Working paper for the IOTC Billfish Working Party July 2009

An Exploratory Age-Structured and Spatially Disaggregated "Stock Synthesis" Assessment of the Indian Ocean Swordfish Fishery 1950-2007

Dale Kolody
(Uglyface.Inc@gmail.com)
Hobart, Tasmania, Australia

Abstract

An exploratory Indian Ocean swordfish (*Xiphias gladius*) stock assessment using "Stock Synthesis 3" (SS3) software is presented for the Indian Ocean Tuna Commission (IOTC) Working Party on Billfish (WPB). The fish population is disaggregated by age, sex, and 4 regions, and iterated on a quarterly time-step from 1950-2007. The assessment attempts to integrate the available fisheries data from the Indian Ocean (catch in mass from 24 fleets, standardized CPUE from 3 nations (disaggregated into 9 longline fleets), size composition data from 18 fleets) and biological data from local and global biological research (e.g. on growth rates, stock structure, and migration rates). The assessment is described as "exploratory" because the Indian Ocean swordfish data and population biology are poorly understood at this time (plus the author is not an experienced user of SS3, and anticipated software problems given the short-time frame for the assessment). SS3 represents only one of several models that have been proposed for the 2009 WPB, but is unique in that it explicitly examines spatial questions. A series of specifications were compared to examine the implications of several key modelling assumptions:

- Conflicting influence of the different data sources,
- Stock-recruitment relationship steepness (and the implications of estimating recruitment deviates),
- Alternative growth rates, natural mortality, and maturity schedules,
- Different selectivity functions (number and functional form).

The main text of the paper describes model specifications that were explored prior to the WPB. These models identified a number of sensitivities and data conflicts for further discussion. Two attachments describe additional work that was undertaken during and after the WPB, in which preliminary stock status estimates are presented.

Attachment 1 describes additional models that were fit during the WPB. In these latter models, the Japanese CPUE series prior to 1995 in the south-west (SW) region was eliminated. Discussions at the WPB concluded that the drastic drop in Japanese CPUE around 1995 was likely an artefact. The drop was not very consistent with other fleets in the SW (La Reunion), and seemed to coincide with a strong shift in Japanese effort into the Mozambique channel (a spatial change that was not explicitly described in the catch rate standardization). A subset of these models was selected for representation in the WPB 2009 stock status summary. However, a couple of potential implementation errors were also identified at that time that could not be resolved during the WPB.

Attachment 2 describes a series of models that were developed after the WPB. Two important issues are discussed and resolved (errors related to reference point definitions and fleet-area assignments). The two errors had somewhat offsetting effects, which means that the corrected results are reasonably similar to the results reported to the WPB (slightly more optimistic in terms of MSY-related reference points, and slightly more pessimistic in terms of depletion estimates). A corrected and expanded set of models is presented in attachment 2. These results supersede the main text and attachment 1, and should be considered the most appropriate reference when future assessments are considered.

Introduction

The 2008 WPB report describes three different modelling approaches that were used for the assessment of the Indian Ocean swordfish fishery:

1. Spatially-aggregated surplus production model (age-aggregated, deterministic recruitment) – this model was used to estimate the depletion and production characteristics of the whole Indian Ocean, and provided continuity with the methods applied in 2006 (Nishida and Semba 2008).

2. Spatially disaggregated Pella-Tomlinson production model (age-aggregated, deterministic recruitment) – this 4 area model attempted to describe the apparent differential depletion across different regions in the Indian Ocean (and used life history considerations to bound the surplus production 'shape' parameters) (Kolody 2008). The better fit models were consistent with lower migration rates and higher depletion in the SW, but the data were also reasonably consistent with a highly mixed population. Numerical issues left some question about the general applicability of the model.

3. Spatially-aggregated, age- and sex-structured integrated analysis – this model represented the first attempt to include the swordfish size composition data from the Indian Ocean into an assessment (Sheng-Ping Wang, National Taiwan Ocean university, pers. comm.). Results were preliminary, and generally suggestive of a much less productive stock than the other two methods. The use of size composition data potentially adds another means of estimating fishing mortality effects on the population, and assists in the estimation of year-class strength. However, it also includes strong assumptions about temporally stable selectivity and the randomness of size sampling. The spatial aggregation prevents examination of area-specific effects.

The models described in this paper attempt to combine the best features from the models above to i) examine the impact of the fishery in different sub-areas, ii) use the size composition data to potentially extract additional information about recruitment variability and fishing mortality, and iii) compare the effects of alternative CPUE series from fleets that operate in the same area. This assessment is also undertaken with the understanding that additional models resembling 1 and 3 above will also be presented to the WPB in 2009.

There is a debate in many stock assessments about the value of spatial disaggregation. In this case, a spatial structure was adopted as shown in Figure 1. Some conceptual models are illustrated in Figure 2 and Figure 3 to contrast situations in which there may or may not be any advantage to using a spatially-disaggregated model. At the moment it is not clear which of these models is most appropriate for the Indian Ocean swordfish population, or whether the alternatives would have different management implications. The actual spatial structure represented by SS3 is shown in Figure 4 (e.g. there is spatial partitioning of the population in 4 regions, however, it is assumed that spawning and recruitment events are shared, with a constant proportion of young fish recruiting to each region regardless of the relative contribution of spawners from ach region).

There are some potential downsides to adopting the more complicated assessment framework, but the most obvious ones do not seem to be debilitating:

- Reliance on third party software – this is probably not a problem in this case because the developer (Rick Methot, NMFS) actively supports the Stock Synthesis software, it is widely used, very flexible and core features have presumably become more reliable over time as bugs have been identified and fixed. Source code is also distributed.

- Additional overheads for the user – the software is reasonably well documented, with many example applications available, and to date has proved easier to use than other similar products.

- Increased computation time – the function minimization uses the highly efficient AD Model Builder software. Initial applications of the highly disaggregated swordfish assessment suggest minimization in ~5 minutes for the simpler specifications (3.0 GHz processor).

- Over-parameterization – there is not enough data to estimate all of the potentially important processes. This is inevitable with virtually any complicated stock assessment, and swordfish data is limited relative to many other fisheries (i.e. no tags, stock structure poorly understood, growth rates uncertain, M unknown). The "realism" provided by extra detail should result in reduced model bias relative to simpler models, however reduced bias comes at the cost of increased parameter estimation variance, such that complicated models may not perform better than simpler models (depending on the purpose). However, with a sufficiently complicated model, one can represent important details, examine the sensitivity of the model to the inestimable quantities, prioritize further research for reducing these uncertainties, and develop management strategies that are robust to these uncertainties to the extent possible.

This assessment was undertaken with the ultimate goal of quantifying the impact of the fishery on the swordfish population, and illustrating the effects of alternative management actions which will help the IOTC meet its management objectives. However, this has not been achieved to satisfaction at this time. Given our current understanding of the fishery and data, this assessment should be viewed primarily as an exercise in exploration and prioritization of future work. Stock status estimates were generated in the process of developing this paper, but the greater value is in the groundwork for future assessments.

Software

The current model was implemented with Stock synthesis SS V3.03 (Methot 2009). Graphics are mostly from the R functions SSv3_plots (Google code: BETA May 13, 2009, Ian Stewart and Ian Taylor, NWFSC).

Data

Total catch in mass and catch length frequency distributions were provided by the IOTC secretariat, and disaggregated with the spatial structure shown in Figure 1. Catch and size composition are disaggregated into a total of 24 fleets (Table 1). Catch by area over time is shown in Figure 5.

Standardized Japanese CPUE series were provided by Nishida and Wang (2009, with undocumented updates) and Taiwanese series by Wang and Nishida (2009) for the period 1980-2007 (not all years are available for all series). Standardized CPUE series from La Reunion were also used (Francois Poisson, IFREMER, pers. comm.) for the period 1993-2000. These series are shown in Figure 6 and Figure 7.

The size composition data summed over all time periods is shown in Figure 8 and mean size over time is shown in Figure 9.

Other sources of biological data are described under model assumptions below.

Qualitative inferences from the data

Often the appropriateness of different approaches to stock assessment modelling are evident from simply inspecting the data, and certain unproductive paths can be avoided.

There are some concerns about the total catches for some of the fleets. This data is fundamental to almost every stock assessment, and it is almost pointless to try to quantify the impact of a fishery on the stock without good estimates of the magnitude of the catches. While the total catch data are not perfect, they are derived primarily from the industrial fleets in the Indian Ocean and are thought to be more reasonable than for the other billfish species. The mean size estimates derived from the quotient of the catch in mass divided by the catch in numbers, indicates that there are problems with some of these data for some fleets (not shown).

There appear to be some conflicts among the catch rates from different fleets that operate in the same area (Figure 6, Figure 7). If selectivities vary dramatically among the fleets, this could account for some of the conflicting trends, but since they are all longliners, this does not seem like an adequate explanation. The problem is probably due to catchability changes in one or more fisheries that cannot be accounted for properly in the catch rate standardization. If we choose to believe that all series are valid then the models will attempt to average the results. An alternative approach is to accept that one series is probably closer to the truth, and the other is likely to be wrong (e.g. Schnute and Hilborn 1993). Of course both series could be grossly wrong. In this case, (noting only the case 4 Taiwanese series is shown) the data suggest:

- consistent downward trend of JPN and TWN fleets over the last few years in the NW
- dramatic increase and decrease in the JPN series (particularly in the SW region), which could indicate good recruitment in the 1980s, followed by fishery impact or poor recruitment. However, it is worth noting that the La Reunion CPUE series shows only a very modest decline when the JPN series

declines most dramatically. This suggests that one or both fleets are not indexing abundance very reliably over the whole SW region.

- the JPN series suggests a slow downward trend in the NE region, while the short TWN series is very noisy and may or may not be consistent with a downward trend
- In the SE region, there are two periods of relative stability in the CPUE series, with a sudden drop in the JPN series around 1990.
- In addition to the conflicting trends, the means of the area-weighted CPUE series do not align perfectly either (Figure 7), with the Taiwanese fleet suggesting higher abundance in the West (particularly NW) than the Japanese fleets. This probably reflects different areas of operation, and brings into question the appropriateness of the area-weighting factors.

There are also concerns with respect to the catch size composition data. There is not much modal information for discerning relative year-class strength (e.g. Figure 8), and most of the fleets seem to be catching quite similarly sized fish on average. ALGI_NE is one clear exception, and it seems plausible that the gillnet fleets would catch smaller fish. However, the temporal trends in mean size among fleets do not appear to be very consistent (Figure 9). In many cases, this is probably related to very small sample sizes. However, the conflicting trends probably also indicate that the size sampling is non-random, or the fleet (or swordfish population) may be changing its distribution in space or time (i.e. non-stationary selectivity). Either way, it raises doubts about the usefulness of the model for resolving year-class strength.

Model Assumptions

Most of the biological relationships and assumptions were adopted from the recent southwest Pacific swordfish assessment described in Kolody et al. (2008).

Spatial Structure

The model is disaggregated into 4 areas corresponding to those used in the catch rate standardization analysis of Semba, Nishida and Wang (2008, IOTC-2008-WPB6-Info1) (**Figure 1**). Some evidence suggests that there may be genetic distinction within the IO (Muths et al. 2009), and this is the subject of ongoing investigation. Given the vast size of the Indian Ocean, and the migration rate inferences that have been made from tagging studies (particularly in other oceans), it seems unlikely that there would be rapid mixing processes across the whole basin, even if the population was genetically homogeneous. As such, localized overfishing could result in negative local consequences even if the overall stock is not overfished and there is a low risk of declining genetic diversity. The 4 area structure seems reasonably consistent with spawning area hypotheses (e.g. Poisson and Fauvel 2009), and also conveniently partitions most of the national fleets.

Migration Dynamics

There are very few direct observations of swordfish migration in the Indian Ocean. The few conventional tag recaptures near the Australian coast provided no indication of large scale movements.

However, we can indirectly infer that there are probably some relatively large seasonal migrations. Swordfish are caught in the temperate waters south of 35S, however, the spawning regions (and larval distributions) tend to be in the tropical regions. At least in the southern hemisphere this suggests directed seasonal migrations. The quarterly standardized CPUE series (not shown) suggest that there are strong seasonal trends that could be explained by migration, and they are out of phase in the northern and southern regions. At the resolution currently available in the standardized CPUE series, it is not clear whether this represents a single migration from the north to the south, or whether distinct populations independently move between lower and higher latitudes in each hemisphere.

A number of migration situations such as those described in Figure 2 and Figure 3 can be described using Stock Synthesis 3. However, only the relatively simple low (less than 1% per year) and high (>35%/year) mixing scenarios (e.g. Figure 2) have been explored to date, and it is recognized that these scenarios could be reasonably represented with spatially-aggregated models (i.e. either 1 combined or 4 separate models).

Fishery Definitions

In the example model, 24 fleets were defined, corresponding to the data aggregation units of the catch data as supplied by the IOTC Secretariat (Table 2). Each fleet resides in a single area only. If the same nation operates in more than one region, these operations are described as a separate fleet. Many of the small fleets were aggregated in the data as provided by the IOTC secretariat. It would probably be sensible to further reduce the number of fleets on the basis of the general similarity in catch size composition.

Time Period

The model was run from 1950-2007 using a quarterly time-step. The particular models explored could be run on an annual time-step, but the quarterly time-step does allow the model to potentially resolve seasonal migration characteristics.

Age and Sex Structure

The swordfish population is age- and sex-structured with cohorts of 0-40+ years, for each of two sexes. Sex-specific characteristics invoked in the model include:

- Growth curves
- Age-based selectivity (when derived from a length-based function that is the same for both sexes)

Natural mortality can also be sex-specific, but no distinction was made. SS3 also supports the fitting to catch size composition disaggregated by sex, and this data is available for some fleets, but this was not yet implemented.

Age and Size

There is strong evidence for sex dimorphism in swordfish, and it is likely that aggregating data across heterogeneous units (sexes) can lead to statistical biases in these sorts of models. However, it is not clear that this is a high priority for the assessment because there is currently considerable uncertainty about all swordfish age estimation methods, and a number of other related assumptions (e.g. M), which are likely more important than the sex aggregation biases.

Two sets of growth curves were explored to bracket two cases that might be plausible given the scarcity of age validation data for swordfish (**Figure 10**):

- CSIRO curve, derived from South-East Indian Ocean fin rays samples (Young and Drake 2004).
- NMFS curve, derived from Hawaiian samples (DeMartini et al 2007).

The uncertainty in age estimation resulting from different methods is described in Young et al. (2008). The biology of the Hawaiian swordfish may differ considerably from the Indian Ocean, however, if the NMFS age estimation method is more accurate than the CSIRO method, then the Hawaiian growth curve is probably preferable. e.g. Young et al. (2008) illustrate that the Hawai'ian growth curve is probably very similar to the East-coast Australian growth curve that would have been estimated if the NMFS fin-spine reading method had been employed. However, the Australian growth curves should also be compared with others derived for the Indian Ocean. It seems likely that the NMFS growth curve cannot adequately account for the largest fish in the Indian Ocean very well (but it still might accurately reflect the high growth rates for young individuals).

Maturity and Spawning Stock Biomass

SS3 can use age-specific vectors of female maturity or fecundity for biomass spawning calculations. While a number of studies quantify the relationship between size and maturity, the uncertainty of age estimation that undermines the growth relationships also undermines the maturity/fecundity by age relationship. Two relationships were explored, roughly corresponding to extremes associated with the growth curves:

- 50% maturity ~age 10, corresponding to the CSIRO study (mostly based on SW Pacific samples).
- 50% maturity ~age 4, corresponding with one of the youngest age at maturity schedules used in swordfish assessment.

The old maturity schedule was associated with the slow growth curve, young maturity with the fast growth curve. The age/size-dependent fecundity relationship was not included.

Natural Mortality

For the slow growth curve scenarios, M was assumed to be 0.2, constant across all ages. For the fast growth curve scenarios, M was assumed to be 0.4. These values resemble those used in Kolody et al. (2008), in which several age-specific, and growth-rate specific M vectors were tested.

Selectivity

The catch data indicate similar size composition for most fleets. SS3 supports length- and age-based selectivity with numerous functional forms. In the models presented here, three different selectivity options were explored:

- Two different size-based "double normal" selectivity curves were shared among fleets, one for longline fleets, and one for the gillnet (and related) fleets.

- Two different age-based “double normal” selectivity curves were shared among fleets, one for longline fleets, and one for the gillnet (and related) fleets.
- A unique size-based “double normal” selectivity was estimated for each of the 18 fleets with size composition data.

The “double normal” selectivity has considerable freedom to represent a dome-shape, or an approximately logistic curve that either reaches a plateau or is monotonically-increasing. Fleets without size data were assumed to have the same selectivity as the most similar fleet with size data.

Catchability

Catchability was assumed to be constant over time for the fleets with standardized CPUE. It was further assumed that catchability could be shared among areas for all of the Japanese fleets, and all of the Taiwanese fleets (but it was assumed that catchability was independent among nations). The sharing assumption means that CPUE is interpreted as a consistent measure of density among areas, and relative abundance by area is the product of area and density. However, the validity of this assumption really depends on how representative the fleet coverage is within each area. Without more information about the standardization process, it is not clear that this is the most appropriate way to use this data, and the differing abundance by area estimated for the Japanese and Taiwanese fleets led to some scenarios being tested in which catchability was shared for only one nation.

Catch in mass observation errors

Total catch in mass was assumed to be known essentially without error for all fleets.

Catch-at-Size sampling characteristics

Some of the sample sizes are very large for some fleets. In the context of the current model, this might cause a misleading overfitting to the size composition data for a number of reasons, including: i) sampling is probably not truly random for all fleets, ii) selectivity is probably not truly constant for any fleet. To partially account for these problems,

- each length distribution with fewer than 10 fish was discarded.
- all sample sizes were reduced by a factor of 10.
- a somewhat arbitrary (and relatively large) constant (1%) has been added to each of the predicted and observed length bins to reduce the influence of outliers on the catch-at-size likelihood term.

CPUE characteristics

The standardized CPUE was assumed to be directly proportional to selected abundance (in numbers) and unrealistically informative (see dot points). The annual indices were assumed to correspond to abundance in quarter 1 (the implications of abundance changes within a year are negligible relative to the noise in the CPUE series). Several different CPUE weighting options were considered, and only those models in which the Japanese series was given higher weight than the Taiwanese fleet were considered to be successful.

Stock Recruitment Relationship

A Beverton-Holt stock recruitment relationship was assumed, with a fixed steepness of 0.9 (0.65 for uncertainty exploration). In SS3 there is a single annual spawning biomass calculation, but the recruitment can be partitioned in various ways. For this application:

- Recruitment was assumed to occur once annually (quarter 1),
- Area-specific parameters were estimated to distribute the recruitment among regions (these proportions are constant among years),
- Each model specification was fit with and without the estimation of annual deviations from the stock recruitment curve. When estimated, the deviations were highly constrained (SD(log) = 0.1), to reflect the fact that the size composition data does not seem to be very informative with respect to recruitment.

Fishing Mortality

The “hybrid” fishing mortality parameterization was used, in which SS3 starts with Pope’s approximation and then conducts a fixed number of iterations to approximate instantaneous F from the Baranov catch equation.

Initial Population

The population was assumed to be in unfished equilibrium in 1950, the start of the catch data series.

Model fitting

The models fit to date involved minimization of an objective function with the following terms:

Likelihoods:

- CPUE – lognormal observation errors
- Length frequencies – multinomial distribution (downweighted sample sizes by a factor of 10 with 1% added to each bin)
- lognormal annual recruitment deviates

Prior distributions:

- parameters are either fixed or extremely uninformative priors were used

Penalties:

- smooth penalties for parameters approaching bounds (at this time, no action was taken for parameters approaching bounds).

The estimated number of parameters varies by model specification (number):

- catchability for CPUE series (3-6)
- mean virgin recruitment (1)
- 5 selectivity parameters for each independent function (10 or 90)
- recruitment proportion by area (3)
- annual recruitment deviations from the stock recruitment relationship (0 or 57)

Additional estimable parameters that were fixed in the current application:

- natural mortality
- migration rates - two sets of two age-specific parameters (with linear interpolation between ages) for every combination of two adjacent areas
- stock recruitment steepness

Uncertainty Quantification

At this time, no attempt was made to quantify the statistical uncertainty associated with any of the models, and only the Maximum Posterior Density (MPD) estimates are presented. With these models it generally proves to be the case that the uncertainty associated with model specification is greater than the statistical uncertainty associated with any individual specification.

Model definitions

A core set of preliminary models is defined in Table 2.

Results and Discussion

This section describes the WP results obtained prior to WPB 2009, while the following attachments describe results in relation to discussions and additional model specifications set up during and after WPB 2009.

Table 3 lists a number of diagnostics for a selection of models, and a number of key reference points. The diagnostics include the RMSE for each of the 9 CPUE series, and the maximum gradient of the objective function. RMSE describes the quality of fit between predictions and observations and ideally should approximately equal the input SD(log) for each CPUE series. Convergence was acceptable for almost all of the models, though there remains a question of whether the global minimum was identified.

Two models are included which exhibit representative dubious behaviour (Figure 11, Figure 12). In all of the models in which CPUE of Taiwanese fleets was weighted equally or more highly than the Japanese series (and in which recruitment deviates were not estimated), the population is extremely large, and the impact of the fishery is negligible (swo15p4). Many cases in which annual recruitment deviates were estimated from the stock recruitment relationship resulted in a similar problem, in which the population increases and decreases purely due to recruitment variability (swo11p6). These are common pathological problems in these types of models and often indicate that the global minimum has not been identified. Initial attempts to guide the function minimization with a carefully constrained phased approach have so far not been successful.

The detailed results are based on one of the more pessimistic models (swo11p4) which qualitatively illustrates the general characteristics of the models that do show a fishery impact on the population.

Figure 13 shows a reasonable fit to the northern Japanese CPUE, but the declining trends in the southern regions are poorly fit. This is probably not due to the effect of the Taiwanese and La Reunion fleets, because it occurs in other models where these series are even more highly down-weighted (including initial trials that only included the Japanese data). In the absence of recruitment variability, the model must average through the suspicious increase in CPUE in the 1980s in the SW region, so the biomass cannot begin its decline from the height observed in the CPUE.

Figure 14 illustrates the agreement between the predicted and observed mean catch-at-length time series.

Figure 15 illustrates typical dome-shaped size-based selectivity estimates. It is plausible that the distribution of the swordfish stock might leave the larger/older individuals less vulnerable to most fisheries. However, dome-shaped selectivity can also arise as an artefact in relation to improper specification of natural mortality (i.e. through ignorance about the variability in M by age).

Conclusions:

1. Stock Synthesis 3 software provides a powerful, flexible and numerically efficient framework for integrating a diverse range of structural features and statistical assumptions for the assessment of Indian Ocean swordfish. While the flexibility to explore alternative assumptions about natural mortality, stock recruitment curve steepness, migration rates, etc. is useful, it is unlikely that estimates from the model would be very meaningful with the type and amount of data that is available. It is, however, useful to show how sensitive the results are to alternative plausible values of these parameters, and provide a reasonable illustration of the stock status uncertainty.

2. A couple downsides to the SS3 software have been identified which are relevant for this assessment:
   i. SS3 cannot resolve time-area interactions in recruitment variability. This is probably not serious in so far as the size data does not seem to be very informative about year class strength anyway. When global recruitment deviates are estimated, there is a tendency for the model to explain the large increase and decrease in Japanese CPUE (particularly in the SW region) as primarily a recruitment effect (as opposed to a fishery effect).
   ii. Many of the common stock status reference points generated by SS3 cannot be partitioned by area, so separate models would need to be formulated to examine individual areas under the assumption of discrete populations.
   iii. (additional comments in Attachments 1 and 2 relate to the author’s mistaken interpretation of the software documentation.)

3. The explorations undertaken to date are not entirely satisfactory in that a number of seemingly plausible models did not converge to produce sensible parameter estimates. Further work is required to determine if this is a global minimization problem or model specification problem that can be easily resolved.

4. At the moment, this analysis does not constitute an assessment. However, with insight and feedback from the IOTC WPB, it may be possible to produce an interim assessment. A number of issues can be considered within the timeframe of the WPB:

    i. There are a number of analyses that might be conducted with respect to the conflicting trends in the standardized catch rates. It would be worth quantifying changes in spatial and temporal coverage of the fleets to consider how appropriate that it is to i) assume that the standardization can account for the important changes in catchability, and ii) assume that the catchabilities can be shared across areas.
    ii. Prior to the WPB, it was suggested that a two-tiered approach to the assessment should be undertaken, with the whole Indian Ocean being one option and the most depleted of the 4 regions being considered as a second localized option (under the assumption that some regions might be reasonably isolated from the others). At the time, it was assumed that the SW region was the highest priority. However, comparison of the Japanese, Taiwanese and La Reunion CPUE series in this region raises the question of whether the steeply decreased Japanese series reflects a change in abundance or a change in catchability unique to the Japanese fleet (in which case the SW region might not be the most depleted).

5. There are a number of longer term paths of investigation that might help reduce the assessment uncertainties, including:
    - in depth analyses of catch rates to identify temporal patterns in effort, CPUE and operations that might be affecting catchability in ways that cannot be corrected for in the standardization approaches employed to date.
    - review the size composition data for the purposes of identifying homogenous fleet characteristics that are more likely to conform to constant selectivity (and shared selectivity) assumptions.
    - Include additional CPUE series, e.g. from the Spanish and Seychelles fleets
    - review stock structure in relation to new genetics and tagging work
    - direct age validation of fin spine annulus counts
    - incorporation of sex-specific catch data
    - examine implications of catch uncertainty

6. If this type of modelling is to be continued in future years, it would be very useful to have an agreed timeline for data exchange, with sufficient time allocated to prepare the assessment in advance of the Working Party.

References

DeMartini, E. E., James H. Uchiyama ,Robert L. Humphreys Jr. ,Jeffrey D. Sampaga, Happy A. Williams (2007)Age and growth of swordfish (Xiphias gladius) caught by the Hawaii-based pelagic longline fishery Fish. Bull. 105:356–367.

Kolody, D. 2008. A Flexible Spatially-Disaggregated Production Model for Exploratory Assessment of Indian Ocean Swordfish - updated. Working paper IOTC-2008-WPB-09-add1.

Kolody, D., Campbell, R. and N. Davies. South-West Pacific Swordfish (Xiphias gladius) Stock Assessment 1952-2007. Final report for the funding bodies, available as Working Paper WCPFC-SC5-2009/GN-IP-2.

Nishida, T. and Semba, Y. 2009. Preliminary stock assessment of swordfish (*Xiphias gladius*) in the Indian Ocean by A Stock-Production Model Incorporating Covariates (ASPIC). Working paper IOTC-2009-WPB-12.

Nishida, T. and Wang, S.P. 2009. Estimation of the abundance index of swordfish (*Xiphias gladius)* in the Indian Ocean based on the fine scale catch and effort data in the Japanese tuna longline fisheries (1980-2007). Working paper IOTC-2009-WPB-08.

Poisson, F. and Fauvel, C. 2009. Reproductive dynamics of swordfish (*Xiphias gladius)* in the southwestern Indian Ocean (Reunion Island). Part 1: oocyte development, sexual maturity and spawning. Aquat. Living Resour. 22, 45-58.

Wang, S.P. and Nishida, T. 2009. CPUE standardization of swordfish (*Xiphias gladius*) caught by Taiwanese longline fishery in the Indian Ocean for 1980-2007. Working paper IOTC-2009-WPB-12.

Schnute, J.T. and Hilborn, R. 1993. Analysis of contradictory data sources in fish stock assessment. Can. J. Fish. Aquat. Sci. 50: 1916-1923.

Young, J. and A. Drake. 2002. Reproductive dynamics of broadbill swordfish (Xiphius gladius) in the domestic longline fishery off eastern Australia. Final report for project 1999/108, Fisheries Research Development Corporation, Canberra, Australia.

Young, J. and A. Drake. 2004. Age and growth of broadbill swordfish (Xiphius gladius)from Australian waters. Final report for project 2001/014, Fisheries Research Development Corporation, Canberra, Australia.

Young, J., Humphreys, R., Uchiyama, J. and Clear, N. 2008. Comparison of swordfish maturity and ageing from Hawaiian and Australian waters. WCPFC-SC3-BI SWG/WP-1.

Table 1. Fishery definitions and characteristics. Suffixes denote regions within the Indian Ocean as indicated in Figure 1: NW – North-West; NE – North-East; SW – South-West; SE – South-East.

ALGI	Contains data for gillnet, trolling and other minor artisanal fisheries
AUEL	Contains data for the longline fishery of Australia (target is SWO)
EUEL	Contains data for EU longliners (from Spain, Portugal and the UK) targetting SWO plus other longliners assimilated to EU longliners (generally owned by Spanish nationals)
ISEL	Contains data for the semi-industrial longline fleets operating in Reunion(France), Mayotte(France), Madagascar, Mauritius and the Seychelles, which also target SWO
JPLL	Contains data for the longline fishery of Japan plus other fleets assimilated to the Japanese fleet (e.g. South Korea, Thailand, Oman)
TWFL	Contains data for the fresh-tuna longline fleets of Taiwan and Indonesia, plus other fresh-tuna longline fleets assimilated to those and all sport fisheries and fleets operating hand lines
TWLL	Contains data for the large scale tuna longline fleet of Taiwan,China, plus other longline fleets assimilated to the Taiwanese fleet (a component of those fleets may target SWO)
ALGI	Contains data for gillnet, trolling and other minor artisanal fisheries

Fleet Number	Name	CPUE Series Number	Numbers of quarters with Size Composition Observations
1	ALGI_NW		
2	EUEL_NW		12
3	ISEL_NW		40
4	JPLL_NW	1	97
5	TWFL_NW		12
6	TWLL_NW	5	112
7	ALGI_NE		74
8	EUEL_NE		5
9	JPLL_NE	2	130
10	TWFL_NE		36
11	TWLL_NE	6	112
12	ALGI_SW		
13	EUEL_SW	9	40
14	ISEL_SW		43
15	JPLL_SW	3	108
16	TWFL_SW		
17	TWLL_SW	7	110
18	ALGI_SE		
19	AUEL_SE		20
20	EUEL_SE		28
21	ISEL_SE		
22	JPLL_SE	4	127
23	TWFL_SE		
24	TWLL_SE	8	109

Table 2. Distinguishing features of some of the assessment models considered to date.

Model ID	Japanese CPUE SD(log)	Taiwanese CPUE SD(log) (case 4)	La Reunion CPUE SD(log)	Growth and Mortality	Number of Selectivity functions estimated	Movement Rate	Beverton-Holt Steepness (*h*)	Recruitment deviates estimated
ioswo-11p4	0.05 q shared	0.3 q shared	0.3	Growth Slow M=0.2	18	<1%/year	0.9	No
ioswo-11p6	0.05 q shared	0.3 q shared	0.3	Growth Slow M=0.2	18	<1%/year	0.9	Yes
ioswo-15p4	0.1 q shared	0.1 q shared	0.1	Growth Slow M=0.2	1=longline 2=gillnet	<1%/year	0.9	No
ioswo-16p4	0.05 q shared	10 q shared	10	Growth Slow M=0.2	1=longline 2=gillnet	<1%/year	0.9	No
ioswo-40p4	0.05 q shared	0.3 q not shared	0.3	Growth Slow M=0.2	1=longline 2=gillnet	<1%/year	0.9	No
ioswo-43p4	0.05 q shared	0.3 q not shared	0.3	Growth Fast M=0.2	18	<1%/year	0.9	No
ioswo-44p4	0.05 q shared	0.3 q not shared	0.3	Growth Slow M=0.2	1=longline 2=gillnet	<1%/year	0.65	No
ioswo-46p4	0.05 q shared	0.3 q not shared	0.3	Growth Fast M=0.4	1=longline 2=gillnet	<1%/year	0.9	No
ioswo-	1995+ 0.1	0.1	0.1	Growth	1=longline	<1%/year	0.9	No

49p4	q shared	q not shared		slow M=0.2	2=gillnet			
ioswo-50p4	1995+ 0.1 q shared	0.1 q not shared	0.1	Growth Fast M=0.4	1=longline 2=gillnet	<1%/year	0.9	No
ioswo-51p4	1995+ 0.1 q shared	0.3 q not shared	0.1	Growth Slow M=0.2	1=longline 2=gillnet	<1%/year	0.9	No
ioswo-52p4	1995+ 0.3 q not shared	0.1 q shared	0.1	Growth Slow M=0.2	1=longline 2=gillnet	<1%/year	0.9	No
ioswo-49p4	1995+ 0.1 q shared	0.1 q not shared	0.1	Growth slow M=0.2	age-base sel WRONG! 1=longline 2=gillnet	<1%/year	0.9	No

Table 3. Quality of fit diagnostics for the 9 CPUE series, and key reference points for the assessment model specifications from Table 2.

Model	CPUE RMSE JPN-NW	CPUE RMSE JPN-NE	CPUE RMSE JPN-SW	CPUE RMSE JPN-SE	CPUE RMSE TWN-NW	CPUE RMSE TWN-NE	CPUE RMSE TWN-SW	CPUE RMSE TWN-SE	La Reunion
swo11p4	0.46	0.46	1.09	0.75	0.85	0.58	0.35	0.78	0.08
swo11p6	0.40	0.44	0.83	0.56	0.77	0.59	0.39	0.73	0.09
swo15p4	0.79	0.58	1.39	0.77	0.58	0.52	0.41	0.59	0.12
swo16p4	0.56	0.44	1.32	0.75	0.84	0.56	0.41	0.71	0.10
swo40p4	0.56	0.44	1.32	0.75	0.25	0.50	0.39	0.53	0.10
swo43p4	0.47	0.42	1.25	0.75	0.25	0.51	0.38	0.53	0.08
swo44p4	0.54	0.44	1.30	0.77	0.24	0.51	0.38	0.53	0.10
swo46p4	0.58	0.45	1.33	0.76	0.24	0.50	0.40	0.53	0.10

Model	Max. Gradient	F2007/FMSY	B2007/BMSY	B/B0	MSY (1000t)	N Parameters
swo11p4	8.79E-04	2.92	0.57	0.28	17	97
swo11p6	8.79E-04	0.02	2.06	1.00	1163	154
swo15p4	3.22E-04	0.00	2.05	0.99	130025	17
swo16p4	4.63E-04	0.78	1.30	0.63	33	17
swo40p4	1.67E-04	0.79	1.30	0.63	33	20
swo43p4	3.34E-01	1.46	0.72	0.35	25	100
swo44p4	7.97E-04	0.89	1.41	0.60	27	20
swo46p4	2.96E-04	0.63	1.24	0.60	41	20

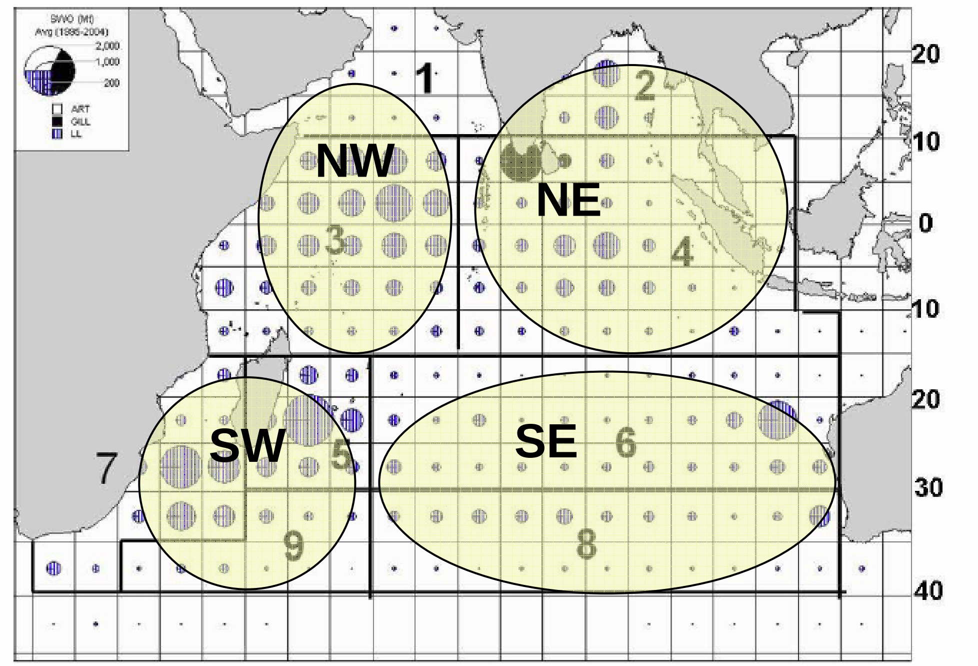
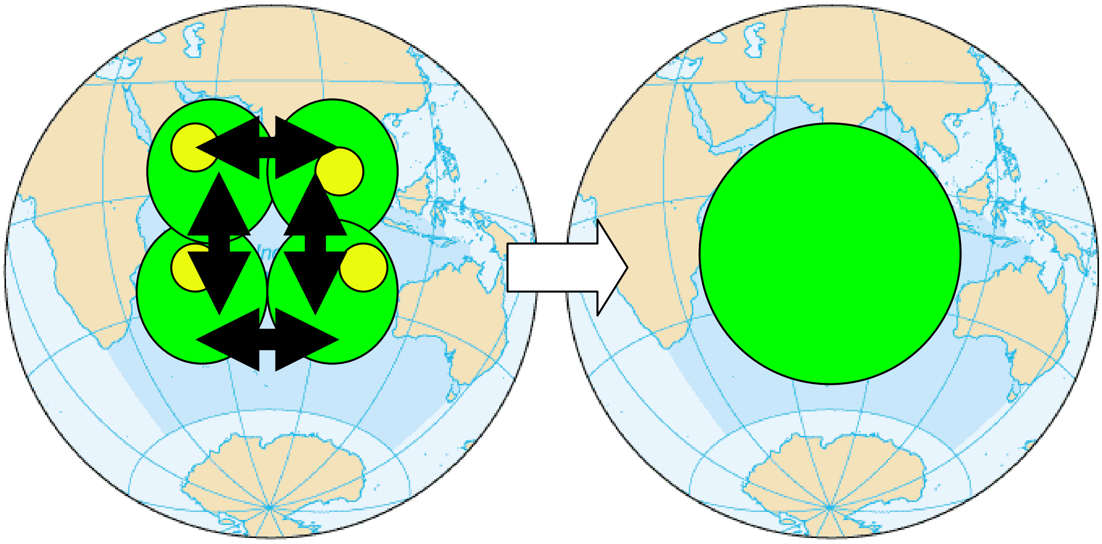
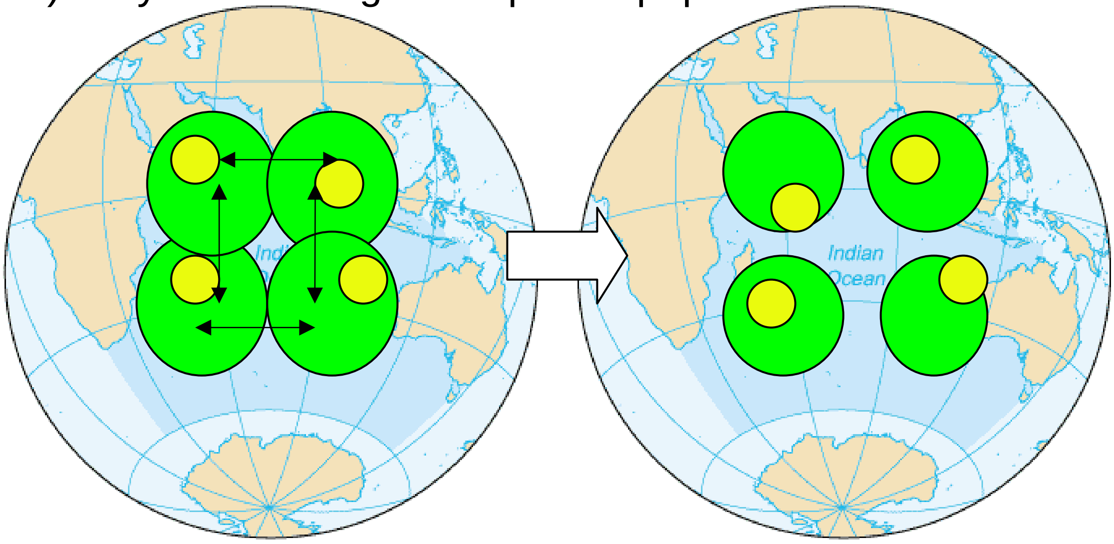


Figure 1. Spatial structure showing the 4 areas used for the exploratory model, superimposed on the IOTC statistical areas and the swordfish catch distribution aggregated over 1995-2004.

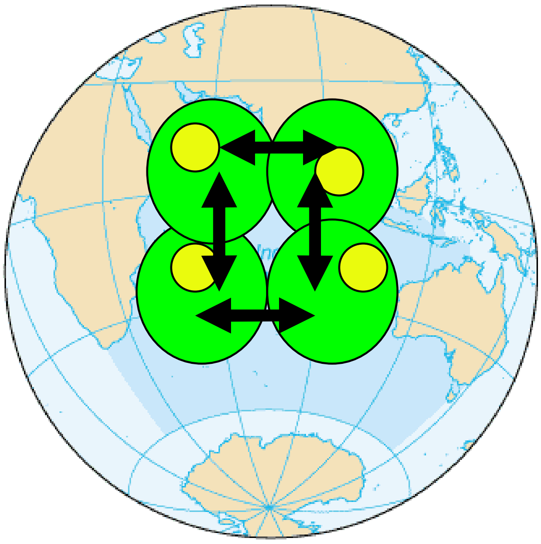


A) Very fast mixing ~ single population model



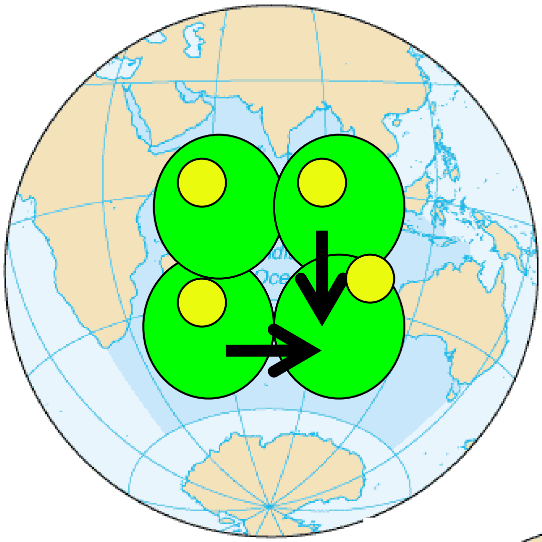
B) Very slow mixing ~ 4 separate population models

Figure 2. Conceptual models indicating situations where a spatially linked assessment is probably of no real advantage. In (A) movement rates are fast and random, such that depleting any one area affects all areas. The other extreme is (B) in which populations are almost isolated, and can simply be treated independently for assessment purposes. Large green circles represent foraging areas, small yellow circles indicate spawning areas, and arrows indicate migration



C) Intermediate Mixing rates within a genetically homogenous population

D) Source and sink dynamics



E) Genetically distinct but partially overlapping populations

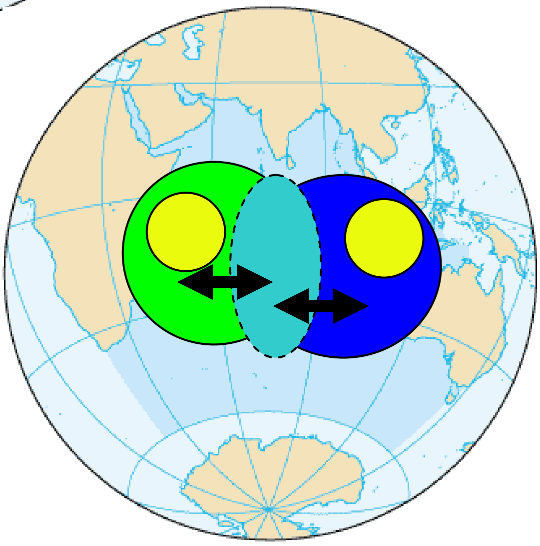


Figure 3. Conceptual models indicating situations where a spatially linked assessment may be useful. In (C) populations mix at a moderate rate. D is similar to C except that movement is non-random, with preferred habitat regions. In (E) spawning stocks are distinct, but they are caught together in a central mixed-stock fishery.

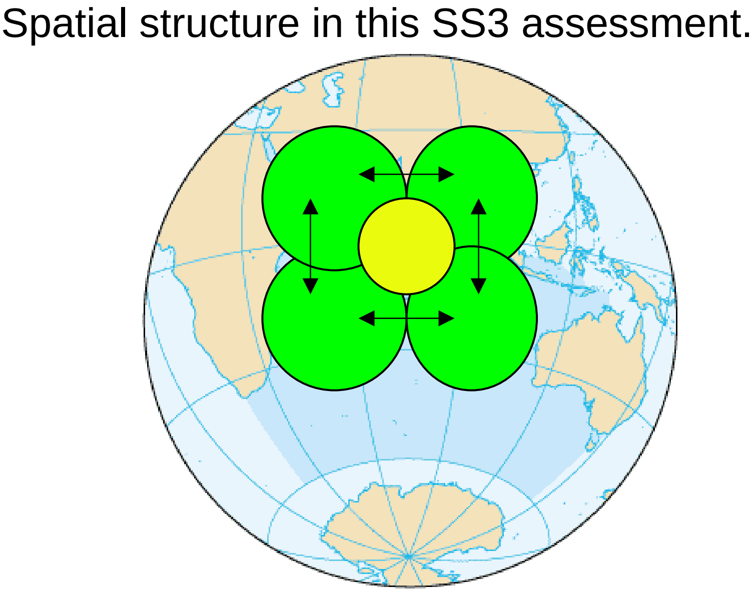


Figure 4. More precise description of the spatial representation of the swordfish population in this SS3 asssessment. Large green circles represent foraging areas, small yellow circles indicate spawning areas, and arrows indicate migration

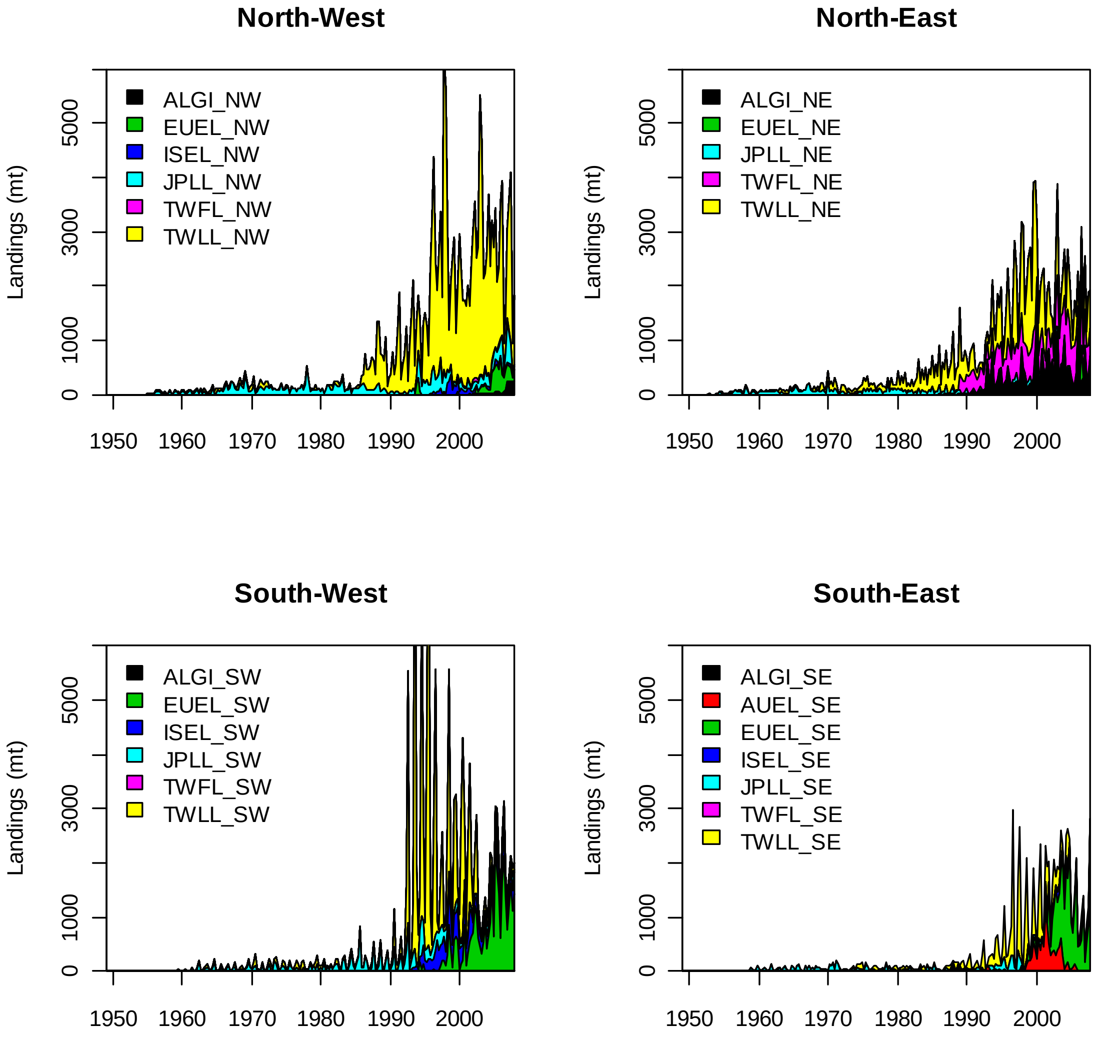


Figure 5. Total catch in mass over time (quarters) by assessment area and fleet. Note that 3 observations with quarterly catches approaching 10000 t are off the scale in the southwest.

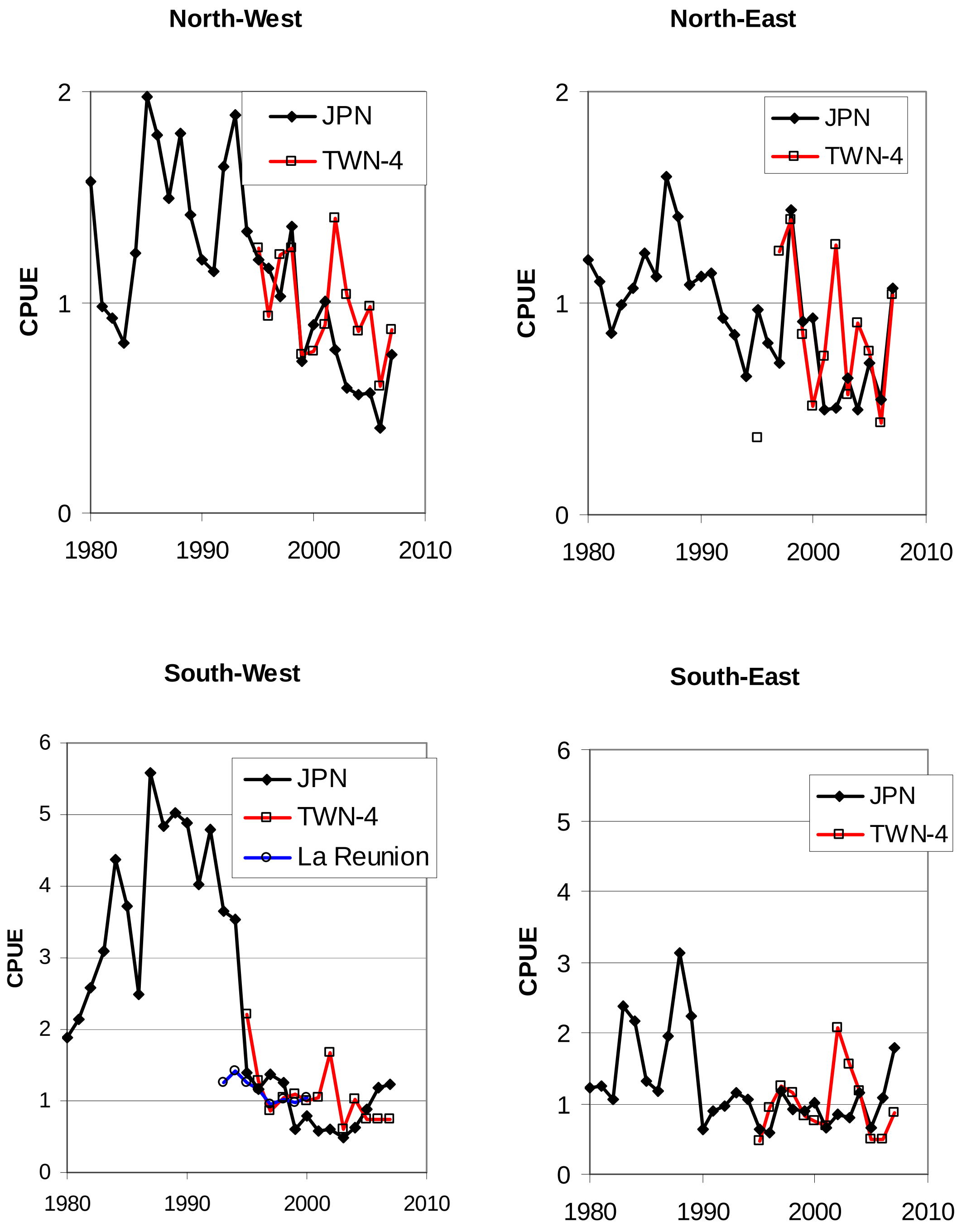


Figure 6. Standardized CPUE by area for Japanese, Taiwanese and La Reunion longline fleets. All series have been rescaled to have a mean of 1 over the interval 1997-2000 (note that this re-scaling is undertaken for the purposes of comparing trends among fleets within each area, and does not reflect the relative weighting across areas that is used in the assessment).

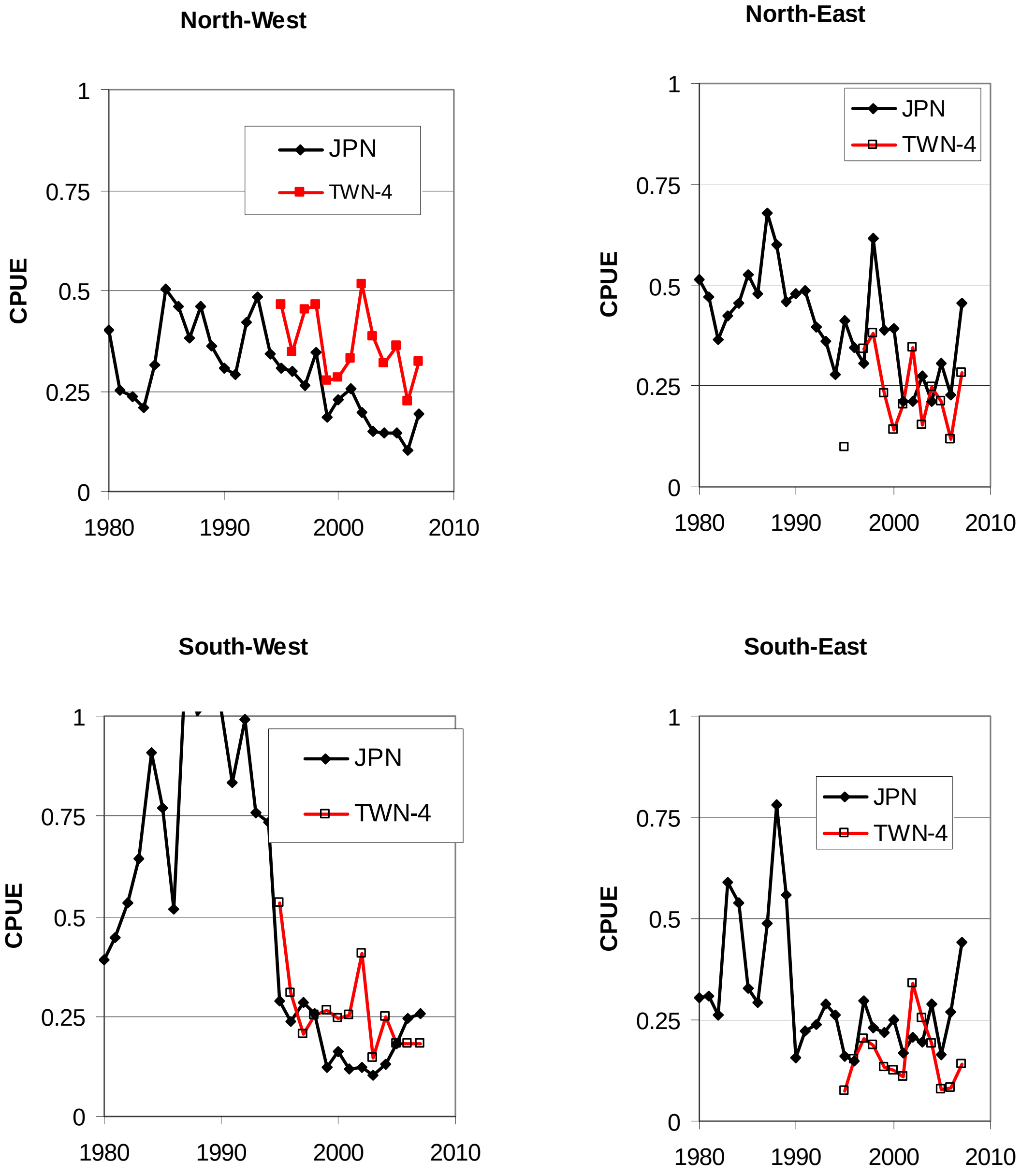


Figure 7. Standardized CPUE by area for Japanese and Taiwanese longline fleets. All series have been rescaled relative to the area-weighted Indian Ocean aggregate series for each nation over the interval 1997-2007. This re-scaling is undertaken to compare the consistency in relative abundance by area between the two fleets. The Taiwanese series suggest that there is higher abundance in the western regions than the Japanese series (particularly the NW).

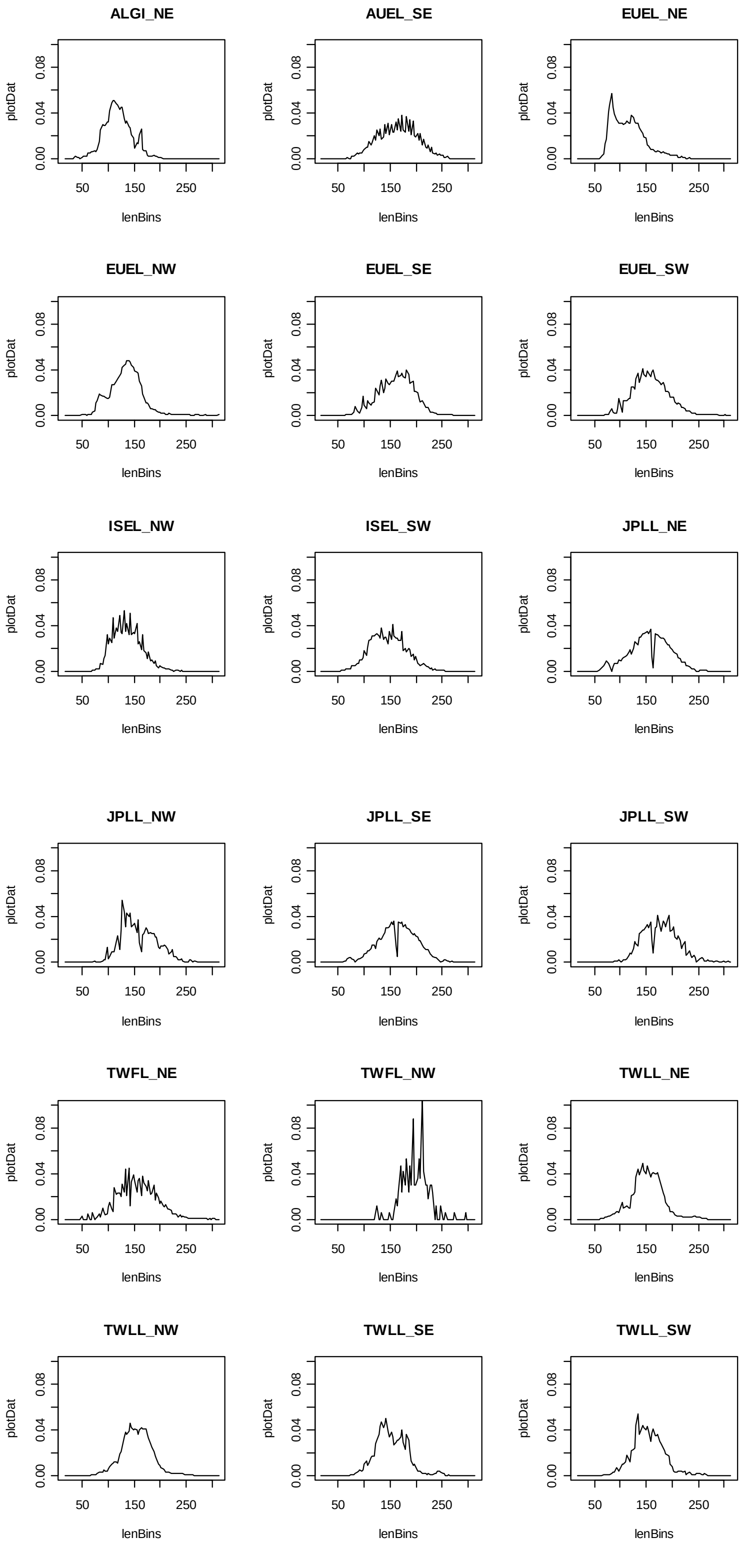


Figure 8. Indian Ocean swordfish length frequency distributions for each of the 18 fleets with size composition data (aggregated over all time periods).

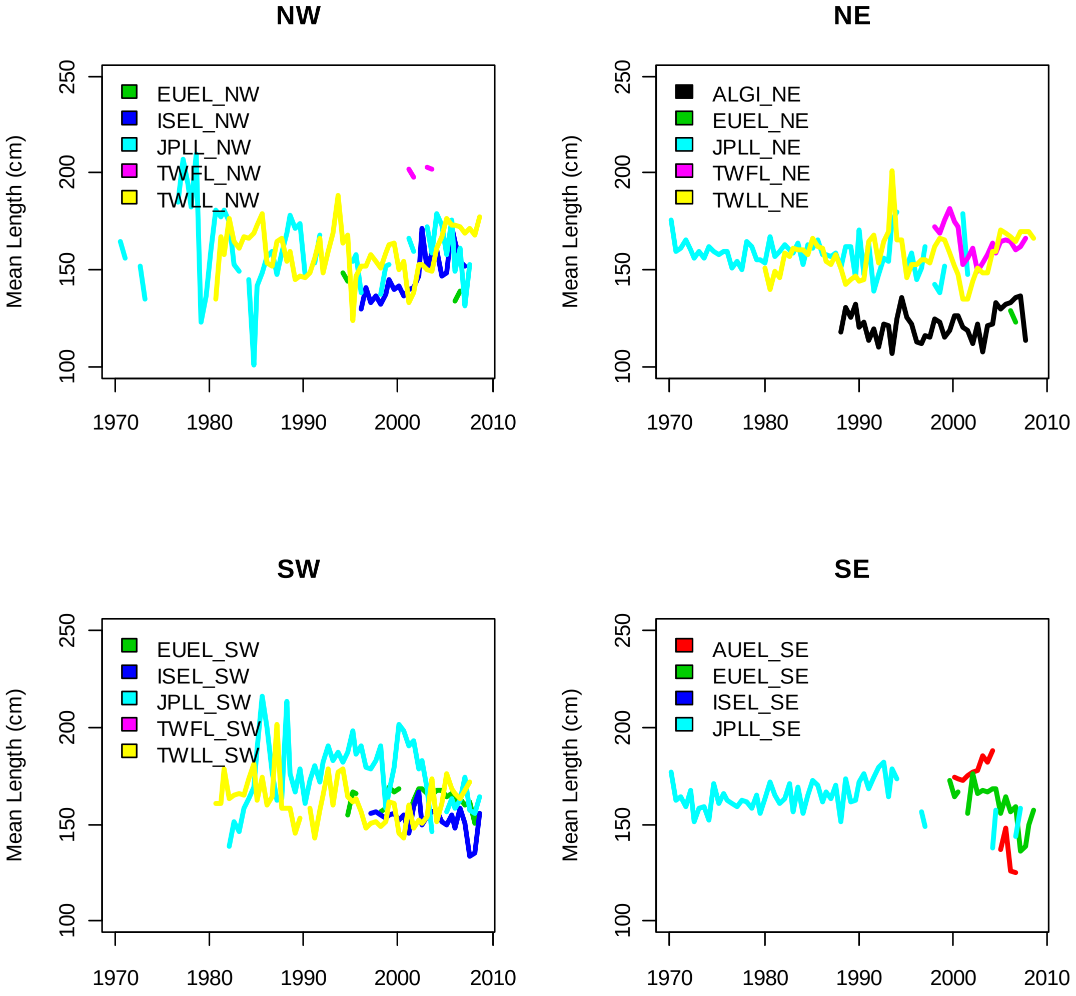


Figure 9. Indian Ocean swordfish mean length over time for each of the 18 fleets with size composition data, partitioned into the 4 areas used in the assessment model.

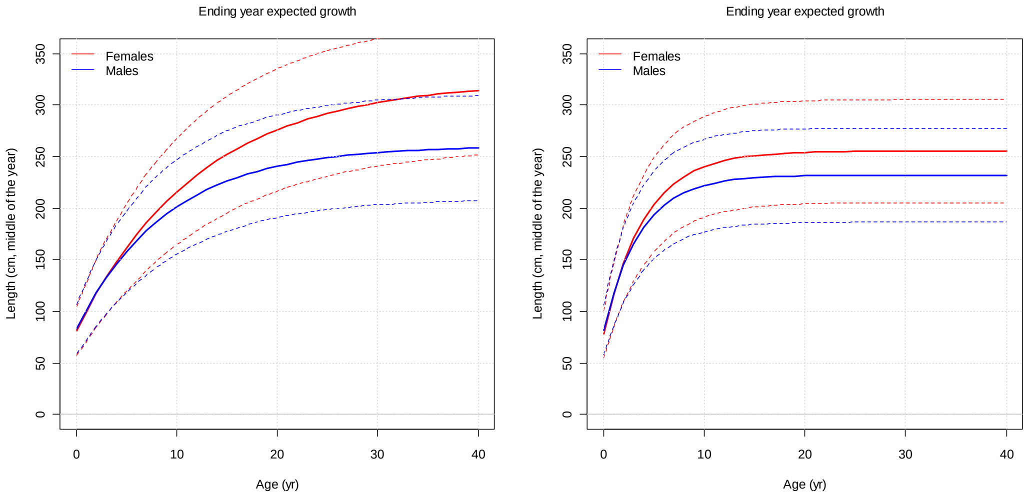


Figure 10. Growth curves (Lower jaw fork lengths) estimated for the Eastern Indian Ocean (CSIRO, top panel, corresponding to example model 1), and north-central Pacific (NMFS, bottom panel, corresponding to example model 2). Age estimation comparisons revealed differences of fin ray annulus interpretation when readers from both labs read the same fin rays. Real differences in biology might account for some of the estimated growth rate discrepancy, but most of it is seems to be attributable to the unresolved problem of age estimation (Young et al. 2008, WCPFC-SC3-BI SWG/WP-1**).**

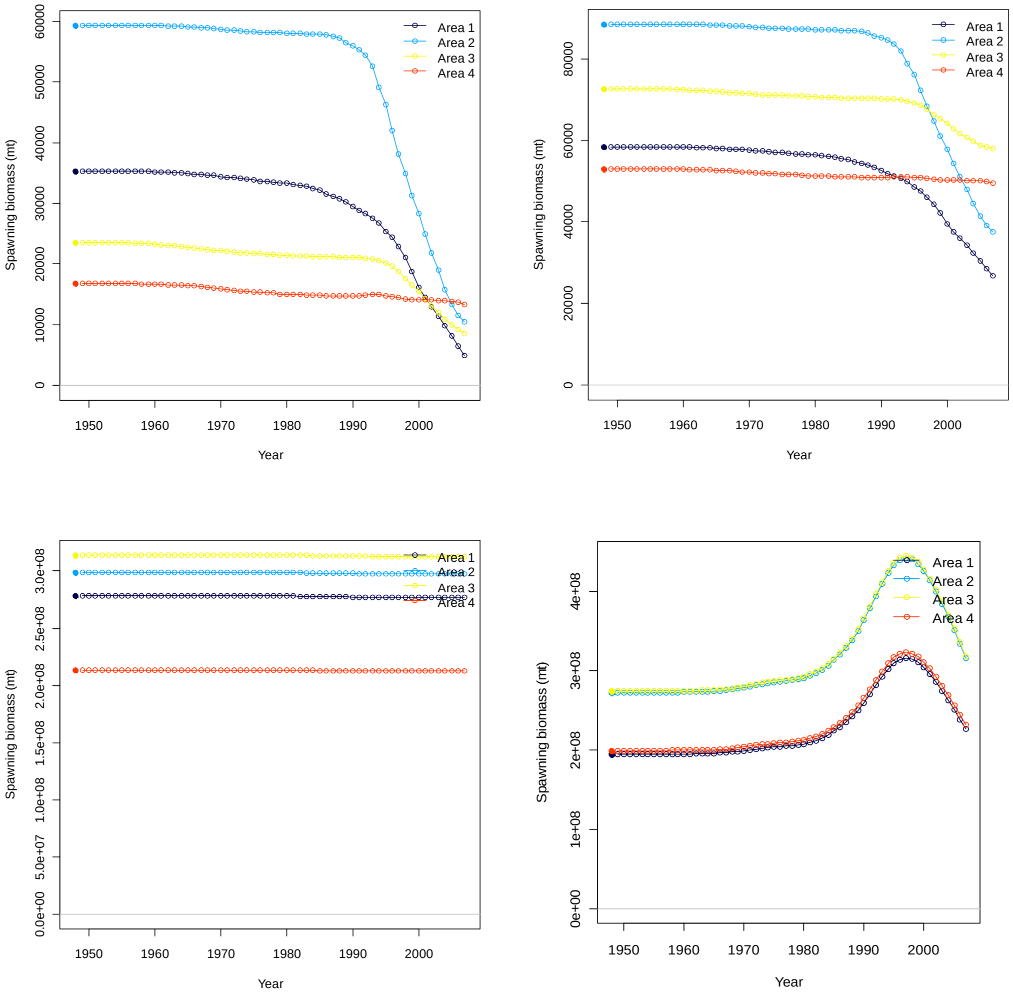


Figure 11. Estimated spawning biomass time series by region for 4 models. The top panels show two models with plausible dynamics swo11p4 (top left) and swo16p4 (top right). The bottom left panel is typical of the models that do not have high weighting on the Japanese CPUE. The bottom right panel is typical of the model results when recruitment variability is estimated (i.e. recruitment variability explains the increase and decrease in Japanese CPUE.

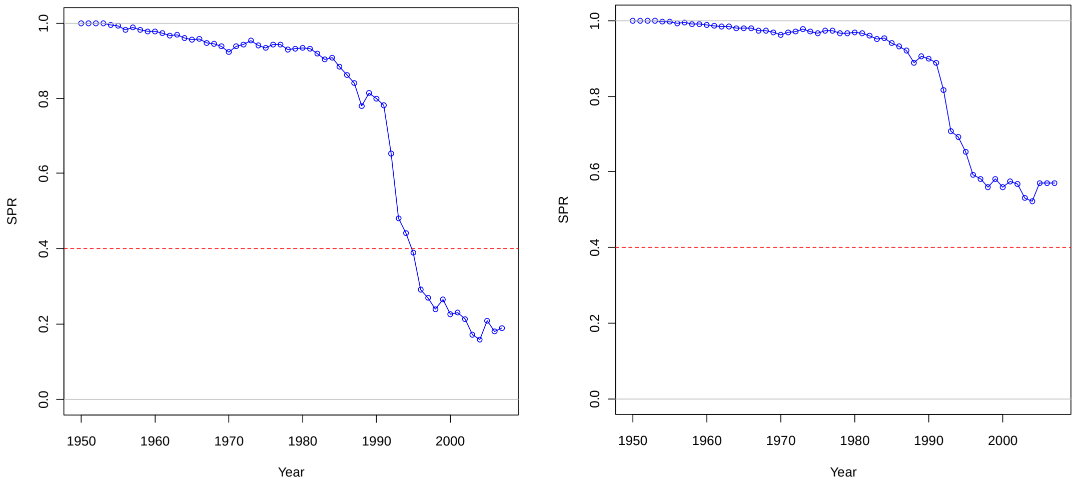


Figure 12. Estimated Spawning Potential Ratio (egg production (t) / virgin egg production) for model swo11p4 and swo16p4. The figures indicate the general implication of decreasing the relative weighting between the Japanese and other CPUE series (left panel has higher weight on the Japanese series)

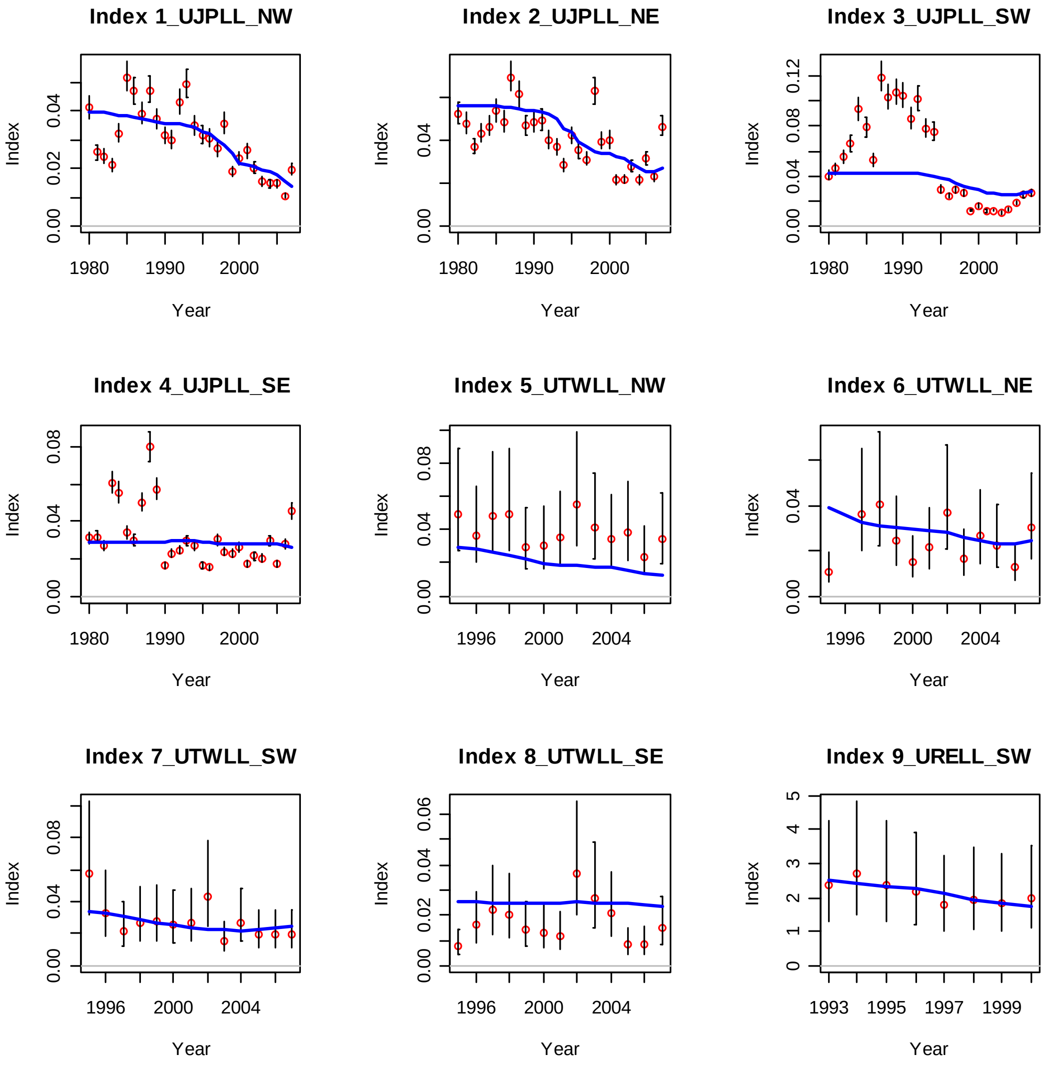


Figure 13. Predicted (lines) and observed (circles with 95% error bars) standardized Japanese longline CPUE for swo11p4.

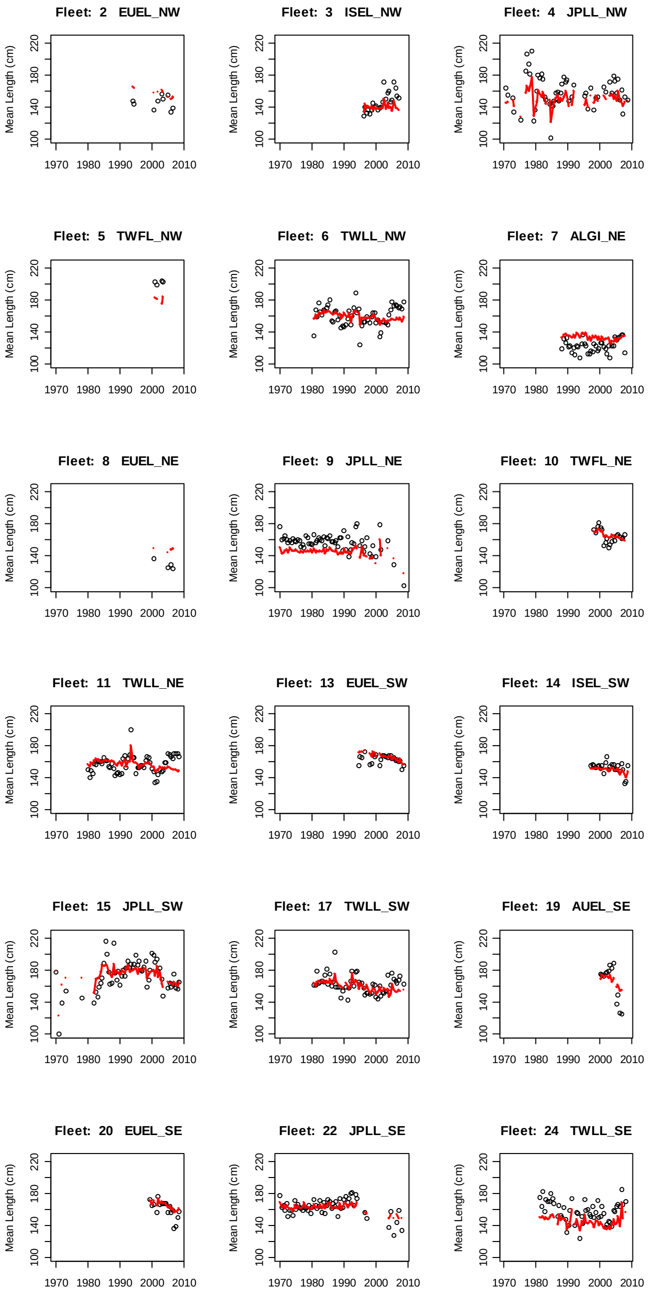


Figure 14. Predicted (red lines) and observed (black circles) mean catch for each of the 18 fleets with size composition data, for model swo11p4. Note that the SS3 predicted mean catches are not quite right because only predicted size bins with corresponding observation bins are included in the calculation (i.e. it is particularly problematic when sample sizes are small).

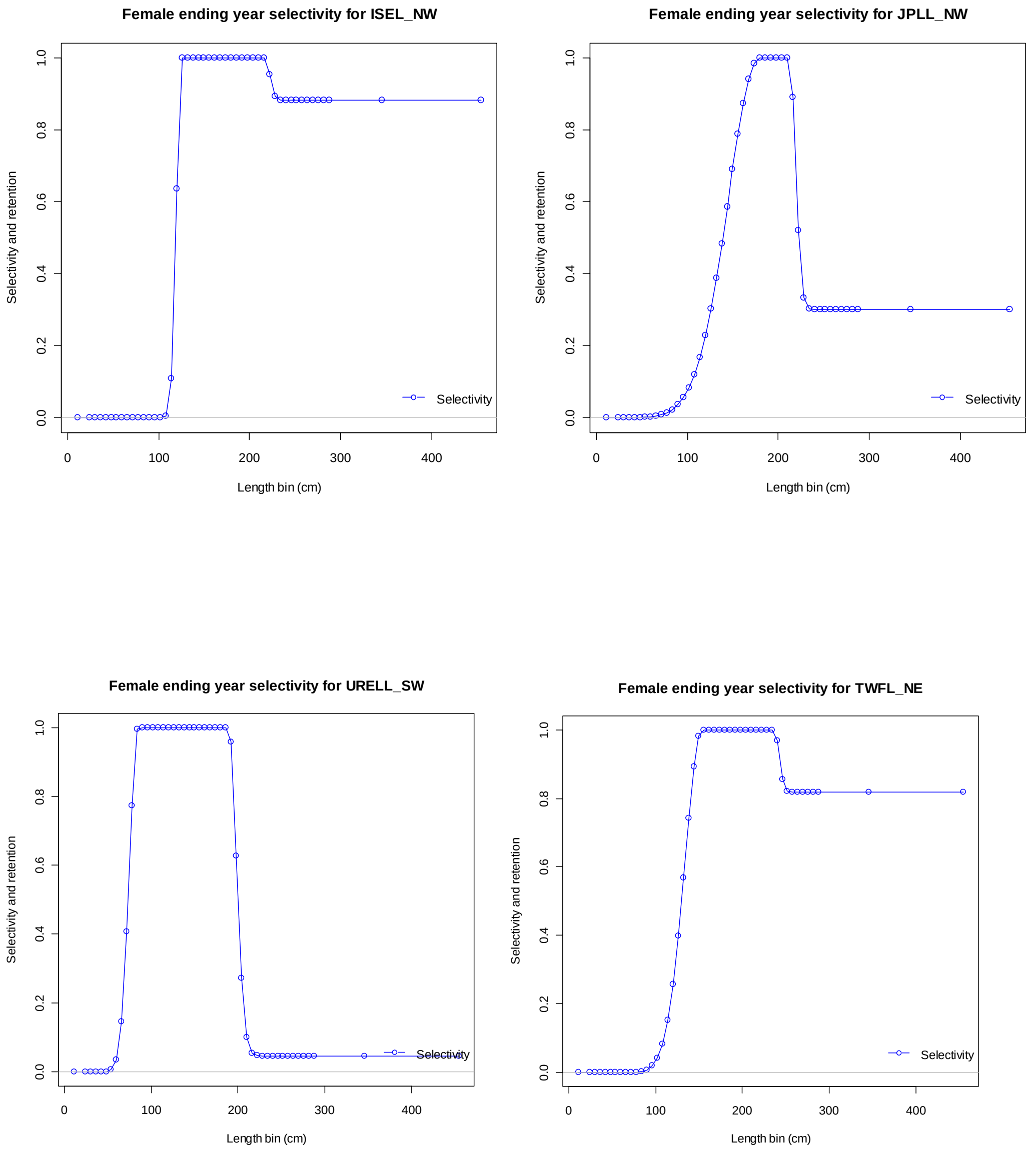


Figure 15. Example size-based selectivities estimated for swo11p4.

Attachment 1. Additional IO swordfish assessment model specifications developed during the IOTC WPB 2009.

Note that Attachment 1 was produced during the WPB 2009, and summarized in the WPB report, however, two important errors were subsequently corrected as described in Attachment 2.

During the WPB, it was agreed that the relative abundance indices in the SW seemed to be particularly problematic, and there was an effort to examine the conflict between the model, the Japanese CPUE, and the La Reunion CPUE. Based on the catch history, it seems very unlikely that abundance actually dropped by ~60% between 1994-1995 as suggested by the Japanese CPUE (e.g. Figure 6). This sharp drop is not observed in the more localized La Reunion series. There is also evidence that a number of operational changes took place in the Japanese fleet in that short period, including i) a spatial shift of effort into the Mozambique channel, and ii) a widespread change from traditional to monofilament line. This provided a justification to drop the pre-1995 part of the SW Japanese CPUE series from subsequent versions of the model. However, it also brings into question the CPUE series from other areas, and a more detailed analysis of operational factors is probably warranted. In principle, if the proper factors are included in the CPUE standardization then the major effects should be adequately quantified. In practice, if there is not adequate contrast in the data (e.g. different gear types rarely operate in the same time/area strata), or heterogeneous spatial factors are not described in the model at the appropriate resolution (e.g. fleet movements in and out of the Mozambique Channel), then the confounding between abundance trends and catchability trends cannot be resolved.

Five additional model specifications (using only the Japanese SW CPUE 1995+) were explored during the WPB (swo-61 to swo-65, Table A1-1). CPUE RMSE and reference points are summarized in Table A1-2. To expedite a troublesome minimization, the size-based indices were down-weighted by a factor of 0.1, and the Taiwanese CPUE series were down-weighted by a factor of 0.0001. Figure 16 and Figure 17 show the reference case (swo62p4) fit to the CPUE and mean size composition data. Figure 18 illustrates the quality of fit to the CPUE data and Figure 19 provides an indication of the consistency between assumed sample sizes and the quality of fit to the size composition data. Qualitatively, the models all seemed to capture the general features of the data, however, given the substantive problems with the input data assumptions, it was not clear that any goodness of fit diagnostics should be considered adequate to choose the “best model” at this time.

Figure 20 shows the biomass and fishing mortality time series associated with the 10 model specifications. Figure 21 compares a number of current stock status estimates. The models with deterministic recruitment all have reasonably similar and pessimistic results, which suggest that (relative to BMSY and FMSY reference points), the stock is either in an over-fished state, or it is currently being overfished, or both. In contrast, the stochastic recruitment time series models are highly variable, and more optimistic than the deterministic models. The stochastic models attribute a considerable amount of the recent biomass decline to a recruitment effect, rather than a fishing mortality effect. For the purposes of the WPB, only results from the deterministic recruitment case were included in the stock status synthesis, because:

- the size composition data do not seem to be very informative for determining year class strength:
    - there is limited modal information in the catch size distribution to reliably distinguish recruiting year classes
    - mean size trends seem to differ among fleets operating in the same region
- There is additional evidence for shifting targeting, and inappropriate aggregation of size data across non-homogeneous fleets. Both of these effects can invalidate the stationary selectivity assumptions.
- The increasing Japanese CPUE trend in the 1980s may be an artefact of increasing catchability due to targeting shifts, rather than increasing abundance due to recruitment variability.

While the author considers the five deterministic recruitment models submitted to the WPB to be plausible, they are also recognized to be at the more pessimistic range of results that might be considered plausible, for two main reasons:

- They rely primarily on the Japanese CPUE series, which is a by-catch fishery, with evidence for substantial operational shifts in the 1990s which may not be properly accounted for in the standardization, and which seem to be more pessimistic that other fleets operating in the same areas. In contrast, the Taiwanese CPUE series (those that extend back earlier than the 1990s), suggest much less decline in abundance. However, the Taiwanese fleets are often targeting swordfish, and might be expected to be overly optimistic if catchability increases (improved targeting efficiency) are not properly described.
- If recruitment deviates are actually substantive and responsible for both increases and decreases in the Japanese CPUE series, then the stochastic recruitment scenarios are probably more realistic.

The results suggest that there may be cause for concern at this time, if the worst case scenarios are close to reality. In future assessments, additional effort needs to be taken to evaluate the plausibility of the different model specifications, particularly the reliability of the different data sources and the implications of the conflicting inferences. Items of concern and priorities are outlined in the conclusions stated in the main text.

Table A1-1. Distinguishing features of some of the assessment models considered to date. Yellow highlights indicate models that seem to have converged to a solution that does not trivialize the impact of the fishery.

Model ID	Japanese CPUE SD(log)	Taiwanese (case 4) CPUE SD(log)	La Reunion CPUE SD(log)	Growth and Mortality	Number of Selectivity functions estimated	Migration rate	Beverton-Holt Steepness (*h*)	Recruitment deviates estimated
swo61p4	SW 1995+ 0.1 q shared	0.1 q not shared	0.1	Growth Slow M=0.2	1=longline 2=gillnet	<1 % per year	0.7	No
swo62p4	SW 1995+ 0.1 q shared	0.1 q not shared	0.1	Growth Slow M=0.2	1=longline 2=gillnet	<1 % per year	0.9	No
swo63p4	SW 1995+ 0.1 q shared	0.1 q not shared	0.1	Growth Fast M=0.4	1=longline 2=gillnet	<1 % per year	0.9	No
swo64p4	SW 1995+ 0.1 q shared	0.1 q not shared	0.1	Growth Slow M=0.2	18	<1 % per year	0.9	No
swo65p4	SW 1995+ 0.1 q shared	0.1 q not shared	0.1	Growth Slow M=0.2	Age-based 1=longline; 2=gillnet	<1 % per year	0.9	No
swo61p6	SW 1995+ 0.1 q shared	0.1 q not shared	0.1	Growth Slow M=0.2	1=longline 2=gillnet	<1 % per year	0.7	Yes
swo62p6	SW 1995+ 0.1 q shared	0.1 q not shared	0.1	Growth Slow M=0.2	1=longline 2=gillnet	<1 % per year	0.9	Yes
swo63p6	SW 1995+ 0.1 q shared	0.1 q not shared	0.1	Growth Fast M=0.4	1=longline 2=gillnet	<1 % per year	0.9	Yes
swo64p6	SW 1995+ 0.1 q shared	0.1 q not shared	0.1	Growth Slow M=0.2	18	<1 % per year	0.9	Yes
swo65p6	SW 1995+ 0.1 q shared	0.1 q not shared	0.1	Growth Slow M=0.2	Age-based 1=longline; 2=gillnet	<1 % per year	0.9	Yes

Table A1-2. Quality of fit diagnostics for the 9 CPUE series, and key reference points for the assessment model specifications from Table A1-1.

Model	CPUE RMSE JPN-NW	CPUE RMSE JPN-NE	CPUE RMSE JPN-SW	CPUE RMSE JPN-SE	CPUE RMSE TWN-NW	CPUE RMSE TWN-NE	CPUE RMSE TWN-SW	CPUE RMSE TWN-SE	La Reunion
swo61p4	0.52	0.43	0.39	0.74	0.24	0.51	0.37	0.53	0.08
swo62p4	0.54	0.42	0.38	0.75	0.25	0.51	0.38	0.54	0.08
swo63p4	0.59	0.44	0.39	0.75	0.25	0.50	0.41	0.54	0.08
swo64p4	0.48	0.42	0.36	0.74	0.25	0.51	0.37	0.54	0.08
swo65p4	0.51	0.44	0.38	0.75	0.25	0.52	0.37	0.53	0.08
swo61p6	0.46	0.37	0.36	0.63	0.28	0.51	0.36	0.57	0.08
swo62p6	0.46	0.37	0.36	0.63	0.28	0.51	0.36	0.57	0.08
swo63p6	0.43	0.34	0.32	0.57	0.29	0.49	0.39	0.58	0.10
swo64p6	0.41	0.37	0.30	0.58	0.28	0.51	0.40	0.60	0.08
swo65p6	0.46	0.39	0.36	0.64	0.27	0.52	0.37	0.55	0.08

Model	Max. Gradient	F2007/FMSY	B2007/BMSY	B/B0	MSY (1000t)	N Parameters
swo61p4	9.59E-05	1.27	1.11	0.49	23	20
swo62p4	2.77E-04	1.28	0.94	0.45	26	20
swo63p4	9.45E-06	0.92	0.95	0.46	33	20
swo64p4	2.52E-04	1.96	0.69	0.33	21	100
swo65p4	6.39E-04	1.64	0.86	0.42	22	20
swo61p6	9.59E-05	0.16	1.95	0.86	128	77
swo62p6	4.37E-05	0.14	1.79	0.87	154	77
swo63p6	3.03E-04	0.15	1.29	0.62	171	77
swo64p6	1.13E-04	0.84	1.23	0.60	34	157
swo65p6	6.39E-04	0.80	1.23	0.60	35	77

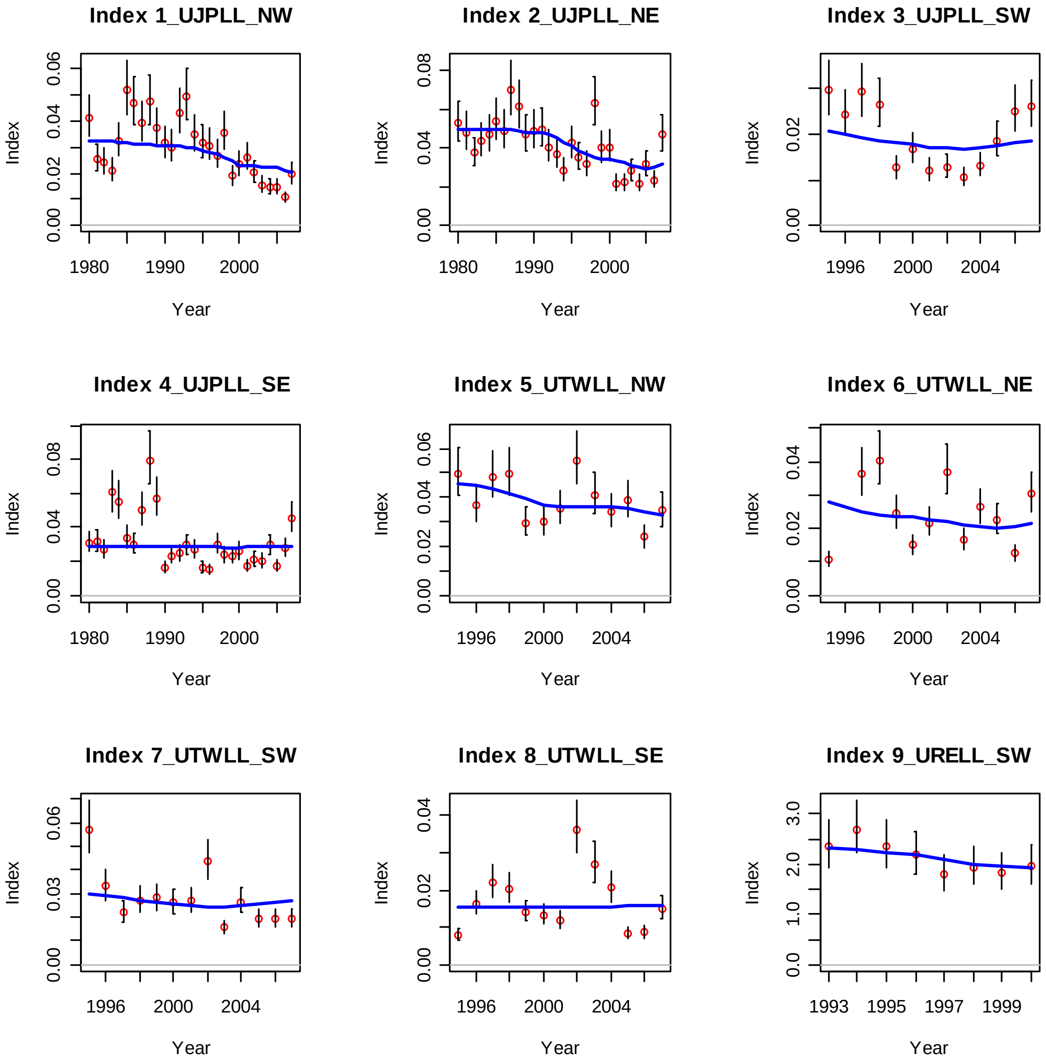


Figure 16. Predicted (lines) and observed (circles with 95% error bars) standardized Japanese longline CPUE for swo62p4.

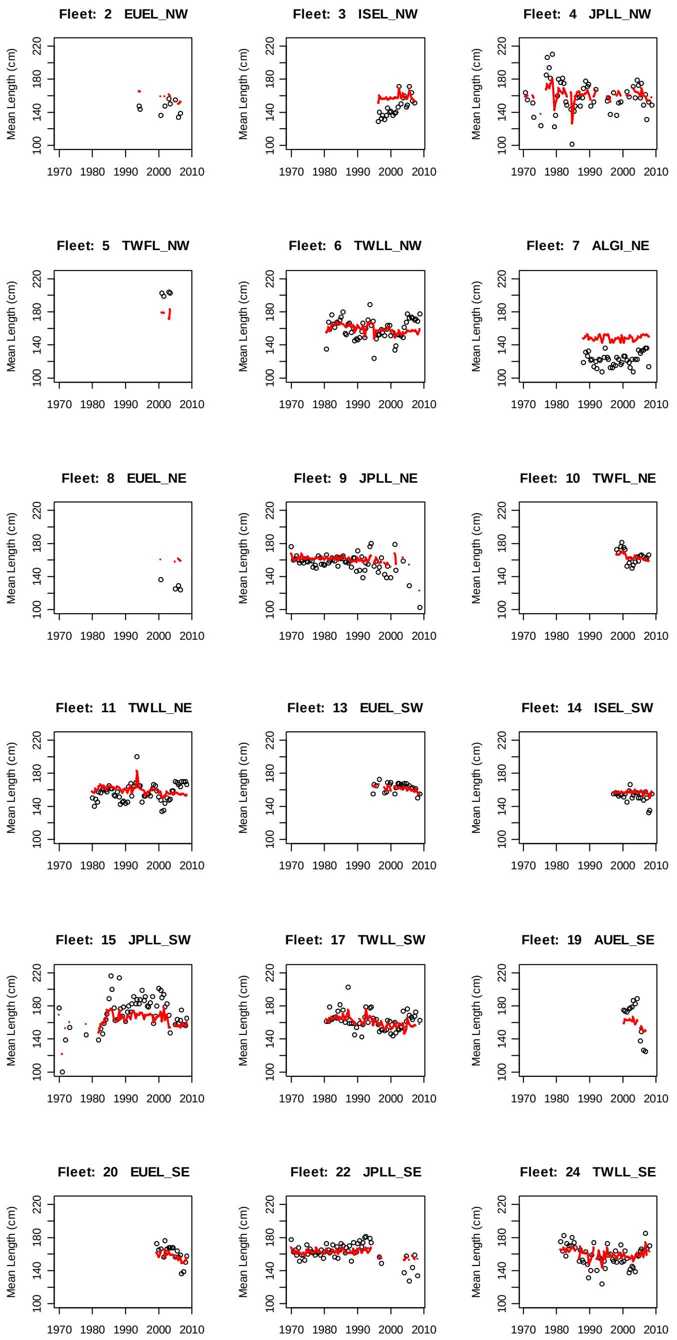


Figure 17. Predicted (red lines) and observed (black circles) mean catch for each of the 18 fleets with size composition data, for model swo62p4. Note that the SS3 predicted mean catches are not quite right because only predicted size bins with corresponding observation bins are included in the calculation (i.e. particularly when sample sizes are small, the plotted predicted values will closely resemble the observed values, even if the full distributions are quite different).

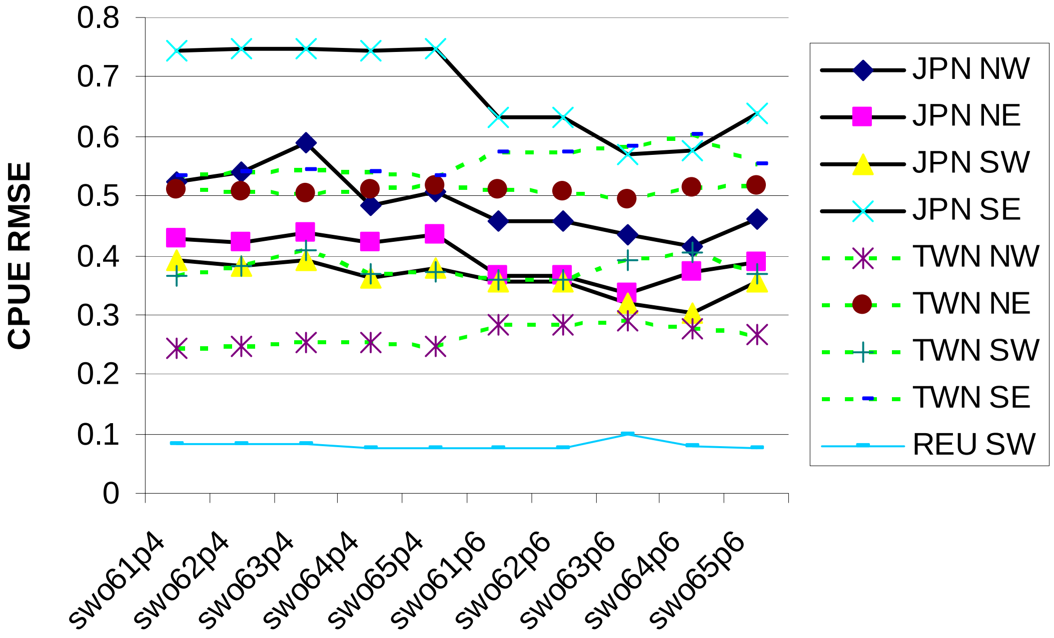


Figure 18. Comparison of the CPUE quality of fit indices across the 10 models. Note that CPUE RMSE provides an index of the quality of fit that is independent of the input CPUE CV assumptions. Main points from this plot: 1) qualitatively, the differences among models in the CPUE fits are not really large, 2) the most important model feature influencing the fit to the CPUE is whether or not the recruitment deviates are estimated, and 3) The Taiwanese CPUE series fits are generally similar or better than the Japanese CPUE series even though they are highly down-weighted relative to the Japanese series (but they also cover a shorter time period).

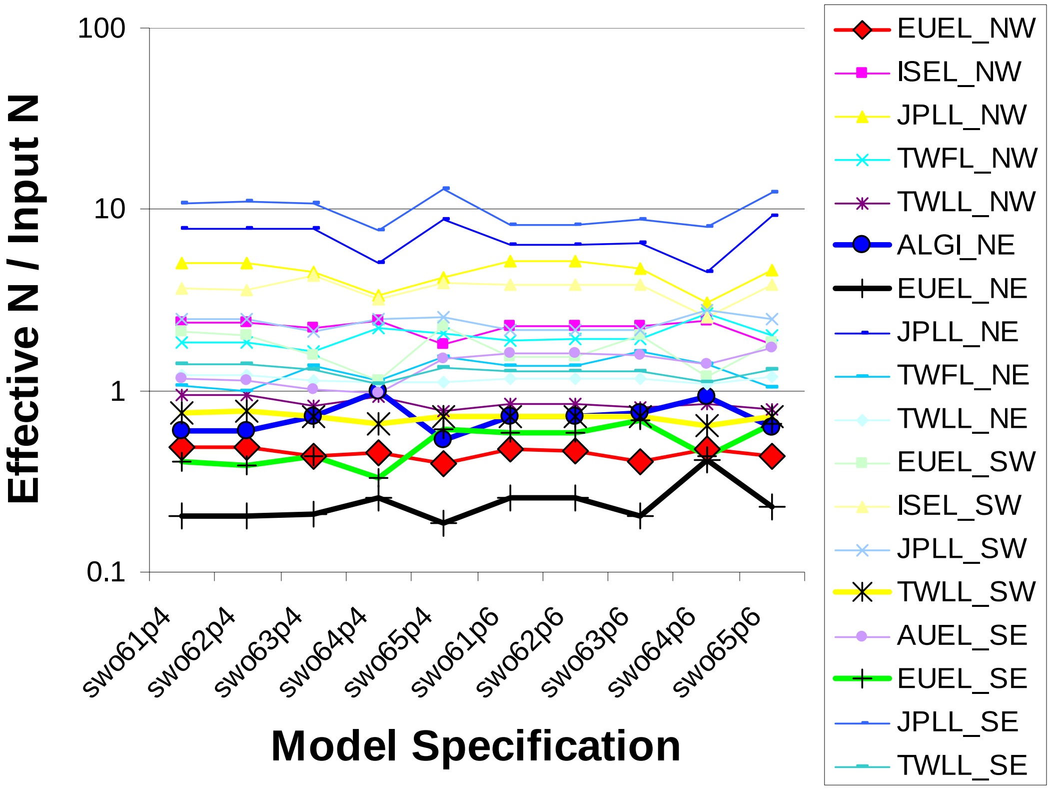


Figure 19. Comparison of the size composition consistency between predictions and observations across models. Note that the index "Effective N / Input N" is quantifying the degree to which the model fit is consistent with the sample size assumptions. Ideally, all values should be near to 1 if a model is completely internally consistent. If the ratio is much lower than one, it does not necessarily mean that the model fit is very bad, but it does reflect that the fit is worse than expected if the input sample size is truly random and the other model assumptions are met (e.g. selectivity constant over time). The key messages from this plot include: 1) some of the size composition fits are considerably worse than expected given the model assumptions, and 2) the consistency of fits are not very dependent on the alternative structural assumptions across the 10 models, hence they do not provide much basis for selecting preferable models among them.

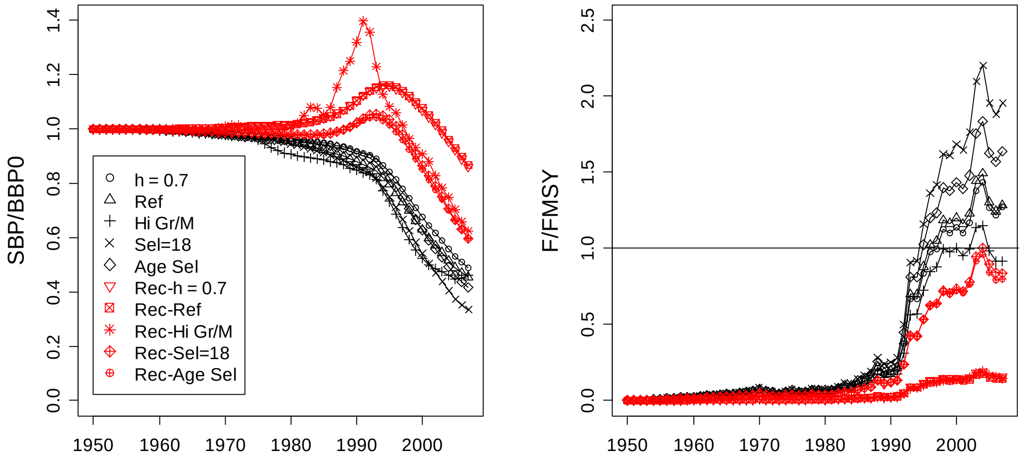


Figure 20. Spawning biomass and fishing mortality time series for the 5 SS3 models fit during the WPB 2009. The first 5 time series in the legend (black) have no recruitment deviations estimated, the latter 5 time series (red) estimate annual recruitment deviations (but otherwise the two sets of model specifications are identical).

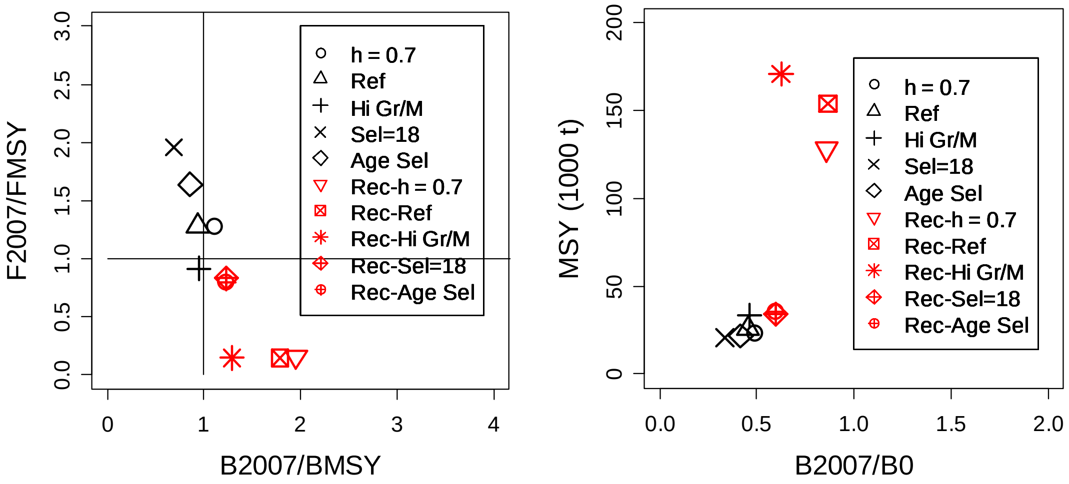


Figure 21. Maximum Posterior Density Stock status summary plots for the 10 Indian Ocean swordfish models fit during the WPB 2009. The first 5 time series in the legend (black) have no recruitment deviations estimated, the latter 5 time series (red) estimate annual recruitment deviations (but otherwise the two sets of model specifications are identical).

Attachment 2. Resolution of SS3 IO swordfish assessment issues identified during the IOTC WPB 2009

Two important issues were identified prior to and during the WPB 2009, that the author could not resolve during the meeting. This attachment illustrates that the consequences of these issues are potentially important for the inferences made at the WPB 2009. Results tabled at the WPB are broadly consistent with the revised results, but these latter results are subtly different and more correct. Attachment 2 should be taken as the starting point for the next iteration of the IO swordfish assessment.

Problem 1

The SS3 report files provided an MSY estimate that differed (and was consistently lower than) the mode of the equilibrium yield curve. It turns out that the MSY value was actually defined using a switch in the SS3 forecast.SS file which used SPR(50%) as a proxy for MSY. The effect on the 5 SS3 models reported to the WPB 2009 is shown in Figures A2-1 and A2-2. The revised MSY-related reference points suggest a dramatic shift, from somewhat pessimistic stock status to reasonably optimistic. While it goes without saying that this is a cautionary warning about being familiar with all of the switches in the software, it is also a general indicator of the sensitivity of MSY-related reference points (presumably somebody considers SPR-50% to be a reasonable MSY proxy in some circumstances). Note that the depletion estimators remain unchanged between Figures A2-1 and A2-2.

Problem 2

There seemed to be ambiguity in the instructions related to the ordering of fleet and survey area specifications in the SS3 set-up files, with differences between the software documentation and the labelling in the graphics routines. As a consequence, some of the size composition data and CPUE series were allocated to the wrong area in the WPB specifications. This is potentially a serious problem, and would ordinarily have been expected to cause obvious pathological problems in the model. However, the diagnostic problems that were evident, were actually relatively subtle (poor fit to the relatively minor gillnet-associated fleets, and minimization sensitivity to some sources of data that were expected to be relatively uninformative).

The stock status results estimated for the incorrect and correct area allocations are shown in fig. A2-3 (for the models defined in Table A1-1). The area assignment problem obviously had an impact on the stock status estimates, however, a number of factors seemed to prevent major and obviously detectable problems: i) the fishery development history is similar for most areas, ii) the catch-at-size data are similar for most fleets (or at least assumed to be equivalently uninformative), and the selectivity parameters are shared for most fisheries (in most scenarios), iii) the gillnet fleets (which have substantially different selectivity), were not well fit in any of the original scenarios, but they represent a very small proportion of the total catch, and iv) the

Japanese CPUE trends provide most of the informative contrast in depletion, and are rather similar among most of the areas with substantial depletion.

The levels of depletion in the different areas are broadly similar with the correct and incorrect fleet-area assignments, with the exception of the SE (fig. A2-5). The depletion in the SE region was estimated to be minimal in the WPB specifications, while the corrected models suggest that depletion in the SE is similar to that observed in the other regions.

The combined effects of the MSY-reporting and area assignment errors offset one another to some extent. Relative to the results reported to the WPB, the corrected MSY-related stock status estimates are somewhat more optimistic, while the depletion-based estimates are somewhat more pessimistic. Given that the WPB stock status summary represented a qualitative attempt to integrate results across numerous model formulations, and these revised results are not particularly alarming relative to the other approaches used, there is no urgent need to amend the WPB report to reflect these corrections.

New Sensitivity Tests

In the available timeframe it was not possible to conduct a meaningful exploration of the model uncertainty (and it is not clear what the results would have meant anyway, given the two problems identified above). This section describes a series of new models in which the (known) SS3 specification problems were resolved, and looks at a few interactions among core data sets and assumptions.

Seventeen models are described in Table A2-1. Each model was fit with and without the estimation of recruitment deviates. The definitions reflected a cursory attempt to:

- Resolve the poor fit to the gillnet-associated size composition data
- Understand the stock status implications of the relative weighting of the CPUE and size composition data
- Repeat the preliminary sensitivity analyses, including
    - alternative steepness
    - alternative growth and mortality
    - alternative selectivity assumptions

Figures A2-5 to A2-7 show some quality of fit indices for the Japanese CPUE series, and the size composition data for a representative longline and gillnet fishery for the full suite of models. The interpretation of these qualitative fit indices requires further consideration, however at least one index does map directly onto an obvious problem that is discernible by eye. The relatively low effective sample size for the gillnet-associated fishery (fig. A2-7) is associated with an obvious bias in mean catch size (see the contrast in the ALGI_NE fishery in figs. A2-8 and A2-9). The contrast in quality of fit for the longline fishery size composition data is more subtle (i.e. there is less evidence for a systematic bias in estimated mean size, but there are time trends in the quality of fit (particularly toward the end of the time series). In the results below, all of the models with badly biased gillnet mean size indices were removed.

Three models with contrasting fit to the CPUE series are shown in figures A2-10 to A2-12. Fig. A2-10 shows that the more pessimistic models can fit the declining

CPUE series very well, even if the recruitment time series are deterministic. Fig. A2-11 shows a typical fit to the CPUE series when the catch-at-size data are not down-weighted and recruitment deviates are not estimated. These models tend to be much more optimistic, but they generally fail to describe the magnitude of the downward Japanese CPUE trend. However, if recruitment deviates are estimated, the pessimistic models (without down-weighted longline size data) can also fit the CPUE trends reasonably well (Fig. A2-12).

Stock status summary plots (from the models without obvious biases in gillnet fishery mean size) are shown in Figures A2-13 to A2-15. These models span a large range of stock status possibilities. The more pessimistic models are the ones in which the size composition data for the longline fleets is downweighted, and they tend to be the ones that fit the Japanese CPUE decline most closely. There are a number of reasons for questioning the assumptions related to both data sources. However, in the absence of additional information, we would tend to have more confidence in the models that adequately fit the CPUE series. If the CPUE series is reliable, it is imperative that the trend should be properly described. If the CPUE series is poor then it is unlikely that the model will be very useful even if the size composition data is reliable (even if the constant selectivity assumption is met).

Further work is clearly required to understand the interactions of the data sources and the model assumptions, including:

- There are obvious problems with the catch rate assumptions in some regions. Can mechanistic explanations be found for the discrepancies in series observed (e.g. in the South-West). Can we be confident that either the Japanese or Taiwanese catch rates are able to be appropriately standardized to remove the effects of species targeting, technological changes and systematic spatial shifts in effort?
- Does it make sense to assume that fishery selectivity is constant over time? Can the fleet aggregations and spatial/temporal resolution of the catch-at-size data be processed in a way that makes the constant selectivity assumption more defensible? Or can temporally variable selectivity be invoked instead?
- What are the implications of the interactions of the alternative assumptions in the model? (i.e. one dimensional deviations in assumptions from an arbitrary reference case do not describe the real uncertainty)
- It would be useful to quantify the uncertainty associated with individual models (i.e. using the inverse Hessian approximation, or bootstrapping, etc.), however, at this time the model specification uncertainty is expected to dwarf the statistical uncertainty conditional on any specific model.

Table A2-1. Distinguishing features of a suite of Indian Ocean swordfish assessment models that were fit subsequent to the WPB 2009. Yellow highlights indicate model differences from a nominal reference case (swo102). Model ID labels with blue shading indicate models that had obvious problems fitting the mean size of the gillnet fisheries.

Model ID	Japanese CPUE SD(log) (LLH-weighting)	Taiwanese (case 4) CPUE SD(log) (LLH-weighting)	La Reunion CPUE SD(log)	Number of Selectivity functions estimated	Gillnet catch-at-size LLH-weighting	Longline catch-at-size LLH-weighting	Growth and Mortality	Beverton-Holt Steepness (*h*)
swo81-85	corrected versions of swo61-65 as defined in Table A1-1							
swo101	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (1)	0.1	1=longline 2=gillnet	1	1	Growth Slow M=0.2	0.7
swo102	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (1)	0.1	1=longline 2=gillnet	1	1	Growth Slow M=0.2	0.9
swo103*	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (1)	0.1	1=longline 2=gillnet	1	1	Growth Fast M=0.4	0.9
swo104	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (1)	0.1	18	1	1	Growth Slow M=0.2	0.9
swo105	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (1)	0.1	Age-based 1=longline; 2=gillnet	1	1	Growth Slow M=0.2	0.9
swo107	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (1)	0.1	1=longline 2=gillnet	0.01	0.01	Growth Slow M=0.2	0.9
swo108h7	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (0.01)	0.1	1=longline 2=gillnet	1	1	Growth Slow M=0.2	0.7
swo108h9	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (0.01)	0.1	1=longline 2=gillnet	1	1	Growth Slow M=0.2	0.9
swo109	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (0.01)	0.1	1=longline 2=gillnet	0.01	0.01	Growth Slow M=0.2	0.9

swo110	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (1)	0.1	1=longline 2=gillnet	0.1	0.1	Growth Slow M=0.2	0.9
swo111	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (1)	0.1	1=longline 2=gillnet	0.001	0.001	Growth Slow M=0.2	0.9
swo112	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (1)	0.1	1=longline 2=gillnet	1	0.01	Growth Slow M=0.2	0.9
swo113	SW 1995+ 0.1 q shared (0.01)	0.1 q not shared (1)	0.1	1=longline 2=gillnet	1	0.01	Growth Slow M=0.2	0.9
swo121	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (1)	0.1	1=longline 2=gillnet	1	0.01	Growth Slow M=0.2	0.7
swo123	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (1)	0.1	1=longline 2=gillnet	1	0.01	Growth Fast M=0.4	0.9
swo124	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (1)	0.1	18	1	0.01	Growth Slow M=0.2	0.9
swo125	SW 1995+ 0.1 q shared (1.0)	0.1 q not shared (1)	0.1	Age-based 1=longline; 2=gillnet	1	0.01	Growth Slow M=0.2	0.9

* the stochastic recruitment version of model swo103 estimated an extremely large biomass, trivial fishery mortality, and explained declining recent biomass trends purely through recruitment variability.

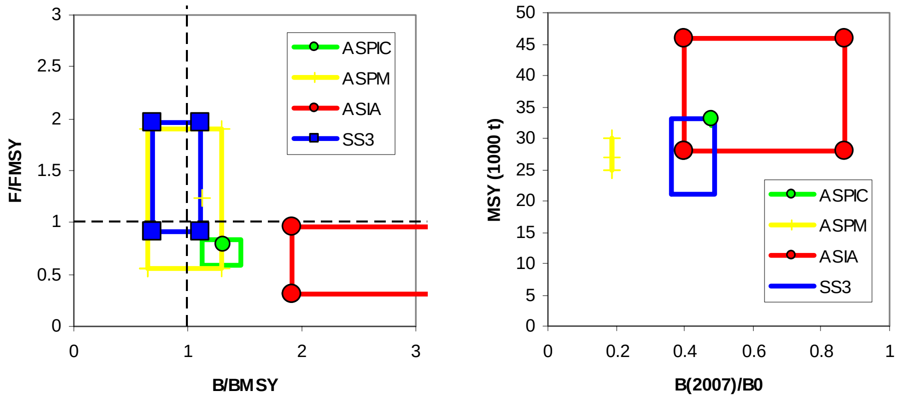


Figure A2-1. MSY-related Reference points reported to the WPB 2009. Note that the SS3 results (blue boxes) are all based on a proxy for MSY (SPR-50%), in contrast to the true MSY values (below).

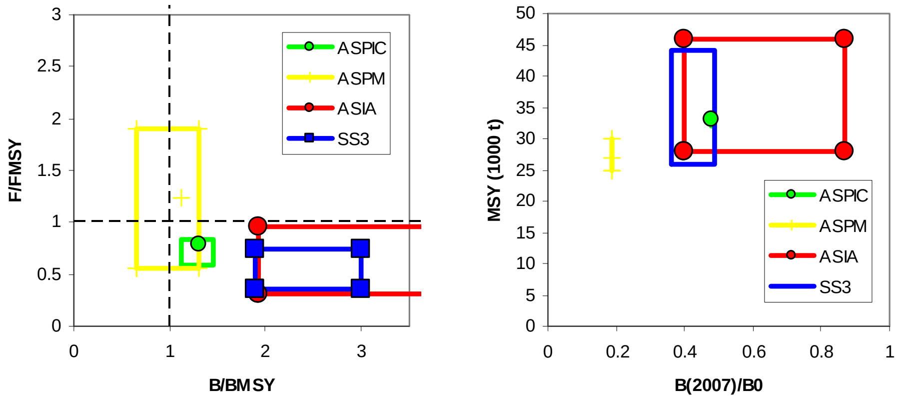


Figure A2-2. Comparison of stock status reference points for the various models represented in the WPB 2009 report, as Figure A2-1, except that the SS3 results use the actual MSY-related reference points instead of the SPR proxy.

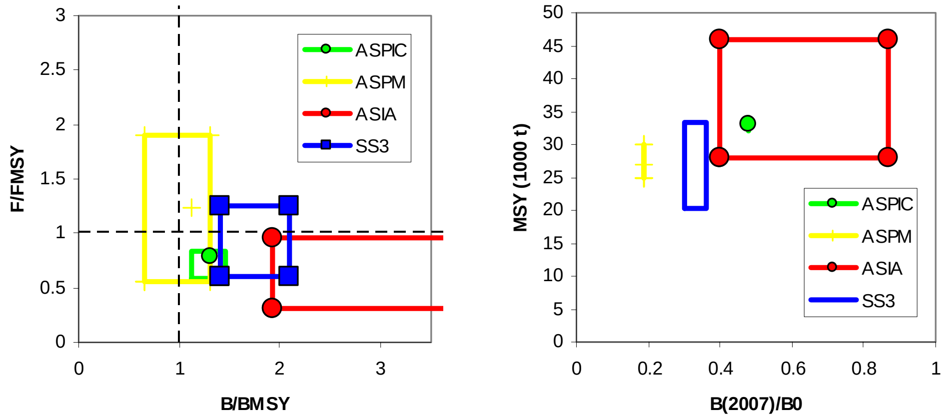


Figure A2-3. Represents the same figure as the above two, except that the fishery area definitions are corrected (and the correct MSY definition is employed).

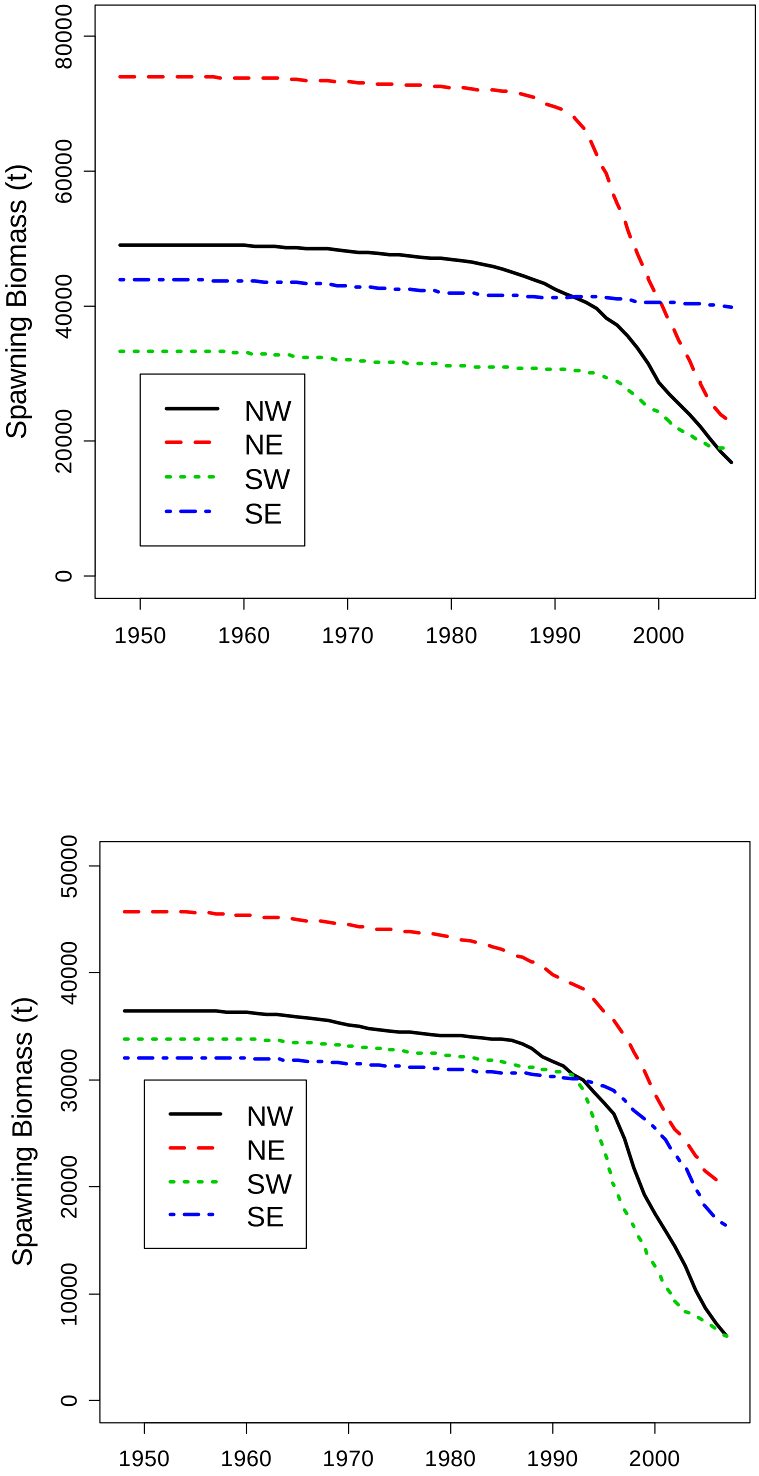


Figure A2-4. Comparison of the relative depletion across areas for the reference case models with the incorrect (top – swo62p4) and correct (bottom 82p4) fleet-area definitions.

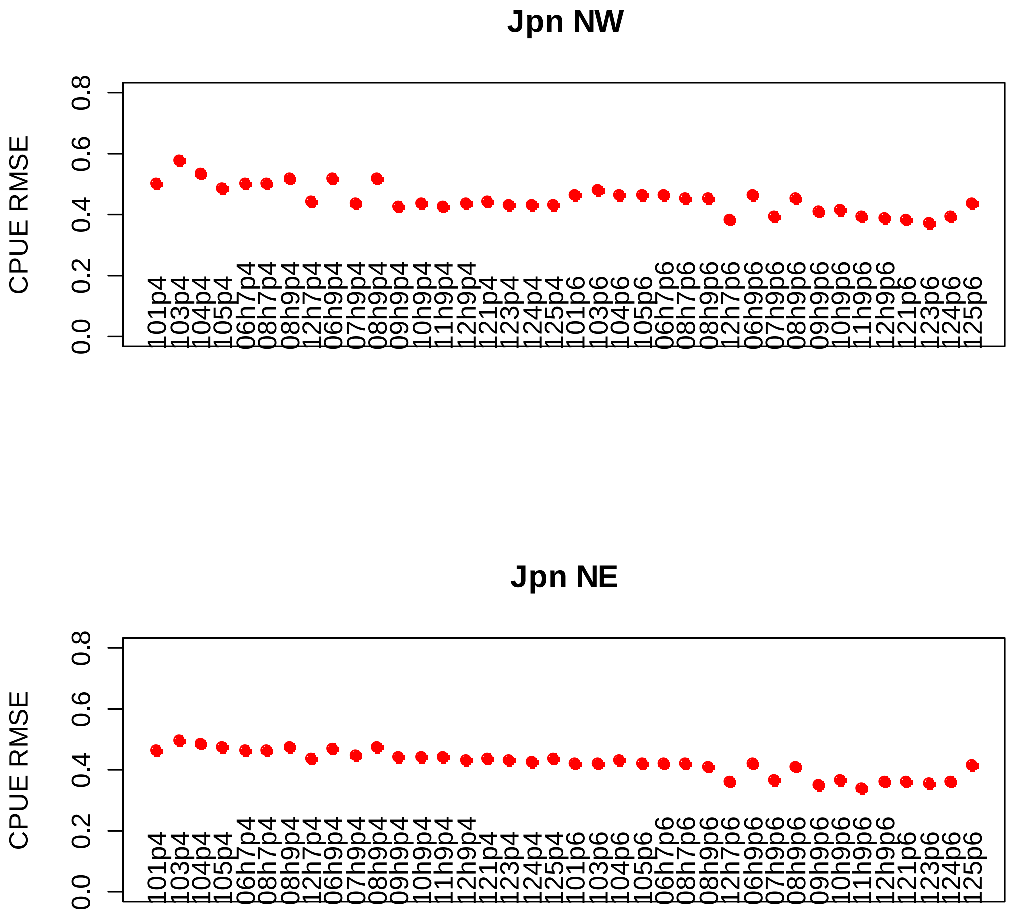


Figure A2-5. Quality of fit (RMSE) indices for the Indian Ocean swordfish Japanese CPUE indices for a range of models.

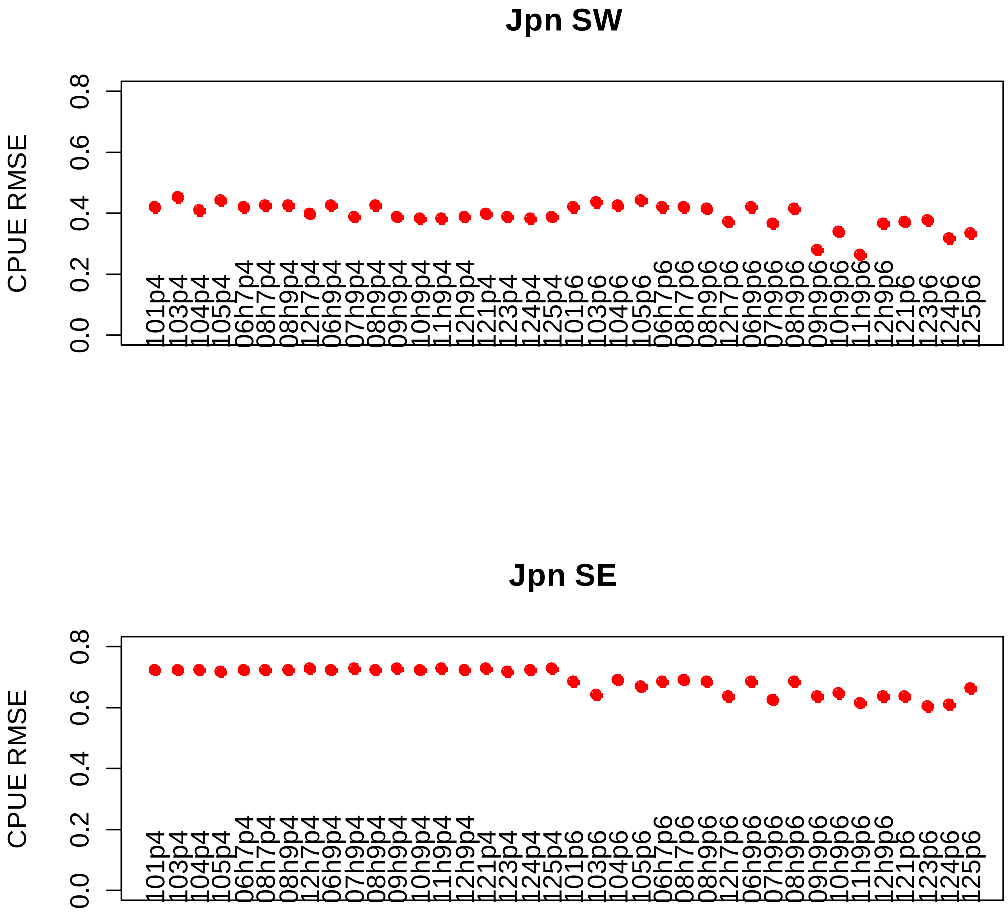
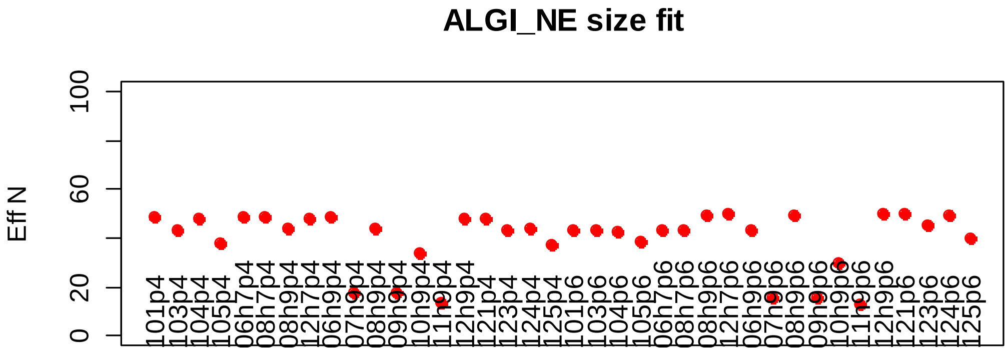


Figure A2-6. Quality of fit (RMSE) indices for the Indian Ocean swordfish Japanese CPUE indices for a range of models.



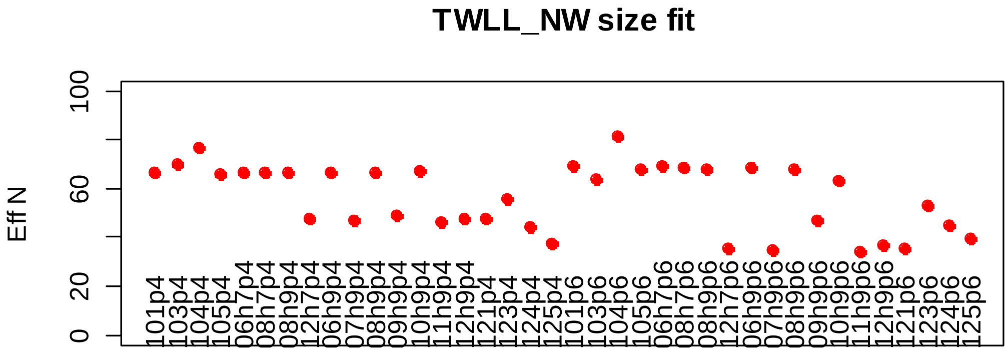


Figure A2-7. Quality of fit indices for the catch-at-size data (effective N) for the main gillnet fishery (top), and a typical longline fishery (bottom).

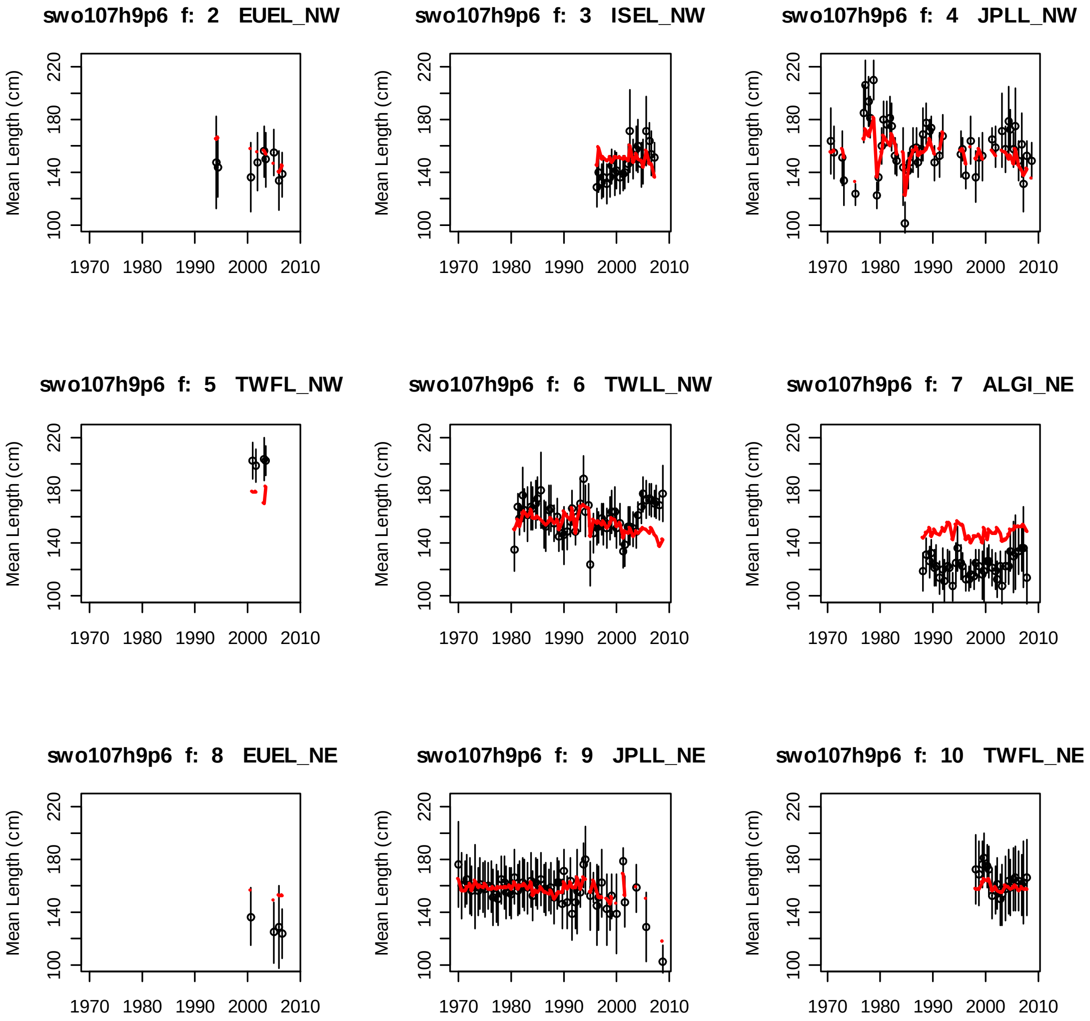


Figure A2-8. Predicted (lines) and observed (circles with 95% confidence intervals) mean size indices for 9 fisheries. Note the consistent bias in fishery 7 ALGI_NE. Note that predicted means are an approximation based on only the size bins with corresponding observations (approximation is very poor for very small sample sizes).

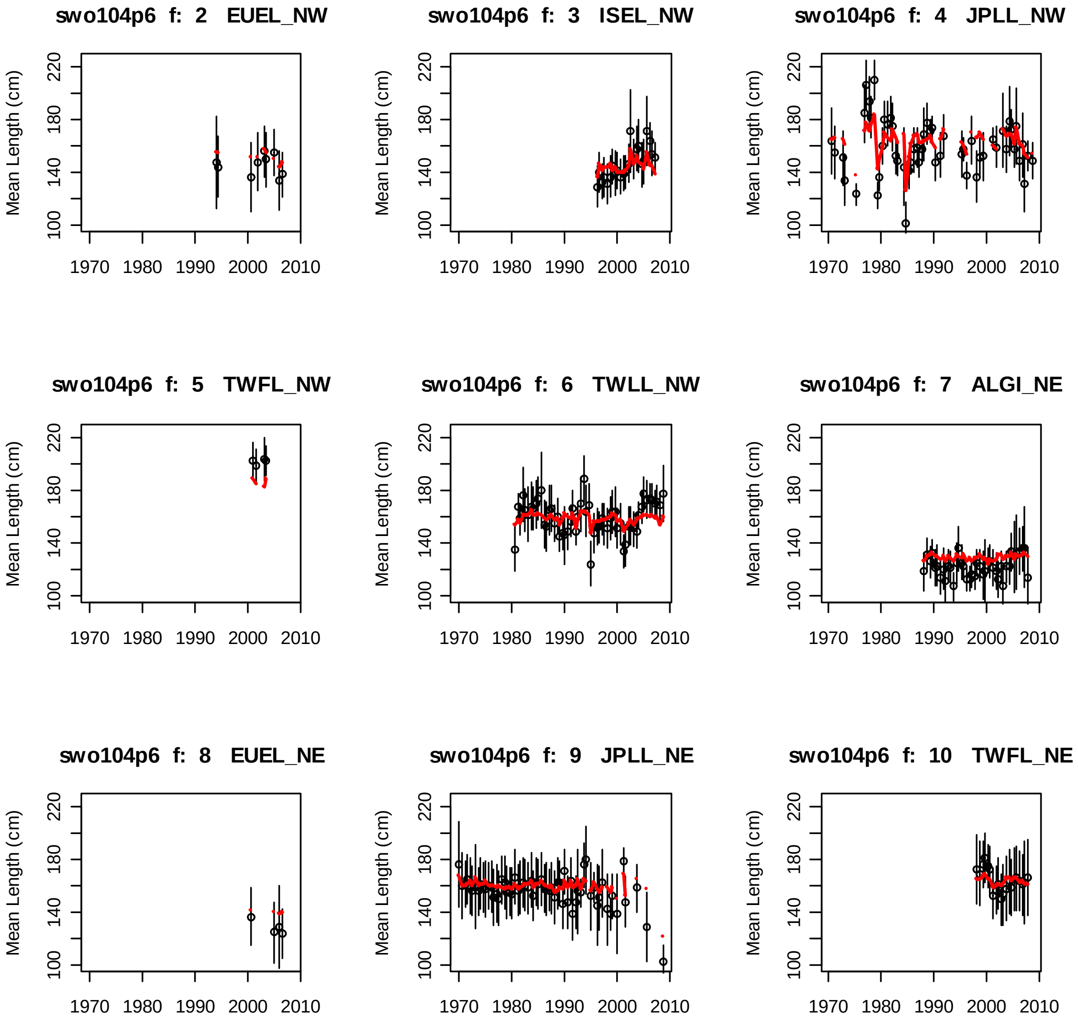


Figure A2-9. Predicted (lines) and observed (circles and 95% confidence intervals) mean size indices for 9 fisheries. Note the bias in fishery 7 ALGI_NE is much reduced relative to Figure A2-8. Note that predicted means are an approximation based on only the size bins with corresponding observations (approximation is very poor for very small sample sizes).

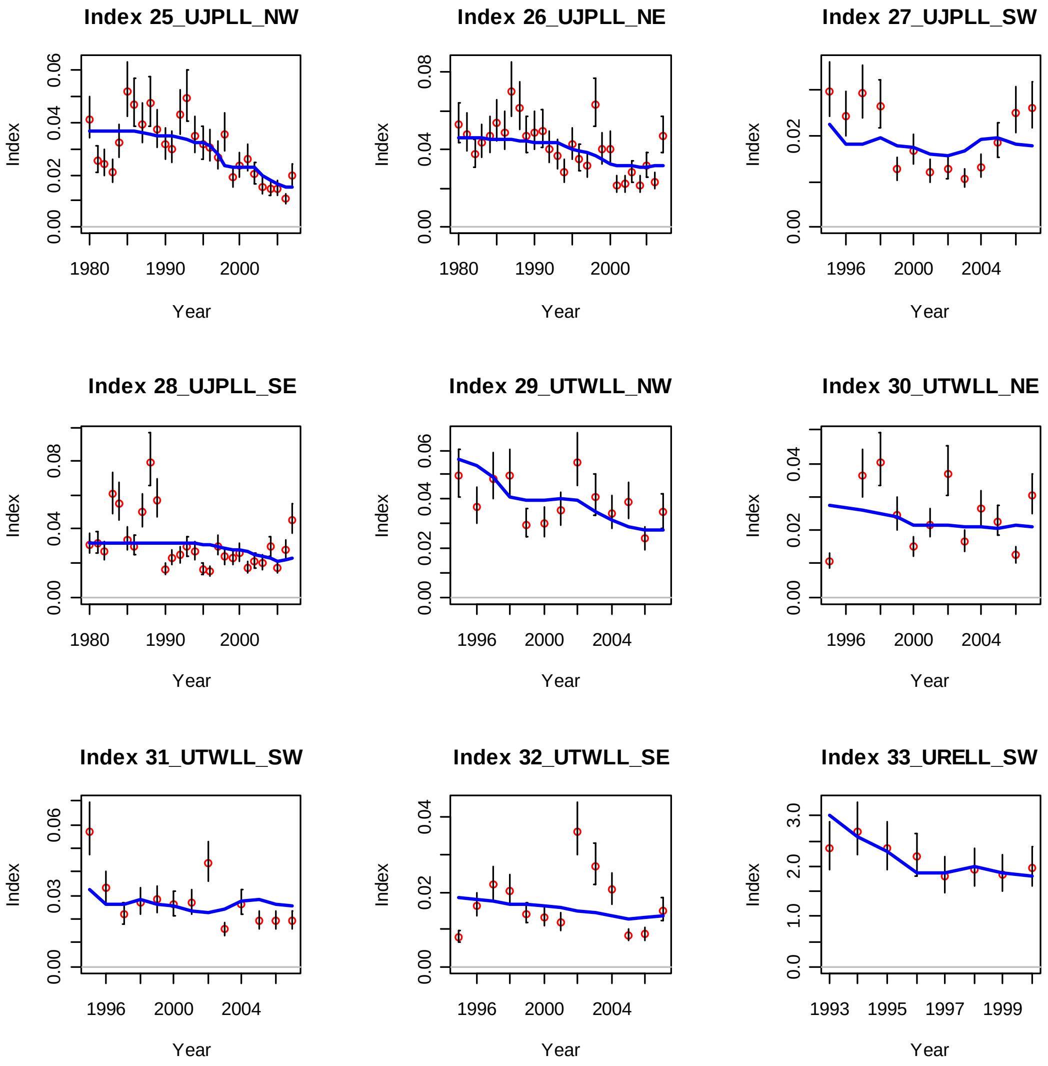


Figure. A2-10. Predicted (lines) and observed (points with error bars) CPUE for model swo123 (recruitment deviates not estimated).

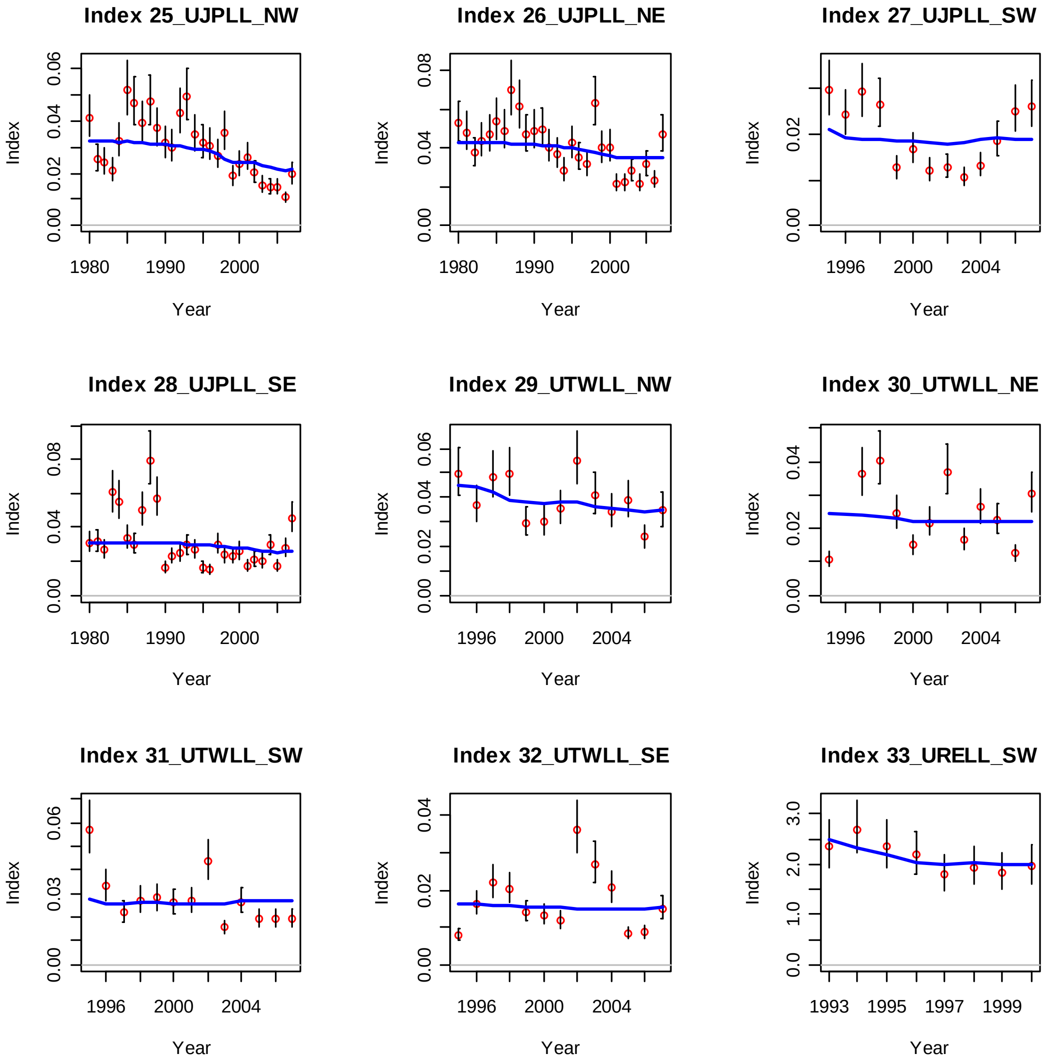


Figure. A2-11. Predicted (lines) and observed (points with error bars) CPUE for model swo104 (recruitment deviates not estimated).

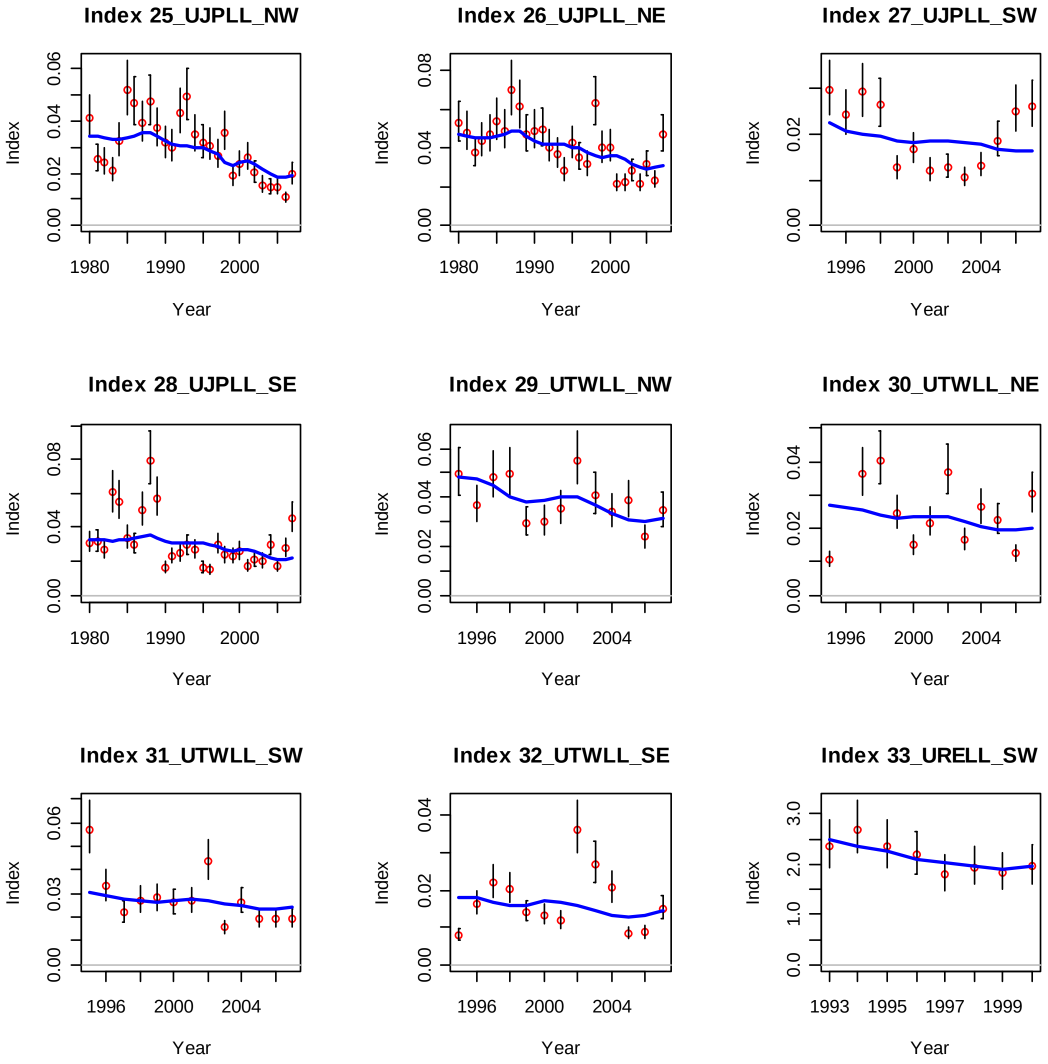


Figure. A2-12. Predicted (lines) and observed (points with error bars) CPUE for model swo104 (recruitment deviates estimated).

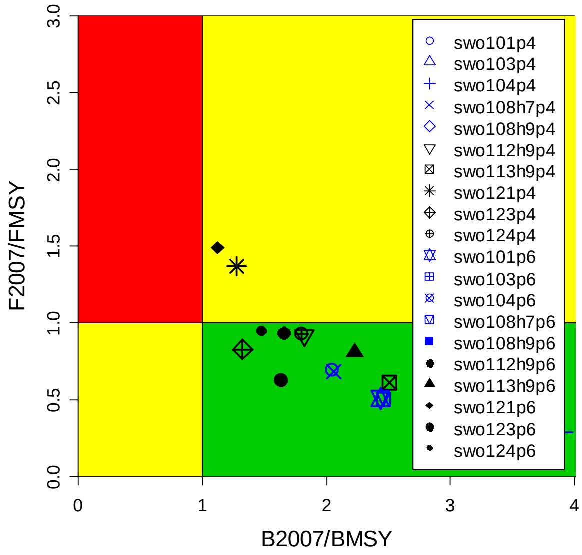


Figure A2-13. Kobe plot for a selection of SS3 models fit after the WPB 2009 (defined in Table A2-1). Black symbols (the last 4 in the legend) represent models with down-weighted size composition data for all longline fleets (blue symbols represent models without down-weighted size data). Several models without down-weighted size data are off of the scale.

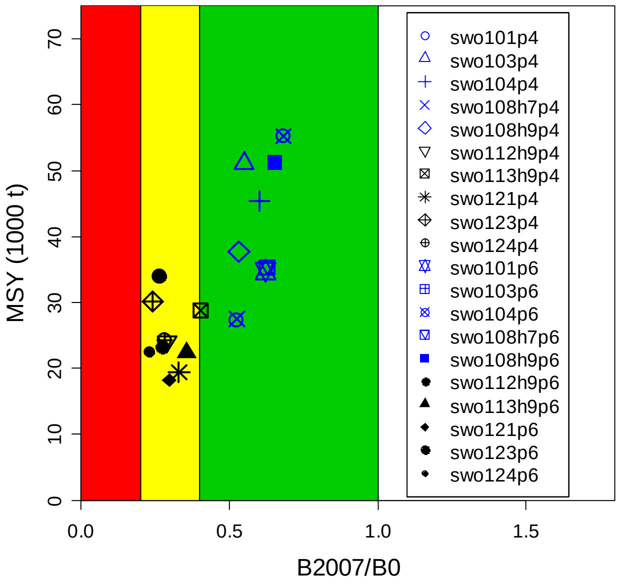


Figure A2-14. Depletion and MSY for a selection of SS3 models fit after the WPB 2009 (defined in Table A2-1). Black symbols (the last 4 in the legend) represent models with down-weighted size composition data for all longline fleets (blue symbols represent models without down-weighted size data. One of the models without down-weighted size data is off of the MSY scale. Note that the yellow region reflects the zone of depletion from B/B0 = 0.2 – 0.4. These values are often used as limit and target reference points in Australian and U.S. fisheries.

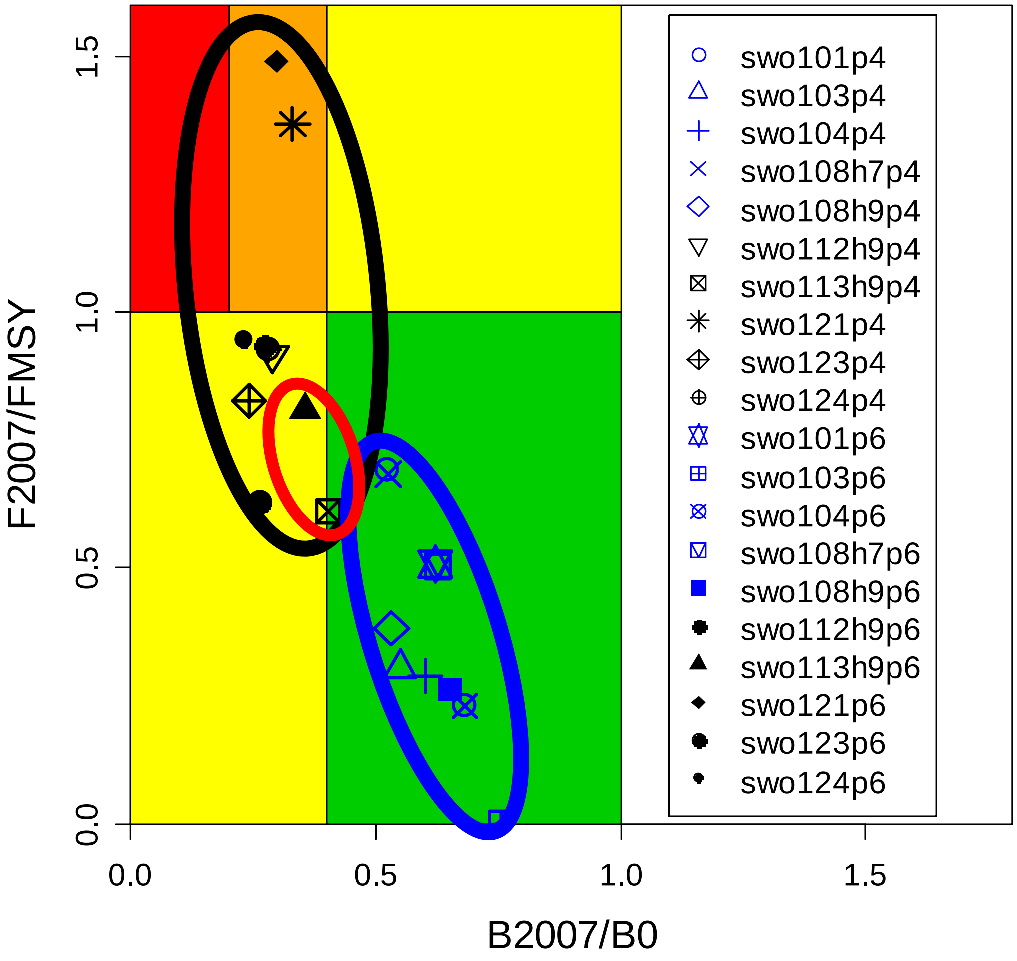


Figure A2-15. Illustration of the main factors influencing the Indian Ocean swordfish stock status uncertainty. Symbols in the upper left (black) ellipse indicate models with down-weighted size composition likelihood terms. Symbols in the bottom right (blue) ellipse indicate models without down-weighted size composition data (although sample sizes are still reduced by a factor of 10 relative to the raw data). The symbols in the central (red) ellipse indicate models with Japanese CPUE and longline size composition down-weighted.