Review Report on the 2014 Stock Assessment of the Chilean Sea Bass

(Patagonian toothfish, *Dissostichus eleginoides*)

Tom Polacheck
Running Tide
PO Box 184
Woodbridge, Tasmania
Australia 7162

e-mail: runningtide.tom@gmail.com

January, 2015

Table of Contents

Executive Summary

The University of Concepción was contracted by the Undersecretariat of Fisheries of Chile to undertake a review of the stock assessment of the sea bass or toothfish which had been performed by the Instituto de Fomento Pesquero (IFOP) in September 2014. The review process was coordinated by Dr Billy Ernst. The author was requested to participate in the review and contracted to provide an independent review report. The review consisted of a one week workshop from November 11-15, 2016 along with the provision of a large quantity of background documents and material. The current report summarizes the reviewer's independent review of this stock assessment and addresses the terms of reference specified for the reviewer's report.

The September 2014 assessment included a number of substantial and significant additions and innovations in its development. These included joint modelling of the Argentinean and Chilean catches, development of CPUE indices for the Argentinean longline fishery, the integration of the IFOP log book data with the scientific observers data belonging to the Centro de Estudios Pesqueros (CEPES S.A.) and incorporating a correction model for age readings obtained from scales (although this latter was mistakenly not utilized in the results presented in the assessment). The reviewer considers that all of these were important factors that need to be considered within the assessment for this resource. These combined with the large amount of effort expended required to accomplish them need to be recognized and acknowledged.

Patagonian toothfish is a deepwater demersal species found circumpolar in the shelf regions of the southern oceans. It is slow growing with a relatively late age of maturity (8-14 years of age) and lives in excess of 30 years. Numerous fisheries exist in Antarctic and around near Antarctic islands and on the shelf waters of Patagonia and Chile (extending into Southern Peru). Substantial uncertainty exists about components of its biology and life history which are important for assessing its status. In particular for the fisheries in Chile, the uncertainty about the stock structure, spatial dynamics and the relationship among fish caught in the areas where the Chilean artisanal and industrial fisheries operate and also among the areas where the Argentinean and Falkland Island fisheries occur is of critical importance. All the data and information are consistent in indicating that a single population exists across these areas and that this population is distinct from other toothfish populations. The assessment constitutes a major advance in this regard. It is the first assessment that utilizes data from more than a single area (i.e. catches from southern Chile and southern portion of the Argentinean fishery) and is based on a conceptual model that explicitly acknowledges and accounts for catches from all areas in terms of the consequences for the overall population dynamics and status. This conceptual model has two fundamental implications for management in terms of catches and the sustainability of the resource: (1) that there is no need to regulate catches outside of southern Chilean and Patagonia shelf except possibly in terms of yield per recruit considerations (this includes the Chilean Artisanal catches north of 47°S) and (2) that assessment advice in terms of overfishing and sustainability can only be provided on the consequence of the combined total removals by the Chilean and Argentinean fisheries. The allocation of catches among the Chilean and Argentinean fisheries is almost exclusively a management issue for which the stock assessment is basically uninformative (the former issue is not discussed in the Assessment Report).

Although the conceptual model underlying the current assessment accounts for the consequences of all historical catches from all areas, it is only one plausible model. Other (and in the reviewer's view more plausible models) exist. Other models would have substantially different implications for the assessment and management advice. In particular, the reviewer is concerned about the assumption that the substantial catches north of 47°S in the Pacific and 54°S in the Atlantic come from sink populations and do not need to be taken account of in assessing or managing the resource. The supporting evidence for this hypothesis is limited (i.e. comments about physiological constraints on toothfish ability to migrate). The hypothesis is not supported by the limited tagging data or evolutionary considerations. Other conceptual models need to be considered if the assessment results and management advice are to be considered robust and precautionary.

The results for the conceptual model developed and presented in the Assessment Report are limited to a single model run. The Assessment Report contains no consideration of model uncertainty. The issue of the selection of the base case is not adequately addressed. No sensitivities are provided to alternative structural hypotheses, data inputs or weightings. There appears to have been arbitrary selection and insufficient understanding of the input data and fixed parameter values as well as inadequate consideration of alternatives or evaluation of the consistency among the inputs. The reviewer considers these to be major deficiencies in the assessment. In addition, a substantial number of errors in the inputs, implementation and running of the model were detected. These latter problems combined with the lack of detail and errors in the documentation make it difficult to make definitive conclusions about the specific results and recommendations in the Assessment Report. They also undermine confidence in the reliability and accuracy of the actual numerical results. Overall, in the reviewer's view, the current assessment could not be considered as an adequate and robust basis for the determination of stock status and provision of management advice.

Stock assessments and management decisions for this resource need to be cognizant and take into account the catches from all areas and be based on consideration of a full range of plausible assumptions for the interactions among the fish and fisheries found in the various areas (i.e. the artisanal and industrial fisheries in Chile, the longline and trawl fisheries in Argentina including those north of 54S, and the fisheries around the Falkland Islands). At the stock assessment level, this is a significant challenge because of the spatially disjoint fisheries and incomplete mixing of fish among areas compounded by data access problems. It is also necessary to recognize that there are significant political constraints and issues in terms of how to deal with management advice which requires addressing allocation of catches among different areas and fisheries while there is little or no scientific basis for the provision of advice on the allocation of catches among areas. Nevertheless, there needs to be an understanding of the limitation of the advice science can provide (e.g. allocation among fisheries). Scientists should not be requested or attempt to provide advice outside of these limits.

In spite of the extensive data that exists from the fisheries for toothfish in the South American /Patagonian shelf area, it needs to be recognized that there are severe limitation in these data and the information available for assessing this resource. There is a need to improve the information base underlying the assessment as well as the procedures and approaches to the assessment. These are discussed in more detail within the report and include:

- Comprehensive review and development of the biological and fishery inputs available for the assessment;
- Substantial international cooperation in the development and implementation of stock assessments; in the collection. standardization and sharing of data; and in basic research;
- Institution of internal rigorous procedures to be used when developing, implementing, reviewing and documenting sock assessments and stock assessment models;
- Developing an integrated and comprehensive approach to stock assessments and research;
- Utilization of an operating model approach and management evaluation procedures
- Implementation of an on-going and comprehensive tagging program capable of providing fishery independent estimates of abundance and quantitative data on spatial movements and interactions. This requires that reporting rates are estimable.

Finally, the current stock assessment indicates that the stock is in a highly depleted state and that the magnitude of the stock is low relative to current catches (i.e. estimates of fishing mortality rates are very high). The trends in CPUE, which dominates the trend estimates in the stock assessment, clearly suggest that stock size has been substantially and more or less continuously reduced. As such, this is likely to be a conclusion (with varying degrees of severity) of any analytical stock assessment with the currently available data unless strong assumptions are invoked about hyper-depletion or temporal trends in marine mammal depredation effects (i.e. assumptions that would allow the model to disregard the trends in CPUE). However, the issues and problems discussed in more depth in the report mean that a large amount of uncertainty exists and will persist (at least in the short term) about the precise state of the stock, exploitation rates and specific numeric estimates. Without improvements to the information base and expansion of the scope of the assessment, substantial risk either to the stock and/or the fishery (in terms of foregone catches) will be embedded in any management advice and decision.

Background

The reviewer was invited and contracted to participate in the review process of the 2014 stock assessments of Chilean Sea Bass (Patagonian toothfish, *Dissostichus eleginoides*) and Nylon Shrimp (*Heterocarpus reedi*) stocks which had been completed by the Instituto de Fomento Pesquero (IFOP). The review was requested by The Undersecretariat of Fisheries of Chile. Its implementation was contracted to the University of Concepción and coordinated by Dr. Billy Ernst. The review consisted of a one week workshop in which the information used and the results of the stock assessments for these two species were presented and discussed along with the provision of background documents and material previous to the workshop. The reviewer was specifically contracted to provide an independent, standalone report summarizing his findings and conclusions with respect to Chilean Sea Bass stock. The current document constitutes this report. The content and format of the report follow the specification for this in the Terms of Reference provided for the Review (Appendix 2). It should be noted that the review was of the Chilean Sea Bass assessment completed by IFOP in August 2014 (Tascheri et al 2014). There has been further assessment work and reports since then.

It should be noted that the reviewer had no previous involvement in stock assessments for Chilean Sea Bass and no direct, first hand involvement in the fishery or management of this stock. The reviewer has had previous involvement with IFOP in terms of collaborative research and reviews – some of which involved some of the same personnel responsible for the stock assessment and data collection/analyzes for the Chilean Sea Bass. The reviewer does not consider any of this latter involvement to constitute a substantive conflict of interest or to compromise the independence of the findings and conclusions.

It should be further noted the reviewer had no more than a general knowledge of the literature and information available on this stock and fishery prior to undertaking this review. There is a rich and extensive scientific literature and data available on the biology and fisheries for this species world-wide and for the resource exploited by Chilean fisheries. In the time available for the review, it was not possible to undertake a separate and exhaustive review of this literature. The reviewer's findings and conclusions are based on the documents provided to him for the review (see Appendix 1), the presentations provided at the review workshop, discussions among the experts present at the workshop (both formal and informal) and a few additional published documents that the reviewer sourced independently. Simultaneous translation was available for all presentations and formal discussions during the workshop. The translation greatly facilitated the functioning of the workshop and the review process. The reviewer considers that this was indispensable for being able to complete his work.

Description of Role in the Review Activities

During the weeks period prior to the review workshop which was held during the week of November 10, 2014 extensive background material was provide on Patagonian toothfish. A number of peer reviewed background publications in the primary scientific literature were provided and translations in English of the 2014 stock Assessment Report (Tascheri et al. 2014)[1] and other key reports on the basic data and inputs used in the stock assessment of

[1] In the rest of this report this will be referred to as "the Assessment Report". The English translation of this document was provided ~2 weeks prior to the review workshop.

Patagonian toothfish (Appendix 1). Computer code for fitting the stock assessment model was also provided. Additional documents were provided during the workshop including a somewhat different version of the computer code. In the period prior to the workshop, the reviewer undertook to read and review all of the material which had been provided. In total, the amount of material available was voluminous. While all documents were examined, the limited timeframe required focusing on the material that the reviewer considered contained the key and most important information in the context of the Terms of Reference for the Review (Appendix 2).

During the workshop, the reviewer attended all sessions. The sessions relevant to toothfish were conducted primarily in the morning and those for Nylon shrimp in the afternoon. Participation in the workshop involved listening to all of the presentation (for both species) and taking a lead role in the discussion of the material provided on toothfish and active participation in the discussions on Nylon shrimp. This included seeking clarification and providing questions and feedback on the technical and more general aspects of the assessment and data collection process.

Written documentation was often incomplete and not clear. In addition, the numeric results presented from the stock assessment model were inaccurate in terms of the specification provided for it in the Assessment Report. There were also inconsistencies between the input data as described and specified in the Assessment Report and those actually used in fitting the model. These problems meant that a large amount of the review process and the reviewer's time in preparing this report was devoted to clarifying what was actually intended to be done and what was actually done. Consequently, in the time available, it was neither possible to fully clarifying some of the technical details in the Assessment Report nor to fully explore and/or clarify some of the more fundamental issues related to the assessment and the terms of reference.

As part of the workshop, additional runs or scenarios of the model for toothfish were requested by the reviewer and performed by Mr. Tascheri. The purpose of these was to provide improved understanding of the behaviour of the assessment model, its sensitivity to alternative assumptions, the uncertainty in the results and the appropriateness of some of the model assumptions and specifications. The results for these runs were not available until the last day of the workshop, which did not allow for detail examination of them at the time. After the workshop while preparing this report, the reviewer found problems in the results provided for these runs (see below). Additional runs were conducted by Dr Tascheri and provided to the reviewer on December 16[2]. All of the additional runs were an important component of the review process as they provided important insight into the assessment model, computational issues, the analyses and the conclusions. Results from these additional model run, as appropriate, have been incorporated into the current report. It must be stressed that the purpose of these alternative runs was not to conducted an alternative or obtain an improved stock assessment. This was outside of the terms of reference for the Review. They are presented for illustrative purposes of technical issues in the current assessment and as examples of the sorts of scenarios that are important to undertake within the assessment process and that would be useful to consider in the future.

After the review workshop, the reviewer undertook substantial additional review of the documents and material provided, including analyses of the results from the additional model

[2] This resulted in delays in the finalization of this review report.

runs performed during and after the workshop. Finally, substantial time was devoted to production of the current review report.

Overview and General Comments

Extensive information in terms of data, scientific literature and presentations were provided to the review workshop on the underlying biology, the Chilean fishery for the Patagonian toothfish (Chilean Sea Bass/Bacalao - *Dissostichus eleginoides*)) and the current stock assessment model. An impressive amount of fishery and biological sampling work has been undertaken on this species in Chile. Such work underlies any assessment and is critical for the evaluation of stock status and the provision of management advice.

Patagonian toothfish is a deepwater demersal species found circumpolar in the shelf regions of the southern oceans. It is slow growing with a relatively late age of maturity (8-12 years of age) and can live in excess of 30 years. Numerous fisheries exist in Antarctic and around near Antarctic islands and on the shelf waters of Patagonia and Chile (extending into Southern Peru). The fisheries tend to be relatively low volume but high value. Fisheries for this species are fairly recent (principally since the 1980s). In spite of the extensive scientific literature on this species, there is substantial uncertainty about components of its biology and life history which are important for assessing its status. In particular for the fisheries in Chile, the uncertainty about the stock structure, spatial dynamics and the relationship among fish caught in the areas where the Chilean artisanal and industrial fisheries operate and also among the areas where the Argentinean and Falkland Island fisheries occur is of critical importance both for the assessing the status of the resource and provision of management advice.

The fishery for toothfish in Chile has been divided into and managed as two disjoint but contiguous spatial components (i.e. an artisanal component north of 47 degrees south and a commercial component south of this – all of this takes place within the exclusive economic zone (EEZ) of Chile). The area fished by the Chilean commercial component is contiguous with the distribution of fish and areas fished by Argentina vessels within Argentina's exclusive economic zone (EEZ). In turn the distribution of fish and area fished in Argentina is also contiguous with the distributional area of the fish and fisheries for it within the Falkland (Maldives) Islands. A large amount of genetic and related information exists about the fish across these areas. All the data and information are consistent. They indicate that there is a single population across these areas and that this population is distinct from other toothfish populations, including the closest other one around South Georgia. Some recent results suggest that there may be some spatial structuring within this South American/Patagonian Shelf population, but with substantial mixing and interchange (Ferrada, 2014).

Until the current stock assessment being reviewed here, all previous stock assessments that the reviewer is aware of for toothfish in the South American/Patagonian Shelf area have been myopic in that they have only considered the catches and catch rates from only one of the three EEZ where fisheries exist (e.g., catches south of 47 degrees within Chilean EZZ, catches within Argentina's EEZ and catches within the Falkland Islands EEZ). The implicit, if not explicitly stated, assumption embedded in these assessment is that the catches from each of these area comes essentially from a single isolated population. As this assumption is not supported by any of the genetic, micro-chemistry, spawning ground and dispersal information, the interpretation and validity of the stock status conclusions is unclear, at best. The resulting management advice, in the reviewer's view, does not contain reasonable or reliable measures of risk in terms of either the conservation of the stock or under-utilization of the resource.

The current assessment constitutes a major advance in this regard. It is the first assessment that utilizes data from more than a single area (i.e. catches from southern Chile and southern portion of the Argentinean fishery) and is based on a conceptual model that explicitly acknowledges and accounts for catches from all areas in terms of the consequences for the overall population dynamics and status. The conceptual model is based on the assumption that all spawning occurs only in the areas where fisheries exist in southern Chile and southern Argentina and that toothfish in areas outside of these represents sink populations. These sink population cannot and do not contribute to the spawning and future recruitment. The sink populations arise as a result of larvae and juvenile dispersing from the spawning areas to regions that are impossible for them as adults (or sub-adults) to be able to make the return journey to the spawning areas. This conceptual model has two fundamental implications for management in terms of catches and the sustainability of the resource (1) that there is no need to regulate catches outside of the southern Chile and Patagonia shelf areas (except possibly in terms of yield per recruit or other economic considerations) and (2) that assessment advice in terms of overfishing and sustainability can only be provided on the consequence of the combined total removals by the Chilean and Argentinean fisheries. The allocation of catches among these two fisheries is almost exclusively a management issue for which the stock assessment is basically uninformative[3].

While the conceptual model underlying the current assessment does account for the consequences of all historical catches in terms of the current status and longer term dynamics of the resource, it is only one plausible model. Other (and in the reviewer's view more plausible) models exist. Other models would have substantially different implications for the assessment and management advice. This issue is discussed in more depth below. It is raised here because of its fundamental importance. As such and independent of any of the technical and computational issues raised below, in the reviewer's view the current assessment could not be considered as an adequate and robust basis for determination of stock status and provision of management advice.

Stock assessments and management decisions for this resource need to be cognizant and take into account the catches from all areas and be based on consideration of a full range of plausible assumptions for the interactions among the fish and fisheries found in the various areas. At the stock assessment level, this is a significant challenge for this resource because of the spatially disjoint fisheries and incomplete mixing of fish - i.e. non-homogenous distribution among areas (this is discussed further below). It is also recognized that there are significant political constraints and issues in terms of how to deal with management advice which requires addressing allocation of catches among different areas and fisheries while there is little or no scientific basis for the provision of advice on the allocation of catches across all areas. Nevertheless, there needs to be an understanding of the limitation of the advice science can provide (e.g. allocation among fisheries). Scientists should not be requested or attempt to provide advice outside of these limits.

[3] To the extent that selectivities are different among the fisheries, there would be implication in terms of the consequences on total catches for different relative allocations among different fisheries components. However, for any specific allocation equally robust advice in terms of sustainability and risk could be provided – i.e. there is no "scientific" basis for advising on the "best" allocation.

Within the constraint of the conceptual model, substantial and significant problems, limitations, potential errors and concerns were found with the data/parameter inputs, the implementation of the model and the results presented in the Assessment Report. These included (1) unrealistically high estimates of the level of precision associated with the assessment results and management advice, (2) lack of exploration of the assessment model behaviour, (3) arbitrary selection and insufficient understanding of input data and parameters without consideration and exploration of alternatives, (4) inconsistency and errors in the input data and CPUE standardization and (5) model implementation, software computational and documentation problems. The first three of these means that there is an underestimation both in the level of uncertainty in the resulting estimation of stock status and in the degree of risk associated with the management advice. In addition, one is not able to evaluate robustness of the conclusions and whether the assessment is centred around the most appropriate or "likely" region (i.e. the base case). Finally, the latter issues undermine confidence in the actual numerical results presented and the validity of their use for determining specific management decisions (e.g. quota levels) even if there were no other concerns and issues with the model structure and implementation. Overall, these implementation and computational issues, by themselves, in the reviewer's view, would also be sufficient for concluding that the results as presented in the Assessment Report are an inadequate basis for the determination of stock status and provision of management advice.

It should be noted that there was a change in the lead scientist performing the toothfish assessment for IFOP between 2013 and 2014. There was clearly a large amount of effort and hard work dedicated to this assessment. The assessment included a number of substantial and significant additions and innovations. These included joint modelling of the Argentina and Chilean catches, development of CPUE indices for the Argentinean longline fishery, the integration of the IFOP log book data with the scientific observers data belonging to the Centro de Estudios Pesqueros (CEPES S.A.) and incorporating a correction model for age readings obtained from scales. The reviewer considers that all of these were important factors that need to be considered within the assessment for this resource. The large amount of effort needs to be recognized. In the long term, these additions should result in improved and more robust assessments. However, it appears that there was little or no overlap between the current lead assessment scientist and the previous one. This meant that almost all of the components of the assessment were developed from scratch (*de nova*) including the model code, data compilations and the standardization of CPUE. Consequently, the workload to complete the assessment within a relatively short time was tremendous. Also, while the assessment report is a joint authored publication, it appears that most, if not all of the analyses, model development and computational work was performed by the lead scientist. Moreover, the lack of continuity both in personnel and analyses prevents understanding and evaluating the appropriateness and consequences of numerous changes in the assessment. Such discontinuities present difficulties for management and review processes. A more structured transitional approach with appropriate resources should be utilized when there are major changes in assessment personnel or model structure/data and even more so when both take place at the same time. In addition, there appears to be a lack of rigour and sufficient internal review in performing the stock assessments and in the preparation of assessment results. This is a major weakness in the overall assessment process, which needs to be addressed. This lack of rigour compromises the quality of the results and the large efforts that have been expended to produce them.

In spite of the extensive data that exists from the fisheries for toothfish in the South American -Patagonian shelf area, it needs to be recognized that there are severe limitation in these data and the information available for assessing this resource. These include (1) non-spatial overlap among the major fisheries combined with potentially a large amount of non-homogeneity in the distribution of adults and juveniles, (2) major temporal discontinuities in both the operations of the fisheries and the data collection, (3) lack of critical information for meaningful standardization of effort data (particularly historically and for the artisanal and non-Chilean industrial components), (4) major changes in gear in the last six years (5) large, but basically, unquantifiable interactions with marine mammals which may have been increasing over-time (5) uncertainty about historical catch levels and (6) no fishery independent estimates of abundance or fishing mortality rates. These combined with the biological characteristics of the species (e.g. slow growing, depth and spatial stratification with size/age, apparently long residence times) mean that the two basic pieces of information (i.e. CPUE and age/size composition) for assessing the stock are not very informative. The slow and variable growth, changes in aging method, the late age of recruitment to the fishery and the lack of any direct aging data from the Argentina component of the stock means that the aging data contain little information about the relative strengths of different cohorts, particularly the most recent (e.g. at least the last 5 – see below). Thus, essentially equally good fit to these data appear to be possible for a wide range of recruitment scenarios, Similarly, the interpretation of the trends in the CPUE data as trends in abundance are highly dependent upon assumptions about changes in catchability and selectivity. They are also potentially highly confounded by interactions with marine mammals. In particular, the most recent CPUE trends, particularly for the Chilean longline fishery, show a steep and continuous decline. However, information is lacking for adequately determining the extent to which these recent trends should be attributable solely to changes in abundance and not, at least in part, to operational changes in the fishery and marine mammal interactions.

Finally, the current stock assessment indicates that the stock is in a highly depleted state and that the magnitude of the stock is low relative to current catches (i.e. estimates of fishing mortality rates are high). The trends in CPUE, which dominate the stock trend estimates in the assessment, indicate that stock size has been substantially and more or less continuously reduced since the inception of the fisheries. Resolving the data input, implementation and computational issues is essential in order to ensure confidence in the results within the context of the conceptual model used in the Assessment Report. However, without broadening the range of conceptual models considered and the approach and methods used for dealing with uncertainty (particularly model uncertainty), there will remain an insufficient basis for the determination of stock status and provision of management advice and associated risk. These latter issues would probably be most effectively addressed within the context of an operating model and management procedure evaluation approach (see below). In a longer term framework, additional information, particularly on the spatial dynamics of the resource and fishery independent measures of abundance, is essential if the basic underlying and structural uncertainties involved in assessing this resource are to be reduced. Tagging experiments, well designed and properly implemented, are probably the most effective, if not the only, method for obtaining this information. Currently, and until these issues are dealt with, substantial risk either to the stock and/or the fishery (in terms of foregone catches) will be embedded in any management advice and decisions.

Summary of Findings for each Terms of Reference

1. Stock assessment approach

The Assessment Report presents a detailed and structured description for the conceptual framework or model underlying the analytical stock assessment model. The reviewer considers this a valuable component of the Assessment Report, which is frequently not found within many stock assessments. It allows for evaluating whether the mathematical model reasonably represents key components in terms of the biology of species and operation of the fishery.

As stated in the Assessment Report "the most remarkable features of the conceptual model refer to the interpretation of the immigration and emigration processes". Thus, the conceptual model is based on the premise that all recruitment to the stock is derived from the one large contiguous spawning area located in the Magellan Region that extends from the austral southern zone of Chile (below 47°S) eastward to the westward part of the Burwood Bank. Within this spawning area, individuals are all part of one common stock. Recruitment to the north of this area both in the Pacific and Atlantic is the result of diffusion of eggs and larvae and emigration of juveniles. However, individuals that are advected or immigrate to the north of 47°S in the Atlantic and to the north of 54°S in the Atlantic are assumed to be unable to return to the spawning area to spawn once mature (i.e. are "sink populations"). As such, individuals in these sink populations have no potential to contribute to future generations and the sustainability of the resource. This is a strong assumption with important implications for both the stock assessment and management of the resources. Under this assumption, catches to the north of 47°S in the Atlantic and 54°S in the Atlantic (i.e. from the sink populations) nor trends in abundances in these area are of no consequence or need to be considered within the stock assessment[4]. This includes the substantial catches by the Chilean artisanal fleet, the Argentinean catches north of 54°S and the catches in the Falkland Islands. The magnitude of these catches in 2013 exceeded the catches from the main area or source population as defined in the Assessment Report. In terms of conservation and sustainability of the stock, there is no reason to manage the catches in any of these northern areas. The only real management consideration would be in terms of selectivity/yield per recruit or other economic considerations. However, the fisheries in the northern area are totally dependent upon the conservation and management in the southern areas. The critical management decisions for the sustainability of the stock are dependent only upon the catches and population status in the southern area. It is from this population that all recruitment is derived and from which estimates of maximum sustainable yield (MSY) can be determined[5].

---

[4] Indices of recruitment in these areas could possibly be useful as indices of the spawning stock sizes that produced them given an assumption that the proportion advected was similar over time and spawning stock size.

[5] Note that embedded in the Assessment Report is the implicit assumption that the population level that yields MSY for the southern areas will also yield MSY for the entire fisheries (i.e. those in the source and sink populations). While possibly not an unreasonable assumption, it does depend upon recruitment to the sink population being a linear function of recruitment to the source population independent of spawning biomass levels or else independent of spawning stock sizes. It also does not consider differences in selectivities between the fisheries in the sink and source populations. The risks under this implicit assumption about MSY levels are very asymmetrical (e.g. over-catches in the southern area risk the sustainability of catch in both the sink and source areas while under-catches in the southern area should not risk obtainable or sustainable catches in the north and could potentially actually increase them).

The underlying justification in the Assessment Report for this strong and "remarkable" assumption about movement between southern and northern areas being one-directional comes from statement in the discussion paper in Ashford et al (2012)[6] that concluded that "adult toothfish are not physiologically capable of large sustained counter-current migration, and, although neutrally buoyant, there are no return pathways in the large circulation model." It should be noted that the Ashford et al (2012) paper is about dispersal of larvae and juveniles based on a large scale circulation model and has no data nor does it provides any references to support the statement that toothfish are not physiologically capable[7]. In fact the statement about being physiologically incapable of such migrations is contradicted by tag return data in which individual toothfish have travelled extensive distances (e.g. over 2,800km in a period of less than 14 months). Thus, the critical question in terms of the plausibility of the immigration/emigration assumption underlying the conceptual model underlying the stock assessment is not whether toothfish are capable of such large distance and counter-current migrations, but the frequency with which such migrations actually occur, particularly for the purpose of spawning. While most of the recovered tagged fish have been recaptured within 50 miles of their point of release, almost all of these tagged and recaptured fish have been juvenile. Thus, these data provide little direct evidence whether or not individuals actually can and do undertake migrations to the spawning area when they mature[8]. From an evolutionary perspective, the selection pressure to undertake such migrations from these "sink" areas would be large, particularly given the large sizes of these sink populations based on the catches taken from them.

Consequently, the underlying sink/source conceptual stock structure model underlying the assessment is only one plausible hypothesis. In the reviewer's opinion, it is not necessarily the most plausible. Currently available direct information is insufficient to determine the actual structure or even the most likely model. Since alternative conceptual models may potentially have very large and different implications for the stock status and management, it is critical that a range of alternative assessment models be considered that represent the range of plausible alternatives. Basing management decisions based only on the single conceptual stock structure model in the Assessment Report could not be considered precautionary without understanding what the implications would be for other plausible alternatives.

With respect to sink populations, in the reviewer's view, the data are inconclusive as to whether in fact they do exist and, if so where the southern and eastern limits for sink populations would be. In the assessment, the assumption that a sink population in the Pacific exists beginning at 47°S appears to be arbitrary. No biological or physical justification is provided for this line. It appears to have been chosen because it is the management boundary between the industrial and artisanal fishery. As such, it means that the artisanal catch and

[6] Note that while this is a key reference for the Assessment Report it was not included in the list of cited references provided in the report.

[7] No additional supporting evidence for this statement was supplied during the workshop although the question of support for this statement was raised.

[8] Fenaughty (2006) expresses similar concerns about the interpretation of tagging data for toothfish in the Ross Sea. A similar situation with respect to stock and size segregation exists in this area. He notes that the predominance of sexually mature fish in the north and the general absence of fish smaller than 100 cm and concludes that an appropriate interpretation of the available data is that "*D. mawsoni* builds up condition in the south over one or more seasons and then migrates north for spawning. He further notes that "Some preliminary work carried out by the New Zealand National Institute of Water and Atmospheric Research (NIWA) indicated that large geographical movements may be limited to larger fish. If this is the case, major movements of tagged fish will not be apparent until increased numbers of larger toothfish are recaptured from the tagged pool over coming seasons".

effort data need not be considered in the assessment, although these artisanal catches are substantial in terms of the overall catches of toothfish in Chile[9]. Similarly, the assumption that a sink population exists north of 54°S in the Atlantic Patagonian shelf also seems arbitrary. In the area of this assumed sink population, substantial catches of toothfish are taken both by Argentina and the Falkland Islands. As such, the implication of alternative boundaries needs to be explored within the context of the conceptual model underlying the current assessment.

In terms of the Chilean fishery, there are a range of alternative hypotheses that need to be considered both in terms of stock structure and possible sink populations (e.g. immigration/emigration). With respect to stock structure, the limits taking into account available information would seem to be:

1. That the stock with respect to the Pacific component (Chile) is essentially closed both with respect to recruitment and spawners. In other words, all recruits/juveniles found in the Pacific from adults that spawn in the Pacific and all adults that are found in the Pacific are from the juveniles were recruited into the Pacific (i.e. no significant immigration of larvae, juvenile or adults into the Pacific occurs from individuals from the Atlantic side). Supporting this would be the strong and persistent ocean currents structure current around the southern tip of South America combined with the limited movement shown by juveniles based on tagging data and possible physiological swimming limitations in toothfish. Note that for the "stock/s" found in the Atlantic, the hypothesis of its being closed with respect to the fish and fishery in Pacific does not seem plausible as substantial recruitment from spawning in the Pacific would be expected due simply to currents (Ashford et al. 2012). Under this hypothesis, the fishery in Chile could be assessed and managed without regard to catches in Argentina or the Falklands (but not vice-versa).

2. That the stock in terms of fish found in the Pacific and Atlantic are open to each other largely as a result of adult mixing within the spawning grounds and perhaps with more limited mixing of juvenile/sub-adult. Thus, recruitment in both the Pacific and Atlantic is dependent upon the combined spawning biomass from both oceans. Supporting this hypothesis is the fact that there is one contiguous spawning ground with no definitive barrier between the Pacific and Atlantic combined with the limited tagging data that suggests reasonable interchange of individuals in the area around the border between the Atlantic and Pacific as well as some longer scale movements. Under this hypothesis, the fishery in Chile could neither be assessed nor meaningfully managed without consideration of the fisheries in the Atlantic. Note that mixing maybe (and likely is) incomplete.

3. That the stock in terms of the Pacific is essentially closed (the same as under hypothesis 1) and that the more northern areas where fish are found constitute sink populations (e.g. the implicit hypotheses used in past Chilean assessments). The most border between the source and sink population is unclear. The division at 47°S between the artisanal and industrial fishery appears arbitrary in terms of the biology of the species. As such, the implication of alternatives and possibly more appropriate division need to be explored.

[9] Misreporting of the location of artisanal catches is the one complicating exception. This issue is discussed in Annex 4 of the Assessment Report.

Consequently, in addition to the stock structure/sink population hypothesis in the current stock assessment, at least three additional alternatives assessment models would appear to be necessary to ensure a reasonable level of robustness from the perspective of the fishery and management of the resource that exists within Chile. These are:

1. The entire Pacific shelf and Patagonian fisheries should be assessed as a single stock (i.e. inclusion of all catches from Chile, Argentina and the Falkland Islands);
2. Only consider catches by the Chilean industrial Fishery;
3. Only consider catches by the Chilean industrial and artisanal fisheries.

Variants of the above would also be important to consider (e.g. excluding catches from the more northern regions of the Chilean artisanal fishery, excluding the Falkland Island catches, etc). The implications in terms of stock status and management advice are likely to be substantially different. The development of a robust and precautionary approach is essential for integrating across results for such a range of alternative hypotheses. Moreover, some form of international cooperation is required for (1) the collection, standardization and sharing of data required for conducting the stock assessment and (2) the implementation of management advice (e.g. allocation of catches). In the longer term, research (which will also require a framework for international collaboration) should be undertaken to provide an improved understanding of the population structure and spatial dynamics of the toothfish within the Pacific shelf and Patagonian area and the inter-relationship between the various fisheries. The results of such research should allow for the development of more realistic and spatially explicit approaches for modelling of the stock and the fishery dynamics.

2. Life history parameters used in the assessment

Natural Mortality

Natural mortality, in general, is perhaps the most difficult biological parameter to estimate for fish stocks and direct estimates are unavailable with a few exceptions. Commonly, the values used in stock assessments are derived from theoretical consideration in conjunction with comparative studies of life history parameters (e.g. longevity). This is the situation in the current assessment. While such indirect estimates provide a guide as to what could be considered a reasonable value for a species, they can hardly be considered to be precise or accurate. There is a tendency for assessments to arbitrarily adopt the value used previously with "tradition" and "continuity" assuming the rationale. The sensitivity of assessment results to uncertainty in the parameter value used for natural mortality rate is often not considered. More importantly, there is little consideration of how uncertainty in M affects estimates of uncertainty in estimates of stock status and of risk in the derived management advice. Assessment results (particularly absolute values) and the implication for management are frequently sensitive to the value used. This is the case with the current assessment and is one factor that needs further consideration in any assessment of this resource.

It is worth noting that there is one published direct estimate of natural mortality for toothfish derived from tagging data, although not discussed in the stock assessment report (Candy et al, 2011). While not for the same stock, the estimate of 0.155 does lend support for the value of 0.15 used in the current assessment. However, the confidence intervals for this estimate are wide, which also supports the importance of considering uncertainty in the value used for natural mortality.

Historically, natural mortality was not directly estimable parameter within stock assessment models. However, with the development of statistical catch-at-age models, natural mortality can be included as an estimable parameter and model estimated values can be obtained. Nevertheless, there is generally little actual information on natural mortality in the data provided to the model. The model derived estimates are generally not well determined. They can be confounded with other parameters and underlying structural assumptions in the model (e.g. fishing mortality rates and selectivity). Such model derived estimates can be useful in providing a guide to what is a reasonable range to consider. Assumed values that are well outside the range estimated by the model indicate an inconsistency with the model and observed data used in the model (e.g. either that the assumed value is unlikely or that structural problems exist in the model and/or data). Thus, if advice is made conditional on a specific model structure, it would not be appropriate in most situations to use estimates for natural mortality that were widely inconsistent with those estimated from the model.

Uncertainty in natural mortality (M) was not considered within the Assessment Report. During the workshop, model runs were made in which alternative values for M were assumed and also two runs in which M was estimated by the model. In the latter case, different estimates were obtained for different initial values for M (i.e. the resulting estimates were 0.146 and 0.178). The resulting fit in the latter case was clearly not a global minimum (i.e. the objective function was greater than when M was fixed at 0.15) although the maximum gradient (a standard criterion) indicated that the model had converged[10]. This indicates that the "likelihood" surface is quite flat with local minimum when M is allowed to be a free parameter. Further, it indicates that M is highly confounded with other parameters. This indicates that there is insufficient information in the data to provide a reliable estimate of M directly from the model. When alternative values of M were assumed, the results, not surprisingly, were sensitive to the value used (Appendices 4 and 5). The degree of sensitivity to the value for M was in part affected by whether the aging error matrix was applied (see below). For example, a value of M=0.10 yielded an estimate of 0.08 for the current depletion of level of the spawning stock while M=0.20 yielded an estimate of 0.24 when the aging error matrix was applied (Appendix 4). Values of 0.06 and 0.16 respectively when the aging error matrix was not used (Appendices 5). The consequence of these differences for management advice would be substantial and emphasize the importance of integrating uncertainty about the value of M into the advice provided, particularly within the estimates of risk.

Even more difficult than the average value of natural mortality is the question of whether it changes with size or age. Outside of theoretical considerations (e.g. differential vulnerability with size to predation, senescence, etc.), there is no basis for assuming a functional form or determining actual values in most cases. Assessment model generally assumes constant natural mortality rate with age with parsimony underlying the choice of this assumption. This assumption could be considered reasonable approach here. Nevertheless, the longevity and large size of adults of this species combined with changes in habit with size/age suggests that natural mortality may likely be age/size dependent. Additionally, the fact that the model estimates that full selectivity is at a late age could be the result of confounding with age-

[10] In the revised set of runs provided after the workshop (see below), this problem of not finding a global minimum still remains although not obviously apparent in the results as presented in Appendix 5. This is because there was a miss specification in the control file for the run in which M was set as an estimable parameter (Run 12). The aging error matrix was actually turned on in this run. As such the results for this run should be compared to Run 2 in the tables and figures in this appendix and not Run 1 (see below for more information).

specific changes in natural mortality. While not among the highest priority, exploration of possible age-specific natural mortality rates would be valuable and informative.

Growth and Weight-Length Relationship

The Assessment Report notes that "growth is not explicitly modelled" within fitting of the stock assessment parameters. As such, the estimation of growth was not considered within the stock assessment context. However, implicit in the estimate of sexual maturity is an estimated growth curve. This is because the studies of sexual maturity cited in the report are based on length with no direct estimate of age for the individuals sampled. In addition, the cohort splicing of the Argentinean length data to yield estimates of the catch-at-age data requires an explicit growth curve as noted in the Assessment Report. The Assessment Report provides no documentation of the growth curves used in either case. Based on the table of estimated growth curves provided in the stock assessment, there remains a lot of uncertainty about growth (e.g. Table 2 in the Assessment Report and Appendix 6 of this report). It is not clear whether the variability seen reflects actual variability in time and space or reflects sampling and age estimation issues. The problem of growth is further complicated by the sexual dimorphism in growth that exists in this species and that the two sexes are not caught in equal proportions. To the extent that the uncertainty reflected in various estimated growth curve reflects actual variability in growth it could have implications for the assessment and management advice (e.g. if there has been a temporal trend or growth is density dependent).

Integral to the stock assessment are the estimates of weight-at-age (e.g. the actual number of removals from the stock as a result of fishing when fitting the model are dependent upon this inputted relationship). The Assessment Report utilizes "empirical data of average weight-at-age" and uses a single vector of average weight-at-age for this purpose. No documentation was provided on the source of this or how it was calculated (e.g. a simple average of all observed weight-at-length data across all years). While more detail on this was requested, these were not able to be provided at the workshop. Further, it is not clear whether the empirically derived weight-at-age relationship used in the stock assessment only used data since 2007 onward (when aging has been done with otoliths), or used data from the entire time series of catch-at-age data. This issue needs to be clarified[11]. Moreover, it is important that the annual data are examined to ensure that there has been no substantive temporal trend reflecting changes in growth that need to be accounted for within the stock assessment.

While using an empirical derived weight-at-age relationship does not require the explicit use of a parameter based growth curve, there is still an implicit growth curve within it. This is in terms of growth in weight. There is also an implicit growth curve in terms of length which depends upon the weight-length relationship. The reviewer is concerned that the empirical derived weight-at-age relationship used in the assessment appears to be inconsistent with what might be expected in terms of the estimates of growth curves for this species provided in the Assessment Report. There was insufficient time within the workshop to explore this. However, after the workshop, the reviewer undertook some preliminary examination of this question (see Appendix 6). These preliminary analyses suggest that the mean weight-at-age used in the assessment may substantially miss estimate the actual mean weight-at-age (Figure 1). However, this is confounded by the fact that most of the estimated growth curve from

[11] Confounding will exist between mean estimates of the weight-at-age and age reading errors from scales. Thus, if the actual true weight-at-age estimates are used (e.g. based on data since 2007 and assuming no change), then the actual number of removals in assessment model and not only the age distribution of the catches will be distorted. Whether and to what extent these two factors would compensate for each other is not obvious.

Chile and the Patagonian shelf were estimated prior to 2006. As such, it is not clear whether the estimates of age used in estimating these curves was based on scale or otolith readings. The one growth curve in Table 2 of the Assessment Report estimated after this year (i.e. completed in 2013) is perhaps the most consistent with the empirically derived weight-at-age relationship used in the Assessment Report. Nevertheless, this curve still suggests a substantial mismatch (particularly for younger ages) with the empirical relationship predicting higher values for the weight at any given age (Figure 1).

In addition, the reviewer calculated estimates of the total weight of the catch based on the estimated catch-at-age data time the empirical weight-at-age relationship. These estimates were always greater than the total weight from the landing data for the Chilean longline fishery since 2007, when aging has been based on otoliths (Figure 7 below). The difference is around 60% in three years and always greater than ~20%. This is discussed in more detail below in the context of total catch estimates. In the current context, an over-estimation of the mean weight-at-age would lead to such a discrepancy.

There is a need for a thorough and complete analysis of the data and the relationships between age, length and weight.

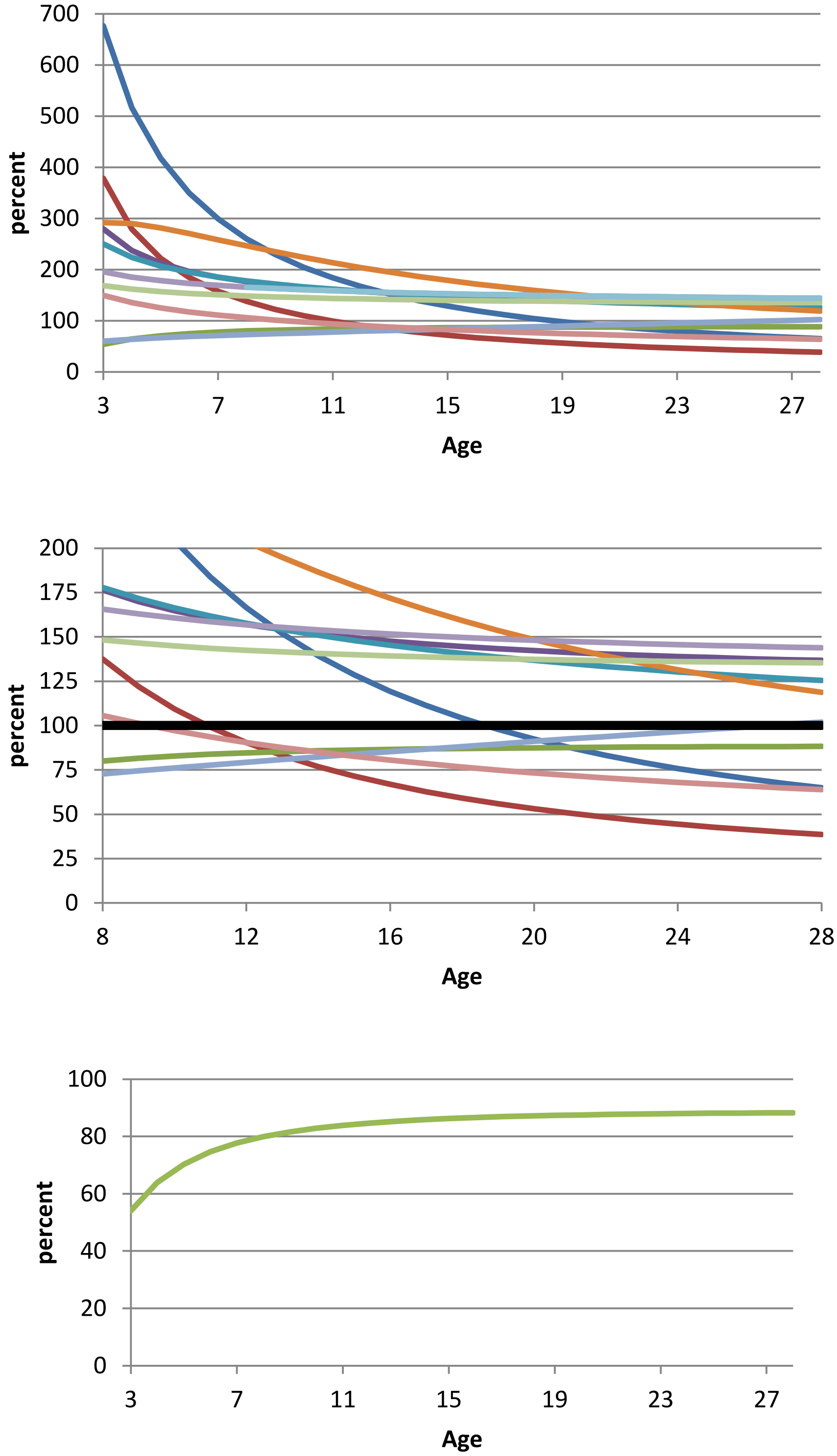


Figure 1: Estimates of the mean weight-at-age based on the growth curve parameters provide in Table 2 of the Assessment Report as a percent of the mean weight-at-age values used in fitting the stock assessment report. The top panel is for all ages, the middle panel provides more detail for older ages and the bottom panel provides the result for the one growth curve estimated after 2006 (see Appendix 6 for more detail).

Size/Age at Maturity

The stock assessment considers only a single maturity ogive based on a logistic function of age. In the Life History Section of the Assessment Report (Section 3.1.2), estimates of 82.3cm for males and 83.7cm for females are provided for $L_{50\%}$ for the size of maturity. The Assessment Report also provides a table with a large number of estimates for $L_{50\%}$ which shows a wide range of values for both the Chilean/Argentinean Patagonian Platform area and the species in general. In the other main document provided for the review (Gálvez, 2014), 110 cm is used consistently as the reference size for the age of maturity. The ogive used as input when fitting the stock assessment model has an $A_{50\%}$ equal to an age of ~12.5 (Figure 2). This age of 12.5 does not appear to be internally consistent with the growth and the size of maturity estimates stated in the Assessment Report. Thus, 83.7 cm generally corresponds to a mean length-at-age of less than 8 over the wide range of estimate of the growth curves provide in Table 2 of the Assessment Report and never exceeds 10 (Figure 3 and Appendix 6). Even a size of 110cm generally does not exceeds 10 years of ages over the range of estimated growth curves and when it does it corresponds to a greater age and does not appear to correspond to an age of 12.5. In addition, if one estimates the weight of a fish corresponding to 82.3 or 83.7cm using the weight-length relationship provided in Gálvez, 2014 for the 2013 catches, then the resulting weights are between ~5.8 to 6.2kg. Based on the empirical weight-at-age, this would correspond to a fish between 8 and 9 years of age. As such, the age at maturity ogive used in the assessment appears to be inconsistent and inappropriate with the information provided. There is a need for a more complete analysis of existing data along with thorough documentation in order to be able to evaluate what would be appropriate values to use.

Estimation of size or age of maturity is not straight forward for a species such as toothfish in which there is spatial segregation in size and age. Estimates from any given sample are dependent on the location and timing of where the samples were obtained and are likely to show high variability. Most frequently estimates of maturity are derived from samples taken on or near the spawning grounds. However, samples taken on or near the spawning grounds will likely show a disproportionally high fraction of sexually mature individuals for smaller or younger animals. This is because only those individuals that are mature are likely to move to the spawning grounds. As such, estimates taken from the spawning ground will tend to bias upward the estimates of the proportion mature for younger/smaller animals. What is needed to obtain an unbiased estimate is a representative sample from the entire population. Obtaining such a sample is not simple as it requires sampling in proportion to abundance for each age/size class in space, but their relative abundance is not known. Mixing sample from across the spatial range of a population and then weighting by relative catch rates would be one approach. Generally, samples and data are insufficient or unavailable for even doing this. Fortunately, assessment results generally are not highly sensitive to the precise value used, if the discrepancy is not too large. Nevertheless, it is important to consider this sampling problem when estimating age/size of maturity (e.g. to evaluate samples from on and off the spawning grounds) and also to check the sensitivity of the assessment results to the specific value used.

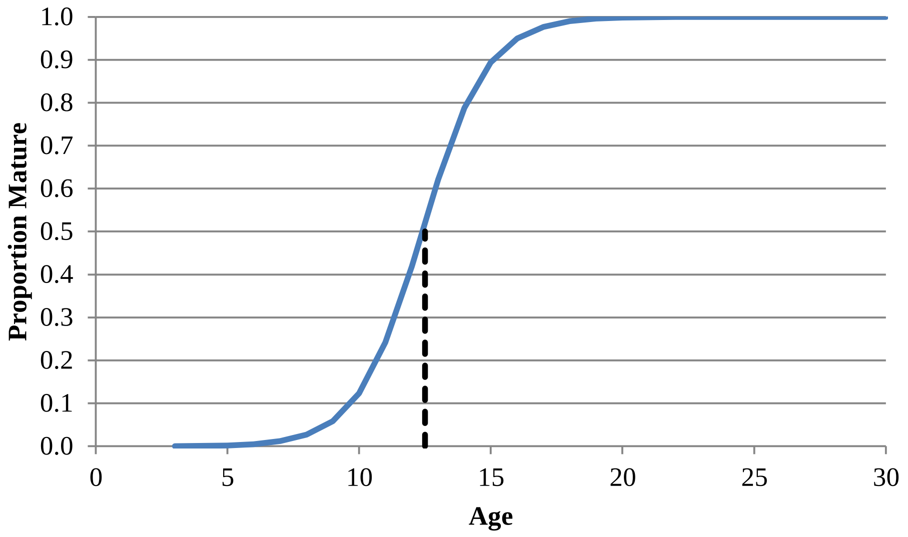


Figure 2: The maturity ogive used in the estimation of spawning biomass in the Assessment Report and in the alternative runs conducted as part of the Review Process. The vertical dashed line corresponds to an age of 12.4.

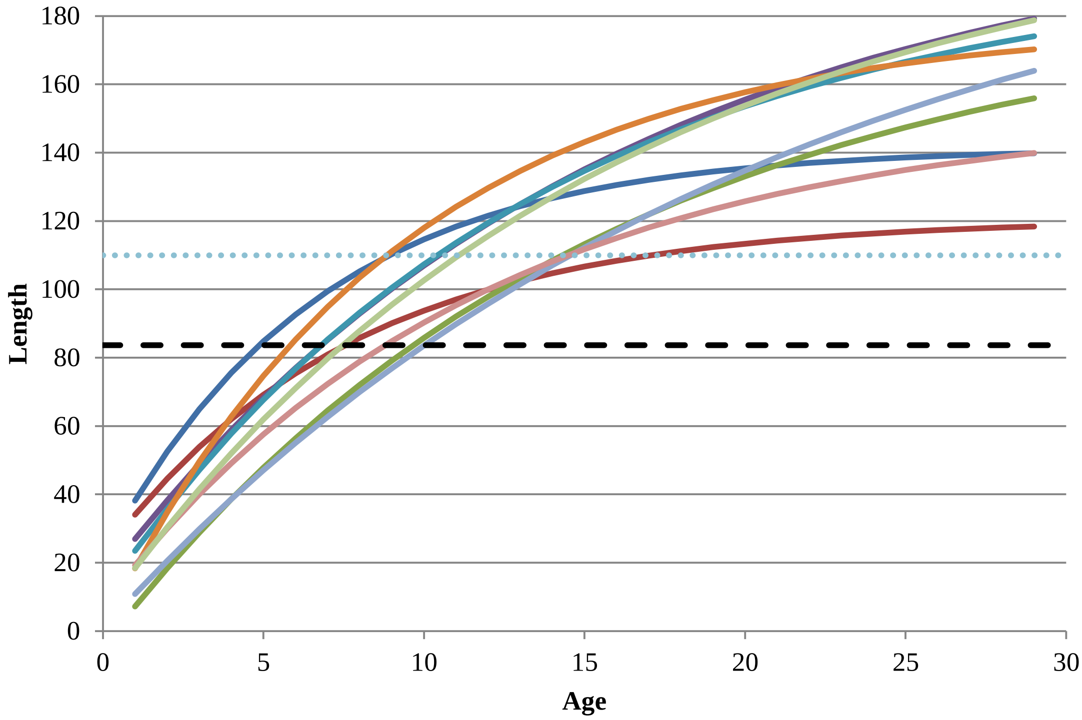


Figure 3: Estimated mean length-at-age for the range of growth curves considered in Appendix 6. The horizontal dashed-line corresponds to a length of 83.7cm. This is the length given in the Life History Section of the Assessment Report (3.1.2) for L50% (see Appendix 6 for more detail).

Sex Ratio and sexual dimorphism

The assessment model does not model the male and female segments of the population separately. Available data indicates that there is sexual dimorphism in growth based on the size-at-age estimates from direct aging and unbalanced sex ratios in the catches from the industrial fishery (biased towards males). While information on this is presented in the two main assessment documents, it is not discussed in the context of the conceptual model or the specification of the stock assessment. The implicit assumption in the model is that there is no differential in growth or sex ratio within the catches. As a first approximation, this is not an unreasonable assumption. The additional complexity that would be required to develop a sex specific population model may not be supportable by the available data and is not among the higher priorities given the other issues in the stock assessment and data. Nevertheless, the implicit assumption in the stock assessment with respect to sex, as well as possible implications, should be discussed and, in the fullness of time, explored within a modelling framework. In this regard, it should be noted that the results may be precautionary in terms of implications for the spawning stock. The differential and higher exploitation rate of males means that the model would likely be an underestimation of the reduction in the female component of the stock, which is the critical component with respect to the spawning biomass.

2. Use of age information coming from scale and otoliths reading

The age information used in the assessment comes from three different sources. For the Chilean fishery, aging of the catch from 1991-2006 are based on reading of rings from scales while since 2007 aging has been based on readings from otoliths. Although initial comparative studies of readings from the two approaches suggested no difference, latter studies showed significant differences. Scale readings provide under estimation for older ages when compared to estimates from otolith readings. However, readings from the two approaches still appear to be consistent for younger ages. Differences of this sort are not uncommon between otolith and scale readings. Otoliths are generally considered to be more reliable. Otolith readings have been adopted as the standard internationally for age determination for this species. As such the switch to otolith reading appears to have been a reasonable decision even though, as noted in the workshop, uncertainty exists as to whether the bands read in the otolith corresponded to annual increments[12].

The switch to otolith aging produces a bias in the pre-2007 catch-at-age (i.e. an under estimation of the age structure) and also induces a discontinuity in the time series between 2006 and 2007 (i.e. older cohorts will tend to be over-represented after 2006 relative to their proportion in the catch prior to this time). As such, it is important to account for the error in the age data based on scale readings. The current assessment is commendable in that it attempts to account for this error and appears to be the first time that any attempt to correct for the error has been included within the stock assessments[13]. The approach used was to create a "reading error matrix" which was meant to estimate the probability that for a given true age of a fish what its perceived age based would have been when derived from scale

[12] While not discussed in the workshop or the Assessment Reports, there is some work which does provides direct evidence for the bands are laid down annually (Horn et al. 2003). This further supports the decision to use otoliths for age determination.

[13] Note that because of a miss-specification in the input control file used for producing the results in the Assessment Report, the correction was not actually applied. Thus, the results provided in the Assessment Report contain no correction for age reading errors from scales (see below and Appendix 5).

readings. These estimated probabilities were then used to correct the model predicted catch-at-ages from the Chilean fleet when fitting the model to the observed catch-at-age data from scale readings[14]. For any true age, the probabilities need to sum to one. However, two problems exist in the actual approached used.

- The first is that any correction needs to be applied to the age-length keys and not directly to the resulting catch-at-age matrix (i.e. the actual error in the estimated observed catch-at-age data will depend not only upon the error induced in the age-length keys but on the on the proportion of individuals within each length class in the catch. This is because age-classes overlap in length)[15].

- The second and equally, if not substantially more, important than how the reading error matrix was applied is the actual reading error matrix used (or intended to have been used) in the current stock assessment. While it is not clear which of two possible version of the reading error matrix were intended to have been used in calculation of the result presented in the Assessment Report (see below), either version would have resulted in substantially over-correcting for the biases in the scale reading data based on data on the extent of the reading error from scales provided during the Review Workshop. The Assessment Report provides no documentation as to the source of reading error matrix or how it was constructed. As reported in the Review Workshop, the matrix was constructed based on the judgement of the stock assessment scientist of the likely amount of error there may have been and was not based on actual observed data. However, actual data exists from which an error matrix could have been derived. It is not clear why these observed data were not considered or used.

Figure 4 illustrates the results from comparative scale and otolith age reading from the same fish. This figure indicates that by beginning around age 15 there is a very strong tendency for scale readings to under-estimate age compared to estimates based on otoliths. In contrast for younger ages, it is not clear if in fact a bias does exist. A statistical analyses would be required (the figure does not allow for this as it is possible to determine the actual number of readings represented by each point on the graphs). If a bias exists for younger ages, the figure suggests that it would be relatively small and shows that for any age less than ~15 that there exists some positive probability that the some of the scale derived age readings would be greater or equal to the age read from otoliths. However, in either version of the possible reading error matrix intended for used in the stock assessment, it is assumed that there is zero probability that the scale derived age could equal or exceed the true age (i.e. its otolith read

[14] The assessment report provides only a reading error matrix "used in modelling age compositions obtained by reading growth rings on scales" but provides no documentation on how this matrix was actually used within the assessment model. This is only determinable from examination of the computer code.

[15] Two approaches could be used within the current modelling framework. One would be to correct the observed catch-at-age data used in fitting the model by deriving estimates what the probability of what the true age of a fish was given its perceived age based on scale readings (i.e. the reverse probability used in the assessment document), then applying this to correct the actual age length keys from scale readings and then multiplying these corrected age-length keys by the observed catch-at-length data to obtain corrected catch-at-age data to use as input data in fitting the model. The other approach would be to continue to model within the stock assessment, itself, what the observed data would have been based on modelling expected age-length keys and catch-at-length data and comparing these to the observed data. While from a likelihood estimation approach the latter might be considered preferable (i.e. fitting to observed data), the former would be more straightforward and simpler to implement and would not require the additional complexity of modelling lengths within age (i.e. growth) within the assessment model. Also, the advantage in terms of a likelihood approach is theoretical, in any case, as the error associated with the observed catch-at-age when fitting the model is based on assumed effective sample sizes.

age) with the exception of age three fish (which are almost non-existent in the catch data). For example, all age 4 fish are assumed to have been aged as 3 when aged with scales, all age 5 are assumed to have been aged as 4, all age 6 are assumed to have been be equal to age 5 or less, etc. (see Table 8 in the Assessment Report and Tables Appendix 5.1 and 5.2 in this report.) Thus, the reading error matrix would have resulted in an over-correction and the introduction of substantial biases, particularly for younger ages. These biases are likely to be greater in extent and affect (but opposite in direction), then the bias resulting from simply not including any aging error correction.

In addition to the problems and biases in the Chilean catch-at-data resulting from the scale readings prior to 2007, the basis for the catch-at-age data for the Argentinean catches are also potentially a substantial problem and concern, No direct aging of the Argentinean catch data has been undertaken. The catch-at-age estimates for the Argentinean catches are derived from estimates of the catch-at-size data based on some form of cohort slicing based on estimates of the mean age-at-length. The latter is derived from an estimated growth curve. However, no information was available in the Assessment Report on the actual growth curve employed, how it was estimated or the actual cohort slicing method used. Cohort slicing methods will tend to smooth out the variability in the catch-at-age matrix resulting from variability in recruitment and are also likely to over-estimate the proportion of the catch among the older age classes. These effects will depend upon the estimated growth curve and the associated variability in length-at-age along with the actual approach used for the cohort slicing. Comparison of the age structure for the older ages in the catch from the Chilean longline catches based on direct otolith age estimation with those for the Argentinean longline catches based on cohort slicing suggests that this second effect does exist in the “observed” data for older ages (e.g. Figure 5). Moreover, both of these cohort sliced effects can produce inconsistencies and conflicts when a stock assessment attempts to fit simultaneously to data from cohort sliced and direct estimates of the catch-at-age from different fisheries components. Such effects are apparent in the current stock assessment based on the residuals for catch-at-age data (see below). A more appropriate approach for incorporating the data from the Argentinean fisheries would be to have the stock assessment model predict the expected size distribution for these catches based on either an estimated growth curve or estimates of the size-at-age derived from the Chilean fishery.

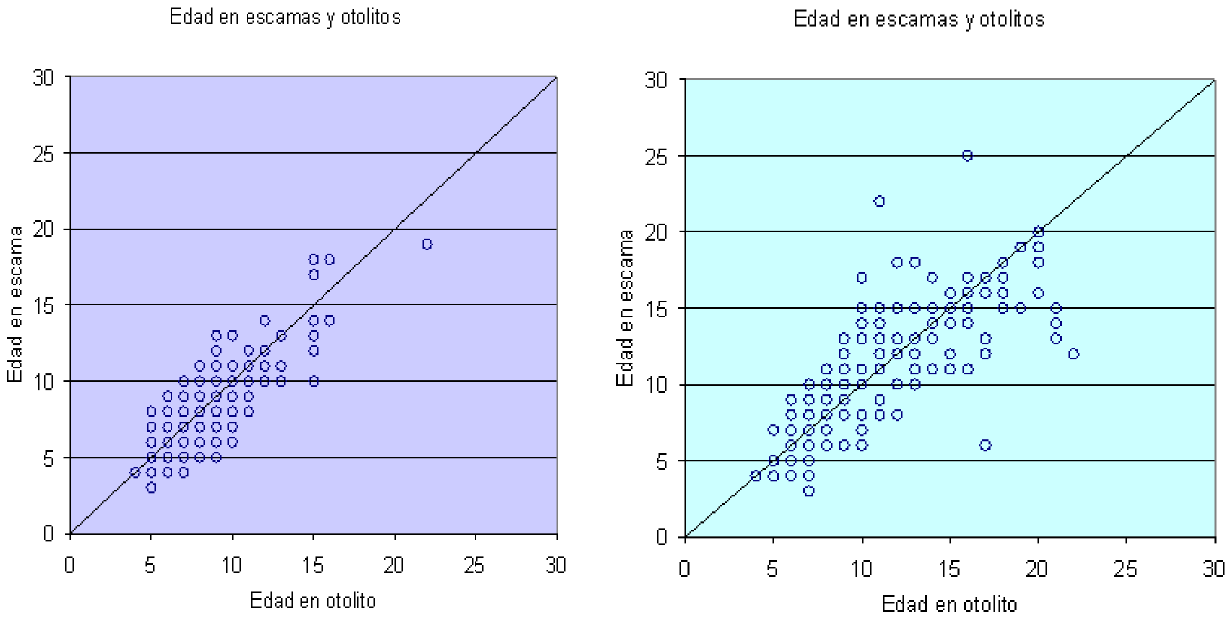


Figure 4: Comparison of age estimates obtained from reading of scale and otolith from the same fish from an experiment conducted in 2007. The figure on the left is from a 198 samples collected from the Artisanal fishery in 2007 and the figure on the right is from a 194 samples collected in the Industrial fishery. (Figures taken from the presentation "EDAD BACALAO DE PROFUNDIDAD (*Dissostichus eleginoides*)" by Vilma Ojeda C. at the Review Workshop).

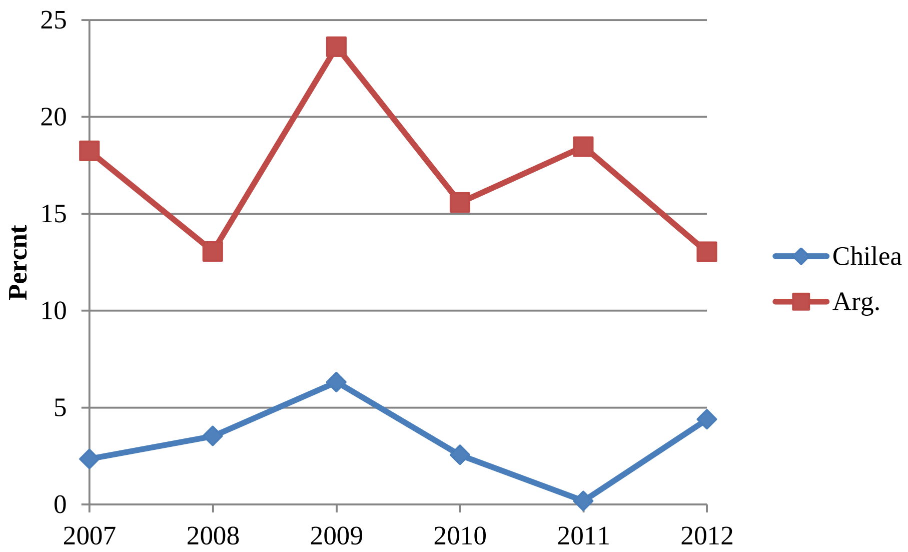


Figure 5: Percent of the "observed" catch-at-age data used in the stock assessment which is in the plus group (plus 29) out of the total for ages greater than 20. Values are shown for the Argentinean and Chilean longline fleet since 2007. Note that prior to 2007, this percentage for the Chilean catch is always zero but the observed percentage is confounded by the use of scale readings for age determination.

3. Quality and reliability of different pieces of information and estimation approaches – CPUE and Catches

CPUE

Given the available information for assessing this resource, the CPUE indices will be a dominant component, if not the principle driver, in determining the results. Thus, it is critical that complete and thorough analyses of the existing data be undertaken. This includes exploring alternative model structures and hypotheses when conducting the standardization. The statistical standardization of the CPUE data included in the stock assessment report utilizes widely accepted and well developed methods. Within the context of the specific analyses preformed, the resulting abundance indices probably provide valid and reasonable statistical analyses (but see below for a number of technical problems). However, the reviewer found that the analyses were limited in their exploration of the model structure, the terms included in the model and in the interpretation of the results as indices of abundance. In addition, the actual documentation of the CPUE analyses was also limited, unclear and contains errors. All of these made reviewing the CPUE standardization difficult.

In terms of inaccuracies and incompleteness in the documentation and errors in the calculations, some basic examples of these are:

- It was not clear why month was excluded in the factors included in the final model standardized model for the IFOP CPUE series (i.e. equation given in Annex 7) when it was significant in the analysis of deviance results provided in Table 2 of Annex 7 of the Assessment Report and the Assessment Report notes that month was found to be significant;

- In the standardization of the Chilean CPUE 1989-2006 series[16]; it is not explained which of the sub-assembly of vessel was actually chosen and why;

- Years 1989 and 1990 were included in the standardization of effort when neither year is used in the assessment. No explanation for this is provided. Presumably the data for these years are very limited. The estimates are clearly unrealistically low (possibly reflecting exploratory and learning). If this is the case, they probably should be excluded in the actual analyses;

- No interaction terms were included or appear to have been explored in the standardization of the Chilean post 2006 CPUE series;

- It is not explained why boat and year-area interactions were treated as random effects in the analyses of the Argentinean longline fishery[17] or vessel in the final CPUE index for the index for the Chilean longline fleet since 2006 ;

- If the number of vessels in CRT 1 equals 33 (Table 1 of Annex 7 in the Assessment Report) for the Chilean CPUE standardization is a correct, then there should be 32

[16] Based on the degrees of freedom in Table 2, one can surmise that it was the second assembly that was used.

[17] Based on Figure 13, it is clear that the analysis would not support boat as a fixed effect because of non-overlap in time across vessel. In particular, there was only one vessel that operated in 2012 and 2013 and it was a different vessel in each year. Moreover, neither of these vessels operated in any other year. Thus, there was no overlap from which the model could estimate fixed effects.

degrees of freedom for the boat effect in the analysis of deviance (Table 2 of Annex 7 in the Assessment Report) instead of the values of 31 listed and used to calculations. In any case, one of the two numbers is incorrect;

- There are 22 year terms estimated for the Argentinean longline CPUE series but the analysis of variance table for the standardization of this series has only 20 degrees of freedom for the year effects (Table 5 of Annex 7 in the Assessment Report). It should be 21. The incorrect value was used in the calculation of the means square and F values.

- The number of areas for the Argentinean vessel data as implied by the degrees of freedom (135) in Table 5 of Annex 7 in the Assessment Report seems excessive and inconsistent with Figure 32 in the Assessment Report. This figure suggests that the number should be ~23. The figure of 135 is not simply a typo in the table as it was the value used to calculate the mean square error value.

Most, if not all of these, issues if resolved would probably have little effect on the realized standardize CPUE as calculated. However, they do raise concerns about whether the data were actually correctly coded and inputted and whether the statistical models were run as specified.

Beyond these detailed issues of documentation and computation, there are more fundamental issues and concerns about the CPUE series and their standardizations. In both the Chilean and Argentinean series, a single very high value is estimated near or at the beginning of the time series (i.e. in 1992 for the Chilean series and 1994 for the Argentinean). Such extreme values as an indication of actual changes in abundance seem unrealistic given the overall dynamics of the resource and will undoubtedly be problematical in attempting to fit to them within a stock assessment, particularly for a long-lived species like toothfish. While such points should not be excluded as outliers simply based on their extreme value, they do warrant further investigation as to what in the analyses are generating them and to ensure that they are not they results of the omission of relevant factors in the standardization, of confounding in the data due to incompleteness in the data coverage relative to the statistical model or of ignoring relevant interaction terms.

With respect to the high value in the Chilean series in particular, it is worth noting that such an extreme high value is not seen in the nominal CPUE series (Figure 6 of Annex 7 of the Assessment Report). This is somewhat surprising as it would suggest that vessels were either concentrating in the least productive areas/time periods or that the vessels fishing in this year were among the least efficient or perhaps that the level of spatial and temporal resolution is insufficient. With regards to it being a vessel effect, this seems unlikely as there is a large overlap among the vessels that were included in 1992 and 1993 based on Figure 3 of Annex 7 (i.e. 10 vessels were the same in both years, two vessels that fished in 1992 did not fish in 1993 and 3 new vessels fished in 1993).

The filtering of the raw Chilean data to exclude apparent errors in recording or coding of the data resulted in the exclusion of a large number of records. According to the Assessment Report only 25% of the records remained after the filtering for basic outliers in the data. Filtering for obvious recording and coding errors is appropriate. However, when basic filtering, such as was applied here, results in the exclusion of a high percentage of the actual records, it raises substantial concerns about the validity and accuracy of the data collection and processing system and whether the remaining data can be considered representative of the actual fishing that took place. Given the large percentage of data that were excluded, analyses are needed to understand the main reasons for rejection (e.g. missing data, recording or coding errors), which filter resulted in the major lost of records (e.g. do the records contain valid CPUE information and only incomplete in terms of operational data for CPUE) and to assess the likely representativeness of the remaining records in terms of space, year, vessels, etc. Such analyses are important to assess potential biases in the resulting data and to determine if substantial amount of the data filtered can be recovered or corrected. The high apparent errors in these data does cast doubt on the reliability of the actual data collection/processing system for these log book data and suggest that there is a need to review and improve the procedures involved.

It should be emphasized that the actual information available for standardization are limited. Data are insufficient or simply not available to account for many of the factors that can induce changes in CPUE besides change in abundance. In particular, data for the Argentinean longline fiehery are limited and inadequate, while this series has a strong influence on the assessment, particularly recent trends. No real consideration was given to the relative reliability of the different CPUE series or how their reliability/variances over time either in the analyses or when they were used within the stock assessment. All of the CPUE series were considered of equal reliability. Within each series, no temporal differences in reliability were considered when fitting the stock assessment model. Note that the Assessment Report provides figures showing estimates of the confidence intervals for the standardized CPUE series, these estimates are undoubtedly unrealistically smalls. This is because of the unrealistically high degrees of freedom in the standardization of catch rates, which assumes that each individual set is a representative, independent observation (e.g. random sample) of the area and time from which it came. In fact, there is both a high co-variance between operations and a high degree of selectivity (i.e. non-representativeness) in the detailed choice of location, area, depth, time period, etc. fished based on the skill and knowledge of the captain. There is no straight forward resolution to this problem.

Interpretation and standardization of CPUE data as indices of abundance for this fishery is problematical and confounded by the segregation that exists by size/age with area and depth in this species irrespective of the additional complication of the effects of marine mammals and change to cachlotera gear (see below). Moreover, no data are available (or at least were utilized) on factors that would be expected to affect increases in catchability over time (e.g. GPS, plotter, learning, etc). Spatial/depth segregation means that substantial changes in nominal CPUE can occur simply due to changes in the distribution of effort (e.g. a change to shallower depth would yield increases in catch rates in numbers, not necessarily in weight, simply due to the fact that younger ages are more abundant). Including depth and area within the statistical standardization is intended to account for such area and depth segregation. However, as different age/size components of the population would not be expected to change in equal proportions over time (e.g. due the recruitment of a large year-class), significant year-depth and year-area interactions would be expected. Thus, not unsurprisingly when year-area interactions terms[18] were included in the statistical standardization they were found to be significant[19]. (The statistical analyses did not consider possible year-depth interactions and any such interaction may be highly confounded with year-space interactions. Nevertheless, this should have been explored). Perhaps more important are that the relevant

[18] Note in the statistical standard analyses different areas are referred to as fisheries.

[19] No year-area interaction terms were included or apparently explored in the CPUE analyses of the industrial Chilean longline fishery data since 2006. It is not clear why.

spatial, depth and temporal factors including localized targeting and depletion are likely to be occurring at scales of resolution finer than are documented in the data or able to be accounted for the in the analyses.

When the data coverage is poor for factors which are important within the standardization of effort, interpretation of year effects can be highly confounded by the incomplete statistical "design" of the data. For example, when year interactions terms exist in statistical standardization models and the design is incomplete (e.g. all areas are not sampled in all years as is the case here), the resulting year effects do not yield straight-forward interpretable measures of abundance because the model has no information to estimate what the CPUE should have been in areas with no data (it essentially uses the average for those cells for which data exist). There is no statistical "solution" to this problem as there is no data for the cells without data. The potential seriousness of this problem depends upon how incomplete the "sampling" design is and the strength of the interaction terms. Similarly, if two main fixed effects are represented in the data by no overlap at some levels, then these two effects will be completely confounded in the area of the data where there is no overlap (e.g. if all observations in a year come from a single fishing vessel which did not fish in any other year, then there is no statistical way to produce separate estimates for this boat and, more importantly for CPUE, this year). Both of these situations appear to occur within the Argentinean CPUE data but are not fully addressed in the standardization within the Assessment Report (see below). In undertaking standardization of CPUE data, it is critical that a thorough exploration of the data, potential factors and alternative is undertaken to ensure that the results are robust and not confounded by missing data or factors not included.

The most appropriate measure of effort also needs to be considered (e.g. set, number of hooks, hooks-soak time and which of these might best be considered as effects or off-sets). Finally it is critical that sufficient and accurate documentation of these is provided so that the reader is able to understand and be convinced about the robustness and completeness of the results.

Within the stock assessment, the size/age segregation with area and depth also induces confounding and conflict between the standardized CPUE indices with assumptions and estimation of selectivities. The selectivities estimated within the stock assessment are a combination of gear selective and local availability of differing size/age fish. Since the actual gear used has been relatively standard (with the exception of the switch from Spanish longline to the protection bell system), most of the selectivity changes in the Chilean fishery would be expected to arise from availability changes due to differing size/age composition found in different area/depth strata. However, the standardization of the CPUE data is meant to account for such availability effects. For example, if there is a shift to more shallow and northerly areas (as has occurred in the most recent year's in the Chilean fishery), then there would be expected to be a proportional increase in the number of small fish caught as well as in increase in total numbers caught if the population was stable (i.e. the standardized CPUE would be constant in number and appropriate compensated for in weight). If selectivities are assumed fixed with such a change, then the model would need to estimate an increase in recruitment in order to match the observed change in the size/age data. However, this would be in conflict with the observed no change in the standardized CPUE. The model would attempt to find a balance between these two depending upon the relative weight given to the catch-at-age/size data and CPUE indices within the model and the persistent of the selectivity changes and abundance trends.

Cachalotera-Marine Mammal Interactions

The effect on total catches, catch rates and standardized estimates of CPUE as the result of depredation by marine mammals of catches from longlines while being hauled and the subsequent switch to Cachalotera is a fundamental and key-uncertainty in the stock assessment. There is abundant evidence for interaction and depredation by sperm and killer whales (e.g. see discussion of this in the Assessment Report and background documents provided for the review process - Appendix 2). However, insufficient data are available for quantification of the magnitude of the effect. More importantly in terms of the estimation of CPUE as indices of abundance, there appears to be little or no data for estimating trends over time and space. Interactions with marine mammals became apparent soon after longlining for toothfish started in the Southern Ocean and mostly likely occurred early in the development of the Chilean and Argentinean longline fisheries (e.g. data recording such interactions obtained by scientific observers are available since the early 2000's (Hucke-Gaete et al, 2004)). Although depredation rates for individual hauls can be as high as 100%, statistical analyses that have attempted to estimate the overall effect (including results presented in the Assessment Report) suggest that it is relatively small (e.g. the analyses in Annex 7 of the Assessment Report).

However, there a number of factors which confound attempts to quantify the magnitude of the effect. Observations of the actual number of fish affected are dependent upon depredation being incomplete (e.g. lips, heads or damaged bodies remaining on the line). If substantial numbers of fish are totally removed from the hooks, observed numbers would be substantially underestimated. Alternatively, comparisons of catch rates with and without marine mammals observed in the vicinity of the longline are confounded by all the other factors that affect catch rates and potentially by factors affecting the marine mammal behaviour, which may be difficult to account for when undertaking the comparisons (e.g. time of haul, presences of other food sources or the possibility marine mammals may be more attracted and likely to feed on lines with high density of catches, etc). Estimation and interpretation of the effects of marine mammals is also confounded by the behaviour of fishermen in response to the depredation. The most obvious of these is the development and shift from traditional longline gear to cachalotera or trotline gear as described in the assessment. However, as reported by industry during the workshop, there are also other behavioural responses that fishermen use to avoid and reduce depredation. These include delaying hauling when marine mammals are present[20] and moving fishing grounds. This latter potentially may have a large effect on overall catch rates as it has been suggested that areas of highest toothfish densities may be associated with highest depredation rates. As such avoidance of such areas could artificially decrease catch rates as a measure of abundance. It may also affect selectivities as vessels may move to areas or depths with different size/age distributions of fish. During the workshop, this was suggested by Industry participants as the explanation for the increased fishing effort in the north and the increase in the relative proportion of smaller fish in the catch in the most recent year.

Within the current assessment, the effects of marine mammal depredation are not directly accounted for. The implicit assumption is that it has been negligible and non-changing over time[21]. As noted there are few direct observations with which to estimate depredation effects

[20] This has implications for the use of soak time within the measure of effort which has been employed that should be explored.

[21] The switch to cachalotera gear is dealt with in the standardization of CPUE in terms of catchability but not in terms of depredation effects (see subsequent paragraph).

or trends in time or space and those which do exist would suggest that they may be small. As such, the implicit assumption is not unreasonable. However, there is a large amount of anecdotal reports by Industry of the substantial loses of catch due to marine mammals. This combined with the effort to develop cachalotera as a means to prevent depredation and its rapid adoption by the fleet does suggests that depredation could in fact be substantial in terms of overall catch rates and total catches[22]. Of particular concern for the assessment is the possibility that cachalotera gear when introduced was initially effective in preventing depredation but that with time its effectiveness has diminished as the result of learning by killer whales as reported during the workshop. This could be a possible explanation (or partial explanation) for the increase seen in the Argentinean CPUE in 2007 and 2008 and also the steep decline after 2008 seen in the Chilean CPUE series based on cachalotera gear. This latter is supported to some extent by the GLM analysis in Annex 7 of the Assessment Report. The analysis suggests that the reduction in CPUE due to depredation increased to20-30% in 2012-2013, although overall the depredation effect was not significant overall.

It is difficult to predict how not accounting for depredation, if substantial, may affect the assessment and current perception of the status of the stock. This is because of the conflicting effects. Thus, actual fishing mortality rates (i.e. human induced mortality or total actual catches) would be greater as would catch rates (CPUE). As such, depredation remains a substantial source of unaccounted uncertainty in the current assessment. There is no simple or totally satisfactory solution for dealing with this. Nevertheless, it would be worth considering some alternative hypothesis (e.g. for magnitude and trends) in the context of an operating model (see below) for testing the potential consequences of ignoring depredation in the stock assessment model.

Cachalotera-standardization of Chilean CPUE

As noted in the Assessment Report, the number of hooks per set used as a unit of measure with standard longline gear is not directly comparable with the number of hooks used in a set with cachalotera gear. This is because of the clustering of hooks within one bell or cachalotera such that the hooks are clearly not independent[23]. Information presented at the workshop based on paired trials of traditional and cachalotera longline gear showed that for the same number of hooks that the cachalotera gear was more efficient (i.e. had higher catch rates). As such, simply using the number of hooks set as a measure of effort without accounting for whether the gear used was standard or cachalotera gear would be inappropriate as the CPUE with cachalotera gear would yield over estimates relative to standard gear.

For the Chilean longline fishery, the Assessment Report dealt with this problem by creating two separate, non-overlapping temporal CPUE series – one based on only sets with traditional gear and the other based on sets with cachalotera gear. As the transition between the two gears took place rapidly over a two year period, the Assessment Report broke the series from the beginning of 2007. While this approach adequately accounts for the non-direct comparability of the two gears, it does result in a loss of information for the standardization (e.g. substantially more data exists for estimating effects such as space,

[22] It is not clear to what extent the rapid adoption was due to its effectiveness in preventing depredation or to the overall improved catch obtained with cachalotera gear even in the absence of marine mammals (see below).

[23] Even within traditional longline gear, hooks within a longline are not independent. Factors such as number of hooks per set, hook spacing, number of hooks between weight and saturation are factors that introduce non-independence but are not considered in standardizing CPUE.

season and boat in the combined data series than in each series separately[24]). More importantly, this approach results in a discontinuity in the time series of relative abundance indices. This allows the assessment model to adjust the catchability coefficient (q) for the two series to best fit other components within model, irrespective of the relative efficiency of the two gears. In the current assessment, this is of particular concern as discontinuity occurs at the same period in which the switch from scale to otolith aging occurred. In addition, it is in this period when the Argentinean CPUE series shows an increase in its otherwise continuous decline. As such the assessment is likely to attempt to consider this as a true increase in abundance in the absence of other information.

An alternative to breaking the Chilean CPUE into two independently estimated series would be to use data from the full time series but allowing for the effect of the change in gear to be estimated as a factor in the model (e.g. using set as a unit of effort with type of lonline gear as a factor as well as number of hooks per set and number of hooks per cachalotera). There are two years of overlap when both gears were being used by some components of the fleet as well as the data from the paired experiments which could be used to inform the model about the relative efficiency of the two types of gears. While it is not possible in advance to know for sure whether there is sufficient overlap in the data to allow for such a joint analysis, it should be investigated as a more preferable alternative to having two discontinuous series. In investigating this alternative, it would be worth considering catch as the independent variable with number of sets and other effort factors as an off-set (e.g. number of hooks, soak time). This may yield better performance.

Argentinean CPUE

The standardized Argentinean longline CPUE series is given equal weight in the assessment as that of the Chilean CPUE. It is the only available continuous relative abundance series[25]. As such, it is an important and influential component in the assessment. The documentation for this series and its standardization is very sparse and incomplete within the Assessment Report. Time available in the Workshop for review and discussion of this index were limited. Consequently, a number of outstanding concerns and issues exist about this series that were unable to be adequately reviewed and addressed. These include:

- The Assessment Report states that it is “only possible to count the number of trips as an effort measure” and similar statements were made at the workshop. Trip as a unit of measure is problematical as the length of a trip is likely to vary in response to actual catches (e.g. vessels will tend to compensate for lower catch rates by extending the length of trip so as to maintain a more stable catch per trip). However, it is actually not clear after detailed consideration of Annex 7 whether in fact trip or sets was the actual unit of effort. The standardized catch rates seem too low to represent catch rates per trip as are the number of observations . These are more consistent with what might be expected if catch rates were in terms of sets. Additionally, the catch rates used in the Argentinean stock assessment (Hanchet et al 2014) based on data from the same data collections are reported in terms of catch per set.

[24] It would be worth examining the estimates of the fixed effects between the two analyses for consistency.
[25] Within the model as estimated in the Assessment Report, this continuity is lost by allowing catchability to change in 2006.

- Area-year interactions are treated as a random effect. Treating area-year interactions may be acceptable if in fact coverage across years is substantially broad and the interaction is relatively small (i.e. it can be considered as noise in terms of the larger fixed effects – particularly year). However, it seems likely that temporal/spatial coverage may be very incomplete particularly in the most recent year when data was only available from a single vessel. In this case, treating year-area interactions as a random effect is problematical as the areas fished are neither likely to be representative nor random samples of the overall areas (see discussion above). Large temporal trends in areas fished may occur due to concentration of the resource into the most favourable habitats or localized depletion effects.

- The vessel composition of the fleet (or at least as represented in the data) has changed dramatically over time and coverage is in recent years is very incomplete. Thus, only a different single vessel fished in 2012 and 2013 and neither of these vessels fished in any other year. If boat were treated as fixed effect then the model would not be able to estimate a year effect. Similarly, between 2008 and 2011, only a single vessel provided data. This vessel did overlap a single other vessel in 2007. As such, boat and year effects would be separable but the estimates for 2008-2011 would be totally dependent on the reliability which the effect or relative efficiency of the boat that fished in these years could be estimated from the single year of overlap in the data. In the standardization, boats were treated as random effects (like year-area interactions). This avoids the year effect from not being estimable[26] but begs the question of the extent to which the recent trends in these data could be boat and not abundance effects.

- Treating boats as random effects in this situation confounds interpretation of the year effect as it assumes that there is only random noise induced by these effect and that are no systematic temporal or spatial effects (e.g. only the most efficient boat are able to economically survive when catch rates decline). Of particular concern is that the catch rates in the last two years may be simply reflect a boat effect and, similarly, the general increase in catch rates that occurred around 2007. Given that the estimate of the standard deviation for boats when treated as a random effect is 0.70, this would imply that the CV for the abundance index in the last two years is on the order of 2 or greater (i.e. the estimated index was 0.20 and 0.38 respectively and the estimates are derived from a single vessel). In essence, there is very little information content relative to changes in abundance.

- The data used in this series is derived from “fishing reports recorded by the Secretary of Agriculture, Fishing and Aquaculture”. No information is provided on how these data were collected, their coverage or reliability. Coverage appears to be incomplete[27] at least in recent years. Incomplete coverage is a concern as there may be substantial non-representativeness in terms of the overall fleet (e.g. effects of management regulations on cooperation and willingness of vessels to report).

[26] The reason for this is not discussed.

[27] The table for the results of the standardization of the Argentinean observer data (Table 6) has only 1 degree of freedom listed for boat (i.e. two boats) but covers all year between 2003 and 2013. However, based on Figure 13, data from four vessels would be required in order to have data for every year if the vessel reported data were complete.

- Neither the analysis nor the discussion of the Argentinean CPUE data considers the effect of the switch to cachalotera gear within this fishery. Information in Hanchet et al 2014, not provided as part of the review, reports that the Argentinean fleet also employs cachalotera gear. However, not all of the fleet was using this gear at least through 2011 (Hanchet et al). Given the concern about the mixture of regular and cachalotera gear for the Chilean longline CPUE, this lack of any consideration of the effect of cachalotera gear in the Argentinean fishery is a substantive omission. It raises concerns about the reliability of the standardization and comparability of the time series relative abundance indices within this time series (e.g. is the increase in the series between 2006 and 2010 in part due to the use of cachalotera gear).

Catches

The Assessment Report contains two annexes dealing with the estimation of total catches for the Chilean fisheries (Annexes 3 and 4). The first deals with catches between 1984 and 2001 for both the artisanal and industrial fleets. The second deals only with artisanal catches between 2004 and 2012 and is focused on the fraction of the catches that occurred north or south of 47°S (this is the line of division for the stock used in the Assessment Report). Both annexes result in revised estimates to the historic official catch series.

Annex 3 reviews historical landing data for the period prior to 2002. It suggests that there are a number of substantial issues with respect to total landings in terms of miss reporting of location of catches (i.e. in international waters and north/south of 47°S - in or out of the tendered area) and potential under-estimation of the industrial catches. It provides a revised set of landing data for the industrial fleet for the period 1989 to 2001 but exactly how this was derived is not clearly documented nor how it compares to the official reported statistics. In any case, it does suggest that there is considerable uncertainty about the overall catch levels. For example, this Annex provides estimates of the historic landings in international water that needs to be reassigned to Chilean austral south catches (Box 6 in the Annex) for years 1991 through 1995[28]. For the years 1992-1994, the values in this Box exceed 50% of the total corrected estimated catch for the Chilean fleet used in the assessment (i.e. the values in Box 6 relative to Table 2 of Annex 3 in the Assessment Report). It is not clear whether after 1995 miss reporting of catches as in international water is not an issue or the data do not exist.

Annex 4 deals with the question of miss reporting the location of catches by the artisanal fleet in the period from 2004 to 2012 due to regulations making it illegal for them to fish south of 47°S. This annex contains several alternative plausible catch series based on different interpretation of the available data and concluded:

> "Consequently, according to expert judgment, it is suggested to employ information from Table 12 as reference to explore landing scenarios located southwards parallel 47° S and first, to especially consider the information provided in the column that shows ≥20 dfp (dop) and ≥4 t/trip criterion; secondly, to consider the information published in the column that shows ≥20 dfp (dop) and ≥7 t/trip criterion; criterion describing the combination of day out of port and landing in t per trip covers in a

[28] Note that he values in Box 5 (which are not discussed in the text of the Annex) for estimated total landings in Region X-X11 and international waters are not totally consistent with the estimates in Box 6 (i.e. the estimates in Box 6 for the years 1992-1994 exceed those in Box 5 although these should be a sub-set of the values in Box 5 based on the titles for these tables).

better way landings of artisanal fishing boats that performed fishing operations in tendered area. (Annex 4, Assessment Report)."

A copy of Table 12 from Annex 4 in the Assessment Report is provided here as Table 1. The Assessment Report states in relationship to the results in Annex 4 that:

> "Optionally, input data to the model provides with the possibility to study scenarios with different "correction" degrees, to employ criteria to approximate catch magnitudes of fishing volumes that were extracted from the industrial fishing area but were declared as catch coming from the artisanal fishing area located north to 47˚ S latitude (Annex 4)."

However, the stock assessment results provided in the Assessment Report are bases on a single catch series. The Report does not in fact discuss sensitivity to possible alternatives or provide any results for any alternative catch series. In addition, it does not document which catch series was actually utilized. It is only from examination of the actual computer code that one can determine the series which was used. The input data file provided to the reviewer for running the assessment models contains four alternative series for the total Chilean landings south of 47° (these are provided here as Table 2). Hardwired into the code is that the first of these series is used to fit the model. Unless one modifies the actual computer code or the input data file there was no possibly of considering the alternative series. More significantly, the values for the total landings in the data input file for the years 2004-2012 do not correspond to any of the values in Table 12 of Annex 4, although Annex 4 is cited for the source of the landing data for these years (i.e. compare Tables 1and 2). The reviewer could not ascertain from where the values used in the input file for fitting the assessment model came based on the available documentation. In addition, the differences between the actual catch series used in the assessment and the most recommended catch series in Annex 4 are quite large in several years (i.e. over 50% - see Figure 6). This discrepancy as well as sensitivity to the alternate series needs to be resolved.

Under-reporting and miss reporting of catch location is commonly founded in quota managed fishery unless there is extensive verification of catches and location of fishing operations (e.g. high observer coverage, VMS reporting). Based on the information in Annexes 3 and 4, this seems to have occurred in the Chilean toothfish fisheries and the amount of miss reporting appears substantial. While it is encouraging that the assessment does allow for miss reporting of catches, the reviewer is concerned that the evaluation that have been done are not complete (e.g. there is no consideration of miss reporting for the years 2002 and 2003 and no consideration of miss reporting by the industrial fleet after 2001)[29]. Given the estimated magnitude of the miss reporting, there is likely to be large uncertainties associated with the actual values used in the assessment (this certainly is the case for the estimates in Annex 4 which only deals with one component of possible miss reporting). Alternative catch series reflecting the full range uncertainty should be developed and the sensitivity of assessment results to such uncertainty needs to be evaluated.

In terms of the Argentinean catches, no information is provided on their reliability or potential problems of miss reporting. However, similar problems of under and miss reporting

[29] Perhaps the introduction of the catch documentation scheme by CCAMR since 2000 has eliminated this as an issue. If this is in fact considered the case, some discussion of this should have been included in the Assessment Report.

would not be unexpected. This is an issue that should be pursued in collaboration with Argentinean researchers.

Finally in terms of the total catch estimates used in the assessment, the reviewer calculated estimates of the implicit annual total catches based on the catch-at-age matrices and average weight-at-age used in the assessment. Comparison of these implicit estimates and those inputted into the assessment reveals very large discrepancies between these in some years (Figure 7). For the Argentinean landings, 3 of the annual estimates based on the catch-at-age data for the longline fishery and for the trawl fishery are over 200% (i.e. over double) that of the reported catches and in one year in each series they are exceed it by a factor of 4. Unless there has been a coding error in these data in the input data files used for the stock assessment, such massive differences are indicative of either large under-reporting or serious problems in the sampling/compilation of the catch-at-size/catch-at-age data[30]. For the Chilean data, the estimates of the total catch in the years prior to 2007 are either below or around the values inputted for the total catches. That they are below is not surprising because of the biases in the aging reading in these years due to basing the aging on scale readings (see above). For the three years beginning in 2007, the estimates of the total catch derived from the catch-at-age are 60% greater than the inputted corrected “official” values and subsequently are ~20% or greater. These discrepancy may be indicative of underreporting – particularly in the years 2007-2009, but alternatively the may stem from possible problems with the mean weight-at-age vector used in the assessment (see above). In any case, these large discrepancies and inconsistencies are a major concern. The source of the difference needs to be investigated, understood and resolved.

[30] Such large discrepancies appear inconsistent with text in the assessment report which states “Catch-at-age data of Argentinean trawl and long line fishing fleet obtained south to 54˚S latitude during 2003 to 2012, were provided by INIDEP; they correspond to annual compositions in number per length class that are afterwards extended to total catch and transformed to ages using growth parameters from von Bertalanffy´s equation.”

Table 1: Official corrected landing (t) and corrected in the official tendered area (south of 47° S) and north of that obtained for the 2004-2012 period through application of both criteria developed in Annex 4 of the Assessment Report. Source SERNAP. (Copy of Table 12 in Annex 4 of the Assessment Report)

Landing

	Oficial (SERNAP)		Corrected 1 port		Corrected 2 ≥ 20 dfp y ≥ 4 t		Corrected 2 ≥ 20 dfp y ≥ 7 t	
Years	NORTH 47°	Tendered area (south of 47° S)	NORTH 47°	Tendered area (south of 47° S)	NORTH 47°	Tendered area (south of 47° S)	NORTH 47°	Tendered area (south of 47° S)
2004	3.419	1.651	1.759	3.311	2.636	2.434	3.083	1.987
2005	2.236	1.809	1.546	2.499	1.938	2.107	2.177	1.867
2006	2.091	2.455	1.143	3.403	1.423	3.123	1.708	2.838
2007	2.090	2.358	1.061	3.387	1.400	3.048	1.763	2.685
2008	1.558	2.883	806	3.635	1.087	3.354	1.421	3.020
2009	1.681	3.018	896	3.803	1.238	3.461	1.541	3.158
2010	1.467	3.293	669	4.091	1.014	3.747	1.306	3.454
2011	2.189	2.298	1.099	3.389	1.585	2.902	1.972	2.516
2012	2.069	1.934	983	3.020	1.471	2.532	1.814	2.189

Table 2: The four catch series for the Chilean landings south of 47°S that were included in the input file used in fitting the assessment model. Note that catch series 1 was used in the estimation of the assessment results in the Assessment Report and in all the alternative runs performed during the Review Process.

Year	Series 1	Series 2	Series 3	Series 4
2004	2079	3739	2862	2415
2005	1991	2681	2289	2050
2006	2004	2952	2672	2387
2007	1974	3003	2664	2301
2008	2154	2906	2625	2291
2009	2345	3130	2788	2485
2010	2988	3786	3441	3149
2011	2298	3388	2902	2515
2012	2382	3468	2980	2637

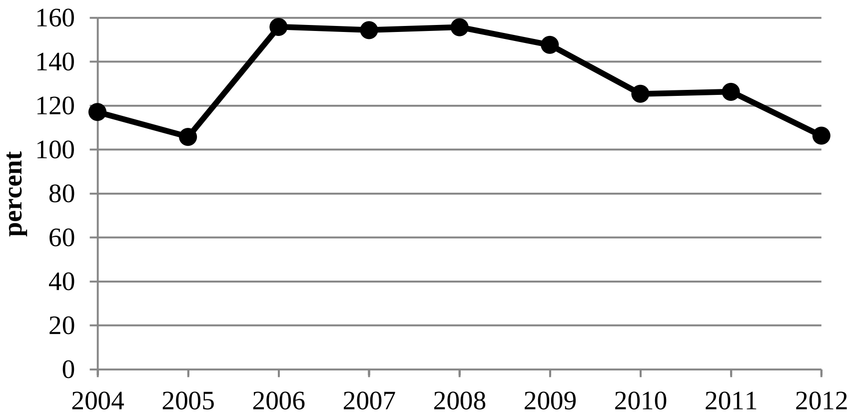


Figure 6: The preferred or recommended corrected total catch estimates for the Chilean Industrial fishery for the years 2004-2012 (i.e. the column labelled ≥20 dfp (dop) and ≥7 t/trip criterion in Table 2 above) as a percent of the actual total catch figures used in fitting the stock assessment.

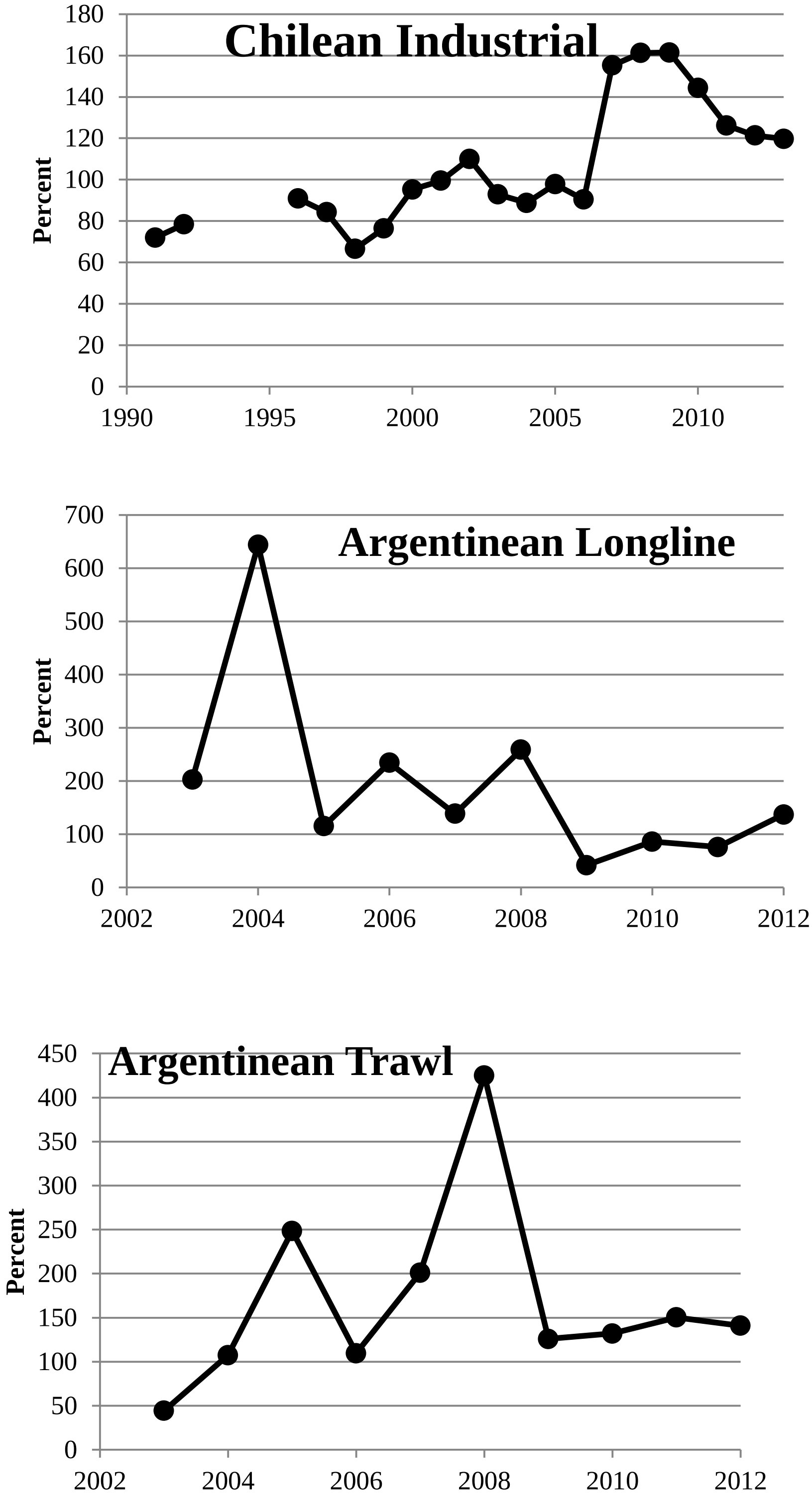


Figure 7: Comparison of the "official" estimates of the total catch in tonnes (i.e. those used in fitting the stock assessment model) and estimates of the total catch calculated from the product of the numbers in the catch-at-age matrix times the mean weight-at-age vector used in the assessment and then summed by year (see Appendix 6). Shown in the graphs are the estimated total catches derived from the catch-at-age matrix as a percentage of the official catches.

4. Base Case and Uncertainty

Estimating the amount of uncertainty associated with stock assessment results is difficult and challenging. Nevertheless, estimating uncertainty is essential for providing a measure of risk[31] associated with any advice. There are two basic forms of uncertainty in stock assessment results: estimation and model uncertainty. Estimation uncertainty is the uncertainty associated with output parameters estimates from a model conditional on the specific model and input data. These inputs include priors, penalties, assumed fixed parameters and the estimated or assumed variances associated with the observed data. A change in any of these constitutes a different model which will yield different estimates for the output parameters and their estimated variances. Uncertainty associated with the model structure including priors and penalties is referred to as model uncertainty. In stock assessments, model uncertainty is generally much greater than the estimation uncertainty.

While often complex, there are well developed objective methods for obtaining uncertainty estimates (i.e. variances and co-variances) associated with the output parameter estimates for a particular model run. The estimation of model uncertainty is a much more complex problem. In some cases, statistical methods can be used (e.g. the uncertainty associated between a linear and a curve-linear model), but in most instances there is an element of expert judgement (i.e. subjectivity) involved in determining which alternative models to consider/evaluate and the relative weight to assign to different models. Based on my experience and despite these difficulties, it is essential that careful consideration of model uncertainty be undertaken and incorporated into the overall assessment results and management advice in order to provide any reasonable measure of the overall uncertainty and risk.

The Assessment Report contains basically no alternative case studies. Consequently, neither what are the important "axes" of uncertainty nor the general question of model uncertainty was not even considered, much less represented in the report. As such, the main axes are not represented in the Assessment Report. The reviewer considers this a major deficiency.

The Assessment Report does contain very limited results for an update of the 2013 assessment with "updated" data. The Assessment Report provides no documentation of the actual model, its conceptual structure, the parameters which were estimated, the input data used and priors/penalties and their specific values used in this update of the 2013. It contains only two figures with any relevant results about current stock status (Figure 47 and 48 of the Assessment Report) – both relative to reference points. There is no information on which to evaluate the goodness of fit of this updated model much less to what degree both in terms of model structure and data inputs were in this update comparable to the 2013 assessment (e.g. were the selectivity and catchability blocks the same, was CPUE standardized comparable, was an aging error matrix applied, were the priors and penalties similar, etc). Similarly, there is no meaningful way to evaluate and compare the relative appropriateness of the new assessment model results compared to either the updated model 2013 results or to the results in the 2013 assessment. The Assessment Report makes no attempt to do this and provides no information for the reader to be able to do this. Based on the information provided, the only real conclusion in terms of stock status that can be drawn is that an alternative model with a different conceptual basis for stock structure yields different estimates.

[31] "Risk" is used in the common meaning of probability of making an error and not in the formal decision making context of probability times utility.

Additional Runs

A number of alternative runs of the assessment model were requested to be run by the reviewer near the beginning of the Workshop. Technical problems were detected in the initial results provided to the workshop and a final set of runs was not available until the last day of the workshop. Subsequent to the workshop, additional problems were detected in the runs provided at the workshop. A request for re-running the three runs which specified no changes in catchability[32] was made to the Review Coordinator as well as a query as to why the results from Run 1 provided at the workshop did not match those in the Assessment Report. Run 1 was supposed to be a re-run of the one model results in the Assessment Report (see below for more discussion of this problem). It was run in order to be able to directly compare numerically the results from the alternative runs and also to be able to evaluate in more detail the fit and results of the base case than was possible with the limited documentation provided in the Assessment Report.

In mid-December, the reviewer was provided with an explanation for the discrepancy between the results in the Assessment Report and Run 1 conducted at the Review Workshop. A new set of run results based on the requests made at the workshop was also provided (Appendix 5). As stated in the explanation accompanying these runs (Appendix 5), the results in the Assessment Report had been mistakenly run with the correction for age reading errors from scales turned off. Thus, the “base case” as specified in the Assessment Report did not correspond to the actual numeric results presented. This creates some awkwardness in dealing with this item in the reviewer’s report. It is not clear whether the base case should be considered as the one intended and specified in the Assessment Report or the one for which numerical results were presented and on which management recommendation were based. In one sense, it is not material. Many of the sensitivities and behaviour of the model are independent of whether the aging error matrix was applied or not. Moreover, the reviewer considers that neither the version of the model that was intended to be implemented or the one actually ran constitute an adequate choice within the context of the conceptual model underlying the assessment much less in the broader context of possible conceptual models for the population structure (see above). However, the difference between the intended and realized “base” case do have implications for the status of the stock and the consequences for future catches if either one of them was to be used as a basis for management recommendations (e.g. Figures 8-10).

[32] There was confusion in the specification of these runs because the Chilean CPUE indices before and after 2006 are treated as two independent series in the standardization process and in the input files for the model. However, within the computer code when generating the predicted values of CPUE in fitting the model, they are treated as single series but with a change in the catchability coefficient q between them. Thus, when no changes in q were specified in some of the additional runs, this was intended to be for no change in q within a series. Given that the two Chilean series do not even have a common unit of effort, linking the two series with a common q makes no sense. Thus, the reviewer requested that these three runs be re-calculated with distinct q’s being estimated for the two Chilean CPUE series.

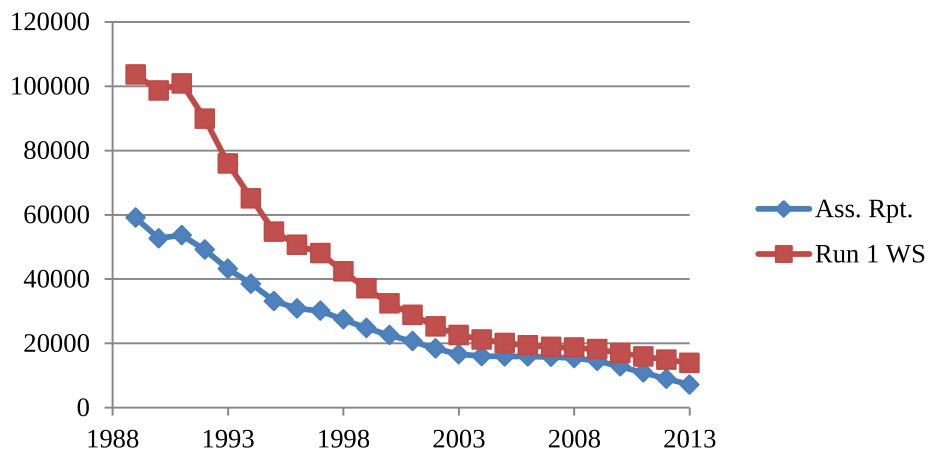


Figure 8. Comparison of estimated spawning biomass trends contained in the Assessment Report (Run 1 Appendix 5) and that intended to be the base case in the Assessment (Run 1 of Appendix 4)..

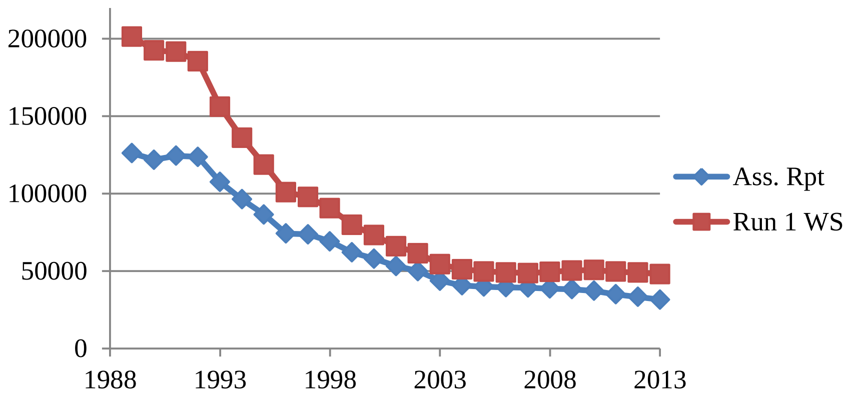


Figure 9. Comparison of estimated total biomass trends contained in the Assessment Report (Run 1 of Appendix 5) and that intended to be the base case in the Assessment (Run 1 of Appendix 4).

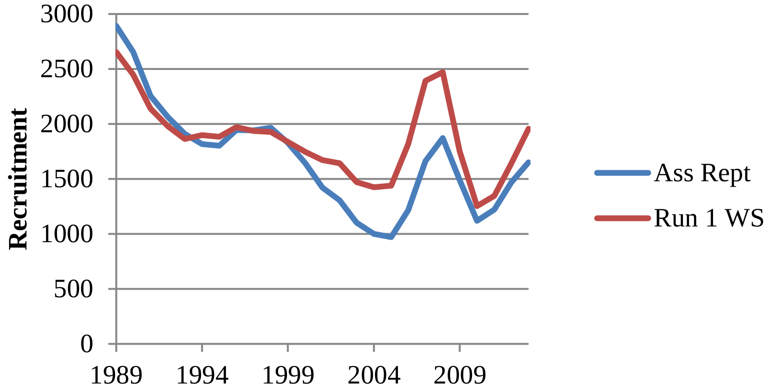


Figure 10. Comparison of estimated recruitment (age 3) trends contained in the Assessment Report (Run 1 of Appendix 5) and that intended to be the base case in the Assessment (Run 1 of Appendix 4).

Base Case Issues

Fundamental to the specification of a base case or a set of base cases is resolving the various data input issues raised above. Without this, it is not clear to what degree features of the model performance are due to these, especially data conflicts seen in fitting the model. In particular, it is critical to ensure consistency (1) between aging across time and fleets, (2) between the mean weight-at-age vector, implicit growth and maturity vectors and (3) between input total catch weights and implicit total catches based on catch-at-age and mean weight-at-age is critical. In addition, it is essential to ensure that the data in the input files corresponds to the actual "intended" data. As such, evaluation of possible detail specifications for the current model structure are premature and any such specifications may not be the most appropriate when the data input issues are resolved. Nevertheless, there are a number of more general concerns in the specification of the base case in the Assessment Report that should be noted. These include:

- The precision assigned to the CPUE and catch-at-data seem overly precise. For CPUE, estimates of variances or CV's from the standardization will almost inevitably yield inappropriately too small values because of the degrees of freedom are highly inflated (see above). Similarly for the catch-at-age data, the appropriate sample sizes cannot be simply determined based on the sample sizes involved in the collection of the size-frequency and age-at-length data. Not only is there unaccounted for aging estimation error in the aging data, the error in fitting the catch-at-age data within the model is a combination of sampling and process error (i.e. the model assumes fixed selectivities within a time period and that no process error occurs). (Note also that in the context of the current model that model generated posterior estimates of the effective sample sizes are not a reliable guide.) The model has a high degree of flexibility in fitting to both types of data. Large changes in model specification appear to have little effect on the actual fit, particularly for the Argentinean catch-at-age data (e.g. in almost all of the runs the resulting value of the objective function for the fit to these data were quite similar even as long as the effective sample sizes were kept constant – see Appendix 5).

- In terms of their absolute values, the specification of the CVs for the CPUE series and the sample sizes for the catch-at-age data impact the resulting estimates of precision associated with model estimates of stock sizes, status relative to reference points and future projections. The estimates in the current Assessment Report appear unrealistically too precise (e.g. Figure 46 and 47 in the Assessment Report which suggest almost negligible uncertainties in the absolute estimates of biomass and their trends overtime).

- Not only is the absolute value of precision assigned to the different input data important but the relative value for the different components is critical. This determines the weight that the model gives to different pieces of information when fitting. This also includes the values assigned to penalty terms. If all of the input data and model structure components are consistent, then changing the weights assigned to different input components will not affect the estimates derived from the model but only the resulting variance estimates. However, if the results are sensitive to changes in relative weighting than this indicates that there are conflicts amongst the different data inputs and/or the underlying model structure. Thus, exploration of alternative weights (i.e. CVs and sample sizes) should be undertaken routinely as part of the evaluation of a stock assessment model of this sort and results from these taken into account when determining a base case. When conflicts do exist, it is important not to simply average across them, one or the other may more likely be correct. Consideration needs to be given to the possibility that either one or the other is correct. This needs to be reflected in the assessment results and their associated uncertainty.

- There exist objective approaches for assigning relative weights to different relative abundance indices and catch-at-age data (e.g. Francis, 2011). Evaluation of these should be undertaken as a possible source of guidance. However, the usefulness of such approaches may be limited for this assessment given the limitation of the existing data.

- In the set of alternative runs (Appendices 4 and 5), the CPUE series and catch-at-age data appear to be in conflict. Thus, decreasing the weight assigned to either all of the CPUE series (Run 14) or only the Argentinean series (Run 4 and 8) substantially affects the results. For example, perception of current stock status for in Runs 14 and 4 relative to the "base case" is improved (e.g. current depletion is about 50% less - ~15% versus 10%). The main conflict appears to be between the CPUE series and the Chilean catch-at-age data (e.g. the overall fit to the Argentinean catch-at-age data remain relatively unchanged while there is a substantial improvement to the fit to the Chilean catch-at-age data, perhaps reflecting the problem in the scale age readings within the Chilean data).

- Applying the aging error correction matrix to the Chilean catch-at-age data (Run 2), both improves the fit to these data and substantially changes the results (e.g. current depletion increases to 15% in contrast to when no aging error correction matrix is applied). This combined with the preceding dot point emphasize the critical need to appropriately account for the biases in the age-reading data from scales. At a minimum, the older ages should be pooled into a plus group when fitting to the Chilean age data based on scale readings (probably at least from age 15 onwards as this is when there appears to be definite biases).

- There appears to be a conflict between the Chilean and Argentinean catch-at-age data. Thus, above age 10, the residuals for the fit to the Chilean data tend to be negative while those for the Argentinean are almost always positive. This is consistent with the different method used to estimate the ages of the catch (Figure 11). Note that time did not permit, the effect of exploring differential weighting between the Chilean and Argentinean catch-at-age data.

- The base case model allows for unconstrained changes in selectivity and catchability in specified years. The basis for these seems arbitrary. Based on information provided at the Review Workshop, they appear to have been selected largely after examination of the residual fits to the CPUE and catch-at-age data. Introducing selectivity and catchability changes can only improve the model fit to the input data. However, without an external and objective basis for these, the model is likely to be over-fitted and can yield spurious and biased results. If the model was a true likelihood model, then changes in the value of the objective function relative to the increased number of parameters might be used as a guide to whether such changes were appropriate based on statistical significance test. If the changes were non-significant, their inclusion would probably not be appropriate. However, the opposite is not valid. The objective function is not a true likelihood function and the input variances (e.g. CVs and samples) are not reliable estimates of the variance associated with the input data.

Introduction of changes in selectivity and catchability should be done conservatively. If changes are allowed for the resulting estimates need to be carefully evaluated in terms of the whether the resulting estimated parameter values are realistic relative to the fishery. In any case, unless there are very convincing reasons for the changes, alternative case studies should be presented and considered in terms of the overall uncertainty in the assessment results.

- In the additional runs that were preformed, the results were sensitive to inclusion or exclusion of the catchability and selectivity changes (e.g. Figure 12 and the tables of run results in Appendix 5)[33]. The Assessment Report did not provide the model estimates for the changes in q. However, the estimates are very substantial (e.g. q changes by a factor of around two) and the changes are in opposite directions for the Argentinean and Chilean longline fisheries (Tables 3 and 4). Thus, the catchability of Chilean longline fleet is estimated to have decreased by around half between 1997 and 1998[34] while the Argentinean longline fleet increased it catchability by over a factor of two between 2006 and 2007. The Assessment Report not did not present these estimate much less discuss their plausibility[35].

  The estimated selectivity changes were also large – particularly for the Argentinean trawl and Chilean longline fleets. The former does appear to be a dominant feature of the available data as discussed in the Assessment Report (e.g. Figure 39 in the Assessment Report) but the underlying factors behind this are not clear[36]. For the Chilean longline fishery, the Report suggests that the change may reflect the change in age reading being based on scales to otoliths. In which case, including the change as real not surprisingly results in an improved fit but the appropriateness of the results is problematical. An alternative to explore would be to estimate selectivity only based on the aging data since 2007 or using the pre-2007 but with a younger plus group.

---

[33] Note that while the SSB trends are quite similar, the estimates of current depletion vary by 50% (i.e. 0.10 versus 0.15) this is because the estimates of B_o differ reflecting the differences in the estimates of recruitment.

[34] Note that comparison of the last change in q values for the Chilean fleet is meaningless as the effort measures for this last series is on a different scale and the series was standardized independently (see above).

[35] Possibly the change in the Argentinean q could be due at least in part to the change in vessel upon which the estimated CPUE series was based after 2007 (see CPUE section above for discussion of boat effect in the standardization of the Argentinean CPUE series) but this would imply the vessel which began operating in 2006 was very substantially more efficient than previous vessels. In the Chilean CPUE perhaps marine mammal interactions may have resulted in a decrease in q. However, the scale of the change appears to be beyond the range based on observed data. If in fact this is the cause, it would have substantive implications for the overall catches which would also need to be addressed and included in the model.

[36] The Assessment Report does state that there was a change in Argentinean management regulations since 2000 which imposed a minimum size. Perhaps this could be a cause, but the timing of the regulation is not consistent with the timing of the estimated change. Additional, size regulations such as this, particularly in trawl fisheries, often result in discarding rather than real changes in selectivity. Additional information is needed.

- The procedure used for estimating initial numbers suggests substantial disequilibrium when the fishery began. Thus, the spawning stock biomass is estimated to be around 20% below B_o when the fishery began in the base case (Run 1 in Appendix 5). This is because the earliest estimates of recruitment that go into the estimates of N_o are below the mean value for R_o and it is these year classes that form the bulk of the spawning biomass when the fishery begins. Additionally, the estimated recruitment trends prior to and at the beginning of the fishery indicate a period of very high recruitments (well above R_0) followed by a period of declining and low recruitments (Figures 13 and 14). The implication, if this were an accurate reflection, would be that the dynamics of the stock (at least in the first 10 years of the fishery) are governed by temporal trends in recruitment independent of effects of the fishery and that the fishery began just coincidentally after a period of high but falling recruitments. These estimates of the pre-exploitation recruitment and the implicit spawning biomass dynamics raise questions about the validity of using the estimates of R_o as a basis for defining MSY reference points. However, these dynamics more likely are a consequence of the way recruitment has been modelled (see next dot point) and the procedure used to estimate N_o than the actual recruitment trends. The sensitivity to alternative specification for estimating the initially numbers and defining R_0 should be considered as they may have substantial effects, particularly with respect to stock status indicators.

- Recruitment is modelled in the base case and all of the alternative runs as being a random log-normal variable with a large CV (0.60) and as being independent of spawning biomass. This is a common approach. It avoids problems of indeterminacy often found when attempting to fit a stock recruitment internally within an assessment model (particularly the degree of compensation or steepness in the context of a Beverton and Holt relationship). The resulting estimates of recruitment and stock sizes are usually similar. Modelling recruitment as constant but with high variances avoids confounding and convergence problems in the estimation of the stock-recruitment curve parameters and the other parameters. Nevertheless, in the base case run, recruitment show remarkably little inter-annual variability (e.g. have a large temporal co-variance).

  Moreover, they also suggest a possibly defined well stock-recruitment relationship. Thus, for the base case when estimates of recruitment are plotted as a function of the estimated spawning biomass (appropriately lagged), there appears a reasonable relationship with the exception of estimates for two years. In these two years the estimates of recruitment are too large relative to the spawning stock biomass (Figure 15 bottom panel). However, it is around these two years when the switch to cachalotera gear occurred, moreover, the relative magnitude of these two estimates is sensitive to whether selectivity changes are included in the model (Figure 15). Thus, without selectivity changes, these two years of high recruitment disappear and the results exhibit a remarkably strong spawning stock biomass and recruitment relationship. This further emphasizes concerns above about allowing for selectivity changes only to achieve improved fits. Moreover, projection results (see below), estimates of initial numbers and reference points based on R_0 are likely to be sensitive whether and at what point in the assessment a stock recruitment relationship is included (e.g. within fitting the historical data and/or for projection purposes). As such, the issue of the most appropriate way to model recruitment for a base case needs to be evaluated and addressed. Particularly given that the base case estimates indicate a highly depleted stock[37].

- Related to the above dot point, there can be a problem in modelling recruitment as constant and using the mean to estimate pre-fishery recruitments for estimating initial numbers if there has been a substantial, continuing decline in recruitment during the period of exploitation. Since the recruitment deviations have to equal zero when fitting the model, the recruitment levels when exploitation begins will almost inevitably be estimated above the mean (R_0) to compensate for the subsequent decline.

- Age specific selectivities were modelled using a double normal function, which will allow for dome shape selectivity curves to be estimated. However, input constraints on parameter estimates meant that in fact functionally selectivities were modelled as a half normal function and dome shape selectivities were not considered. Assessments in many cases are sensitive to whether selectivities are modelled as asymptotic or domed. Domed selectivities may results in estimates of a "refuge" of old/large spawners that are basically invulnerable to fishing, which can have important implication for estimates of depletion and projection results if a stock recruitment curve relationship is used. Generally, the main reason for considering dome shaped selectivity is a consistent overestimation of the predicted number of older age-fish relative to the number actually observed. In such cases, consideration of dome shape selectivities is warranted but needs to be done with caution as a lack of older fish in the catch can be confounded with inappropriate estimates of M and more important aging issues (aging of oldest animals is generally the most difficult).

  Within the base case assessment, there are lack-of-fit issues in terms of the absences of older fish in Chilean longline catch-at-age data. These, however, are highly confounded with the age-reading problems from scales. As such, invoking dome-shaped selectivity would be inappropriate here until the aging problems are resolved.

  For the Argentinean trawl fishery, there are valid external *apriori* reasons for expecting a dome shape selectivity function. This fishery operates in relatively shallow waters relative to the longline fisheries and toothfish are known to segregated by size/age with depth. One of the alternative runs considered (Run 10) allowed for the selectivties for the Argentinean trawl fishery to be dome shape (i.e. it allowed the right hand normal parameter terms to be free in the model). This run resulted in estimating a dome shape selectivity function for the second selectivity block but surprisingly not for the first block. The resulting estimates of trends in stock sizes and recruitment were quite similar (Figure 16)[38]. However, it appears that the model did not converge to the actual minimum in the case of the parameter estimates for the block 1 selectivities. This is because the first of right hand normal parameter went negative and the second became extremely large. In addition, all of

[37] It is not clear whether in fact some exploration of the effects of estimating recruitment based on a Beverton and Holt relationship had in fact been undertaken as the code for this is included assessment program. However, if these had been carried out, there is no discussion of this within the Assessment Report.
[38] Note there is a tabling error in Table 3 of the Report for the Runs done in December (i.e. Appendix 5 of this Review Report) for the SSB_{2013} estimate for Run 10. The figure given is for Run 1 and not Run 10. The actual figure should be 6563instead of the 7171 figure provided in the table (about 8% less).

the residuals for the older ages still remained negative for this block of years. Thus, it is not possible to conclude that a dome shaped selectivity function for the Argentinean trawl fishery would or would not have substantial effects on recruitment or biomass estimates. Checking the output values of added parameters for these kinds of problems should be part of the standard procedure when alternative model runs are performed.

Table 3: Estimates of the catchability parameter ($x10^5$) for the Chilean and Argentinean longline CPUE for the different catchability blocks in the intended base case model run (Run 2 of Appendix 5) based on results for this run provided at the workshop (ageing error correction).

Fishery	q block 1	q block 2	q block 3
Chilean	0.0482	0.0235	0.6440
Argentinean	0.7306	1.4325	

Table 4: Estimates of the catchability parameter ($x10^5$) for the Chilean and Argentinean longline CPUE for the different catchability blocks in the actual results in the Assessment Report (Run 1 of Appendix 5) based on results for this run provided at the workshop (no ageing error correction).

Fishery	q block 1	q block 2	q block 3
Chilean	0.0572	0.0229	0.5377
Argentinean	1.0246	1.9499	

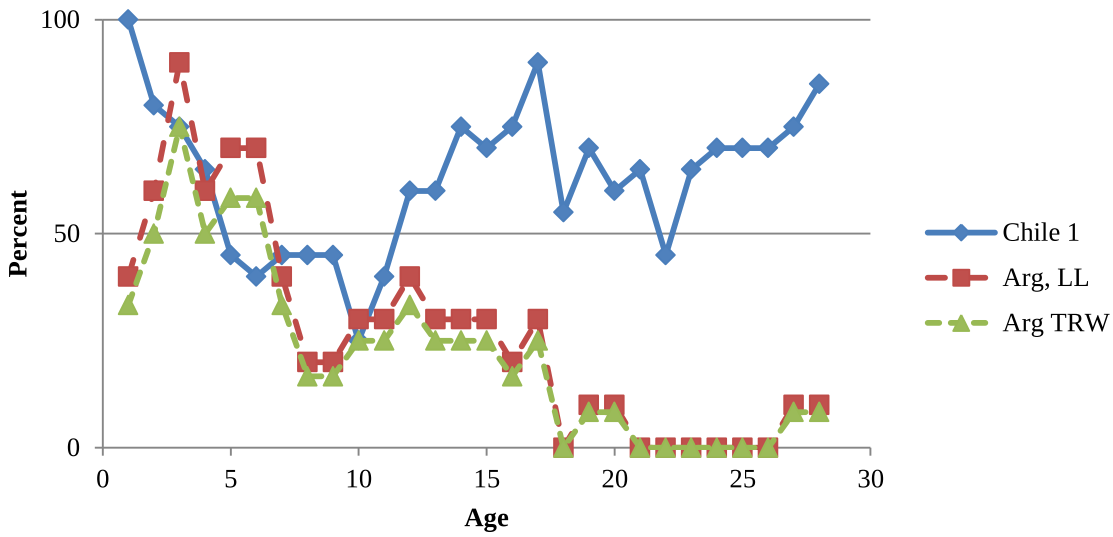


Figure 11: Comparison of the percent of years for each age-class that the residuals for the predicted proportion at age were negative (i.e, the predicted proportion at age exceeded the observed).

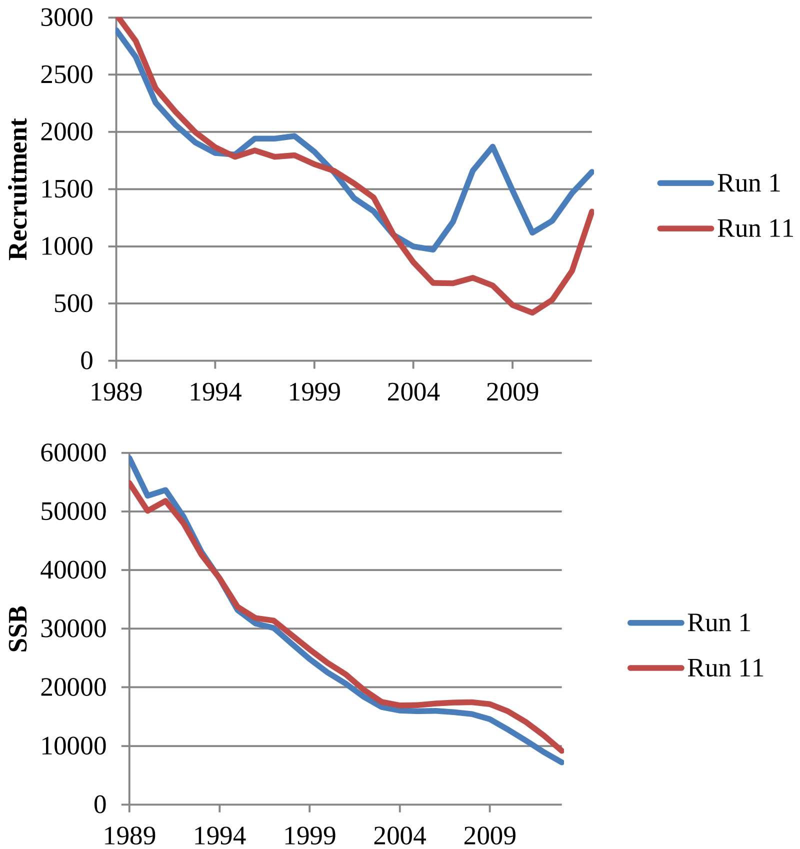


Figure 12: Comparison of estimated recruitment trends for the results presented in the Assessment Report (Run 1 in Appendix 5) and for the same set of parameter values except no changes in selectivity (Run 11 – Appendix 5).

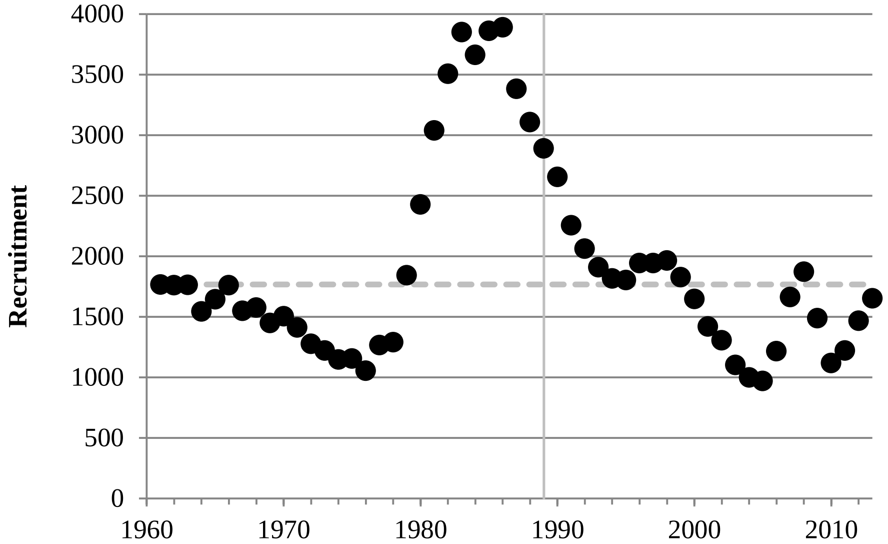


Figure 13: Estimates of recruitment (millions) for the results presented in the Assessment Report (i.e. Run 1 of Appendix 5). Note that the recruitment estimates include those used to estimate the initial abundances in 1989 (i.e. for the years 1961-1988) for which F is assumed to be zero and for the years covered by the assessment (1989-2013). Dashed line is the model estimate of R_0 and the solid vertical line is the year fishing began.

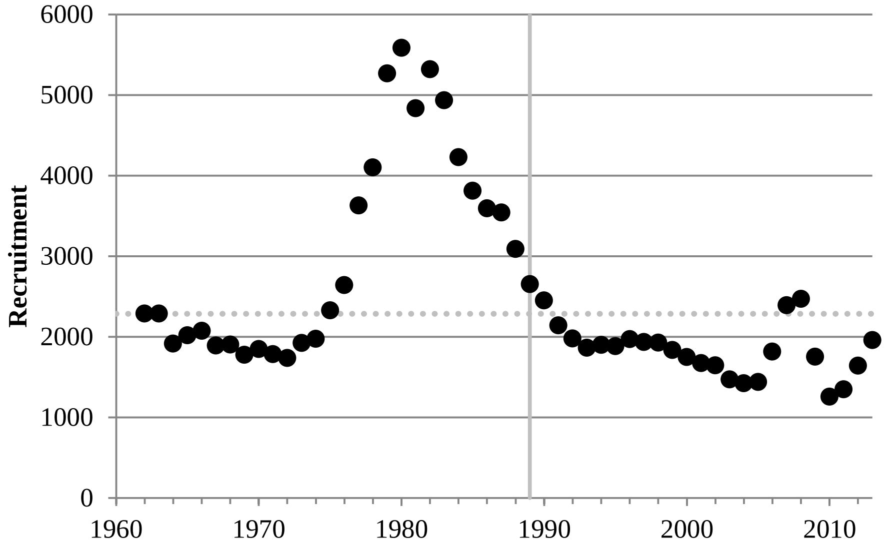
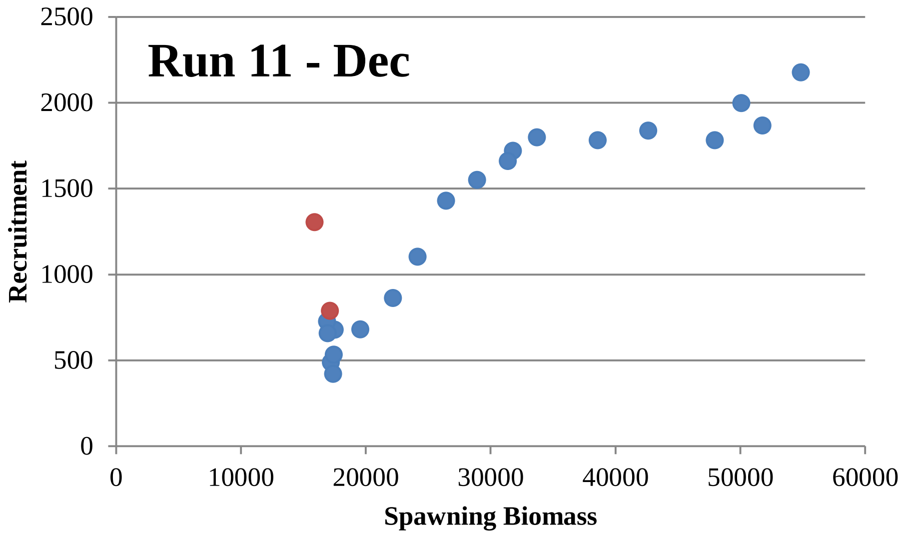
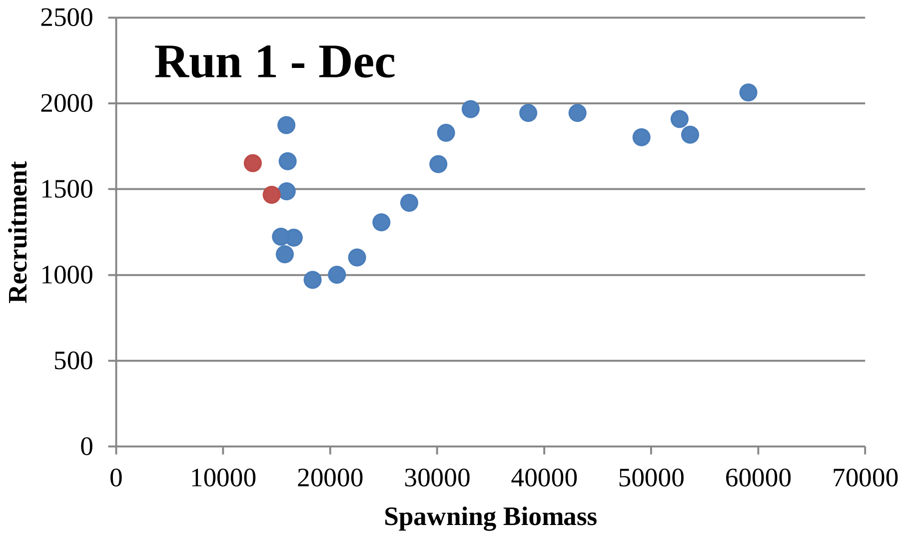


Figure 14: Estimates of recruitment (millions) for the intended base case model run for the assessment (i.e. with the aging error turned on, Run 2 from Appendix 5). Note that the recruitment estimates include those used to estimate the initial abundances in 1989 (i.e. for the years 1961-1988) for which F is assumed to be zero and for the years covered by the assessment (1989-2013). Dashed line is the model estimate of R_0 and the solid vertical line is the year fishing began.





`

Figure 15: Comparison of estimates of the realized stock and recruitment relationship for the results for the "base case" as presented in the Assessment Report (Run 1 in Appendix 5) and for the same run but with no selectivity changes for all fleets (Run 11 in Appendix 5). The red marks are the recruitment estimates for the two most recent years (i.e. number of three year olds in 2012 and 2013) for which there is minimal information (see text for detail).

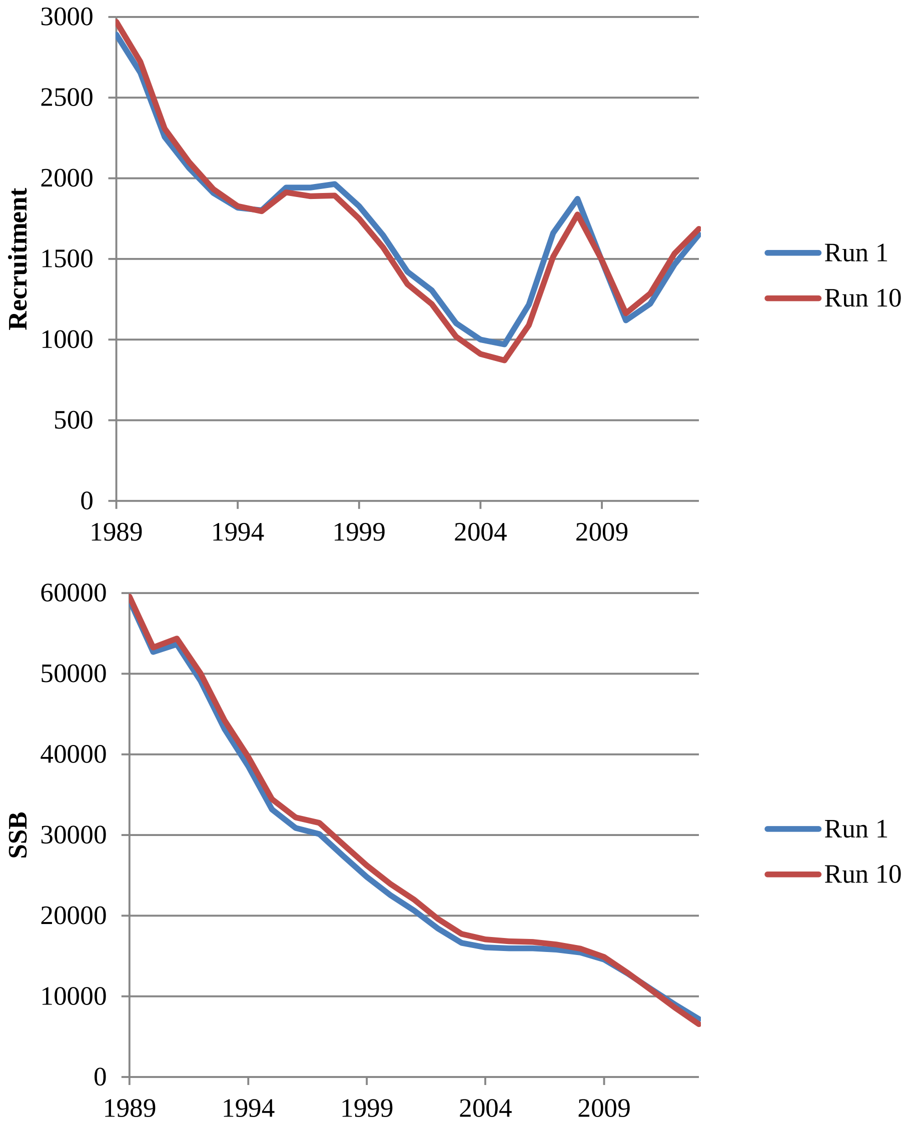


Figure 16: Comparison of estimated recruitment and spawning biomass for the results presented in the Assessment Report (run 1 of Appendix 5) compared to when selectivities for the Argentinean trawl fleet were allow to be dome shape (run10 of Appendix 5). See discussion above about problems in the fit for the latter.

Uncertainty - Alternative Runs
The reviewer spent considerable amount of time examining the results from the various requested runs. This examination yielded a number of issues in terms of conflicts amongst the various input data with potential implications about the model, selecting a base case or a set of base cases to integrate across, and uncertainty in the results. Many of these issues are discussed elsewhere in this report and are not repeated here. In terms of uncertainty, these results indicated that the results are sensitive to various input specifications and (not surprisingly) that considerable model-uncertainty exists (e.g. Tables 3 of Appendices 4 and 5). However, given the issues and problems with the model inputs as well as concerns about the conceptual model and implementation issues (see below), a detailed consideration of these in terms of their implications about the level of uncertainty would not be meaningful. In addition, the number of runs that was able to be explored was limited. As such, the alternative runs that were conducted should be considered as indicative of the sorts of sensitivities that need to be explored and integrated into the assessment advice (particularly in terms of risk). However, they should not be considered as exhaustive.

5. The stock assessment model

Several technical problems were encountered in the technical implementation of model runs and the numeric results presented in Assessment Report. These include:

- **Non-Reproducibility of Code and Results.**
  The numeric results presented in the stock assessment report for the base case model differed substantially from the results provided at the workshop for Run 1, which was meant to be a re-run of the assessment with the same data inputs and model specification (Figures 8-10). The differences are substantial as would be the implications for management advice. It was not possible to determine the source of the discrepancy with the available information and documentation in the Assessment Report or the information provided for the workshop. The reviewer found differences between the computer code for the stock assessment model provided in the Assessment Document, the code provided electronically (i.e. the TPL file) prior to the workshop, which was supposed to be the code used to generate the results in the Assessment Report, and between these two codes and the version of the code that was provided that was used to generate the results for Run 1[39]. There were also differences in the data and control files used for running the model between those supplied prior to the workshop and those used during the workshop to generate Run 1 results (most notably no aging error matrix was included in the former)[40]. These files were not documented in the Assessment Report.

  The actual TPL computer code and data input and control files used to generate the results presented in the Assessment Report were not archived and changes were

[39] For example, differences were in the section of the code where recruitments were estimated (e.g. the code in the Assessment Report has options for either use of a Beverton Holt Function for estimating recruitment or a constant recruitment with log-normal deviates, the code supplied in the file prior to the workshop has only the option to use a Beverton-Holt Function and the code used for Run 1 has both options.
[40] Note that while no aging error matrix was included in the input data file, the source code actually included code for reading this matrix from the input data file. As such, the provided input file was incompatible with the supplied computer code. As such, the combination of model and data input file supplied would not have run.

subsequently made to the TPL Computer Source Code files and the data input and control files. It was not possible for the reviewer to determine which of these combinations of code and input files were used to generate the results in the Assessment Report and the actual the source of the numerical differences. The computer runs made during the workshop were not available until the last day of the workshop and there was limited time to review them within the time available. It was not until after the workshop that it became apparent that there were substantial differences between the Assessment Report and the Workshop Runs.

This issue was raised with the Dr Ernst (coordinator of the review), who communicated the problems to IFOP. Subsequently, Mr. Tascheri provided a document explaining the source of the discrepancy and additional set of results for runs requested during the Workshop (Appendix 5). As reported in this document, the actual results provided in the Assessment Report were run without applying the correction for age determination from scales, although the application of the age reading error was highlighted as one of the new features of the stock assessment model in the Assessment Report. Consequently, the results presented in the Assessment Report do not correspond to the model as described and specified.

Potential implementation, computing and input mistakes are always a concern with any complex assessment model and mistakes do occur (this is only human). Validation and verification is difficult and hundred percent certainty is hard to achieve. For this reason, it is imperative that rigorous evaluation, review, documentation and cross-checking of code, input files and results are undertaken. The reviewer is concerned that this was not the case in the current assessment and appears to be symptomatic. While lack of resources may be a contributing factor, there appears to be lack of attention or importance attached to the validation and verification problem. The reviewer would have expected as a routine matter that whenever a major new feature such as a reading error matrix was added to an assessment model that comparative runs would have been made and results of the comparison included within the assessment document. If this would have been the case here, the problem in the results would have been obvious and prevented[41].

- **Error in the Input Data for the Aging Error Matrix**
  An aging error matrix is provided in Table 8 in the Assessment Report. This matrix is clearly makes little sense and is clearly in error. This matrix is meant to redistribute the model's estimate of the number caught at age in any year to other (younger) age classes in order to account for the bias from age reading from scales. The columns of the matrix are meant to represent the proportion within each age classes that would have been estimated to have been a different age. As such, each column should sum to one. This is clearly not the case in Table 8 in the Assessment Report (e.g. column

[41] In the revised set of workshop runs provided to the review in December, Run 12 contains the reverse input specification error. In this case Run 12 was run with the aging error turned on when it was supposed to have been turned off (i.e. in the control file for this run the parameter "operror" which controls whether the aging error is used was set to 1 – which means active). The reviewer detected this when after examining the plot of SSB trends over time provided with the revised set of runs (Appendix 5). Run 12 has M as an estimable parameter and the resulting estimate is 0.18. As such, the resulting trends would be expected to be similar to Run 4 in which M was fixed at 0.20. However, in the figure of SSB trends over time, Run 4 and Run 12 have substantially different trends while the trend for Run 12 is the only one that resembles the run in which the aging error was meant to have been turned on (Run 2). This re-enforces the concerns about a lack of attention or importance attached to checking results and to validation and verification.

one sums to 0.001 and column 3 sums to 1.598, and most columns appear to sum to values greater than one). Consequently, the models predicted "true" number for the catch-at-age would be inflated. However, it was not clear whether Table 8 in the Assessment Report came from or was in fact ever used in any of the calculations as no aging error matrix was included in the data file that was provided for what was supposed to be the one use in the production of the assessment results. Thus, a different aging-error matrix was included in the data file during the Workshop that was supposed to have been the one used in fitting the model for the Assessment Report. This aging error matrix also had problems in that all of the columns did not sum to one (see above). However, the magnitudes of the errors were substantially smaller (and apparently were due to running errors when copying from an EXCEL file). However, as discussed above, there was a miss-specification in the control file so no aging error correction was actually applied in the calculation of results presented. As such the errors in the aging-error matrix were negated by this second error. Nevertheless, these "compensating" errors appear symptomatic of lack of rigor in the implementation and running of the assessment model.

- **Coding Error**
  The way that the predicted total yield is calculated within the computer code when the aging error matrix is used does not correspond to the expected weight of the catch that the model is removing from the modelled "true/predicted" population over time. This is because the aging error matrix is applied to the model's estimates of the numbers caught at age and multiplying this by the mean weight-at-age in order to obtain its estimate of the predicted weight of the catch-at-age in each year. These estimates of the catch weight by age are then summed to give the model's predicted total catch in weight by year (i.e. total yield). These predicted yields are then compared to the observed yields in the estimation of the parameters of the model. However, the actual observed yields are derived from landing data. They presumably do not entail any use of the catch-at-age data. As such, they should be independent of whether the aging was based on scales or otoliths. However, the effect of the aging error matrix is to make the predicted catch-at-age younger then the model's true catch-at-age. Younger fish weigh less then older fish. Consequently, for any iteration of the model, the predicted total yield would be an underestimation of what the actual model was predicting that the total yield should have been. This is because the actual numbers caught within the modelled population remain the same before and after the application of the aging error[42]. How the model compensates for this in the estimation of the parameters and its effect on the overall results is not possible to second guess. Similarly, it is not possible to determine the extent that the large differences in the results with and without the aging-error correction are the result of this coding error or the actual aging-error correction.

- **Convergence Problems:**

  The Runs in which M was set as an estimable parameter clearly did not converge at a global minimum. Thus, in Appendix 5, the value of the objective function is greater for Run 12 with M as a free parameter than for Run 2 with M fixed at

[42] The fact that there were errors in the specification of the aging error matrix inputted into the model meant that when the aging error matrix was applied that the total modelled numbers in the catch would have changed, although this was not the intention (see above).

0.15[43,44](similar problems are seen in Appendix 4). At least for these runs, the model appears to have multiple minimum and appears to be sensitive to initial conditions[45]. No other runs showed similar convergence problems. Increases in the number of parameters always yielded a lower value for the objective function and the gradient at the minimum was always small). However, it seems that Run 10 which allowed for dome shaped sensitivity in the Argentinean trawl fleet did not converge at least in Appendix 5 (see above). Moreover, the apparent lack of insensitivity of some components of the objective function (most notably the catch-at-age data) seen in many of the runs is somewhat unexpected. It would be worth verifying that the solutions found in general were in fact global minimums (e.g. consider alternative starting values, alternative phasing, and restarting the fitting at the found solution).

The above set of problems combined with the lack of detail and mistakes[46] in documentation makes it difficult to make definitive conclusions about the details of the results and undermines confidence in the reliability of the actual numerical results presented in the Assessment Report and in the review process. Over the course of the review, the reviewer did undertake a reasonably detailed examination of the actual source code and the data input and control files used to produce the results for the runs conducted during and after the workshop (more detailed than might normally be expected for a review of this nature). It appears that the code and files should have provided an accurate and appropriate representation of the model when no aging-error correction was applied. However, this examination was not, nor was it intended to be, a validation/verification exercise. This would have required substantially more time and resources than were available in this review process. In addition, the coding errors noted above were only detected late in the review process (i.e. after the second set of runs was provided in mid-December and after portions of the report had been drafted). These two factors combined with the internal lack of review and rigor in the implementation of the code, the running of the model and evaluation of the numerical results does leave lingering concerns that other undetected bugs or problems may still exists. In all the runs in which the aging-error correction was applied, the coding error noted above is still embedded in the results (most of the runs in Appendix 4 and Run 2 in Appendix 5). There was simply insufficient time to attempt to obtain additional runs where this error could have been corrected.

6. Biological reference points and determination of the stock status

The choice of basis for the fundamental biological reference point for determination of stock status and as management objective is specified by law in Chile so as to correspond to maximum sustainable yield. An independent workshop process is in progress for defining and evaluating appropriate reference points for the various Chilean fisheries. In this context, the

---

43 See Footnote 9.

44 Runs 2 and 15 in Appendix 4 should be numerically identical and they appear to be except that there is a tabling error for the SSB_{2013} figure in Appendix 4,

45 In the runs done during the workshop in which M was an estimable parameter, different results were obtained depending upon the initial conditions specified for M.

46 For example, there are errors in the tables of results provided in the Revised Set of Runs supplied in December. Thus, in the Table 1 which gives the values for the likelihood component for each run, the reported results for runs 7 and 8 are identical and the values for Run 8 are clearly in error. The results for Run 8 are merely a copy of that for Run 7 (which was confirmed by inspecting the detailed Report file for these runs). Also, the value for SSB_0 in 2013 for Run 10 is in error. The reviewer has not cross checked every number in the tables against the report and parameter output files produced by the program. The detection of these types of errors in those checked does add to concerns about the accuracy of the results as reported.

Assessment Report noted this process has recommended that for Chilean Sea Bass that $F_{45\%}$ (fishing mortality that decreases spawning biomass per recruit to 45% of that with no fishing) as a proxy or substitute for F_{MSY}´s. It further recommended that a proxy for the spawning stock level that corresponds to that at maximum sustainable yield be calculated as:

$$BD^{*}_{MSY} = R_{med} * BDR_0 * 0.45 \qquad (1)$$

where R_{med} is the average recruitment through the whole period considered for stock and BDR_o is spawning biomass per virginal recruit.

These recommendations appear to be a reasonable approach for providing proxy reference points for this stock in terms of MSY in the absence of a firm basis exists for estimating a stock and recruitment relations (but see above). Beyond noting this, it is outside the current review to review these recommendations.

Reference points can be useful for translating general management objectives into quantifiable and estimable objectives. However, the critical question is how they are used to provide management advice on stock status and recommendations on quota. Using the most recent estimates of F_{msy} and current stock sizes from a most recent assessment are unlikely to provide reasonable performance or to achieve the longer term management objectives (stocks have been reduced to low levels using this kind of decision making approach). Similarly, basing recommendations on constant catch projections from the base case of the most recent assessment are likely to result in large variability between years in quota recommendations and unsatisfactory behaviour. Such approaches do not take into account (1) the uncertainties in the assessment and the underlying population dynamics and (2) retrospective behaviour and other statistical properties of the assessment model (e.g. possible persistent biases in the most recent estimates). They also do not adequately account for risks (which are generally asymmetric). Management procedure evaluation (also known as management strategy evaluation) has been found to be an effective approach for developing decision rules and management recommendations that addresses uncertainties in both the stock assessment and underlying population dynamics. It has been used for a number of fisheries and a variety of different management fora (Kirkwood, 1997, de la Mare 1986, Butterworth et al, 1997, Smith et al 1999). This approach would be worth exploring here.

Calculation of Reference Point

It should be noted that the Assessment Report did not use equation 1 to estimate its spawning stock biomass reference point. Instead, as stated in the report, it used a substitute proxy based on the following equation:

$$BD^{**}_{MSY} = 0.4*BD_0 \qquad (2)$$

where BD_0 is an estimate of the equilibrium biomass derived from the assessment estimate of R_0.

The Assessment Report notes “that the assessment model produces a BD_0 estimate” as it reason for this. The BD_0 estimate from the stock assessment is derived from a straightforward calculation of the assessment estimate of R_0, natural mortality, the maturity ogive and the mean weight-at-age vector. It assumes that the estimate of R_0 produced by the assessment is a valid estimate of equilibrium recruitment under no exploitation. However, as discussed below, R_0 in the assessment is based on the assumption that recruitment is constant and independent of the size of the spawning stock. If recruitment have declined over the evaluation period as a result of declining spawning biomass, then the value of R_0 from the assessment is likely to be an underestimate of R_0 in the context of an underlying stock and recruitment relationship (see below).

In equation 2, it is not clear why the value of 0.40 was chosen instead of 0.45. It should be noted that if R_0 equals R_{med} in equation 1 (which it should if the assumption of constant recruitment used in the assessment were correct), then BD_0 would equal R_{med} * BDR_0. As such, equation 2 would result in the estimate for the reference point being 11% lower (i.e. 0.40/0.45) than the value that would be obtained using the proxy recommend by the workshop (e.g. the assessment results in the Assessment Report are more optimistic in terms of depletion). Further, it is not clear whether the evaluation period to be used with equation 1 specified by the Reference Point Workshop Process is meant to include only the period during which catches are accounted for or also the estimates of recruitment for the pre-exploitation period used for estimating N_0.. If the former and the evaluation period of the assessment is considered to be 1989-2013 for estimating R_{med} in equation 1, the average recruitment estimates over this period is less than the estimate of R_0. In this situation, this would tend to counteract to some extent the effect of using 0.40 instead of 0.45 as the multiplier in equation 2.

Overall, using equation 1 instead of equation 2 to estimate a proxy for BD_{MSY} for this assessment would yield an estimate that was either ~7 or 11% greater than that used in the Assessment Report depending on how the evaluation period in equation 1 is defined. As such, current depletion levels would be underestimated. In addition, an underestimation of BD_{MSY} would have flow on effects on the catch recommendations in the Assessment Report (i.e. they would need to be lower to achieve the same outcome in terms of the BD_{MSY} reference point). As such, there is a need for clarification of the basis for equation 2 and whether this is an acceptable substitute for the proxy reference point recommended by the Workshops on this issue.

7. Procedures used for Stock Projections

The stock projections are preformed based on constant fishing mortality rates and a constant recruitment based on the estimated recruitment in the last five years from the assessment. A number of concerns exist about the approach and methods used for these projections, particularly in terms of their ability to produce robust advice about management actions and their associated risk. These include:

- The most recent recruitment estimates are biased toward the mean (R_0) and cannot be considered reliable. This is because there is almost no information/data available on the actual strength of the most recent cohorts. For example, there is only one year of catch data for the most recent year and two for the year before that. As such, the catch-at-age data provide no real information on the strength of these cohorts except in relationship to the estimated selectivity curves. These most recruitment estimates will be highly dependent upon the selectivity assumption and the assumption of constant recruitment. Within the current assessment, the selectivity on the first two age classes in the model are small so large variability relative to the predicted catch could be expected (and without contributing substantially to the objective function). In addition, given the estimated selectivities for the younger two ages, the strength of the most recent cohorts are basically not represented in either the observed or predicted catch rates. As such, they will have a minimal contribution to the prediction of CPUE within the model. In the absence of data, the model will tend to predict that recruitment would produce estimates approaching the mean (R_0) in order to minimize the recruitment penalty function.

- The projections are based on an assumption of constant recruitment and take no account of the estimated current size of the spawning stock. The common reason for not estimating a stock-recruitment curve within the assessment are that if stocks have reasonable resilience (i.e. high steepness in terms of a Beverton-Holt function), only small changes in mean recruitment levels would be expected over a wide range of spawning biomass levels. In addition, recruitment is usually highly variable. As such, the assessment models generally have poor ability to estimate the parameters for a functional stock-recruitment relationship. The stock assessment results in terms of current status (e.g. depletion levels) are likely to be similar as long as the variability in recruitment is assumed to be large. However, for prediction into the future, spawning stock levels are important to consider (otherwise why be concerned about depletion levels). Given the current estimated stock by the model (i.e. ~10%), no matter how resilient the stock may be recruitment would be expected to be affected. The current stock assessment results also suggest this in that recruitment has generally declined as the spawning stock has declined (Figure 15).

  Basing the projection recruitments on the mean of the last five years (instead of the overall mean) compensates for this to some extent. However, there is a three year lag between the estimates of recruitment relative to the estimated spawning biomass due to the fact that the first age in the model is three (e.g. the model estimate of recruitment in 2013 were the result of the spawning biomass in 2011)[47]. In these three years, the spawning biomass was estimated to have declined by 30%. In addition, there is large uncertainty associated with the most recent recruitment and potential biases as noted above. Based on the estimated CV's for the individual recruitment estimates there is little reliability in any of the estimates in the last 2-4 years (i.e. the CV associated with them are ~50% or greater, Figure 16). If one only considers the last two recruitment estimates as unreliable, there would be a five year lag. During this period the model estimated that spawning biomass declined by over 50%. Given the low level and continuing decline in spawning biomass during these five years, recruitment would have on average been expected to have declined. As such the projection results are likely to be over-optimistic

  .

- As with the stock assessment model, documentation provided for the projections is minimal and incomplete[48]. The main projection results (e.g. those in Tables 11 and 12 of the Assessment Report) appear to be deterministic, but this is not clear. It is also not clear what the two α levels and catch levels in the upper portion of in Tables

[47] Note that the ADMB model code provided in Appendix 8 of the Assessment Reports allows for the option of utilizing a Beverton and Holt stock recruitment relationship in producing projection results (i.e. the function "Eval_Fcte"). In this function, the lag in recruitment is inappropriately implemented as one. However, it is not clear whether this code was ever used or not.

[48] The computer code in Annex 8 of the Assessment Report contains a function "Eval_Fcte" which appears to have been intended for performing the stock projections. However, this code is not included within any of the source code file provide before the workshop as the program code used for generating results in the Assessment Report or in the versions of program code which were provided during and after the workshop which were used to produce the results for the alternative runs.

11 and 12 are meant to represent or how they relate to or were used in the risk statistics stated in the Assessment Report. It is also not clear how the risk statistics provided in the Assessment Report and the probability distributions in Figures 49 and 50 in the Assessment Report were calculated as there is no documentation provided on these. These appear to have been derived using likelihood methods and normality assumptions based on ADMB estimates of the mean and variance for the depletion ratio. These estimates are more akin to estimates of the variance of the mean depletion level then estimates of the variance of the actual distribution of depletion levels (e.g. those that would be derived from stochastic or bootstrap approach). As such, within the context and assumptions of the one scenario, they are likely to be too small and the risk inappropriately estimated (under estimated).Truly stochastic or bootstrap projection methods should be used. They would provide more appropriate measures of risk within the context of a single model run.

- No consideration of model uncertainty is considered. The estimates of uncertainty associated with the assessment are highly dependent upon the assumed values for the variance and penalty components in the model (e.g. the CVs for the CPUE indices and sample sizes for the catch-at-age data) as well as the structural assumption and fixed input parameter values (e.g. q changes, selectivity blocks, natural mortality rate, etc.). In particular, the low CVs for the CPUE series and the relatively high sample sizes for the catch-at-age data (particularly the Argentinean ones) would suggest that the risks have been underestimated.

- The projection results assume that a constant F level with no changes in selectivity will be maintained for 15 years. No consideration is given to the fact that the catch level corresponding to this fixed F level would need to be estimated every year from an updated stock assessment. The catch level estimated in this case would have error associated with them. As such, the actual F's (i.e. the one experienced by the population) would differ from the specified constant one. Moreover, realized selectivity would also vary over time. How this estimation error in the updated assessments and variability in selectivities would affect the overall performance of management based on such a constant F level would need to be assessed within an operating model/management procedure context to provide a meaningful measure of risk. In any case, the actual variability and associated risk would be expected to be greater. However, it is also unrealistic to assume that a constant F level would be maintained irrespective of the updated stock assessment results.

- The procedural steps, input parameter files and all of the actual code used to generate projection results were not documented. Results from an ADMB model run must have been extracted from the report and parameter output files in order to calculate the $F_{45\%}$ levels using the R code provide in Annex 8 of the Assessment Report. Then this $F_{45\%}$ estimate was apparently feedback into ADMB code to produce the projection results. It is not clear whether this was done as an automated process or manually. In any case, given the numerous problems in the implementation and running of the main assessment relative to the intended specifications, concerns exists about the rigour and internal review that were applied in the running, calculation and tabulation of the projection results. As such, these translate into concerns about the accuracy of the numeric results.

In addition to the above issue with the projection procedures, the numerous issue concerning the data inputs and the implementation of the stock assessment detailed in this review would suggest that little reliability or robustness could be attributed to the risk analyses presented in the Assessment Report. These fundamental and more basic problems need to be resolved before meaningful projection results and risk analyses can be performed.

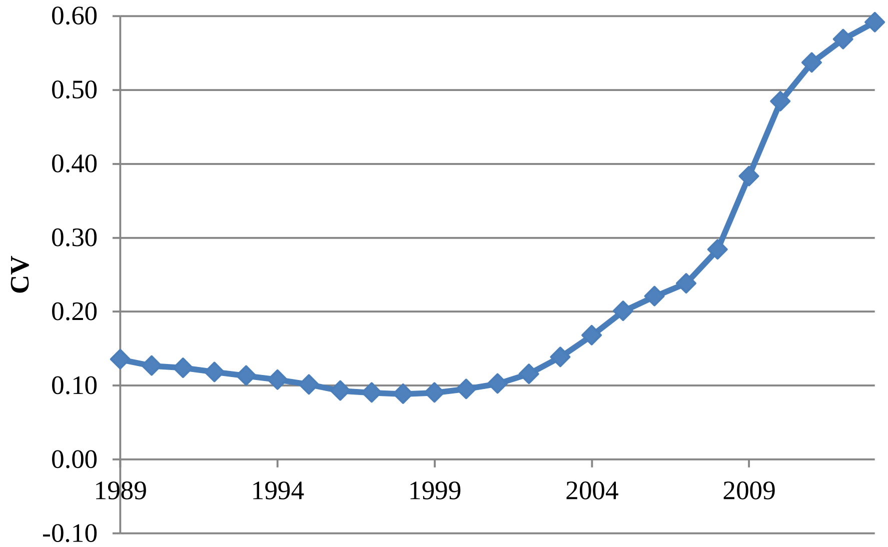


Figure 16: Estimated coefficient of variation (CV) for the recruitment estimates for Run 1 from the revised set of runs provided in December (See Appendix 5) as calculated by ADMB from the Hessian Matrix. (These were calculated from the output files from ADMB provided for the additional runs).

8. Recommend improvements to the assessment process

A large number of issues and concerns with the current stock assessment were identified and discussed above. Except where pertinent to the more general recommendations in this section of the Review Report, these are not re-iterated here. Nevertheless, they are important. These need to be considered and addressed in any future assessment and model development work.

Comprehensive Review and Development of the Biological and Fishery Inputs
Numerous issues are discussed above with respect to data inputs for the assessment. There appear to be errors and substantial inconsistencies (e.g. mean weights at age, age of maturity, catch levels, etc.). These data are fundamental to the assessment. Without reasonable confidence in the accuracy and consistency of the inputs, one can have little confidence in the outputs. A reliable assessment requires that these issues be resolved. A comprehensive and integrated review of the current inputs is warranted. Checking for consistency among the various inputs should be a critical and integral component of this process (e.g. explicit or implicit growth curves used in estimating maturity and weight at ages). Where substantial uncertainty exists (e.g. size of maturity, growth, weight-at-age, etc) an appropriate sets of alternative input data series should be developed that span this uncertainty (but ensuring internal consistency within in any single set). Where possible, relative plausibility or weights should be developed. In any case, the range of uncertainty needs to be carried through into the assessments and integrated into the broader considerations of uncertainty and risk. This

review needs to consider not only data and information from Chile but needs to address the data from all the Patagonian shelf regions.

A critical component of this review needs to focus on how best to estimate the catch-at-age data across time and fishery. For the Chilean fishery data, appropriate approaches for reconciliation of the differences between scale and otolith aging are essential. While for the Argentinean data, approaches to best utilize the length frequency data in the absence of direct aging need to be evaluated (e.g. using the Chilean age-length keys, fitting to size data within the model, etc). There is also a need for a more thorough and comprehensive analysis of the standardization of the CPUE indices, which should be considered either as part of this review or undertaken as a separate process.

This review should be seen as a very high priority. Without resolving these input issues, there is little basis for undertaking any future assessment.

International Cooperation

The best evidence is that the toothfish resource in the Patagonian/Shelf region of South America constitutes a single stock. It is unclear the extent to which it can be considered one homogenous population or the degree of spatial and temporal structuring of recruits and spawners throughout the entire region. Nevertheless, based on the existing information, it seems highly unlikely that the resources in each of the three political jurisdictions (i.e. Chile, Argentina and the Falkland Islands) constitute independent isolated population. Assessments and research require cooperation and interchange of data for their effective development and implementation. Standardized approaches to data collection (e.g. logbooks, observers, age-reading, tagging methods) are critical for ensuring that comparable data are available across the entire region. Cooperation is also important to avoid duplication and for the most efficient use of the limited resources available for basic research and assessments. As such, development of a framework and procedures that will allow for joint and effective collaboration in stock assessment, data collection and basic research (e.g. tagging) is needed.

Meaningful management advice in terms of catch levels and their consequences requires consideration of catch levels across jurisdictions unless the seemingly unlikely result of future research shows the resources in each jurisdiction can be considered as independent units. Meaningful advice on catch levels in any single jurisdiction cannot be provided independent of catch levels in the others. Allocation issues among fisheries/jurisdiction are primarily a management decision, even though different allocations may have implications for overall catch levels. While perhaps outside the terms of reference for this review, a framework for joint management and allocation seems also to be needed.

Procedures for developing, implementing, reviewing and documenting Stock Assessments

Numerous technical problems were encountered in the implementation and presentation of the current stock assessment. Documentation within the report is incomplete and in places inaccurate in terms of what was actually done. Internal documentation within the computer code used for implementing the stock assessment is very sparse and some coding errors were found. Selection of some inputs and decisions on model structures appeared to be *ad hoc* without sufficient evaluation and review of their appropriateness and possible alternatives. A much more systematic and rigorous approach to the development, implementation and documentation of the stock assessment is required. Some of this can be achieved by utilizing formal procedures. Thus, a system of version control for programs and input files that keeps tracks and archives progress in the development of code and runs preformed should definitely be instituted. This should be the standard procedure employed for all stock assessment, not only toothfish. However, there also needs to be substantially more care, detailed examination of results, consideration of alternative scenarios and internal review embedded in the stock assessment process. Formalization of a process to fully deal with this is difficult. Successfully solving this is dependent in part upon on the personal approaches and dedication of those involved. Nevertheless, (1) utilization of a team approach rather than depending upon a single individual for the developing the model, implementing it, and producing results; (2) ensuring that there are adequate resources available (e.g. time and manpower); and (3) instituting rigours internal review of the results and assessment reports should improve the process and should be considered essential.

Integrated and Comprehensive Approach to Assessment and Research

Related to the previous recommendation, assessment scientist need to understand the data going into their models and make sure all appropriate available information is considered. This needs to include a fundamental understanding of the basic data, sampling approaches and analyses of the inputs going into the assessment. Those responsible for providing inputs also need to be aware of how their inputs will be used. These are essential to ensure that the best information is collected and utilized and avoid misinterpretation or inappropriate weighting of data. There appears to be insufficient interaction among those providing the various inputs into the assessment and those performing the actual modelling and estimation. There is a need to develop a more holistic, integrated and interactive approach for the assessment process among all those providing the inputs and those performing the analyses. Those performing the assessment need to be proactive in assuming responsibility for the inputs being used.

Fishery Independent Tunning Information

Stock assessments require some source of external quantitative information on abundance, abundance trends or fishing mortality rates. The current assessment relies totally on CPUE indices as a measure of relative abundance for this. The danger of relying on solely CPUE is well known and documented. Additionally, as discussed above, the interpretation and standardization of CPUE in the current assessment is problematical and there is a lack of data for quantifying critical factors affecting catch rates. There is a critical need to obtain fishery independent information on abundance or fishing mortality rates to use within the stock assessment. Tagging experiments appear to be the most feasible, if not the only, feasible approach for doing this. Ongoing tagging studies form an important component of the routine monitoring for a number of toothfish stocks and results from these studies are incorporated into the analytical stock assessments. The implementation and a long-term commitment to an appropriately designed tagging program with sufficient resources cable of providing on-going estimates should be seen as a high priority.

In this regard, the past and current tagging program in the industrial fleet and its expansion to the artisanal fleet as document in the presentation to the workshop (Rubilar, et al 2014) is an encouraging development. Quality control is critical in any tagging program. Large differences are sometime found between different taggers. Ensuring rigours training and standards is critical. Also, it is important that taggers have no vested interest in the results. In this regards, use of fisherman as taggers has been found problematical in other fisheries.

Critical to the use of tagging data quantitatively within the assessment is the estimation of reporting rates (the fraction of recapture tags which are actually reported). Reporting rates are confounded with estimate of either abundance or fishing mortality rates derived from tag return data. It is essential that data for estimating reporting rates be incorporated as a fundamental part of any quantitative tagging program. PIT tags have been found to be an effective approach for dealing with this problem within toothfish fisheries. The reviewer would recommend that this approach be utilized as it appears to offers the most effective, if not the only practical, method to deal with the reporting issues for this fishery. Results from PIT tag tagging programs have been shown to be able to provide useful overall quantitative estimates for stock assessments. In this regard, automatic PIT tag detection systems are possible to utilize for the industrial longline fisheries. They would likely provide the most effective and cost-efficient approach for obtaining tag return data from this sector. In any case, a strategy that will ensure that reporting rates are high and also estimable needs to be embedded in the design and implementation of any tagging programs.

Catch and Fishery Monitoring

Given the high apparent errors in the Chilean log book data (i.e. basic flittering for CPUE analyses resulted in 75% of them being rejected) and the issues associated with estimates of actual catches, the reviewer is concerned about the reliability of the actual data collection system for the log book and landing data. Hopefully many of the problems are historical and improvements such as VMS, catch certification, collaboration with industry and observers have resulted in substantial improvement in the data. Nevertheless, on-going reviews to confirm that this is in fact the case are important as well as seeking improvements to the accuracy and efficiency of the data collection system. In this regard, there are some issues about the representativeness of sampling in terms of observer coverage and catch sampling. In terms of the former, observer coverage tends to be low and limited to a small sample of vessels. It is important that observer coverage is spread across the fleets to ensure representative sampling. In terms of catch monitoring (including for size), there are recently developed video methods being deployed in the domestic Australian longline fisheries. These methods allow for the automated monitoring of the total catch and it size distribution on a set by set basis. Consideration should be given to adapting and employing this technology within the industrial fleet (and perhaps in larger artisanal vessels). It would not only ensure a high level of accuracy and precision in the catch and size data but would also free up observer time for other important data gathering activities (e.g. marine mammal and bird interaction data collection, tag recovery). It also may reduce the overall level of observer coverage required. In terms of the artisanal fishery, it is important to review that the sampling procedures employed actually result in representative sampling both of the actual catches from individual landings sampled and the set of vessels which are actually sampled. Achieving representative (e.g. random) sampling is often difficult to achieve. In terms of size sampling, obtaining multiple samples from the same landing should be collected as a part of an evaluation of the representativeness of the current procedures.

Aging Issues and Aging Error Estimation

The current direct aging is based on the count of dark bands within the otolith. No consideration is given to the date of the catch relative to the timing of band formation or whether a translucent margin exists. This can result in band counts under or over estimating the actual age by one year. In addition, age readings can be important source of uncertainty in the overall assessment results. Currently, there appears to be no systematic data collected on age reading errors (e.g. from independent blind readings of the same otoliths by different readers or repeated blind estimates by the same reader). The reviewer considers neither of these issues is among the highest priority issues for this stock. Nevertheless, given the on-going commitment to the direct aging (which the reviewer fully supports as continuous time series invaluable and non-repeatable if stopped) and the fact that this process is largely independent of the other data collection components, it would be worthwhile for these issues to be addressed. It is much more efficient and cost effective to deal with these on an on-going basis as part of the age estimation process than to attempt to deal with them retrospectively.

Operating Model and Evaluation of Management Procedures

Management procedure evaluation (also known as management strategy evaluation) has been an effective approach for developing decision rules and providing management advice for a number of fisheries and within a range of fora (Kirkwood, 1997, de la Mare 1986, Butterworth et al, 1997, Smith et al 1999). In contrast, traditional biological reference points and fixed F projections provide little basis for overall estimation of uncertainty and for the robust evaluation of risks and the trade-off among different management objectives. A key part of the management procedure evaluation process is the development and conditioning of an operating model that encompasses the key structural, biological and fishery uncertainties. The development of such a model would provide an approach (perhaps the only approach) for addressing the large uncertainties involving the spatial structure and recruitment dynamics of this resource which are affecting assessment results and management recommendations. As such, independent of the utilization of a management procedure evaluation approach for developing management decision rules, the development of a comprehensive operating model would allow for improved assessment modelling and evaluation of their robustness in terms of stock status indicators. It is recommended that a management procedure approach be pursued as a basis for improving the assessments and for providing management recommendations for this stock.

Acknowledgements

I wish to acknowledge the contribution of all the workshop participants for the extensive information they contributed to the review process. This information was indispensable. Their enthusiasm, openness and willingness to provide clarification and actively participate in discussions were very much appreciated particularly in light of the long session hours, their other responsibilities and my lack of Spanish. Special thanks goes to Renzo Tascheri for his efforts in running and producing results for alternative scenarios during and after the workshop. Discussions with Billy Ernst and Cathy Dichmont provided helpful perspectives on the assessment, technical details and some of the critical issues in this review report. However, the views in this report are those of the author. Additional thanks goes to Billy Ernst for his help with all of the logistical aspects of the review and travel arrangements. Finally, a big thanks is due to the simultaneous translator (Mrs. Milka Rubio) whose task was enormous in light of the volume of material, its technicality and the length of the sessions.

Literature Cited

Ashford, J.R., Fach, B.A, Arkhipkin, A.I., Jones C.M. 2012. Testing early life connectivity supplying a marine fishery around the Falkland Islands.Fisheries Research. (121-122) 144-152.

Butterworth, D.S., Cochrane, K.L., Oliveria, J.A.A., 1997. Mangement procedures: a better way to manage fisheries? The South African experience. In Global trends: fisheries management, pp 83-90. Ed. By E.K. Pikitch, D.D. Huppert, and M.P. Sessenwine. American Fisheries Society Symposium, 20, Bethesda, Maryland.

S.G. Candy, S.G.,, D.C. Welsford, T. Lamb, J.J. Verdouw and J.J. Hutchins. 2011 Estimation of Natural Mortality for the Patagonian Toothfish at Heard and McdDnald Islands using catch-at-age and aged mark-recapture data from the main trawl ground. CCAMLR Science 18: 29–45

de la Mare, W.K., 1986. Simulation studies on whale management procedures. Thirty sixth Report of the International Whaling Commission. Document SC/37/O14.

De Oliveira, J. A. A., and Butterworth, D. S. 2004. Developing and refining a joint management procedure for the multispecies South African pelagic fishery. e ICES Journal of Marine Science, 61: 1432e1442.

Francis, R.I.C.C., 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68, 1124–1138.

Fenaughty, J.M. 2006. Geographical differences in the condition, reproductive development, sex ratio and length distribution of Antarctic toothfish (*Dissostichus mawsoni*) from the Ross Sea, Antarctica (CCAMLR subarea 88.1).CCAMLR Science 13: 27–45.

Ferrada, S. 2014, Chilean Sea Bass Genetics. Presentation at the Chilean Sea Bass Review Workshop. Programa Taller de Revisión Experta de Evaluaciones de stock de Bacalao de profundidad y Camarón nylon 2014. Valparaíso, Chile.

Gálvez, P., Cespedes, R., Chong, L., Checura, R., Ojeda. V., Meléndez, R., Molina, B., López, S., Bravo, R., Muñoz, L., Adasme, L. andGonzalez. J. 2014. Agreement 1: Comprehensive Counseling for Fishing and Aquaculture, 2013. Project 1.9: Follow-up Program for Demersal Fisheries and Deep Waters, 2013. Undersecretary of economy.Section VI: Deep Waters Resources, 2013

Hanchet, S., Morsan, E., Bridi, R. and C. Medina Foucher. 2014. Public comment draft report: Assessment against MSC Principles and Criteria for: ARGENTINE PATAGONIAN TOOTHFISH FISHERY(*Dissostichus eleginoides*). Organizacion Internacional Agropecuria.

Horn, P.L., C.P. Sutton, and A.L. DeVries. 2003. Evidence to support the annual formation of growth zones in otoliths of Antarctic toothfish (*Dissostichus mawsoni*). CCAMLR Science 10: 125–138.

Kirkwood, G.P. (1997) The Revised Management Procedure of the International Whaling Commission. In Pikitch, E.K., Huppert, D.D. and Sissenwine, M.P., eds. Global Trends:

Fisheries Management (Proceedings of the Symposium held at Seattle, Washington, 14±16 June, 1994). Am. Fish. Soc. Symp. 20, 91-99.

Rubilar, P.S. and A. Zuleta. 2014. Toothfish tagging program result and perspectives. Programa Taller de Revisión Experta de Evaluaciones de stock de Bacalao de profundidad y Camarón nylon 2014. Presentation at the Chilean Sea Bass Review Workshop. Valparaíso, Chile.

Smith, A.D.M., Sainsbury, K.J., Stevens, R.A., 1999. Implementing effective fisheries-management systems – management stratey evaluation and the Australian partnership approach. – ICES J. Mar. Sci. 56: 967-979.

Smith, D.C.,K. Krusic Golub, S. Robertson and A.E. Punt. 2008. Quantifying age-reading error for use in fisheries stock assessments, with application to species in Australia's Southern and Eastern Scalefish and Shark Fishery. Can. J. Fish. Aquat. Sci. 65:1991-2005.

Tascheri, R., Canales, C., Céspedes, R. and Chong, L. 2014. STATUS AND QUOTA REPORT.Agreement of Performance 2014: Status and biologically sustainable exploitation potential of main local fishing resources for 2015: Chilean sea bass, 2015. Undersecretary of economy and smaller companies.

Appendix 1: Bibliography of materials provided for review

Main stock assessment documents

Gálvez, P., Cespedes, R., Chong, L., Checura, R., Ojeda. V., Meléndez, R., Molina, B., López, S., Bravo, R., Muñoz, L., Adasme, L. andGonzalez. J. 2014. Agreement 1: Comprehensive Counseling for Fishing and Aquaculture, 2013. Project 1.9: Follow-up Program for Demersal Fisheries and Deep Waters, 2013. Undersecretary of economy.Section VI: Deep Waters Resources, 2013

Tascheri, R., Canales, C., Céspedes, R. and Chong, L. 2014. STATUS AND QUOTA REPORT. Agreement of Performance 2014: Status and biologically sustainable exploitation potential of main local fishing resources for 2015: Chilean sea bass, 2015. Undersecretary of economy and smaller companies.

Background papers provided before the Workshop

Agnew, D., Heaps, L., Jones, C., Watson, A., Berkieta, K., and Pearce, J. 1999. Depth distribution and spawning pattern of Dissostichuseleginoides at South Georgia. CCAMLR Sci. 6, 19–36.

Appleyard, S.A., Ward, R. D., and Williams, R. 2002.Population structure of the Patagonian toothfish around Heard, McDonald and Macquarie Islands.Antarct. Sci. 14, 364–373.

Arana, P. 2009. Reproductive aspects of the Patagonian toothfish (Dissostichuseleginoides) off southern Chile. Latin Am. J. Aqua. Res. 37, 381–394.

Arana, P. M., and Vega, R. 1999. Exploratory fishing for Dissostichusspp in the Antarctic region (sub-Areas 48.1, 48.2 and 88.3). CCAMLR Sci. 6, 1–17.

Ashford, J. R., Jones, C. M., Hofmann, E., Everson, I., Moreno, C., Duhamel, G., and Williams, R. 2005b. Can otolith elemental signatures record the capture site of Patagonian toothfish (Dissostichuseleginoides), a fully marine fish in the Southern Ocean? Can. J. Fish. Aquat.Sci. 62, 2832–2840.

Ashford, J. R., Wischniowski, S., Jones, C., Bobko, S., and Everson, I. 2001.A comparison between otoliths and scales for use in estimating the age of Dissostichuseleginoides from South Georgia. CCAMLR Sci. 8, 75–92.

Ashford, J., Duhamel, G., Jones, C., and Bobko, S. 2005a. Age, growth and mortality of Patagoniantoothfish (Dissostichuseleginoides) caught off Kerguelen. CCAMLR Sci. 12, 9–41.

Ashford, J.R., Fach, B.A, Arkhipkin, A.I., Jones C.M. 2012. Testing early life connectivity supplying a marine fishery around the Falkland Islands.Fisheries Research. (121-122) 144-152.

Belchier, M., and Collins, M. A. 2008. Recruitment and body size in relation to temperature in juvenile Patagonian toothfish (Dissostichuseleginoides) at South Georgia. Mar. Biol. 155, 493–503.

Brandao, A., Butterworth, D. S., Watkins, B. P., and Miller, D. G. M. 2002. A first attempt at an assessment of the Patagonian toothfish (Dissostichuseleginoides) resource in the Prince Edward Islands EEZ. CCAMLR Sci. 9, 11–32.

Candy, S. G., and Constable, A. 2008. An integrated stock assessment for the Patagonian toothfish (Dissostichuseleginoides) for the Heard and Macdonald Islands using CASAL. CCAMLR Sci. 15, 1–34.

Cassia MC. 1998. Comparison of age readings from scales and otolithsof the Patagonian toothfish (Dissostichuseleginoides) from South Georgia. CCAMLR Science 5: 191-203.

Everson, I., and Murray, A. W. A. 1999. Size at sexual maturity of Patagonian toothfish. CCAMLR Sci. 6, 37–46.

Guerrero, A., and Arana, P. 2009. Fishing yields and size structures of Patagonian toothfish (Dissostichuseleginoides) caught with pots and longlines off far southern Chile. Latin Am. J. Aqua. Res. 37, 361–370.

Hillary, R.M., Kirkwood, G.P. and Agnew D.J. 2006.An assessment of toothfish in subarea 48.3 using CASAL. CCAMLR Science (13) 65-95.

Horn, P. L. (2002). Age and growth of Patagonian toothfish (Dissostichuseleginoides) and Antarctic toothfish (D. mawsoni) in waters from the New Zealand subantarctic to the Ross Sea, Antarctica. Fish. Res. 56, 275–287.

Hucke-Gaete, R., C. Moreno y J. Arata.2004. Operational interactions of sperm whales and killer whales with the Patagonian toothfish industrial fishery off southern chile. CCAMLR Science, Vol. 11: 127–140.

Kock, K. H., Purves, M. G., and Duhamel, G. 2006. Interactions between cetacean and fisheries in the Southern Ocean. Polar Biol. 29, 379–388.

Lopez, L.J. 2005. Patagonian Toothfish in international waters of the southwest indian ocean (Statistical Area 51).

Lord, C., Duhamel, G., and Pruvost, P. 2006. The patagoniantoothfish (Dissostichuseleginoides) fishery in the Kerguelen Islands (Indian Ocean sector of the Southern Ocean). CCAMLR Sci. 13, 1–25.

McKinlay, J. P., Welsford, D. C., Constable, A. J., and Nowara, G. B. 2008. An assessment of the exploratory fishery for Dissostichus spp. on Banzare Bank (CCAMLR Division 58.4.3b) based on fine-scale catch and effort data. CCAMLR Sci. 15, 55–78.

Moreno, C., Castro, R., Mujica, L. J., and Reyes, P. 2008. Significant conservation benefits obtained from the use of a new fishing gear in the Chilean Patagonian toothfish fishery. CCAMLR Sci. 15, 78–79.

Nolan, C.P., G.M. Liddle and J. Elliot. 2000. Interactions between killer whales (Orcinus orca) and sperm whales (Physetermacrocephalus) with a longline fishing vessel. Mar. Mamm. Sci., 16 (3): 658–664.

Purves, M.G., Agnew, D.J. Balguerias, E., Moreno, C.A. and Watkins B. 2004. CCAMLR Science (11) 111-126.

Rogers, A. D., Morley, S., Fitzcharles, E., Jarvis, K., and Belchier, M. 2006. Genetic structure of Patagonian toothfish (Dissostichuseleginoides) populations on the Patagonian Shelf and Atlantic and western Indian Ocean Sectors of the Southern Ocean. Mar. Biol. 149, 915–924.

Shaw, P. W., Arkhipkin, A. I., and Al-Khairulla, H. 2004. Genetic structuring of Patagonian toothfish populations in the Southwest Atlantic Ocean: The effect of the Antarctic Polar Front and deep-water troughs as barriers to genetic exchange. Mol. Ecol. 13, 3293–3303.

Additional papers provided during the workshop

Canales-Aguirre, C.B., Ferrada-Fuentes, S., Galleguillos, R. and Hernández C.E. 2014. Population geographic structure, demographic history, and migrations of Patagonian

toothfish (*Dissostichus eleginoides*) in the Southeast Pacific and Southwest AtlanticOcean.PlosOne. (in Review).
Falkland Islands Government, (2013).Fisheries Department Fisheries Statistics, Volume 17, 2012: 72 pp.Stanley, FIG Fisheries Department
Martinez P.A., Waessle, J.A., Wöhler, O.C.and Giussi, A.R. 2014. Síntesis del Programa de Marcado y Recaptura de Merluza Negra (*Dissostichus eleginoides*) en el Atlántico Sudoccidental y de los resultados obtenidos desde su implementación en el año 2004 hasta diciembre de 2013.Informe de investigación.INIDEP.

Presentations

Monday
1.- RenzoTascheri, Life history and parameters used in the stock assessment
2.- Dario Rivas, Chilean Sea Bass Fishery
3.- Sandra Ferrada,Chilean Sea Bass Genetics
4.- Carlos Moreno, Chilean Sea Bass spatial structure

Tuesday
1.- Liu Chong, Chilean sea bass monitoring program
2.- Renato Cespedes, Chilean sea bass monitoring program PDA, industrial fleet.
3.-Vilma Ojeda, Chilean sea bass Aging
4.- Manuel Gónzalez, Chilean fishing statistic, Chilean sea bass

Wednesday
1.-Renzo Tascheri, Stock assessment model
2.- Renzo Tascheri and Cristian Canales, Status and sustainable exploitation of Chilean sea bass.
3.-CiroOyarzún, Comments on stock assessments from 2013 to 2014, Scientific Committee.

Thursday
1.- Alejandro Zuleta, Cpue Standardization Problems in TOP Chilean Fishery.
2.- Pedro Rubilar, Toothfish tagging program result and perspectives.
3.- Renzo Tascheri, Stock assessment model.

Friday
1.- Renzo Tascheri, Runs report

Appendix 2: A copy of the Statement of Work

Scope of Work:

The reviewer is contracted to deliver an independent external review of the 2014 stock assessment of Chilean Sea bass "*bacalao de profundidad*" conducted by the Instituto de Fomento Pesquero (IFOP) and participate in the Chilean Sea bass stock assessment external review workshop.

Project Description:

The Undersecretariat of Fisheries of Chile started in 2011 an independent peer review process to assess the soundness of IFOP stock assessment approach for their major fisheries. This year the Chilean Undersecretariat of Fisheries requested an international independent peer review for Chilean Sea Bass (Bacalao de profundidad) and Nylon Shrimp (camarón nylon).

Location of Peer Review:

The peer review process will be held in Valparaíso (Hotel Diego de Almagro) during a workshop in November of 2014.

Statement of Tasks:

The reviewer shall complete the following tasks in accordance with the terms of reference and specific tasks and deliverables herein.

Pre-review Background Documents: Two weeks before the workshop, the project coordinator, Dr. Billy Ernst will provide the reviewer the necessary background information (working documents in English, stock assessment model, etc.) for the workshop preparation. This material shall consist of stock assessment documents translated to English and other background material. The reviewer/collaborator shall read all documents in preparation for the workshop.

Workshop: The reviewer shall actively participate in the workshop and conduct an independent review of the assessment addressing each Term of Reference.

Specific Tasks for the Reviewer:

The following chronological list of tasks shall be completed by the reviewer in a timely manner.

1. Conduct necessary pre-review preparations, including all the background material and reports provided by the University of Concepcion in advance to the workshop.
2. Actively participate during the workshop in Valparaíso (November 2014), conduct a review and generate a report with recommendations in accordance with the ToRs (**Annex 2**).
3. Participate in the nylon shrimp workshop and provide feedback (afternoons).

3. No later than the December 10th of 2014, submit a report addressed to the "Department of Oceanography, University of Concepción" and sent to Dr. Billy Ernst, Lead Coordinator of the project, via email to biernst@udec.cl. The report shall be written in English addressing each ToR in **Annex 2**.
4. The report (in English) shall include an executive summary with the main conclusions and recommendations about the review process. The main text shall include a description of activities, findings, conclusions and recommendations. An Annex shall include the terms of references, statement of work and the list of references used in the review process. This should be a

Acceptance of Deliverables:

Upon review and acceptance of the independent peer review reports by the Project Lead Coordinator (Dr. Billy Ernst), these reports shall be sent to the Undersecretariat of Fisheries for final approval as contract deliverables based on compliance with the Statement of Work and Terms of reference. As specified in the Schedule of Milestones the University of Concepción shall send via e-mail the contract deliverables (Independent peer review reports) to the Undersecretariat of Fisheries project coordinator.

Schedule of Milestones and Deliverables:

The group shall complete the tasks and deliverables described in this Statement of work in accordance with the following schedule.

October 2014	Project coordinator sends the Reviewer the pre-review documents.
November 2014	The reviewer participates and conducts an independent peer review during the panel review meeting in Valparaíso.
December 2014	Reviewers submit the independent peer review reports to the project lead coordinator.
January 2015	University of Concepción translates and submits Independent peer review reports to the Undersecretariat of Fisheries.
January 2015	Review and approval of Peer review reports upon compliance with Statement of Work and Terms of References

Applicable Performance Standards:

The contract is successfully completed when the Chilean Undersecretariat of Fisheries project coordinator provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:

(1) The report shall be completed with the format and content in accordance with **Annex 1**,
(2) The report shall address each ToR as specified in **Annex 2**,
(3) The reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Support Personnel:

Project Coordinator:

Dr. Billy Ernst
Department of Oceanography
University of Concepción
Barrio Universitario s/n
Concepción, Chile.
Phone: +56-41-2204012
biernst@udec.cl

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

1. Each independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

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 review meeting, providing a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewer should elaborate on any points raised in the Executive Summary that they feel might require further clarification.

   c. The independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.

4. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review

   Appendix 2: A copy of the Statement of Work

   Appendix 3: Panel Membership or other pertinent information from the review meeting.

Annex 2 Chilean Sea Bass

Provide feedback, recommendations and an independent peer review on:

1. To critically review the stock assessment approach, including underlying hypotheses of model structure and the conceptual model (regional versus global assessment) and the structure of the stock assessment model. Comment on potential improvements.

2. To critically review the life history parameters used in the assessment, with special emphasis on recruitment, growth, age-at-first maturity and natural mortality.

3. Comment on the consistency in the use of age information coming from scale and otoliths readings.

4. To review the quality and reliability of different pieces of information and estimation approaches used in the monitoring of the fishery, including CPUE and catch. Assess if the effect of the incorporation of the "*cachalotera*" has been properly addressed in the stock assessment model.

5. Do the selected case studies represent the main axis of model uncertainty and comment on the base case chosen for this assessment?

6. To review the configuration of the stock assessment model, check if it is properly implemented and assessed its performance based on additional model runs requested during the workshop.

7. To comment on the biological reference points, the indicators used for this fishery and the determination of the stock status. Make recommendations.

8. To comment on the procedures used for projecting the stock into the future, in particular to comment on the robustness of the risk analysis to assess the risk of not achieving the desired objectives.

9. Recommend improvements to the assessment process (studies and research programs) which may ultimately lead to a reduction in stock status uncerta

Appendix 3: Panel Membership and other pertinent information from the review meeting.

Review team
Dr. Catherine Dichmont (CSIRO) – Nylon Shrimp
Dr. Tom Polacheck (Independent Consultant) – Chilean Sea Bass

Project co-ordinator and Chair
Dr. Billy Ernst

N°	Name		Institution
1	Adasme	Luis	IFOP
2	Arana	Patricio	PUCV
3	Bucarey	Doris	IFOP
4	Canales	Cristian	IFOP
5	Chong	Liu	IFOP
6	Ferrada	Sandra	UdeC
7	Guzman	Oscar	IFOP
8	Leal	Elson	IFOP
9	Ojeda	Vilma	IFOP
10	Oyarzún	Ciro	CCT
11	Rivas	Dario	SubPesca
12	Rubilar	Pedro	CEPES
13	Tascheri	Renzo	IFOP
14	Uriarte	Manuel	
15	Wiff	Rodrigo	
16	Zuleta	Alejandro	CEPES
16	Mermoud	Nicole	UDEC

Appendix 4: Summary of Assessment Runs performed during the Review Workshop

Reviewer Notes:

This appendix contains the summary tables for the runs performed during the review workshop and provided to the participants on the last day of the workshop. All runs in this set expect for runs 8 and 9 were run with the aging error matrix switched on (see above for how this was applied). Run 1 was intended to replicate the results in the Assessment Report (but in fact the results presented in the Assessment Report as described in Appendix 5 had the matrix switched off). This run used the aging matrix in Appendix Table 4.1. This was apparently the intended matrix which was meant to be applied in fitting the model, although the Assessment Report reports that a different table was the one intended to be used (Table 8 in the Assessment Report). All other runs in which the aging matrix was applied used a corrected version of Appendix Table 4.1 in which the columns have been normalized to sum to 1 (see above). This corrected aging error matrix is provided in Appendix Table 4.2. This same aging error matrix (Appendix Table 4.2) was used in the calculations for the runs in Appendix 5 in which the aging error matrix was switched on.

Note that the specification of “no changes in catchability” in this appendix differs from that in Appendix 5 and as such the runs in which this is specified here (Runs 11, 16 and 19) are not comparable with those in Appendix 5. In the runs here, there was a lack of clarity about what the intended specification was and how changes in q were implemented in the computer code. The intention was that within each of the 3 input CPUE series no changes in q were to be allowed, but that each series would have a separate estimated q. However, for the two CPUE series from the Chilean Industrial Fleet, the code internally treats these as a single series with separate q’s for each. When no changes in catchability was specified, this was implemented so that no change was allowed in q for the two different Chilean CPUE series. The q estimated in this context makes no logical sense as the two series use different measures of effort which are incomparable in scale (hooks versus a set). The results from these three runs should be disregard. The results provide no meaningful estimates of the stock status and are not meaningfully comparable with any of the other run results.

Appendix Table 4.1: The aging error matrix that was intended to have been used in the production of the results in the Assessment Report and the matrix which was used in the production of results in Run 1 in this Appendix. (see above for more detail).

1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.7	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.3	0.8	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.8	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.7	0.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.5	0.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.5	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.5	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.5	0.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.4	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.4	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3	0.4	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.4	0.3	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.3	0.3	0.2	0.1	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3	0.3	0.2	0.1	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3	0.3	0.3	0.1	0.1	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.3	0.3	0.2	0.1	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.3	0.3	0.2	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.3	0.3	0.2
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.2	0.3
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.2
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Appendix Table 4.2: The “corrected” aging error matrix used in all of the runs in this Appendix and Appendix 5 in which an aging error was applied with the exception of Run 1 in this appendix. This Run used Appendix Table 4.1.

1.000	0.999	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.001	1.000	0.034	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.966	0.306	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.694	0.650	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.348	0.794	0.106	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.181	0.788	0.251	0.019	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.106	0.679	0.404	0.060	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.533	0.508	0.131	0.015	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.044	0.401	0.547	0.219	0.040	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.297	0.532	0.302	0.080	0.013	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.219	0.483	0.366	0.132	0.029	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.162	0.420	0.402	0.189	0.054	0.011	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.121	0.355	0.411	0.242	0.087	0.022	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.091	0.295	0.399	0.284	0.126	0.039	0.009	0.002	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.070	0.242	0.373	0.313	0.165	0.062	0.018	0.004	0.001	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.054	0.198	0.339	0.327	0.202	0.089	0.030	0.008	0.002	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.042	0.161	0.303	0.328	0.232	0.118	0.046	0.015	0.004	0.001
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.034	0.132	0.266	0.319	0.254	0.146	0.066	0.024	0.008
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.027	0.108	0.232	0.303	0.268	0.173	0.087	0.036
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.022	0.089	0.200	0.283	0.274	0.195	0.109
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.018	0.073	0.173	0.260	0.273	0.213
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.015	0.061	0.148	0.236	0.266
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.012	0.051	0.127	0.213
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.010	0.043	0.109
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.009	0.036
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.008
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Chilean stock assessment of Patagonian toothfish
Summary of the runs made for <u>Case 1</u> (industrial fleet + Argentinian longline and trawl fleet).

The situations analyzed were:
1.Base case (run without changing anything)
2.Base case 2 (aging-error matrix normalized to sum zero)
3. M=0.1
4. M=0.2
5. CPUE Argentina with a cv=0.4
6. Argentinean catch age compositions n=10
7. All age comps n=10
8 No age correction
9. No age correction and Argentinean catch age compositions n=10
10 Argentinean CPUE with a cv=0.3
11. No changes in catchability
12. Argentinean trawl fleet with dome shape selectivity.
13. No selectivity changes over time. All fleets.
14. Estimated M symmetrical bounds
15 Estimated M asymmetrical bound
16 No catchability or selectivity changes over time
17 All CPUE with cv=0.3
18 All age comps N=25
19 A combination of 16, 17, 18 and 19: No catchability or selectivity changes over time, All CPUE with cv=0.3, All age comps N=25.

Table 1. Likelihood components Model Case 1 (Base case).

	Likelihood Components										
Cases/Like.components	Age comps Chile	Age comps Ar 1	Age comps Ar 2	Cpue Chile 1	cpue Chile 2	cpue Ar.	Catch Chile	Catch Ar. 1	Catch Ar. 2	Recruit . devs.	Initial condition
1	2787	1343	1277	23.2	2.8	24.6	0.152	0.016	0.005	2.94	7.86
2	2786	1343	1277	23.3	2.7	24.6	0.150	0.016	0.005	2.90	7.66
3	2792	1342	1278	22.3	2.3	22.7	0.211	0.023	0.008	1.83	8.91
4	2782	1344	1277	24.5	3.2	26.5	0.100	0.010	0.004	4.34	8.26
5	2783	1342	1278	23.4	4.7	8.1	0.058	0.006	0.002	1.83	6.22
6	2784	269	256	24.0	2.6	23.5	0.110	0.012	0.005	2.94	7.70
7	561	270	255	22.4	1.6	19.3	0.072	0.011	0.004	3.80	4.40
8	2803	1343	1277	25.2	1.5	28.9	0.192	0.033	0.009	2.64	7.61
9	2800	270	263	25.0	1.1	25.6	0.101	0.025	0.010	3.17	8.02
10	2784	1342	1277	23.2	4.0	13.1	0.078	0.008	0.003	2.09	6.68
11	2824	1345	1277	91.0	114.4	54.3	0.974	0.051	0.024	1.10	4.63
12	2785	1343	1272	24.0	1.5	23.7	0.146	0.013	0.007	3.59	7.93
13	2799	1342	1344	23.7	5.8	22.6	0.224	0.027	0.009	8.33	11.24
14	2787	1343	1277	23.2	2.6	24.5	0.154	0.016	0.006	2.80	7.66
15	2784	1344	1277	24.0	2.9	25.7	0.120	0.012	0.005	3.72	7.87
16	2824	1348	1328	112.9	133.6	67.7	0.754	0.050	0.033	2.65	3.48
17	2782	1342	1278	10.4	2.2	14.1	0.049	0.005	0.002	1.88	6.42
18	1397	672	638	23.0	1.9	21.9	0.094	0.012	0.004	3.30	5.85
19	1412	673	665	51.8	60.6	29.5	0.152	0.010	0.006	1.69	1.86

Table 2. Parameters and other relevant quantities. Model Case 1 (Base case).

Case/quantity	Depletion 2013	SSB 2013	Bo	Max.Grad.	Obj.Fun.	M
1	0.15	13955	90632	1.66E-05	5468	0.15
2	0.15	13593	89193	7.11E-06	5468	0.15
3	0.08	10413	123288	3.60E-05	5470	**0.1**
4	0.24	18702	79198	7.88E-06	5470	**0.2**
5	0.23	22303	97488	7.19E-05	5447	0.15
6	0.15	13409	90890	8.77E-06	3370	0.15
7	0.09	8641	94044	4.86E-06	1137	0.15
8	0.10	7171	68971	1.88E-06	5489	0.15
9	0.09	6533	69380	1.56643	3396	0.15
10	0.20	19047	94467	1.28E-05	5453	0.15
11	0.40	42540	107635	3.91E-06	5712	0.15
12	0.12	10331	87444	2.25E-05	5461	0.15
13	0.16	12598	78313	4.98E-06	5557	0.15
14	0.15	13328	90633	9.11E-06	5468	**0.146**
15	0.20	63953	81508	1.03E-05	5469	**0.178**
16	0.58	284162	127487	1.43E-05	5822	0.15
17	0.23	82048	96865	3.29E-06	5437	0.15
18	0.12	45906	91688	1.13E-05	2763	0.15
19	0.55	280731	131618	6.69E-06	2896	0.15

Appendix 5: Summary of Revised Assessment Runs provided to the Reviewer in Mid-December, 2014

Reviewer Notes:

This appendix contains the summary table for the revised assessment runs performed after the review workshop and provided to the reviewer in mid-December. All the runs with the exception of Run 2 were intended to have been run with the aging error matrix switched off (see note from the Assessment Report author below). However, Run 12 was also run with the aging error matrix switched on for the years 1989-2006, as in Run 2. Thus, Run 2, and not Run 1, is directly comparable with Run 12 in terms of allowing M to be estimated within the model. Also the values in Table 1 for Run 8 have been mistakenly tabled and are simply a copy of those for Run 7. The aging error matrix applied in this appendix corresponds to that provided in Appendix Table 4.2.

Note that "no catchability changes" in runs 9, 13 and 16 means that no changes in catchability or the "q" parameter were allow within each of the three CPUE used in fitting the model. However, a separate q was estimated for each CPUE series (see above for more detail).

December 15, 2014

Chilean stock assessment of Patagonian toothfish

Summary of the runs made for <u>Case 1</u> (industrial fleet + Argentinean Longline and Trawl fleet).

Note from the assessment author

A series of 19 runs for Case 1 of the Chilean toothfish stock assessment, as it was informed in the August 2014 report to the Chilean Undersecretary of Fisheries, was performed during the peer review workshop held in November 2014. The first run (Run 1), was intended to be the base case (nothing should be changed, so it will correspond to the original case presented in the August report), this turned out not to be the case.

By the time of the peer review, several new cases had been analyze after the August report, and even a second report was issued in September, in the context of the Chilean process of advice on biological acceptable levels of catch and as could be expected, a number of runs on each of these new cases had already been performed. Thus during the peer review, the author used one of the more up to date versions of the model code, not noticing that the actual numbers in the results for "Run 1" were not matching those included in the August report.

The following is the report of a new rerun of the same cases done in the peer review workshop, this time using a version of the code that do recreates the results of the original report for Run 1.

In the Chilean toothfish 2014 stock assessment, a matrix of empirical probabilities was used to account for the age reading error that stemmed from the fact that until 2006 age readings were performed using the fish scales although this data was produced using the otoliths afterwards. Accordingly when the matrix was used, during run time it was switched on for years 1989 to 2006 and switched off for the rest of the years included in the assessment. It turned out, that when the results of the August report were produced, the reading error matrix was switch off. Therefore, the runs of the model performed during the peer review workshop were rerun and the results are presented in this report. Given the circumstances described here, all the runs were done again with the reading error matrix switched off except for Run 2.

Because of the aforementioned, the original Runs 8 and 9 requiring switching off the reading error matrix were not included in the runs performed for this report.

The request of the reviewer to redo the original Runs 11, 16 and 19, allowing for the two Chilean CPUE series to have different values for catchability, was considered in this new series of runs.

Results

The situations analyzed were:

1. Base case (run without changing anything. Aging-error matrix inactive !)
2. Base case 2 (aging-error matrix normalized to sum zero)
3. M=0.1
4. M=0.2
5. CPUE Argentina with a cv=0.4
6. Argentinean catch age compositions n=10
7. All age comps n=10
8. Argentinean CPUE with a cv=0.3
9. No changes in catchability
10. Argentinean trawl fleet with dome shape selectivity.
11. No selectivity changes over time. All fleets.
12. Estimated M
13. No catchability or selectivity changes over time
14. All CPUE with cv=0.3
15. All age comps N=25

16 A combination of 16, 17, 18 and 19: No catchability or selectivity changes over time, All CPUE with cv=0.3, All age comps N=25.

Table 1. Likelihood components Model Case 1 (Base case).

	Likelihood Components										
Cases/Like.components	Age comps Chile	Age comps Ar 1	Age comps Ar 2	Cpue Chile 1	cpue Chile 2	cpue Ar.	Catch Chile	Catch Ar. 1	Catch Ar. 2	Recruit . devs.	Initial condition
1	2803	1343	1277	25.2	1.5	28.9	0.192	0.033	0.009	2.64	7.61
2	2786	1343	1277	23.3	2.7	24.6	0.150	0.016	0.005	2.90	7.66
3	2809	1343	1277	23.3	1.4	26.9	0.256	0.042	0.011	2.72	10.86
4	2798	1344	1277	27.1	1.6	30.1	0.127	0.021	0.007	3.68	6.96
5	2799	1342	1277	26.4	2.2	8.6	0.076	0.008	0.002	2.18	7.96
6	2800	270	256	25.2	1.3	27.1	0.140	0.031	0.007	2.64	7.60
7	569	271	255	21.1	1.4	18.0	0.051	0.015	0.005	2.29	2.18
8	569	271	255	21.1	1.4	18.0	0.051	0.015	0.005	2.29	2.18
9	2814	1353	1273	52.9	3.0	29.1	0.610	0.157	0.026	1.97	5.14
10	2797	1342	1274	25.7	1.2	29.3	0.180	0.030	0.008	3.25	7.87
11	2810	1341	1335	25.8	3.2	31.5	0.164	0.028	0.011	11.72	10.65
12	2784	1344	1277	24.0	2.9	25.7	0.120	0.012	0.005	3.72	7.87
13	2829	1344	1329	60.1	5.9	38.6	0.565	0.122	0.061	7.93	6.92
14	2797	1342	1277	11.8	1.6	15.6	0.073	0.009	0.002	2.12	8.05
15	1408	673	638	23.5	1.4	24.5	0.113	0.024	0.007	2.28	4.18
16	1416	672	667	27.6	3.3	16.6	0.101	0.022	0.011	4.23	4.33

Table 2. Parameters and other relevant quantities. Model Case 1 (Base case).

Case/quantity	Depletion 2013	SSB 2013	Bo	Max.Grad.	Obj.Fun.	M
1	0.1	7171	68971	0.000005	5489	0.15
2	0.15	13593	89192	0.000005	5468	0.15
3	0.06	5942	101092	0.000008	5494	**0.1**
4	0.16	9010	56365	0.000013	5489	**0.2**
5	0.15	10380	70310	0.000053	5465	0.15
6	0.09	6627	70190	0.000007	3390	0.15
7	0.07	5800	78084	0.000008	1139	0.15
8	0.12	8640	69274	0.000007	5472	0.15
9	0.13	9838	74549	0.000011	5534	0.15
10	0.1	7171	68581	0.000006	5481	0.15
11	0.15	9149	59348	0.000024	5569	0.15
12	0.2	16059	81508	0.00003	5469	**0.18**
13	0.19	12611	67526	0.000033	5623	0.15
14	0.16	11452	70671	0.000008	5456	0.15
15	0.09	6634	74603	0.000009	2774	0.15
16	0.21	15183	73719	0.000009	2811	0.15

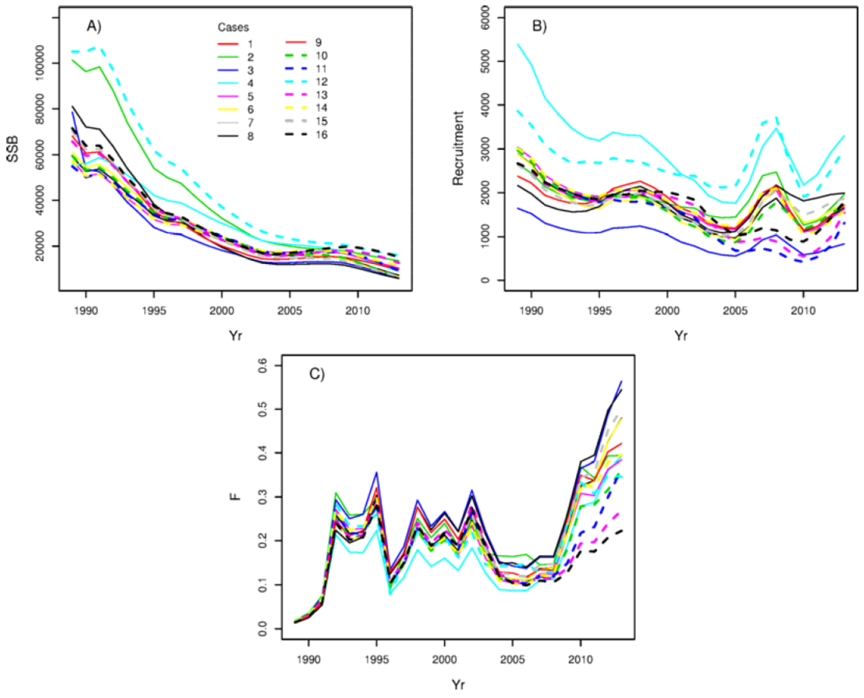


Figure 1. Spawning biomass (A), recruitment (B) and fishing mortality (C) series corresponding to 16 runs of the Chilean toothfish stock assessment model for Case 1, informed in the August 2014 report to the Chilean Undersecretary of Fisheries.

Appendix 6: Comparison of Mean Length-at-Age-and Mean Weight-at-Age Estimates

In this appendix, the reviewer explored the length-at-age and weight-at-age implication for the some of the different parametric growth curves and weight-at-length relationships that have been estimated for toothfish. The estimates for the parameters of the growth curves (VBG) were taken from Table 2 of the Assessment Report for those estimates that had been derived for either Chile or the Patagonian Shelf areas. These parameter estimates are provided here as Table Appendix 6-1. The expected mean length-at-age based on these parameter estimates for ages 1-29 are provided in Appendix Table 6-2. Parameters for the weight-length relationship were taken from Gálvez et al (2013) and provided in Table Appendix 6-3. Table Appendix 6-4 provides estimates of the mean length-at-age for the set of length-at-age estimates in Appendix Table 6-2 crossed with the weight-length parameters in Appendix Table 6-3. Note that the graphs in the main body of the text are based on the weight-length parameters for both sexes combined in Appendix Table 6-3 (i.e. the graph are based on the numbers shown in Appendix Table 6-4c).

Table Appendix 6-1: The set of VBG parameter values from Table 2 of the Assessment Report that were derived from data collected from Chile and the Patagonian Shelf Area. These values were used in this appendix to provide an indication of the range of values for the mean length-at-age and mean weight-at-age that would be reasonable to expect for these within the stock Assessment

Parm.	Parameter Set									
	1	2	3	4	5	6	7	8	9	10
Linf	141.4	120.7	184.7	209.7	195.6	177.8	218	152.2	210.8	212.6
K	0.15	0.13	0.065	0.064	0.0742	0.109	0.048	0.085*	0.064	0.066
Tzero	-1.1	-1.55	0.386	-1.151	-0.721	0.00	-0.066	-0.59	-0.432	-0.477

*In the Assessment Report the tabled value was 0.85. This value was totally inconsistent with any of the other values of k. It also yielded extreme and unrealistic values for the mean length at age. The reviewer assumed that a typo occurred and that the intended value was 0.085

Table Appendix 6-2: Estimate of the mean length-at-age (cm) based on the VBG parameter sets in Table Appendix 6.1.

	Parameter Set									
Age	1	2	3	4	5	6	7	8	9	10
1.0	38.2	34.1	7.2	27.0	23.4	18.4	10.9	19.2	18.5	19.7
2.0	52.6	44.6	18.4	38.3	35.8	34.8	20.6	30.1	30.4	32.1
3.0	65.0	53.9	28.9	48.9	47.2	49.6	29.8	40.0	41.6	43.6
4.0	75.6	62.0	38.7	58.9	57.8	62.8	38.7	49.2	52.1	54.4
5.0	84.8	69.2	47.9	68.2	67.7	74.7	47.1	57.6	61.9	64.5
6.0	92.7	75.5	56.5	77.0	76.8	85.4	55.1	65.3	71.1	74.0
7.0	99.4	81.0	64.5	85.2	85.3	94.9	62.7	72.4	79.8	82.8
8.0	105.3	85.8	72.1	93.0	93.2	103.5	70.0	78.9	87.9	91.1
9.0	110.3	90.1	79.2	100.2	100.5	111.1	76.9	84.8	95.5	98.9
10.0	114.6	93.8	85.8	107.0	107.3	118.0	83.5	90.3	102.7	106.1
11.0	118.4	97.1	92.1	113.3	113.6	124.2	89.8	95.4	109.4	112.9
12.0	121.6	100.0	97.9	119.3	119.5	129.7	95.8	100.0	115.7	119.3
13.0	124.3	102.5	103.3	124.9	124.9	134.7	101.6	104.3	121.6	125.2
14.0	126.7	104.7	108.5	130.2	130.0	139.1	107.0	108.2	127.1	130.8
15.0	128.8	106.7	113.3	135.1	134.7	143.1	112.2	111.8	132.3	136.1
16.0	130.5	108.4	117.8	139.7	139.0	146.7	117.2	115.0	137.2	140.9
17.0	132.0	109.9	122.0	144.1	143.1	149.9	121.9	118.1	141.7	145.5
18.0	133.3	111.2	125.9	148.1	146.8	152.8	126.4	120.9	146.0	149.8
19.0	134.5	112.4	129.6	152.0	150.3	155.4	130.7	123.4	150.0	153.8
20.0	135.4	113.4	133.1	155.5	153.6	157.7	134.8	125.8	153.8	157.6
21.0	136.3	114.3	136.3	158.9	156.6	159.8	138.7	127.9	157.3	161.1
22.0	137.0	115.0	139.4	162.0	159.4	161.6	142.4	129.9	160.6	164.4
23.0	137.6	115.7	142.2	165.0	162.0	163.3	146.0	131.7	163.7	167.5
24.0	138.1	116.3	144.9	167.8	164.4	164.8	149.3	133.4	166.7	170.3
25.0	138.6	116.9	147.4	170.4	166.6	166.1	152.5	134.9	169.4	173.0
26.0	139.0	117.3	149.8	172.8	168.7	167.3	155.6	136.3	172.0	175.6
27.0	139.3	117.7	152.0	175.1	170.6	168.4	158.5	137.6	174.4	177.9
28.0	139.6	118.1	154.0	177.2	172.4	169.4	161.3	138.8	176.6	180.1
29.0	139.9	118.4	155.9	179.3	174.0	170.3	164.0	139.9	178.8	182.2

Table Appendix 6.3: Estimates of the range of weight at length parameters used to estimate the range of mean weight at age values that are consistent with the range mean length at age in Table Appendix 6.2. Values taken from Gálvez et al (2013).

	a	B
Males	0.006754	3.096425
Females	0.005498	3.144316
Both	0.006204	3.116188

Table Appendix 6-4: Estimates of the mean weight-at-age using the mean length-at-age estimates in Table Appendix 6-2 and the parameter set for the weight-length relationship in Table Appendix 6-3.

Table Appendix 6-4a: for males

Age	Parameter Set 1	2	3	4	5	6	7	8	9	10
1	0.5	0.4	0.0	0.2	0.1	0.1	0.0	0.1	0.1	0.1
2	1.4	0.9	0.1	0.5	0.4	0.4	0.1	0.3	0.3	0.3
3	2.8	1.6	0.2	1.2	1.0	1.2	0.2	0.6	0.7	0.8
4	4.4	2.4	0.6	2.0	1.9	2.5	0.6	1.2	1.4	1.6
5	6.3	3.4	1.1	3.2	3.1	4.3	1.0	1.9	2.4	2.7
6	8.3	4.4	1.8	4.7	4.7	6.4	1.7	2.8	3.7	4.1
7	10.3	5.5	2.7	6.4	6.4	9.0	2.5	3.9	5.2	5.9
8	12.4	6.6	3.8	8.4	8.5	11.7	3.5	5.0	7.1	7.9
9	14.3	7.6	5.1	10.6	10.7	14.6	4.7	6.3	9.1	10.2
10	16.1	8.6	6.6	13.0	13.1	17.6	6.0	7.7	11.4	12.7
11	17.8	9.6	8.1	15.5	15.6	20.6	7.6	9.1	13.9	15.3
12	19.3	10.5	9.9	18.2	18.3	23.6	9.2	10.5	16.5	18.2
13	20.7	11.4	11.7	21.0	21.0	26.5	11.0	12.0	19.3	21.1
14	21.9	12.1	13.5	23.8	23.7	29.3	13.0	13.4	22.1	24.2
15	23.0	12.9	15.5	26.7	26.5	32.0	15.0	14.9	25.0	27.3
16	24.0	13.5	17.5	29.7	29.2	34.5	17.2	16.3	28.0	30.5
17	24.9	14.1	19.5	32.6	31.9	36.9	19.4	17.6	31.0	33.6
18	25.7	14.6	21.5	35.6	34.6	39.1	21.8	18.9	34.0	36.8
19	26.3	15.1	23.5	38.5	37.2	41.2	24.1	20.2	37.0	39.9
20	26.9	15.5	25.5	41.3	39.7	43.1	26.5	21.4	39.9	43.0
21	27.4	15.9	27.5	44.2	42.2	44.9	29.0	22.6	42.8	46.1
22	27.9	16.3	29.4	46.9	44.6	46.6	31.5	23.7	45.7	49.1
23	28.3	16.6	31.3	49.6	46.9	48.1	34.0	24.7	48.5	52.0
24	28.6	16.8	33.2	52.3	49.0	49.5	36.4	25.7	51.2	54.8
25	28.9	17.1	35.0	54.8	51.1	50.7	38.9	26.6	53.9	57.5
26	29.2	17.3	36.8	57.3	53.1	51.9	41.4	27.5	56.4	60.2
27	29.4	17.5	38.5	59.7	55.0	52.9	43.9	28.3	58.9	62.7
28	29.6	17.6	40.1	62.0	56.8	53.8	46.3	29.1	61.3	65.1
29	29.7	17.8	41.7	64.2	58.6	54.7	48.7	29.8	63.6	67.5

Table Appendix 6-4b: for females

	Parameter Set									
Age	1	2	3	4	5	6	7	8	9	10
1	0.5	0.4	0.0	0.2	0.1	0.1	0.0	0.1	0.1	0.1
2	1.4	0.8	0.1	0.5	0.4	0.4	0.1	0.2	0.3	0.3
3	2.8	1.5	0.2	1.1	1.0	1.2	0.2	0.6	0.7	0.8
4	4.4	2.4	0.5	2.0	1.9	2.5	0.5	1.1	1.4	1.6
5	6.4	3.4	1.1	3.2	3.1	4.3	1.0	1.9	2.4	2.7
6	8.4	4.4	1.8	4.7	4.7	6.5	1.6	2.8	3.7	4.1
7	10.5	5.5	2.7	6.5	6.5	9.1	2.5	3.9	5.3	5.9
8	12.6	6.6	3.8	8.5	8.6	11.9	3.5	5.1	7.1	8.0
9	14.6	7.7	5.1	10.7	10.9	14.9	4.7	6.4	9.3	10.3
10	16.4	8.7	6.6	13.2	13.3	18.0	6.1	7.8	11.6	12.9
11	18.2	9.7	8.2	15.8	16.0	21.1	7.6	9.2	14.2	15.7
12	19.8	10.7	10.0	18.6	18.7	24.2	9.3	10.7	16.9	18.6
13	21.2	11.5	11.9	21.5	21.5	27.3	11.2	12.2	19.7	21.7
14	22.5	12.4	13.8	24.5	24.4	30.2	13.2	13.7	22.7	24.9
15	23.7	13.1	15.8	27.5	27.2	33.0	15.4	15.2	25.8	28.1
16	24.7	13.8	17.9	30.6	30.1	35.7	17.6	16.6	28.9	31.4
17	25.6	14.4	20.0	33.7	33.0	38.2	19.9	18.0	32.0	34.8
18	26.4	14.9	22.1	36.8	35.8	40.5	22.3	19.4	35.1	38.1
19	27.1	15.4	24.2	39.8	38.5	42.7	24.8	20.7	38.3	41.4
20	27.7	15.9	26.2	42.9	41.2	44.8	27.3	22.0	41.4	44.6
21	28.3	16.3	28.3	45.8	43.8	46.6	29.9	23.2	44.4	47.8
22	28.7	16.6	30.4	48.7	46.3	48.4	32.5	24.3	47.4	51.0
23	29.1	16.9	32.3	51.6	48.7	50.0	35.1	25.4	50.4	54.0
24	29.5	17.2	34.3	54.4	51.0	51.4	37.7	26.4	53.3	57.0
25	29.8	17.4	36.2	57.1	53.2	52.7	40.3	27.4	56.1	59.9
26	30.1	17.7	38.0	59.7	55.3	53.9	42.9	28.3	58.8	62.7
27	30.3	17.9	39.8	62.2	57.3	55.0	45.5	29.2	61.4	65.4
28	30.5	18.0	41.5	64.6	59.2	56.0	48.1	30.0	63.9	68.0
29	30.7	18.2	43.2	67.0	61.0	57.0	50.6	30.7	66.4	70.5

Table Appendix 6-4c: combined

Age	Parameter Set 1	2	3	4	5	6	7	8	9	10
1	0.5	0.4	0.0	0.2	0.1	0.1	0.0	0.1	0.1	0.1
2	1.4	0.9	0.1	0.5	0.4	0.4	0.1	0.3	0.3	0.3
3	2.8	1.5	0.2	1.1	1.0	1.2	0.2	0.6	0.7	0.8
4	4.4	2.4	0.5	2.0	1.9	2.5	0.5	1.2	1.4	1.6
5	6.3	3.4	1.1	3.2	3.1	4.3	1.0	1.9	2.4	2.7
6	8.4	4.4	1.8	4.7	4.7	6.5	1.7	2.8	3.7	4.1
7	10.4	5.5	2.7	6.4	6.5	9.0	2.5	3.9	5.2	5.9
8	12.4	6.6	3.8	8.4	8.5	11.8	3.5	5.1	7.1	7.9
9	14.4	7.6	5.1	10.7	10.8	14.7	4.7	6.3	9.2	10.2
10	16.2	8.7	6.6	13.1	13.2	17.8	6.0	7.7	11.5	12.7
11	17.9	9.7	8.2	15.7	15.8	20.8	7.6	9.1	14.0	15.5
12	19.5	10.6	9.9	18.4	18.5	23.8	9.3	10.6	16.7	18.4
13	20.9	11.4	11.7	21.2	21.2	26.8	11.1	12.1	19.5	21.4
14	22.2	12.2	13.6	24.1	24.0	29.7	13.1	13.5	22.4	24.5
15	23.3	13.0	15.6	27.1	26.8	32.4	15.2	15.0	25.3	27.6
16	24.3	13.6	17.6	30.0	29.6	35.0	17.4	16.4	28.4	30.9
17	25.2	14.2	19.7	33.1	32.3	37.4	19.6	17.8	31.4	34.1
18	26.0	14.7	21.7	36.0	35.1	39.7	22.0	19.1	34.5	37.3
19	26.7	15.2	23.8	39.0	37.7	41.8	24.4	20.4	37.5	40.5
20	27.3	15.7	25.8	42.0	40.3	43.8	26.9	21.6	40.5	43.7
21	27.8	16.1	27.8	44.8	42.8	45.6	29.4	22.8	43.5	46.8
22	28.2	16.4	29.8	47.7	45.3	47.3	31.9	23.9	46.4	49.8
23	28.6	16.7	31.8	50.4	47.6	48.8	34.4	25.0	49.3	52.8
24	29.0	17.0	33.6	53.1	49.8	50.2	37.0	26.0	52.0	55.7
25	29.3	17.2	35.5	55.7	52.0	51.5	39.5	26.9	54.7	58.5
26	29.5	17.4	37.3	58.3	54.0	52.7	42.0	27.8	57.4	61.2
27	29.8	17.6	39.0	60.7	56.0	53.8	44.5	28.6	59.9	63.8
28	30.0	17.8	40.7	63.0	57.8	54.7	47.0	29.4	62.4	66.3
29	30.1	17.9	42.3	65.3	59.6	55.6	49.5	30.2	64.7	68.7