoustic survey series started. The NMFS and ADF&G bottom trawl surveys started in 1984 and 1989, respectively. An effort has, therefore, been made to make information from pre-1984 bottom trawl surveys in the GOA "usable" for the assessment.

Between 1961 and 1983, bottom trawl surveys were also carried out in the GOA, but using a different gear and with a different design from the NMFS series. These earlier surveys normally aimed to cover the whole of the GOA over a period of several years, or to survey a large area to obtain a combined index for a group of groundfish species. Finding ways of using this information in the current assessment is far from straightforward. A procedure based on fitting a GLM to observed CPUE at 4 selected sites, using pre- and post- 1984 data, was developed in the past (details in Dorn *et al.* (2011)) and (some modification of) fitted results for the pre-1984 years used as abundance indices in the current assessment.

As stated in Dorn *et al.* (2011), questions concerning the comparability of pollock CPUE data from historical trawl surveys with later surveys probably can never be fully resolved. It is debatable whether including the pre-1984 indices thus derived in the assessment improves its performance, given all the uncertainties associated with their derivation. Exploratory assessment runs, excluding these indices, were conducted during the Review meeting and results shown below under ToR 2.

ToR 2. Evaluate and provide recommendations on model structure, assumptions, and estimation procedures.

The assessment model is explained in detail in Dorn *et al.* (2011). A clear and detailed presentation was also given during the Review meeting. Many aspects were discussed during the presentation and some of them were subsequently tried in exploratory runs.

As I said earlier in this report, the assessment model is reasonably standard and, from my perspective, it has been constructed sensibly and, on the whole, makes appropriate use of the available data. There are two main points which I feel require further exploration (even after the work conducted during the meeting):

(1) The use of a fixed catchability value q=1 for the NMFS survey. I can not see a convincing justification for using this and my recommendation is that serious consideration is given to estimating this parameter as part of the assessment (with a prior on it, or some kind of constraint, if necessary).

(2) The use of a double-logistic selectivity function to fit the age composition data of the commercial fishery and almost all surveys, without any constraints on selectivity at the older ages. As I understand it, only the ADF&G survey is assumed to have asymptotic selectivity (logistic), but since age composition data for this survey are only available for very few years and the Multinomial sample sizes associated with those data are very low (sample size = 10, considerably lower than for all other age composition data used in the assessment), I imagine the selectivity of the ADF&G survey does not have any noticeable impact on the assessment's results. In my experience, assumptions about selectivity-at-age at the older ages can have considerable impact on the assessment's results (SSB in particular). When the selectivity of the older ages is allowed by the model to be very low, it often tends to be estimated that way, resulting in large population biomass estimates for the older ages. This is a kind of "cryptic" biomass, seemingly not detected by the commercial fishery or the surveys, which makes one wonder whether such biomass really exists or whether the high estimate is just an artifact arising as a consequence of certain modelling assumptions (*e.g.* a value of M that is too low, such that too many old fish remain in the estimated population).

This situation may be occurring to some extent in the current assessment, where the selectivities of the commercial fishery, the acoustic and NMFS bottom trawl surveys are all estimated to be very low for the older ages. I do not think an obvious "solution" exists, but I encourage further exploration by trying (potentially various) ways of constraining selectivities at the older ages. Alternatives to assuming asymptotic selectivities for the fishery or the NMFS survey, could be to assume that selectivities (of fishery and/or surveys) remain constant above a certain age or that the selectivity of the plus group is an average of the selectivities of a predefined set of lower ages. I suggest dealing with this by trying alternative selectivity assumptions on the older ages, exploring the diagnostics from the resulting assessments, and also using expert judgment (ancillary information, discussion with fishery and survey experts, etc.) to make a final choice.

Some of the exploratory runs conducted at the meeting incorporated elements of my points (1) and (2) above, but I still feel additional exploration of those points would be useful. Many of the exploratory runs extended the current modelled range, 2-10+, to 1-13+ or 1-15+. I am not convinced that increasing the range of ages in the assessment, particularly at the older ages, will help. By using a 10+ group in the assessment, an implicit assumption is made that all biological and fishery processes intervening in the assessment remain constant as of age 10. Adding older ages to the assessment means that modelling assumptions have to be made for those ages too. There is, however, very little information about those old ages, which are only present in very small proportions in the commercial fishery and survey data (*e.g.* in the commercial catch, the percentage of 10+ fish is below 10 in most years, and the percentage of 11+ fish is below 5 in most years; similar percentages are found for the NMFS bottom trawl survey and are lower for the acoustic-trawl survey). My feeling is that increasing the range of older ages in the assessment at this stage, given the limited available information about them, and also in line with my comments about selectivity for older ages in point 2 above, is probably an unnecessary complication.

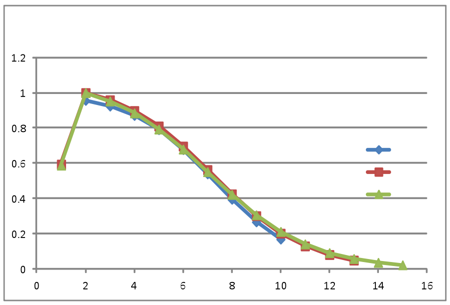
I now summarise the main features and results of the exploratory assessment runs during the meeting.

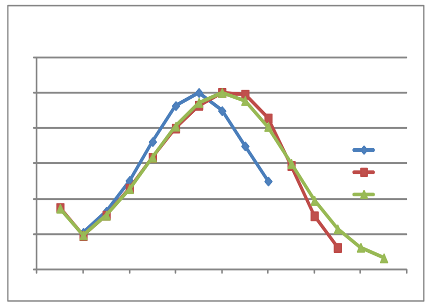
The first set of explorations was conducted in advance of the meeting and consists of the following elements (with respect to the current assessment):

- Increase modelled age range (from 2-10+ to 1-13+ or 1-15+) – see summary below
- Use mean-unbiased likelihoods for survey indices (*i.e.* assume that the Log-Normal observation equation for survey biomass indices has mean, instead of median, equal to the corresponding population quantity) – summary below

- Evaluate bottom trawl survey selectivity and catchability – summary below (closely related to my points 1 and 2)
- Evaluate acoustic survey selectivity – when estimating selectivity-at-age using a separate parameter for each age, the shape of the estimated selectivity-at-age was, overall, in line with a logistic function with negative slope, as assumed in the current assessment; these changes had no significant impact on the population SSB estimated from the assessment.

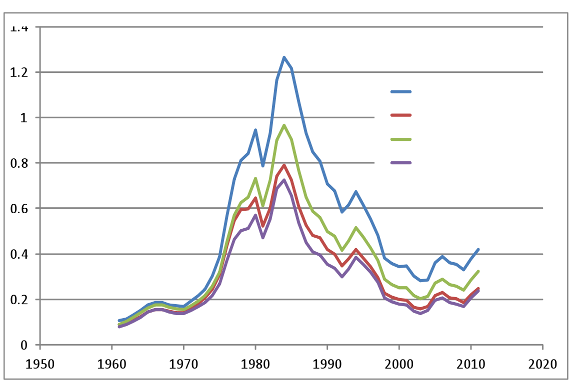
First bullet point: Increasing the modelled age range resulted in some (quite minor) increases in annual SSB and some (also minor) changes in estimates of year class strength, while keeping the same trends through time. Age ranges 1-13+ or 1-15+ led to very similar results. Estimated surveys' selectivities were affected as shown in the following graphs (I note that selectivity at age 1 was estimated as a separate parameter in all models that started at age 1):

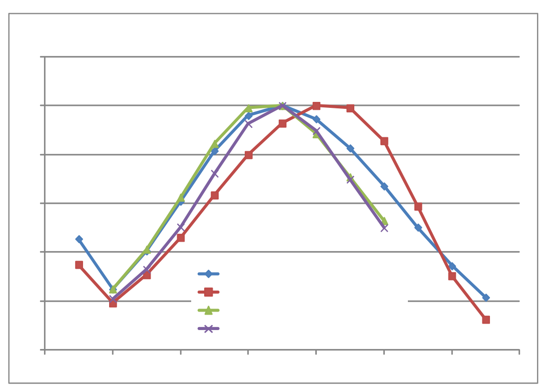




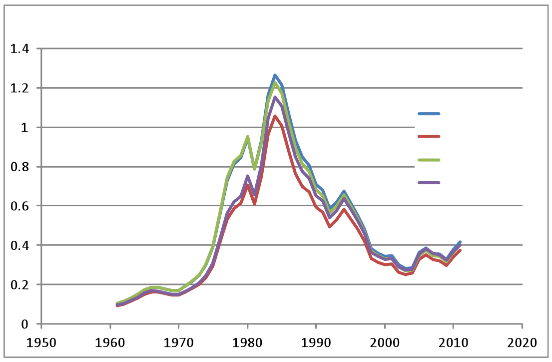
Second bullet point: Using mean-unbiased likelihoods for the survey indices only changed noticeably the SSB estimates before the mid-1980s, lowering them (as would be expected for this type of correction). The fact that only the period before the mid-1980s was affected, could perhaps be related to the fact that only the historical abundance indices (pre-1984) have very large CVs in some years, while the difference between mean and median in a Log-Normal distribution increases as the CV increases.

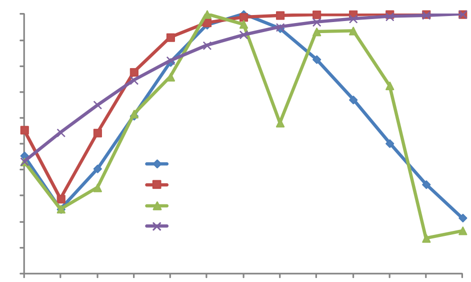
Third bullet point: The following two graphs illustrate the effect of estimating q for the NMFS bottom trawl survey versus fixing it at q=1. Two age ranges are considered (2-10+ and 1-13+). The time series of estimated SSB is shown on the left panel and the selectivity of the NMFS bottom trawl survey (modelled as double-logistic, age 1 separate) on the right panel. Estimating q results in a significant increase in the estimated SSB (as q is estimated to be less than 1) and in an increase of the selectivity of younger ages. When the modelled age range is 2-10+, the selectivities of ages 7 and older are not affected by whether q is estimated or fixed.





The impact of NMFS bottom trawl survey selectivity assumptions was explored under the age range 1-13+ and with q estimated, with some results shown in the following two graphs. Estimated SSB is on the left panel and the selectivity of the NMFS bottom trawl survey on the right panel. All ages in the selectivity function are treated as independent parameters (*i.e.* no functional form is imposed) in the "Saturated model". The shape of the selectivity estimated in the saturated model is overall in line with the double-logistic shape assumed in the current assessment. Asymptotic selectivity forms were also explored and, as expected, led to lower SSB.



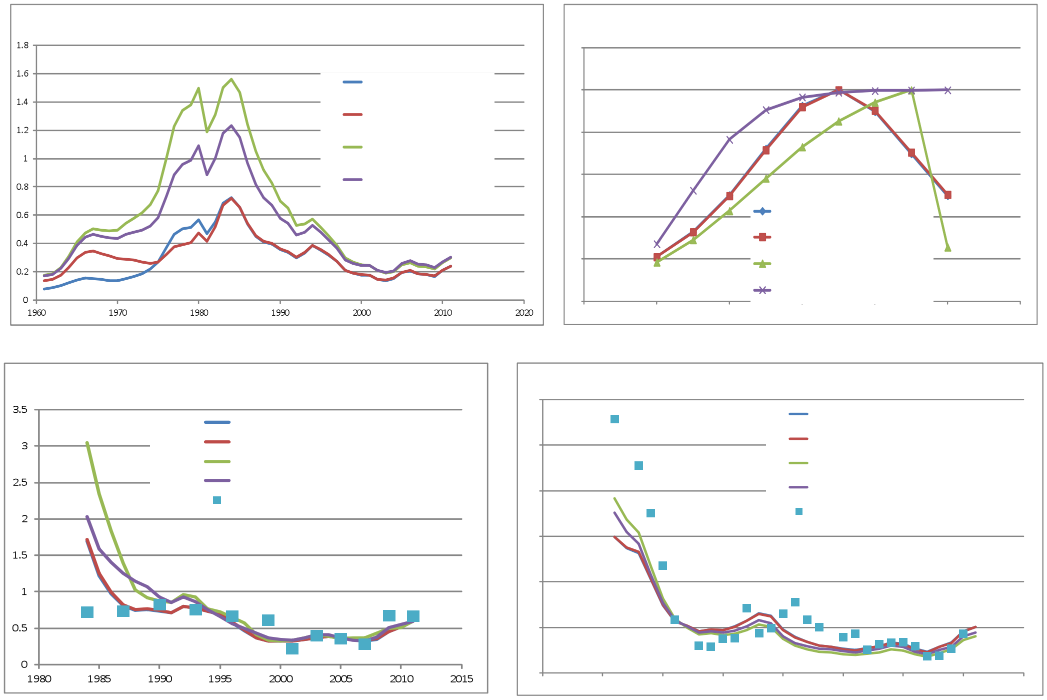


The additional explorations conducted during the meeting are labeled Request 1 to 5 below, as done during the meeting.

Request 1: Apply the following modifications to current assessment model:

- Drop historical (pre-1984) bottom trawl data set
- Drop 1984 and 1987 bottom trawl survey data points
- Estimate trawl survey catchability with asymptotic selectivity

These 3 modifications were implemented cumulatively (each new modification also incorporated the previous ones). Some results are shown in the 4 graphs below this paragraph. "Base" means the current assessment. The SSB graph shows a dramatic increase when the second modification (removal of 1984 and 1987 NMFS surveys) was introduced. The shape of the resulting selectivity function seems quite unrealistic in this case (very big and sudden drop at the oldest age). Assuming asymptotic selectivity for the NMFS survey, even if also allowing q to be estimated (so that q is most likely estimated to be less than 1 – I do not have the estimated q value), leads to lower SSB, as a consequence of the very different selectivity function estimated in that case, with higher selectivity for all ages. It is clear that the removal of the 1984 and 1987 NMFS survey values substantially affects the perceived stock dynamics. I imagine this is due to the discrepancy between those survey values (which essentially indicate biomass stability, bottom left panel) and what the acoustic survey indicates for that period (strong biomass decrease, bottom right panel). When both inputs are in the model, a compromise must be found between them (as in the current assessment), but when one of the signals is removed from the input data the other signal dominates and substantial changes occur in the assessment. Even in this situation, the model is unable to fit the very steep biomass drop indicated by the acoustic survey during the early 1980s (bottom right panel).



Request 2:

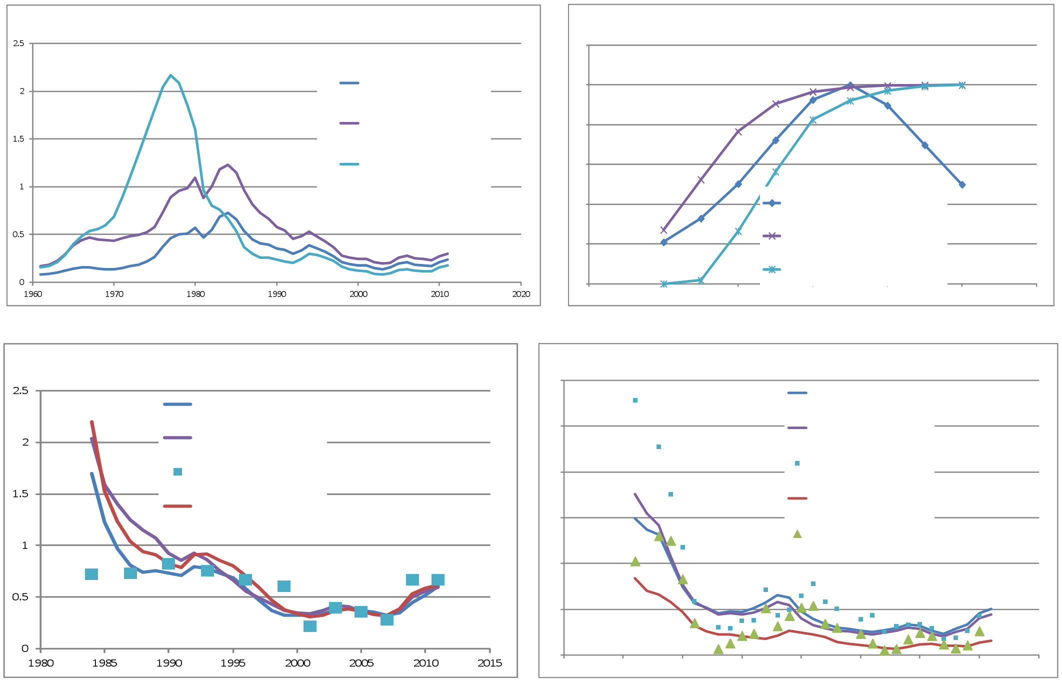
- Settings of final configuration in Request 1 + Remove all acoustic survey composition data and fit to a spawning biomass index computed from the acoustic survey (assuming selectivity equal to the proportion mature at age when fitting the index in the assessment model).

This request was aimed at addressing questions raised about the appropriateness of the acoustic survey index to represent the biomass of the entire population (see discussion under ToR 1). As it was not possible to work during the meeting through the raw acoustic survey data, the spawning biomass index was calculated in the following way, as an interim solution:

$\sum$ {acoustic biomass at length} * {maturity at length},

where the sum is for lengths at or above 43 cm.

Some results are shown in the following 4 graphs. “Base” and “Estimate bottom trawl q...” are the same ones shown in Request 1 (and model fits should be compared to the original observations, marked as blue dots, on the bottom right panel), whereas the new results are labeled as “Fit to SB” (and model fits should be compared to the new spawning biomass observations, marked as green triangles, on the bottom right panel). The NMFS bottom trawl survey selectivity (top right panel), changes substantially with the new model configuration, resulting in lower selectivities for all ages below 7. The resulting SSB estimates show an extreme increase followed by an extreme decrease during the 1970s. It is difficult to ascertain without additional careful checking of different diagnostics what may be the causing this effect on the SSB, it could presumably be related to the commercial catch information during the 1970s and early 1980s, but finding this out would require a detailed cross-check analysis of diagnostics across models to understand where the differences in model fits to all datasets used are. At this stage, the SSB result seems unrealistic. The bottom right panel also indicates that the new model configuration underestimates the acoustic survey spawning biomass indices in the early 1980s.



Request 3: no proper record was kept, hence not reported

Request 4: Apply the following modifications to the current assessment model:

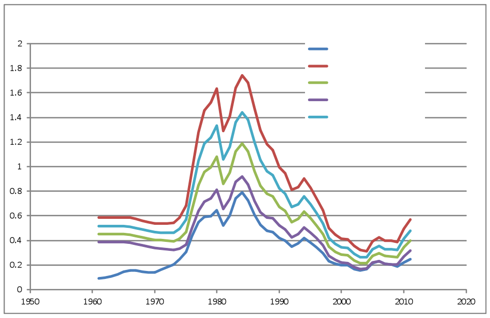
- Model ages 1-13+
- Commercial fishery: Block selectivity into "reasonable" time blocks, during which selectivity-at-age is assumed constant (breaks in years 1975, 1989, 2001)
- Bias-adjust for mean unbiased survey biomass
- NMFS bottom trawl survey: Estimate q; use double-logistic selectivity (with separate parameter for age 1); remove 1984 and 1987; remove all pre-1984 bottom trawl survey data
- Acoustic survey: Take out all acoustic composition data; fit to a spawning biomass index (assuming selectivity equal to the proportion mature at age).
- Include age-1 acoustic index (accidentally left out of the model, but included later, with no appreciable impact)

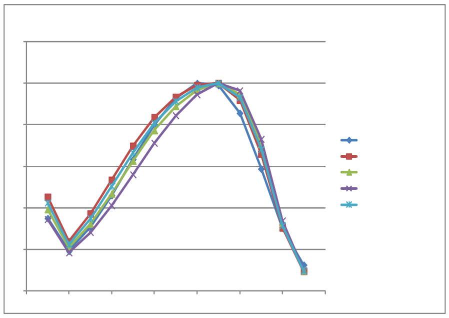
Additional constraints were imposed on recruitment variability as follows:

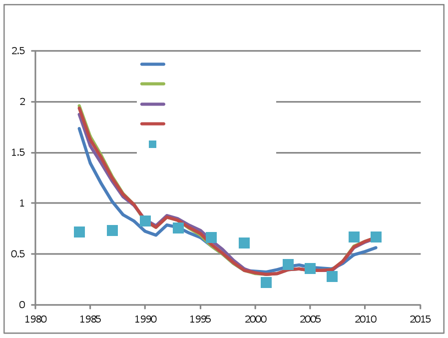
- Initial age composition (in 1961) was estimated using a single parameter, which was decremented by natural mortality to fill in the older ages
- SigmaR was assumed to be 0.1 for the years 1961-67 and 1.0 for the years 1968-75.

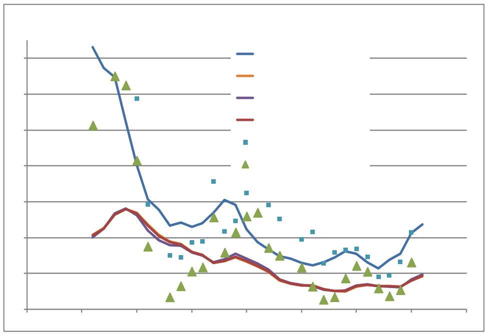
These assumptions enabled the model to estimate the NMFS bottom trawl survey q at a somewhat plausible value (0.61). Runs were also performed with fixed values of q (0.5, 0.75 and 1).

The following 4 graphs display some results from these runs. "Age 1to13 old" corresponds to the current assessment settings and data, but extending the modelled ages to 1-13+. All other runs correspond to the "Request 4" settings and differ between them only on whether the NMFS survey q is estimated or fixed. All runs under the new settings lead to higher SSB estimates than "Age 1to13 old", with SSB becoming larger the lower q. The estimated selectivity function of the NMFS bottom trawl survey is very similar in all cases (top right panel), including "Age 1to13 old". All new runs lead to very similar fits to the survey biomass indices and underestimate the acoustic survey spawning biomass index values in the early 1980s.









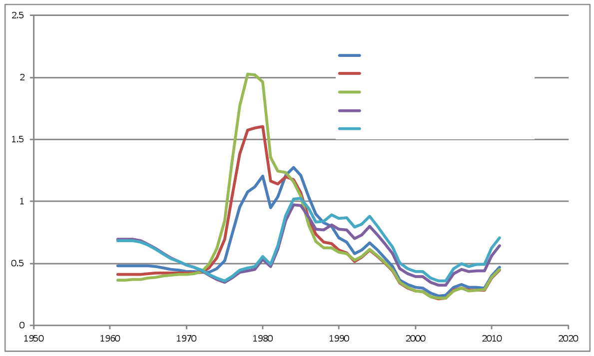
Considering a range of values of the mean(log_recruitment) model parameter around the MLE, the strongest changes in the profile log-likelihood as the mean(log_recruitment) value moved through the range were found for the commercial fishery age composition data and the spawning biomass index from the acoustic survey, and these changes were in opposite directions, suggesting a potential conflict of information between them. This was the reason why Request 5 (see below) was explored.

An additional run under the "Request 4" settings was conducted, re-introducing the acoustic survey age composition data and assuming a selectivity function for those data independent of the selectivity of the spawning biomass index from the same acoustic survey. This resulted in the same SSB trends as the runs under Request 4, but with yet another increase in estimated SSB, which reached a maximum of about 2.2 million t.

Request 5: Starting from the Request 4 settings, make the following modifications:

- Runs 1-2: Down weight commercial fishery age composition data, scaling the Multinomial sample size down by a constant multiplier, so that sample sizes are on average n= 50 or n=100 across years. The commercial fishery length composition data were also scaled down in a similar way. A quadratic prior on NMFS trawl survey q centered on 0.75 and with a CV of 0.3 was used.
- Runs 3-4: Same as previous two runs, additionally dropping the acoustic and egg production time series

SSB estimates before the mid 1980s are strongly changed with these new settings, as the following graph shows ("Working model" corresponds to the Request 4 settings).



The NMFS bottom trawl survey q did not seem to be strongly pulled one way or the other by changing input n for the fishery age composition, so a run removing the prior on q was also conducted. This model estimated q=0.73.

To conclude, after exploratory runs:

The explorations conducted during the Review meeting point to the robustness of the overall stock trends, while also indicating that the stock biomass in an absolute sense is sensitive to the input data used and assumptions made. Estimates of stock development before the mid 1980s are very sensitive to the data used and assumptions made, which is not surprising given that the data before the 1980s are quite sparse.

My recommendations at this stage are tentative. There was only a limited amount of time for analysis of results during the Review meeting and, in my view, further analyses of diagnostics should still be conducted. During the meeting only a small part of the runs' full sets of results could be examined and I feel more analysis is required before reaching definite conclusions.

- I recommend following on my points (1) and (2) (see details at the start of this ToR 2 section), about estimating q in the assessment and further exploration of constraints on selectivity at the older ages (also exploring selectivity constraints for the commercial fishery).

- I would suggest keeping the current age range 2-10+ in the assessment (particularly given the sparseness of data for older ages and my point (2) concerning selectivity at the older ages).

- I can not offer a clear recommendation for how to proceed concerning the historical trawl survey data (pre-1984). Whereas I appreciate the shortcomings these data have, removing them from the assessment also removes a significant amount of information for the 1960s and 1970s (assuming that some reliability is attached to this information). The model fit for the 1960s and 1970s may then become highly dependent on modelling assumptions (we have seen high variability in the assessment results for this period, depending on assumptions made, in the results of Requests 1-5), and there may not be a very clear basis for making some assumptions instead of others. My (tentative) suggestion would be to leave the historical trawl survey data in the assessment.

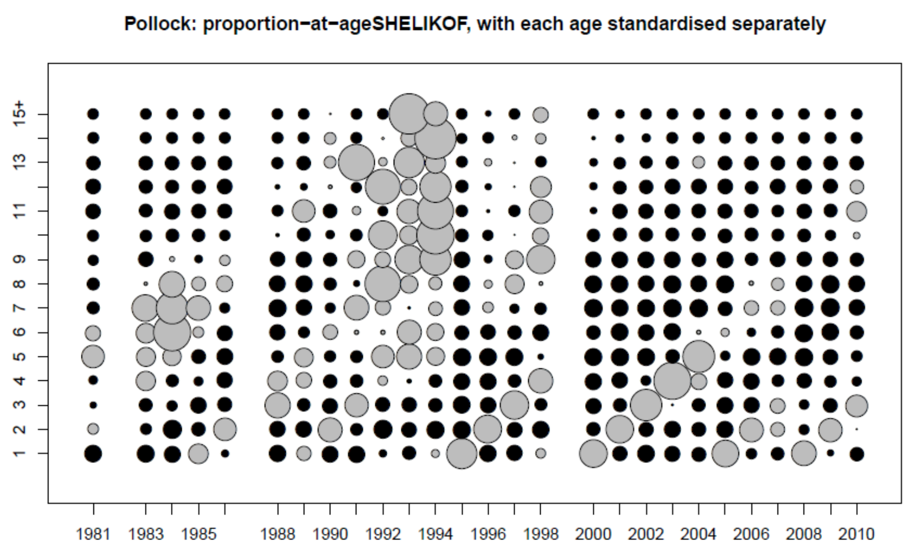
- I recommend that further exploration of the best way to use the acoustic survey in the assessment be done, following on the ideas suggested during the Review meeting.

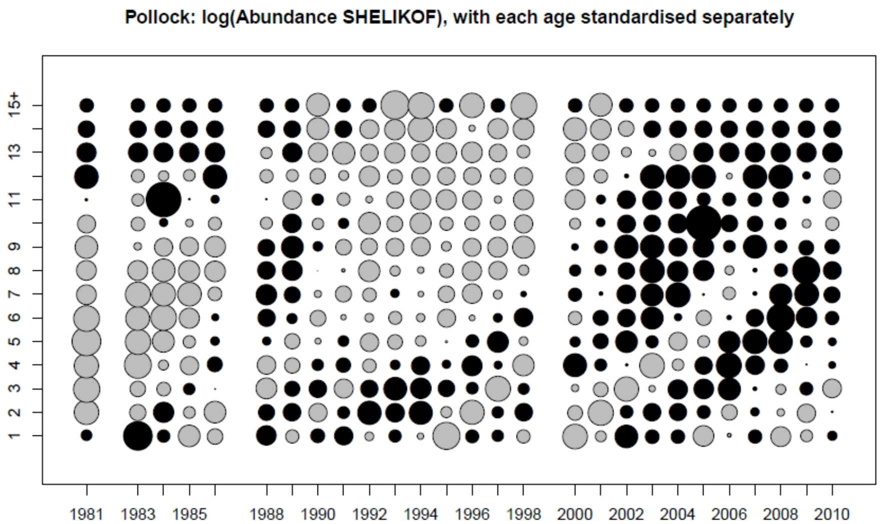
ToR 3. Evaluate and provide recommendations for the reporting of assessment results and characterization of uncertainty.

The assessment model is fitted by maximum likelihood and results are reported in a standard form (point estimates and associated CVs or confidence intervals). I do not find any particular problems with the way assessment results or model fit diagnostics are presented, but I offer some suggestions that I think would improve the presentation. All comments below refer to the Dorn *et al.* (2011) assessment document.

I would have found it useful to have seen a considerably more comprehensive exploratory analysis of input data and, in particular, appropriate plots to help quickly gain an intuition for the particular features of the various datasets going into the assessment. The different datasets that are used in the assessment are indicated in the big centered table in section "Analytic Approach, Model Structure" of Dorn *et al.* (2011). I would have found it useful to see appropriate graphs for each of those datasets, including the early ones. As we have noticed, stock development until the mid 1980s is uncertain, and it would have been useful to have been able to get a better feel for what the data for those early years show – only the historical (pre-1984) trawl survey biomass indices are shown (Figure 1.22). At least for main reviews, the exploratory analysis should be as complete as possible.

As part of the exploratory data analysis, I find it important to explore the extent to which different datasets are able to track cohorts through time. Bubble plots (similar to Figure 1.3), where ages and years are shown simultaneously in a single graph, are very useful for identifying potential patterns. I recommend doing this for all abundance indices used in the assessment, even if the indices are not available in all years. To be able to see properly the older ages in the bubble plots (which can be difficult, as there are few individuals of those ages when compared with the younger ages), I suggest making the bubble plots standardising each age separately, by subtracting the mean (for that age) over the time series and dividing by the standard deviation (for that age) over the time series. By way of illustration, I made the following two graphs to gain understanding of the values shown in Table 1.10 (grey and black bubbles represent values above and below average, respectively; the area of the bubble is proportional to the magnitude of the value):





I interpret the graphs as indicating that the Shelikof acoustic index is reasonably able of tracking strong and weak cohorts through time until approximately ages 7-8, then cohort signals appear to be lost (the proportion of fish at those older ages is very low for this survey, so it is not surprising that the cohort signal is lost for them). The graphs also indicate a predominance of older individuals during the 1990s, when compared with earlier and later decades.

Similar graphs could be displayed for catch-at-age from the commercial fishery (as well as catch curves) and for length frequency data. It is obviously also important to evaluate the consistency of the signals gleaned from different datasets, before incorporating them in the assessment (*e.g.* are the different datasets highlighting the same strong and weak year classes?).

This is, of course, all very basic and merely illustrative, and there are many ways in which exploratory analyses can be conducted. My main point is that this should be done comprehensibly, as it can provide very valuable insights about the features of the data used.

Model diagnostics can be displayed in similar ways. Residuals should be displayed for all datasets (not just the most recent datasets or years). This is particularly important when conducting an exploration of alternative model settings, as done during the Review meeting. Examination of residuals for all datasets, trying to understand where model fits diverge from assumptions made (*i.e.* patterns in residuals), provides very valuable insights about how the model is working and what is causing certain features in the assessment results. This can help in choosing among alternative models. I find bubble plots of residuals (displaying positive and negative residuals in different colours) very useful, because they can highlight in a single graph age, year and cohort effects. Residuals should be computed both in "raw" (observed – expected) and "standardised" (or "Pearson") (observed – expected)/(standard deviation according to the assumed observation equation and its estimated parameters) form, as the two forms indicate different things, with raw residuals showing actual deviations between data and model fits, and standardised residuals allowing to check the assumptions made about the variances (or CVs) of the observation equations for the different datasets. The standardised residuals should all have the same magnitude overall, irrespective of age, year or dataset. So if *e.g.* a dataset shows a smaller overall magnitude of standardised residuals than other datasets, this suggests that the variance (or CV) assumed for that dataset is too large and that consideration should be given to reducing it (this is similar to choosing an effective sample size for Multinomial composition data).

Retrospective plots (where, say, the final 1, 2, 3, 4 or 5 years of data are removed and assessment results compared) are often also useful. If displayed for survey catchability parameters q, or selectivities, they may indicate that systematic departures are happening, even if the assessment model assumes that those parameters are constant through time.

In terms of the displayed assessment model results, I wonder why what I would consider a "conventional" summary of annual F (the arithmetic average of F-at-age over a predefined range of ages), has not been used. It was unusual for me to see only graphs of SSB and Recruitment, without an annual summary of F. The top panel of Figure 1.28 is a measure of the impact of fishing pressure on the population. This figure represents the SPR percentage to which annual F estimates correspond, showing that F is estimated to have been smaller than $F_{40\%}$ in most years (given that the percentage SPR is generally estimated to have been above 40%). This way of representing (the impact of) annual fishing pressure ties in well with the harvest control rule used for this stock, which is based on $F_{40\%}$ or a linear decrease from it, depending on the value of current SSB with respect to $SSB_{47\%}$. I think the meaning of this graph (top panel of Figure 1.28) should be made clearer in the assessment document. A question, however, arises concerning the most appropriate way to compute SPR percentage-based reference points. I discuss this under ToR 4.

ToR 4. Evaluate and provide recommendations on F35% spawning biomass per recruit as an appropriate proxy for FMSY under non-stationarity in vital rates. Also evaluate and provide recommendations on the B35% biomass reference point as a proxy for BMSY.

This ToR was not discussed during the Review meeting and no documentation specific to it was provided for the review. The topic (assuming I have understood the question correctly) is clearly relevant, given the changes observed in the GOA ecosystem since the 1960s and in some biological characteristics of the pollock stock itself. In particular, an increasing weight-at-age has been observed in the last decade in the stock sampled during the acoustic survey (Figure 1.16 of Dorn *et al.* (2011)), which are the weights used in the calculation of the estimated SSB. This raises the question of which are the most appropriate weight-at-age values to use in the calculation of SPR percentage-based reference points ($F_{40\%}$ is central to the harvest control rule). A related question concerns the calculation of reference SSB percentage reference points, which are defined under the current management system as the product of the SPR percentage reference value and average recruitment. Appropriate calculation of SSB percentage reference points is also central to the harvest control rule, in which $SSB_{47\%}$ directly intervenes.

Dorn *et al.* (2011) explains that the average recruitment since 1979 is used in the calculation of SSB percentage reference points, to take account of the regime shift that occurred in the GOA in the late 1970s, after which the pollock stock has exhibited lower productivity.

In addition, average weight-at-age values since 2006 are used in the calculation of the reference points, to reflect the current situation. It is unknown whether the observed increase in weight-at-age is a density-dependent response to low pollock abundance or caused by environmental factors.

As explained in Dorn *et al.* (2011), changes in weight-at-age have potential implications for stock status determination and harvest policy. For example, if lower-valued weights-at-age from an earlier time period were considered representative of an unfished stock, and the currently higher weights-at-age were attributed to a density-density response to low pollock abundance, the SSB percentage reference points would be calculated to be lower than currently done and, therefore, the estimated current SSB relative to reference points would be higher (resulting in higher catch when applying the harvest control rule).

Similar comments could be made about the choices of average recruitment and maturity-at-age, if these were showing changes over time. The issue was encountered in the context of recruitment changes through time by A'mar *et al.* (2009a,b), who evaluated the performance of the current management strategy (MS; assessment model and harvest control rule), as well as alternative harvest control rules (while keeping the same assessment model), using Management Strategy Evaluation (MSE) methods. In A'mar *et al.* (2009a) the consequences of recruitment changes driven by climate were explored, making recruitment (in the simulated "true" population) depend on climate variables and using IPCC datasets to forecast future recruitment. A'mar *et al.* (2009a) used a weighted average of the most recent 25 recruitment values (corresponding to the so-called "dynamic B0") to define "true" SSB percentage reference points each year. Therefore, "true" biomass reference points changed over time in response to changes in recruitment caused by climatic factors. The current MS (using the current assessment and estimated average recruitment since 1979 in the calculation of biomass reference points) was evaluated in the MSE. For the first of the IPCC datasets, the current MS led to future "(true SSB)/(true $SSB_{40\%}$)" values above 1 and increasing over time, even though the "true SSB" was actually decreasing and well below "$SSB_{40\%}$ as calculated in 2006". This clearly illustrated the difficulties and dilemmas one faces when having to set reference points in the context of a changing environment. A'mar *et al.* (2009a) suggest in their discussion that it could be useful also to consider MSs that require stock SSB to be kept above a certain threshold in an absolute rather than a relative sense. Accepting the "dynamic B0" approach for the definition of "true" biomass reference points, the current MS kept SSB above "true" biomass reference points, even if the performance of the MS was far from optimal on other counts. In the context of regime shifts leading to jumps in overall recruitment levels at certain time points, and using the "dynamic B0" approach for recruitment in the calculation of "true" biomass reference points, A'mar *et al.* (2009b) found that the current MS performs at least as well as alternative MSs where the recruitment value used to estimate biomass reference points tried to account for changes in recruitment over time. They indicate, however, that this may not be a general result for all species, as it is probably related to the high variability that pollock recruitment exhibits even within a specific recruitment regime.

The topic appeared in yet another form in Dorn (2004), this time in the context of appropriate computation of the same SPR percentage-based reference points when natural mortality-at-age is changing over time. This is very relevant for pollock, given the very high predation mortality (believed to be well above fishing mortality) and the trends observed in main pollock predators in the last decades. SPR percentage-based reference points could be calculated either based on an average M-at-age through a (long) period of years or on current values of M-at-age. As Dorn (2004) says, the two alternatives represent contrasting philosophies about how fishing mortality should be adjusted in response to changes in natural mortality. Using current values of M for the reference point calculations essentially attempts to match fishing mortality to natural mortality, while using an average value of M over time to compute reference points attempts to adjust fishing mortality so as to compensate for changes in natural mortality.

A'mar *et al.* (2010) evaluate the current MS under time-varying natural mortality due to predation. The simulated "true" pollock population is subject to predation mortality by arrowtooth flounder, Pacific halibut and Pacific cod. The predation mortality rates (modelled as part of the total natural mortality of pollock) are assumed to be proportional to predator biomass, with the proportionality factor corresponding to Holling's Type I, II or III functional responses. The current MS (with constant natural mortality) is tested under scenarios combining each of these three functional responses with different future values of fishing mortality on the predators. A'mar *et al.* (2010) do not explain how "true" SSB percentage reference points were calculated (given the varying natural mortality in the "true" pollock population – see discussion in previous paragraph). The results in the paper concerning the fit of the operating models to the historical data, and how they compare to the results obtained by fitting the current assessment model, are interesting in their own right. Additionally, their MSE results show that the current MS is biased when the operating model incorporates predation mortality, but the bias is in a precautionary direction: the allowable catch is strongly underestimated and "true" SSB remains above "true $SSB_{40\%}$" with very high probability.

I do not have specific recommendations in connection with this ToR. AFSC scientists seem to have the issues of relevance for pollock clearly identified and they have already conducted very interesting work. MSE (as performed by A'mar *et al.*) to test the performance of the current MS or alternatives, when the "true" populations show some of the non-stationary features expected to affect pollock, seems a very useful tool. The MSE work could be expanded to test further the performance of different ways of defining F_{MSY} and B_{MSY} proxies in the context of the changing GOA ecosystem. I also suggest that more realism be incorporated in the current MSE work (or in future extensions of it), by considering an operating model that does not resemble so closely the assumptions of the current pollock assessment model.

ToR 5. Recommendations for further improvements.

My suggestions for aspects that could be further investigated have already been presented in the discussion of each of the 4 previous ToRs. I only add a few thoughts here.

In terms of the current assessment, I suggest that consideration be given to moving it to a "standard" model package, such as Stock Synthesis or CASAL. As far as I can see, the features of this assessment (perhaps with a few minor modifications) could be handled without problems by these packages. Using a model package would minimise the chance of coding errors (which can obviously happen, even with experienced programmers), would mean that a suite of model diagnostics was most likely readily available, and could facilitate communication about the assessment with other stock assessors (assuming they are familiar with those packages).

Environmental and ecosystem effects (species interactions, in particular, predation) are highly relevant for the pollock stock. Predation mortality is believed to be much higher than fishing mortality. Therefore, it is clear that these elements should not be forgotten when conducting the pollock assessment and providing management advice. The problem is how to deal with them formally in the pollock assessment, and this is not straightforward. The following incremental steps seem logical:

1. Start by being aware of the issues and dealing with them in a "soft" qualitative manner (AFSC scientists are doing this already).

2. Management Strategy Evaluation: simulation-test the performance of the current assessment model and harvest control rule under scenarios in which the "true" population is subject to environmental and predation processes (very useful work has been conducted already, and I suggest that it continues to be developed).

3. If considered sufficiently important and, when ready (after enough development and testing has been done), switch to an assessment model incorporating environmental and ecosystem processes.

AFSC scientists have already done a very substantial amount of work on these aspects [MSE by A'mar *et al.*, as well as the work on understanding and modelling ecosystem processes by Gaichas *et al.* (2008, 2010, 2011)]. Attempts to formally include some of these elements in stock assessment models can be seen in Hollowed *et al.* (2000), Dorn (2004) and A'mar *et al.* (2010). I encourage them to continue these lines of work, so that an assessment model incorporating these effects can be developed and tried in phases and, if/when considered appropriate, might become the main assessment model for pollock. I do not claim this is easy, but AFSC clearly has a team of scientists with expertise to progress on this.

Similar issues arise, of course, in most other parts of the world. The Report of the Working Group on Multispecies Assessment Methods (ICES, 2011) discusses many similar issues, including (briefly) ecosystem work on GOA pollock. The report also contains a new key run for the North Sea using the Stochastic Multispecies model (SMS). Natural mortality trends obtained from this run were subsequently used to set time-varying values of natural mortality-at-age in some of the standard single-species stock assessments (see, in particular, the natural mortality and MSY reference points discussions in the North Sea herring stock benchmark assessment, ICES 2012a). Multispecies stock assessment work for the Baltic can be found in the Report of the Workshop on Integrated/Multispecies Advice for Baltic Fisheries (ICES 2012b), with particular focus on the SMS model for joint assessment of cod, herring and sprat.

ToR 6. Brief description on panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

Panel review procedures have been explained in the "Description of review activities" section, at the start of this report, and summary of discussions held during the meeting were provided when covering each of the previous ToRs.

The review process was well organised and ran smoothly. Having a physical meeting (at AFSC) helped very much in gaining understanding of the large body of material to be reviewed and many interesting discussions took place with many aspects clarified during the meeting. Focusing the review on a single topic (the pollock assessment) allowed the reviewers to concentrate on it and to provide a (hopefully) reasonably in-depth review.

Conclusions and recommendations (in accordance with each ToR):

My conclusions, suggestions and recommendations were incorporated in the detailed discussions provided above for each of the ToRs. Therefore, in this section I only highlight main aspects in bullet point form.

ToR 1:

- Consider the technical comments I made for future Bayesian analyses. I recommend that proper priors be used (unless a way can be found to check that the propriety of the posterior, when using an improper prior – this is case-specific, no general result exists), and a sensitivity analysis conducted to explore whether the posterior quantities of interest are unduly impacted by the prior choices.

- Still in the context of Bayesian analyses, I recommend summarising results using posterior quantiles (which always exist if the posterior distribution is proper) instead of posterior moments (which may not exist, even with a proper posterior distribution).

- It could be interesting to redo the Bayesian analysis of length-selectivity mid-water trawls (Williams *et al.* (2011)) with a prior density proportional to $1/\mu_i$, instead of the current uniform prior, and with proper priors on all other model parameters; then check whether the current results are affected by this.

- The mid-water trawl length-selectivity results suggest that trawl selectivity may be highly variable and that further analyses and development would be required before attempting to correct the acoustic-trawl survey series currently used in the assessment.

- I encourage the acoustic survey team further to consider the point raised during the Review meeting in connection with the use of the survey to provide a biomass index for the whole pollock population (as opposed to just the spawning population). As discussed during the meeting (see details under ToR 1, earlier in this report), the main questions relate to whether the current trawl sampling scheme is able to produce reliable LFDs for the whole pollock population. Would it be possible to obtain an index of just the spawning biomass, along the lines suggested during the Review meeting? The impact of making this change in the pollock assessment could be tested in exploratory assessment runs.

- I note that plots of the current LFDs and numbers-at-age from the acoustic survey do not seem too bad at tracking cohorts through time, especially up to about age 7 (see plots under ToR 3, earlier in this report), which to me suggests that the problems with the way the acoustic survey is currently used are possibly not too severe. Nevertheless, I suggest the issue in the previous bullet point is still considered.

- ADF&G bottom trawl survey: Is its spatial coverage sufficient to provide an abundance index for pollock? Is it useful to have it in the assessment? (it probably has very little impact on results). I suggest these issues are explored.

ToR 2:

The explorations conducted during the Review meeting point to the robustness of the overall stock trends, while also indicating that the stock biomass in an absolute sense is sensitive to the input data used and assumptions made. Estimates of stock development before the mid 1980s are very sensitive to the data used and assumptions made, which is not surprising given that the data before the 1980s are quite sparse.

My recommendations at this stage are tentative. There was only a limited amount of time for analysis of results during the Review meeting and, in my view, further analyses of diagnostics should still be conducted. During the meeting only a small part of the runs' full sets of results could be examined and I feel more analysis is required before reaching definite conclusions.

- I recommend following on my points (1) and (2) (see details at the start of the ToR 2 section, earlier in this report), about estimating q in the assessment and further exploration of constraints on selectivity at the older ages (also exploring selectivity constraints for the commercial fishery).

- I would suggest keeping the current age range 2-10+ in the assessment (particularly given the sparseness of data for older ages and my point (2) concerning selectivity at the older ages).

- I can not offer a clear recommendation for how to proceed concerning the historical trawl survey data (pre-1984). Whereas I appreciate the shortcomings these data have, removing them from the assessment also removes a significant amount of information for the 1960s and 1970s (assuming that some reliability is attached to this information). The model fit for the 1960s and 1970s may then become highly dependent on modelling assumptions (we have seen high variability in the assessment results for this period, depending on assumptions made, in the results of Requests 1-5) and there may not be a very clear basis for making some assumptions instead of others. My (tentative) suggestion would be to leave the historical trawl survey data in the assessment.

- I recommend that further exploration of the best way to use the acoustic survey in the assessment be done, following on the ideas suggested during the Review meeting.

ToR 3:

The first two bullet points are recommendations and the third one is a suggestion.

- Conduct and present a comprehensive exploratory analysis of all datasets considered for inclusion in the assessment (my discussion on ToR 3 section, earlier in this report, provides some suggestions).

- Present a comprehensive exploration of model diagnostics.

- Display an appropriate summary of annual F as part of the assessment results.

ToR 4:

I do not have specific recommendations in connection with this ToR. AFSC scientists seem to have the issues of relevance for pollock clearly identified and they have already conducted very interesting work.

- MSE (as performed by A'mar *et al.*) to test the performance of the current MS or alternatives, when the "true" populations show some of the non-stationary features expected to affect pollock, seems a very useful tool.

- The MSE work could be expanded to test further the performance of different ways of defining F_{MSY} and B_{MSY} proxies in the context of the changing GOA ecosystem.

- I suggest that more realism be incorporated in the current MSE work (or future extensions of it), by considering an operating model that does not resemble so closely the assumptions of the current pollock assessment model.

ToR 5:

- I suggest that consideration is given to moving the current assessment to a flexible model package, such as Stock Synthesis or CASAL.

- I encourage AFSC scientists to continue their work on developing models and testing (MSE) for situations involving environmental and ecosystem processes (mainly predation on pollock). An assessment model incorporating these effects can be developed and tried in phases and, if/when considered appropriate, might become the main assessment model for pollock. I do not claim this is easy, but AFSC clearly has a team of scientists with expertise to progress on this.

ToR 6:

The review process was well organised and ran smoothly. Having a physical meeting (at AFSC) helped very much in gaining understanding of the large body of material to be reviewed and many interesting discussions took place with many aspects clarified during the meeting. Focusing the review on a single topic (the pollock assessment) allowed the reviewers to concentrate on it and to provide a (hopefully) reasonably in-depth review.

Appendix 1: Bibliography of materials provided for review

(*) A'mar, Z. T., Punt, A. E. and Dorn, M. W. (2008) The Management Strategy Evaluation Approach and the Fishery for Walleye Pollock in the Gulf of Alaska. In *Resiliency of Gadid Stocks to Fishing and Climate Change. Alaska Sea Grant College Program AK-SG-08-01, 2008.*

(*) A'mar, Z. T., Punt, A. E. and Dorn, M. W. (2009a) The evaluation of two management strategies for the Gulf of Alaska walleye pollock fishery under climate change. ICES Journal of Marine Science 66, 1614–1632.

(*) A'mar, Z. T., Punt, A. E. and Dorn, M. W. (2009b) The impact of regime shifts on the performance of management strategies for the Gulf of Alaska walleye pollock (*Theragra chalcogramma*) fishery. Can. J. Fish. Aquat. Sci. 66, 2222-2242.

(*) A'mar, Z. T., Punt, A. E. and Dorn, M. W. (2010) Incorporating ecosystem forcing through predation into a management strategy evaluation for the Gulf of Alaska walleye pollock (*Theragra chalcogramma*) fishery. Fisheries Research 102, 98–114.

De Robertis, A. and Higginbottom, I. (2007) A post-processing technique to estimate the signal-to-noise ratio and remove echosounder background noise. ICES Journal of Marine Science 64, 1282–1291.

De Robertis, A., Hjellvik, V., Williamson, N. J. and Wilson, C. D. (2008) Silent ships do not always encounter more fish: comparison of acoustic backscatter recorded by a noise-reduced and a conventional research vessel. ICES Journal of Marine Science 65, 623–635.

De Robertis, A., McKelvey, D.R., and Ressler, P.H. (2011) Development and application of an empirical multifrequency method for backscatter classification. Can. J. Fish. Aquat. Sci. 67, 1459–1474.

De Robertis, A. and Williams, K. (2008) Weight–Length Relationships in Fisheries Studies: The Standard Allometric Model Should Be Applied with Caution. Transactions of the American Fisheries Society 137, 707–719.

De Robertis, A. and Wilson, C. D. (2006) Walleye pollock respond to trawling vessels. ICES Journal of Marine Science 63, 514-522.

De Robertis, A. and Wilson, C. D. (2010) Silent ships sometimes do encounter more fish. 2. Concurrent echosounder observations from a freedrifting buoy and vessels. – ICES Journal of Marine Science 67, 996–1003.

De Robertis, A. and Wilson, C. D. (2011a) Silent ships do not always encounter more fish (revisited): comparison of acoustic backscatter from walleye pollock recorded by a noise-reduced and a conventional research vessel in the eastern Bering Sea. ICES Journal of Marine Science 68, 2229–2239.

De Robertis, A. and Wilson, C. (2011b) Underwater radiated noise measurements of a noise-reduced research vessel: comparison between a US Navy noise range and a simple hydrophone mooring. Proceedings of Meetings on Acoustics, Vol. 12, 070003 (2011). Acoustical Society of America [DOI: 10.1121/1.3660550].

De Robertis, A., Wilson, C. D., Williamson, N. J., Guttormsen, M. A. and Stienessen, S. (2010) Silent ships sometimes do encounter more fish. 1. Vessel comparisons during winter pollock surveys. ICES Journal of Marine Science 67, 985–995.

(*) Dorn, M. (2004) Extending separable age-structured assessment models to evaluate trends in juvenile mortality of walleye pollock in the Gulf of Alaska. ICES CM 2004 / FF:31.

(*) Dorn, M., Aydin, K., Barbeaux, S., Guttormsen, M., Spalinger, K. and Palsson, W. (2011) Chapter 1: Assessment of the Walleye Pollock Stock in the Gulf of Alaska. November 2011, Council Draft.

(*) Gaichas, S.K. and Francis, R.C. (2008) Network models for ecosystem-based fishery analysis: a review of concepts and application to the Gulf of Alaska marine web. Can. J. Fish. Aquat. Sci. 65, 1965–1982.

(*) Gaichas, S.K., Aydin, K.Y. and Francis, R.C. (2010) Using food web model results to inform stock assessment estimates of mortality and production for ecosystem-based fisheries management. Can. J. Fish. Aquat. Sci. 67, 1490–1506.

(*) Gaichas, S.K., Aydin, K.Y. and Francis, R.C. (2011) What drives dynamics in the Gulf of Alaska? Integrating hypotheses of species, fishing, and climate relationships using ecosystem modeling. Can. J. Fish. Aquat. Sci. 68, 1553–1578.

(*) Guttormsen, M. A. and Jones, D. (2010) Results of the acoustic-trawl surveys of walleye pollock (*Theragra chalcogramma*) in the Gulf of Alaska, February-March 2010 (DY2010-01 and DY2010-02). AFSC Processed Rep. 2010-05, 85 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.

Guttormsen, M. A. and Wilson, C. D. (2009) In situ measurements of capelin (*Mallotus villosus*) target strength in the North Pacific Ocean. ICES Journal of Marine Science 66, 258–263.

(*) Guttormsen, M. A. and Yasenak, P.T. (2007) Results of the 2003 and 2005 echo integration-trawl surveys in the Gulf of Alaska during summer, Cruises MF2003-09 and OD2005-01. AFSC Processed Rep. 2007-04, 61 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar, Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.

Hjellvik, V. and De Robertis, A. (2007) Vessel comparison on the seabed echo: Influence of vessel attitude. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-171, 34 p.

Hollowed, A. B., Ianelli, J. N. and Livingston, P. A. (2000) Including predation mortality in stock assessments: a case study for Gulf of Alaska walleye pollock. ICES Journal of Marine Science 57, 279–293.

Honkalehto, T., Ressler, P.H., Towler, R.H. and Wilson, C.D. (2011) Using acoustic data from fishing vessels to estimate walleye pollock (*Theragra chalcogramma*) abundance in the eastern Bering Sea. Can. J. Fish. Aquat. Sci. 68, 1231–1242.

ICES (2011). Report of the Working Group on Multispecies Assessment Methods (WGSAM), 10–14 October 2011, Woods Hole, USA. ICES CM 2011/SSGSUE:10. 229 pp. Available at http://www.ices.dk/reports/SSGSUE/2011/WGSAM2011.pdf

ICES (2012a). Report of the Benchmark Workshop on Pelagic Stocks (WKPELA 2012), 13–17 February 2012, Copenhagen, Denmark. ICES CM 2012/ACOM:47. 572 pp. Available at http://www.ices.dk/reports/ACOM/2012/WKPELA%202012/wkpela_2012.pdf

ICES (2012b). Report of the Workshop on Integrated/Multispecies Advice for Baltic Fisheries (WKMULTBAL), 6–8 March 2012, Charlottenlund, Denmark. ICES CM 2012/ACOM:43. 137 pp. Available at http://www.ices.dk/reports/ACOM/2012/WKMULTBAL/wkmultbal_2012.pdf

Mann, D.A., Wilson, C.D., Song, J. and Popper, A.N. (2009) Hearing Sensitivity of the Walleye Pollock. Transactions of the American Fisheries Society 138, 1000–1008.

McKelvey, D.R. and Wilson, C.D. (2006) Discriminant Classification of Fish and Zooplankton Backscattering at 38 and 120 kHz. Transactions of the American Fisheries Society 135, 488–499.

Munro, P.T. and Hoff, R.Z. (1995) Two Demersal Trawl Surveys in the Gulf of Alaska: Implications of Survey Design and Methods. US. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-50, 139 p.

Ressler, P.H., De Robertis, A., Warren, J.D., Smith, J.N. and Kotwicki, S. (2012) Developing an acoustic survey of euphausiids to understand trophic interactions in the Bering Sea ecosystem. Deep-Sea Research II 65-70, 184–195.

Somerton, D. A., Williams, K., von Szalay, P. G. and Rose, C. S. (2011) Using acoustics to estimate the fish-length selectivity of trawl mesh. ICES Journal of Marine Science 68, 1558–1565.

Spalinger, K. (2008) Special project plan: 2008 Bottom trawl survey of crab and groundfish: Kodiak, Chignik, South Peninsula, and Eastern Aleutian Districts. Alaska Department of Fish and Game, Regional Information Report 4K08-09, Kodiak.

Stienessen, S.C. and Wilson, C.D. (2008) Juvenile Walleye Pollock Aggregation Structure in the Gulf of Alaska. In *Resiliency of Gadid Stocks to Fishing and Climate Change. Alaska Sea Grant College Program AK-SG-08-01, 2008.*

Towler, R. and Williams, K. (2010) An inexpensive millimeter-accuracy electronic length measuring board. Fisheries Research 106, 107–111.

(*) von Szalay, P. G., N. W. Raring, F. R. Shaw, M. E. Wilkins and Martin, M.H. (2010) Data Report: 2009 Gulf of Alaska bottom trawl survey. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-208, 245 p.

Walline, P. D. (2007) Geostatistical simulations of eastern Bering Sea walleye pollock spatial distributions, to estimate sampling precision. ICES Journal of Marine Science 64, 559–569.

Walline, P.D., Wilson, C.D., Hollowed, A.B. and Stienessen, S.C. (2012) Short-term effects of commercial fishing on the distribution and abundance of walleye pollock (*Theragra chalcogramma*). Can. J. Fish. Aquat. Sci. 69, 354–368.

Williams, K. (2007) Evaluation of the Macroscopic Staging Method for Determining Maturity of Female Walleye Pollock *Theragra chalcogramma* in Shelikof Strait, Alaska. Alaska Fishery Research Bulletin 12, 252–263.

(*) Williams, K. (2012) Chapter 3. Effect of trawl selectivity on the acoustic survey time series and assessment of Gulf of Alaska walleye pollock (*Theragra chacogramma*). Draft.

(*) Williams, K., Punt, A. E., Wilson, C. D. and Horne, J. K. (2011) Length-selective retention of walleye pollock, *Theragra chalcogramma*, by midwater trawls. ICES Journal of Marine Science 68, 119–129.

Williams, K., Rooper, C.N. and Towler, R. (2010) Use of stereo camera systems for assessment of rockfish abundance in untrawlable areas and for recording pollock behavior during midwater trawls. Fish. Bull. 108, 352–362.

Appendix 2: Copy of CIE statement of work

Attachment A: Statement of Work for Dr. Carmen Fernandez

External Independent Peer Review by the Center for Independent Experts

Gulf of Alaska (GoA) walleye pollock stock Assessment Review

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

Project Description The Alaska Fisheries Science Center (AFSC) requests a Center of Independent Experts (CIE) review of the stock assessment for Gulf of Alaska (GOA) walleye pollock. The walleye pollock stock in the Gulf of Alaska is important to local fishing communities and is a key component of the GOA ecosystem. Walleye pollock stock assessments routinely undergo review by the AFSC, the North Pacific Fisheries Management Council's Groundfish Plan Team and Scientific and Statistical Committee. The assessment model for pollock has been stable for some time, and several significant changes are being contemplated for the 2012 assessment. In addition, the pollock stock assessment has not had the benefit of a CIE review since 2003. Therefore, a CIE review in 2012 would be timely.
The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers must be thoroughly familiar with various subject areas involved in stock assessment, including population dynamics, survey methodology, and estimation of parameters in complex nonlinear models. Reviewers must also have experience conducting stock assessments for fisheries management. Expertise would be desirable in several other areas. First, since the pollock assessment uses AD Model Builder (ADMB) software, expertise in using this software would be desirable. Second, changes being considered for the 2012 assessment include adding ecological interactions and environmental forcing to the assessment model, so expertise in these areas would also be desirable. It is not expected that all three of the reviewers have these specialized areas of expertise, rather that at least one of the three reviewers should be knowledgeable in these areas. Each CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review during the panel review meeting scheduled in Seattle, Washington with dates July 17-20, 2012.

Statement of Tasks: Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Foreign National Security Clearance: When CIE reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewers who are non-US citizens. For this reason, the CIE reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/
http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting at the Seattle, Washington during July 2012 (dates to be determined by Project Contact no later than 15 April 2012).
3) In Seattle, Washington during 17-20 July 2012 as specified herein, and conduct an independent peer review in accordance with the ToRs (**Annex 2**).
4) No later than **August 3, 2012,** each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and CIE Regional Coordinator, via email to Dr. David Die ddie@rsmas.miami.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

18 June 2012	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
3 July 2012	NMFS Project Contact sends the CIE Reviewers the pre-review documents
17-20 July 2012	Each reviewer participates and conducts an independent peer review during the panel review meeting
3 August 2012	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
17 August 2012	CIE submits CIE independent peer review reports to the COTR
24 August 2012	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on changes. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) The CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) The CIE report shall address each ToR as specified in **Annex 2**,
(3) The CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

William Michaels, Program Manager, COTR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov   Phone: 301-427-8155

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net   Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com   Phone: 571-223-7717

Key Personnel:

NMFS Project Contact:

Martin Dorn
NOAA National Marine Fisheries Service
Alaska Fisheries Science Center
7600 Sand Point Way, NE, Seattle, WA 98115-6349
Email: martin.dorn@noaa.gov   Phone: 206-526-6548

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

   a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.

   c. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   d. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work
   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for Peer Review of the

Gulf of Alaska Walleye Pollock Stock Assessment

1. Evaluate and provide recommendations on data collection procedures and analytical methods used to develop assessment model input.
2. Evaluate and provide recommendations on model structure, assumptions, and estimation procedures.
3. Evaluate and provide recommendations for the reporting of assessment results and characterization of uncertainty.
4. Evaluate and provide recommendations on F35% spawning biomass per recruit as an appropriate proxy for FMSY under non-stationarity in vital rates. Also evaluate and provide recommendations on the B35% biomass reference point as a proxy for BMSY.
5. Recommendations for further improvements.
6. Brief description on panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

Annex 3: Tentative Agenda

Review of the Gulf of Alaska Walleye Pollock Stock Assessment
Alaska Fisheries Science Center, NOAA
7600 Sand Point Way N.E., Building 4
Seattle, Washington 98115
Phone: 206 526-4000
17-20 July 2012

The final meeting agenda has not yet been drafted, but will be forwarded by the project contact as soon as it becomes available.

17 July 2012	Presentations by survey and fishery data collection scientists
18 July 2012	Presentation by assessment scientists, Panel discussion and requests
19 July 2012	Panel discussion and requests, Begin drafting reviewer reports
20 July 2012	Draft reviewer reports

Appendix 3: Panel membership or other pertinent information from the panel review meeting

Panel members (alphabetical):

- Patrick Cordue, New Zealand
- Carmen Fernández, Spain
- Ian Jonsen, Canada