Background to the species splitting model adopted for the development of OMP-18.

by

OLSPS Marine

Silvermine House

Steenberg Office Park

Tokai 7945

**November 2018**

# Executive Summary

Potential species splitting models are fitted to either the survey dataset or the combined survey and observer dataset. The observer data shows a noticeable difference to the survey data w.r.t. the relationship between the proportion of ***M. paradoxus*** and depth, and so it is proposed that the final model should be based on fits to the historical survey + observer data, rather than, as previously, solely the survey data. The model also differs from previous models produced for the current OMP because of the inclusion of trawl ID as a random effect, and the consequent use of GLMM techniques.

The BIC is used as a first pass filter to produce a short list of models. This short list includes variants with selected two-way interactions. The cross-validation error for this short list was produced. Based on this, model parsimony considerations and practical considerations (e.g. year effect is an impractical inclusion in the model), it was initially proposed that the preferred model should be a main effects only model with depth, size class and longitude or latitude as main effects. This amounted to modifying the prevailing model to include longitude for the South Coast or latitude for the West Coast, and using the parameter estimates that are fitted to the combined survey and observer dataset.

Following these recommendations the DWG deliberated further and suggested an alternative model based on the following considerations:

* Whether to use the observer data in the development of the model, and the relative weighting of the survey versus the observer data. The decision was to utilise the observer data, and to allow the weighting to be in accordance with the available data (i.e. the data sets were simply merged and analysed as such).
* Whether to use Trawl ID as a random effect in the analyses. The view of the DSWG was that Trawl ID should not be used in the analyses and was therefore omitted from the final model.
* Whether to weight records by the number of fish sampled, or simply allow the value of P to have equal weight across all records. The DSWG chose the latter.
* Whether P should be calculated on a numbers basis or on a mass basis. The DSWG chose the latter.

# Background

The species splitting model presently in use for 2018 and earlier is described in Appendix A (see also OLRAC SPS, 2013a-e). This model uses the following variables:

* Size class as a categorical
* Depth as a covariate

At the time that this model was adopted circa 2013 it was noted that there were differences between the proportion of ***M. paradoxus*** as a function of depth in the survey data and in the observer data (referred to here as SADSTIA and OROP) (see OLRAC SPS, 2013a-e). The DSWG requested that a sensitivity analysis be carried out in which the end-to-end implications of using instead a version of the model in which the statistical weight of the observer data was increased in model fits was tested. Robustness to this sensitivity test gave assurances that the management of the resource was not being biased by the use of the survey based model.

Subsequently, further work was done to try to account for the differences between the observer data and the survey data as regards the relationship between depth and the proportion of ***M. paradoxus*** in catches. This work has included following a number of research directions, involving, inter alia (see OLRAC SPS, 2016; OLSPS Marine, 2017):

1. The preparation of a master database suitable for statistical analysis comprising SURVEY, SADSTIA and OROP data.
2. Extensive efforts directed at data cleaning, as described in OLRAC SPS (2016) and OLSPS Marine (2017).
3. The use of GLMM’s instead of GLM’s because of the modification to use **trawl ID as a random effect** for model fitting purposes.
4. Introduction of the data source indicator, “INPUT”, as an explanatory variable in the GLMM analyses, and pursuing a large number of research directions to try to eliminate the role of “INPUT” in the model fit.
5. Introduction of “Observer ID” as a potential explanatory variable in the GLMMs.
6. The introduction of “CPUE” as an additional potential explanatory variable in the GLMMs.
7. At times, the use of hake size as a covariate instead of the small, medium and large size categories.
8. Defining p as the proportion of ***M. paradoxus*** in a trawl by **weight** instead of **number**.
9. Use of month to represent season instead of quarter.
10. Testing the goodness of model fit for all possible GLMMs including all possible two way interactions (10s of thousands of models).
11. Use of cross-validation methods as an alternative to goodness of fit measures to assess model acceptability.
12. The elimination of outliers.

The last of these, the elimination of outliers proved to be the most productive at minimising the influence of “INPUT” in the final model fits. Nevertheless the influence of “INPUT” was not eliminated and differences remain in the relationship of P, the proportion of ***M. paradoxus***, to depth between the survey data and the observer data. Further work was carried out in 2017 on this topic (e.g. OLSPS Marine, 2017).

The use of trawl ID as a random effect is motivated by the consideration of likely covariance between the proportion of M. paradoxus in the small medium and large size categories for the same trawl, due perhaps to size selective shoaling behaviour. Inclusion of this factor has a material impact on the model parameter estimates, as demonstrated below.

This document reports pertinent technical background information to the revision of the species splitting model for the development of OMP-18.

# Methods

## Data

Three sources of data were used for this study:

1. **Survey data** - data on the species breakdown of trawls from research surveys carried out since about 1985.
2. **Observer data**, OROP - data on the species breakdown of trawls from commercial operations sampled by observers operating under the auspices of the OROP observer programme which was a DAFF run operation.
3. **Observer data**, SADSTIA - data on the species breakdown of trawls from commercial operations sampled by observers operating under the auspices of the SADSTIA observer programme which was and still is an observer programme run by SADSTIA.

## GLMMs

A GLMM with a logit link function and a binomial distribution is used, i.e.

Where depends on which factors are included in the model. The following model description assumes the inclusion of latitude (West Coast) or longitude (South Coast) group, depth, year, season and size class:

where:

; is the proportion of ***Merluccius paradoxus*** by number for a given trawl ID,

; is the model intercept,

; is the mean depth of the trawl in metres, and is the associated parameter for the covariate,

; is a categorical variable, being the latitude bin on the West Coast and the longitude bin on the South Coast

; is a categorical variable for year,

; is a categorical variable designating season,

; is a categorical variable for the small, medium or large size class.

Model variants reported here also explored the use the trawl ID, as a random effect with a normal distribution, an important change to the work reported in OLRAC SPS (2013a-e).

### Depth:

Depth, measured in metres, is included as a covariate in all species splitting models. The **a**verage depth value was utilized when a trawl record had both start and end depth readings, otherwise the value of either the start or end depth was used, whichever was present.

### Longitude or Latitude:

Longitude and latitude are used to measure the effect that geographic location has on the proportion of ***Merluccius paradoxus*** in the total hake catch; longitude is used on the South Coast and latitude is used on the West Coast. Geographic location is binned into 1° bins along the coast and treated as a categorical value. The **a**verage latitude (see LatMean below) or longitude (see LongMean below) reading was used when a trawl record had both start and end quantities, otherwise we used the value of either the start or end, whichever was available. Further details available on request.

### Season:

Seasons are quarterly with Summer in December to February.

### Year Block:

For both observer and survey data, trawls were grouped into 5 year year-block categories starting in 1984.

### Size class:

Hake size is represented in the model as a categorical variable denoting the size classes used for catch recording in commercial operations, i.e. small, medium and large. Since the size composition of commercial catches which will only ever be available in small, medium and large size classes, any such model has to be developed in terms of small, medium and large size classes. The survey and observer data were aggregated into small, medium and large size classes on the basis, 21 cm to 42 cm, 43 cm to 57 cm, and 58 cm+.

### Trawl ID

For the observer data, a unique trawl ID was assigned for each unique combination of Haul\_Date, Trip\_ID and Trawl No. For the survey data, a unique trawl ID was assigned for every distinct Year, Month, Day, Cruise, Station, Trawl No. combination. Only the “abundance” trawl records were used from the Survey Dataset.

### Coast Definition

Records were allocated to the West Coast “W” if their mean latitude and longitude position were within the requisite geographical limits (details available on request). The records falling outside the above-mentioned criteria were designated south “S”.

## Dredge

To overcome the problem of setting up and testing all possible main effect and two way interaction models (for main effects only there are 25 models) this study was selective about the interactions that were considered, and used dredge, an automated model selection tool available in R, to evaluate the performance of models using all feasible predictors, and then to rank the models according to a suitable model selection criterion. The Bayesian Information Criterion (BIC) was used as the goodness of fit criterion for the dredge analysis.

## BIC and Cross-validations

In addition to the dredge search based on the BIC, the cross-validation error variance was calculated for models identified by Dredge as promising. The cross-validation approach adopted here involved running GLMMs on a random 50% sample of the data, the training dataset, Q=30 times over, and using the resultant model to predict the value of the proportion of ***M. paradoxus*** (see VarP) or the value of the number of fish sampled as ***M. paradoxus*** (see VarN) for the complementary 50% of the data (known as the testing dataset). VarDeltoaP and VarDeltaN, shown in, for example Table 5, are then the variance of the difference between predicted and actual values in the complementary testing dataset, averaged over all Q.

## Exclusion of outliers

The outlier exclusion procedure is to omit records which for the best model (i.e. Rank 1 from Table 2 and Table 4) shows a difference between the actual proportion of ***M. paradoxus*** and the predicted proportion of ***M. paradoxus*** of more than 3 standard deviation of the residual.

## Models of Interest

Focussed results were presented for the following models, A, I, ID, IL, LS, L,B:

A = DepthMean + SizeClass + LatCat/LongCat + Season + YearBlocks.

I   = DepthMean + SizeClass + LatCat/LongCat + Season + YearBlocks + DepthMean\*Season + LatCat/LongCat\*Season.

ID = DepthMean + SizeClass + LatCat/LongCat + Season + YearBlocks + DepthMean\*Season.

IL = DepthMean + SizeClass + LatCat/LongCat + Season + YearBlocks + LatCat/LongCat\* Season.

LS = DepthMean + SizeClass + LatCat/LongCat + Season.

L = DepthMean + SizeClass + LatCat/LongCat.

B = DepthMean + SizeClass

# Results

## Dredge

The dredge results for the West Coast are summarised in Table 1 after exclusion of outliers for the combination dataset survey + observers, and for the survey only dataset in Table 2.

The dredge results for the South Coast are summarised in Table 3 after exclusion of outliers for the combination dataset survey + observers, and for the survey only dataset in Table 4.

## Cross-validations

Table 5 shows the available cross validation results for West Coast and South Coast using the Survey + Observer datasets, while Table 6 shows the results for the West Coast and South Coast using only the Survey dataset.

## Other results

Additional outputs are provided to aid in the visualisation of the data and some of the results. The following four figures provide a comparison in the relationship between proportion of M. paradoxus between the survey data and the observer data:

1. Figure 1: PPar\_Mean versus DepthCat line plots for West Coast survey (red) and observer records (blue) pertaining to the different size classes 2, 3 and 4.
2. Figure 2: PPar\_Mean versus DepthCat line plots for West Coast survey (red) and observer records (blue) pertaining to the different size classes 2, 3 and 4 where the DeltaPPar outliers were excluded.
3. Figure 3: PPar\_Mean versus DepthCat line plots for South Coast survey (red) and observer records (blue) pertaining to the different size classes 2, 3 and 4.
4. Figure 4: PPar\_Mean versus DepthCat line plots for South Coast survey (red) and observer records (blue) pertaining to the different size classes 2, 3 and 4 where the DeltaPPar outliers were excluded.

The following four figures give an indication of the differences between model predictions for an LS model fitted to either the survey data only, or to the survey + observer data.

1. Figure 5: Predicted PPar\_Mean Versus DepthCat line plots for the West Coast survey and observer records based on a LS GLMM model which was either fitted initially with only survey records (Mean PredPParS) or survey and observer records (Mean\_PredPParSO) for the different size classes 2, 3 and 4.
2. Figure 6: Predicted PPar\_Mean Versus DepthCat line plots for the South Coast survey and observer records based on a LS GLMM model which was either fitted initially with only survey records (Mean PredPParS) or survey and observer records (Mean\_PredPParSO) for the different size classes 2, 3 and 4.
3. Figure 7: Predicted PPar Versus Depth dot plots for the West Coast survey and observer records based on a LS GLMM model which was either fitted initially with only survey records (Mean PredPParS) or survey and observer records (Mean\_PredPParSO) for the different size classes 2, 3 and 4.
4. Figure 8: Predicted PPar Versus Depth dot plots for the South Coast survey and observer records based on a LS GLMM model which was either fitted initially with only survey records (Mean PredPParS) or survey and observer records (Mean\_PredPParSO) for the different size classes 2, 3 and 4.

## Results related to either including or excluding Trawl ID as a random effect

Table 9 and Table 10 give a comparison between model parameters obtained using GLMM and trawl ID as a random effect, compared to the GLM run without the use of trawl ID as a random effect.

Figure 9 and Figure 10 are plots of the fitted parameter values without trawl ID vs with trawl ID.

Figure 11 compares the catches and CPUE indices that are obtained when the species splitting either includes or exclude the trawl ID as a random effect.

# Discussion

## Graphical features

Note that despite the use of an outlier exclusion approach, the observer data shown in Figure 1 and Figure 2 shows evidence of the metre/fathom unit issue mentioned elsewhere in this document.

## Initial Model Selection Argument

In identifying a preferred model out of initial candidates A, I, ID, IL, LS, L and B, the following points were considered:

1. Size class and depth are non-negotiable inclusions in the model.
2. The statistical criteria are to minimise the number of model factors, the BIC and the cross-validation error.
3. From a practical point of view, it is preferable not to have year as a factor in the model, since the model, if applied to future data, may then have to apply to data for a period for which no corresponding year estimates are available. The same consideration is applicable in the event that interactions are used in the model, since the data may not cover all combinations either possible or that may arise in the future.

With this in mind it was noted that

* Models I, ID and IL do not meet the model parsimony criterion, and their performance did not seem sufficiently superior to outweigh the practical and model parsimony criteria. Accepting this, the choice was narrowed to models A, LS, L, B.
* Model B, the existing model being used and described in Appendix A, is inferior to all others w.r.t. BIC (see Table 5) but does not do too badly overall w.r.t. cross validation errors (see Table 5), although for the South Coast it does quite badly on the proportion based cross-validation error measure VarDeltaP, scoring worst out of all models on this particular criterion. Model B was therefore nor considered further.
* The surviving contenders are therefore Models A, LS and L. Model A is just Model LS with year block included as a main effect. Practical considerations mitigate against the use of year and hence Model A. The results in Table 5 suggest that dropping Model A is not too much of a loss on statistical grounds. For the West Coast it seems that either L or LS will do. Parsimony considerations suggest therefore that the preferred model for the West Coast is Model L. For the South Coast it depends on whether one considers the proportion or numbers based cross validation error measure. On the basis that the N based measure is closer to a catch weighted result, the preferred model for the South Coast would be Model L.

In summary, the initial recommendation was to modify the existing model being used to include latitude as a factor in the case of the West Coast, and longitude in the case of the South Coast. A caveat to this is that, should this lead to the need to exclude a significant proportion of records because their positional information is not available, then Model B should remain the model in use.

### Model parameter estimates

Model parameter estimates are presented for a selection of models, including the initial proposal Model L, for the West Coast in Table 7 and for the South Coast in Table 8.

Despite that the estimation of the model parameters includes a trawl ID random effect, the application of the model does not involve trawl ID consideration, i.e. the trawl ID is implicitly zero for prediction/scoring purposes.

## Subsequent developments

Following the submission of the recommendations outlined above, the DSWG requested a series of additional model results and ultimately came to a different conclusion. In the process the DSWG pointed out certain corrections that were required and which were made by OLSPS. They then focussed on the following considerations:

* Whether to use the observer data in the development of the model, and the relative weighting of the survey versus the observer data. The decision was to utilise the observer data, and to allow the weighting to be in accordance with the available data (i.e. the data sets were simply merged and analysed as such).
* Whether to use Trawl ID as a random effect in the analyses. The view of the DSWG was that Trawl ID should not be used in the analyses and was therefore omitted from the final model.
* Whether to weight records by the number of fish sampled, or simply allow the value of P to have equal weight across all records. The DSWG chose the latter.
* Whether P should be calculated on a numbers basis or on a mass basis. The DSWG chose the latter.

The final model selected by the DSWG is summarised here in Appendix B.

# References

Anon, 2009. Percentage of ***Merluccius capensis*** caught per grid block on the South Coast during Demersal Research Surveys 1985 - 2009. MCM/2009/NOV/SWG-DEM/102. 1 page.

Gaylard J.D. and M.O. Bergh. 2003.An investigation into the procedure used to split commercial catches of hake on the South African South Coast into ***Merluccius paradoxus*** and ***Merluccius capensis***. BEN/JAN04/SAH2b.

Gaylard J.D. and M.O. Bergh. 2004a.A size-dependent species splitting mechanism applied to hake catches off the South African West Coast. WG/08/04/D:H:13

Gaylard J.D. and M.O. Bergh M. 2004b. A species splitting mechanism for application to the commercial hake catch data 1978 to 2003. Marine Coastal Management Document WG/09/04/D: H: 21. 8 pp.

Gaylard J.D. and M.O. Bergh M. 2004c. Comparison of hake species splits from observer data with the survey-generated splitting algorithm. 16 pp.

OLRAC-SPS, 2006. A comparison of the proportions of ***Merluccius capensis*** and ***Merluccius paradoxus*** in survey and observer data. DSSC2006\_6. February 2006, 33 pp.

Gaylard J.D. and M.O. Bergh M. 2006. Comparison of hake species splits from observer data with the survey-generated splitting algorithm. September 2006. 16 pp.

Gaylard J.D. and M.O. Bergh. 2007. Further comparison of hake species splits from observer data with the survey-generated splitting algorithm. November 2007, 24 pp.

Gaylard J.D. and M.O. Bergh. 2009. Update of the hake species split models in the light of more recent survey data and a revision of the large/medium/small size classification. MCM/2009/NOVEMBER/SWG-DEM/…., November 2009, 13 pp.

OLRAC-SPS CC, 2011b, Revised 2012. Progress on client actions for revision of the species splitting model. February 2011 Revised February 2012. 2 pp.

OLRAC-SPS. 2013a. A further update of the hake species splitting model. Unpublished Working Group Document *FISHERIES/2013/FEB/SWG-DEM/12*, and associated ADDENDUM (Updated parameter estimates for the hake species split model). 16pp+1pp

OLRAC-SPS. 2013b. Addendum #1 to “**[FISHERIES/2013/FEB/SWG-DEM/12]** A further update of the hake species splitting model.” – Updated parameter estimates for the hake species split model. 6 pp.

OLRAC-SPS. 2013c. Addendum #2 to “**[FISHERIES/2013/FEB/SWG-DEM/12]** A further update of the hake species splitting model.” 11 pp.

OLRAC-SPS. 2013d. Alternative estimates of the hake species splitting model biased towards observer data. 23 April 2013**. 4 pp.**

OLRAC-SPS. 2013e. Alternative estimates of the hake species splitting model biased towards observer data. 16 September 2013. FISHERIES/2013/SEP/SWG-DEM/48. 11 **pp.**

Leslie R.W., B. Rose and J. Scholte. 1998. Hake grading by Irvin & Johnson and by Sea Harvest. WG/01/98/D:H:03. I do not have a copy of this document.

OLRAC-SPS. 2016. An investigation into the percentage of ***Merluccius paradoxus*** in trawls recorded in surveys and in commercial operations. By OLRAC-SPS, Silvermine House, Steenberg Office Park, Tokai 7945, South Africa. In partial fulfilment of client action plan for maintaining MSc certification. April 2016. 40 pp.

OLSPS-Marine. 2017. Interim results obtained while revising the species splitting formula for hake. By OLSPS Marine, Silvermine House, Steenberg Office Park, Tokai 7945, South Africa. In partial fulfilment of client action plan for maintaining MSc certification. May 2017. 30 pp.

Table 1: Dredge Analysis summary of all the feasible and valid GLMM model combinations using Survey and Observer records pertains to the West Coast.

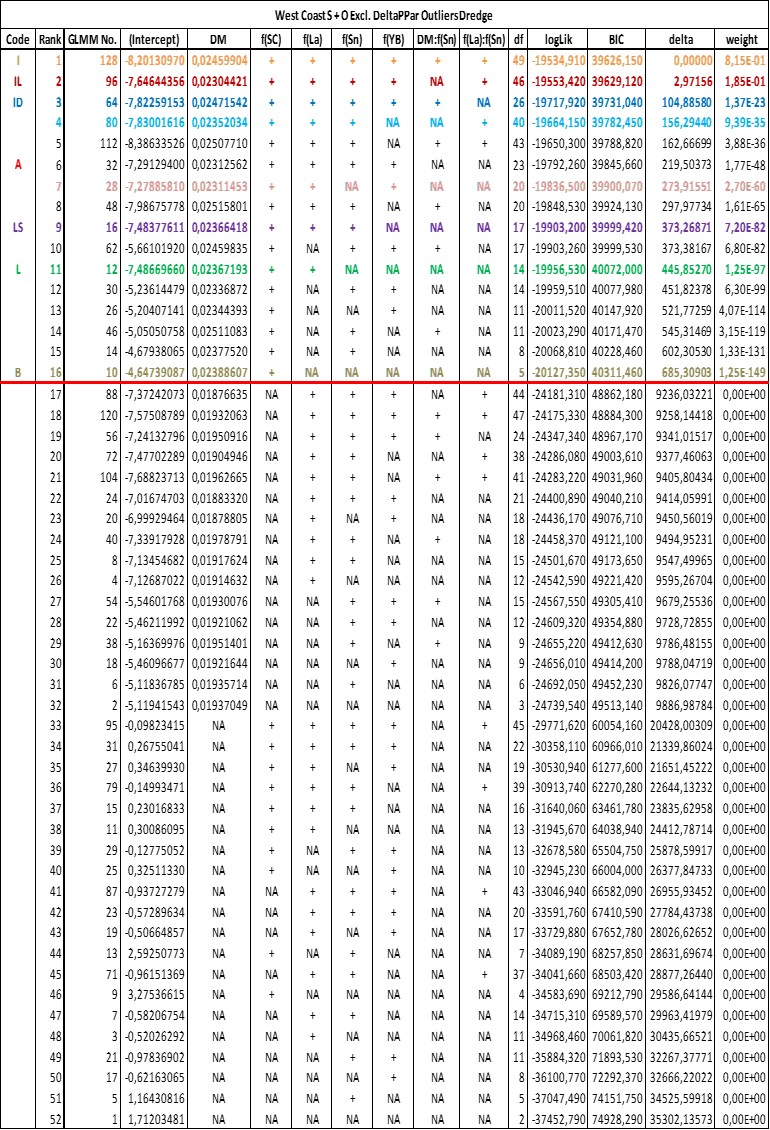


Table 2: Dredge Analysis summary of all the feasible and valid GLMM model combinations using only Survey records pertains to the West Coast.

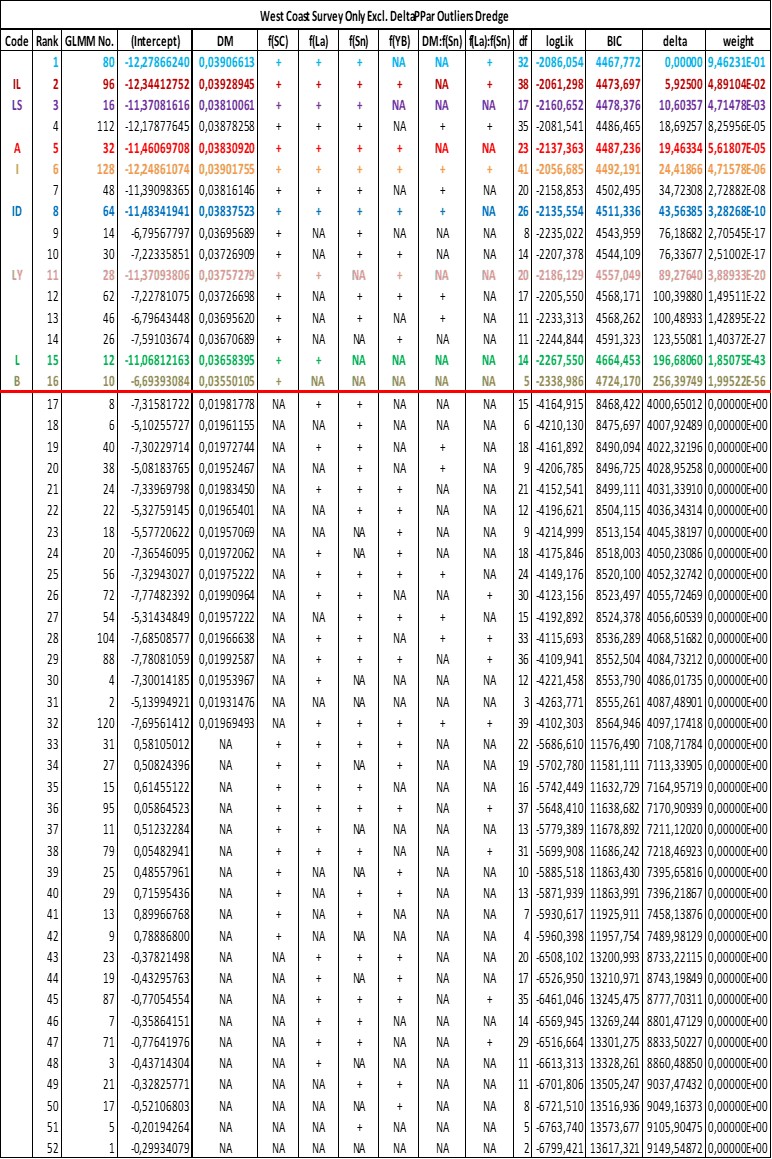


Table 3: Dredge Analysis summary of all the feasible and valid GLMM model combinations using Survey and Observer records pertains to the South Coast.

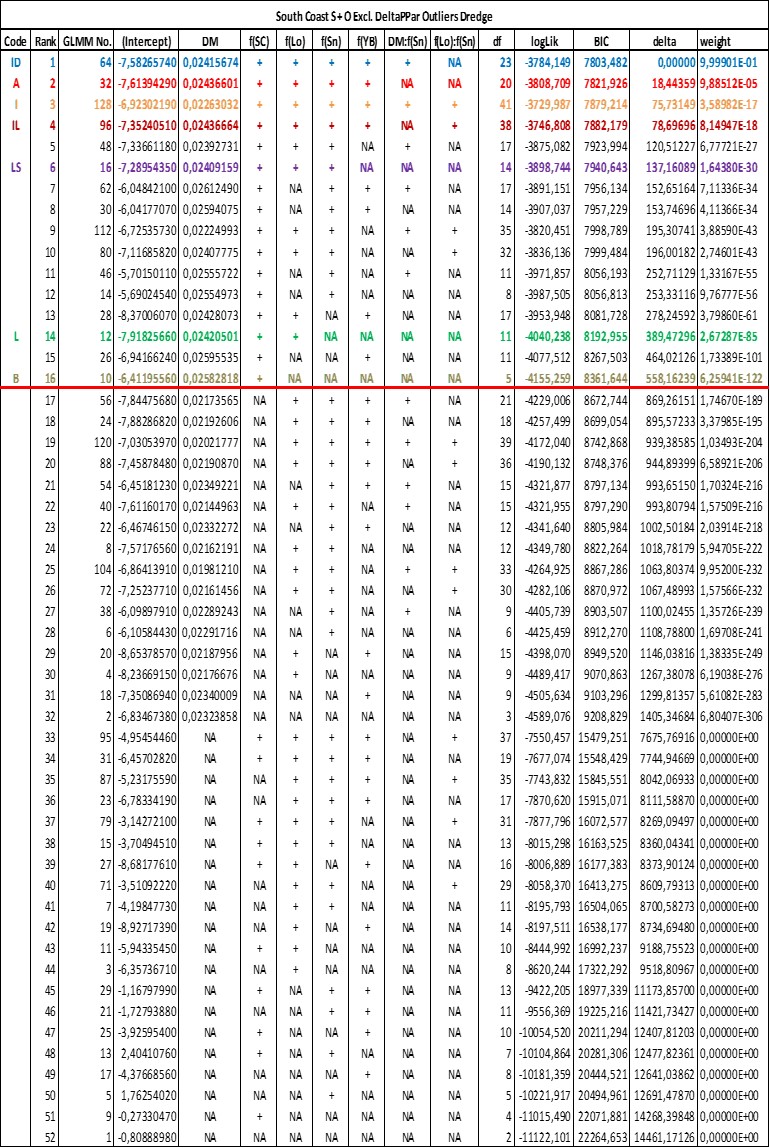


Table 4: Dredge Analysis summary of all the feasible and valid GLMM model combinations using only Survey records pertains to the South Coast.

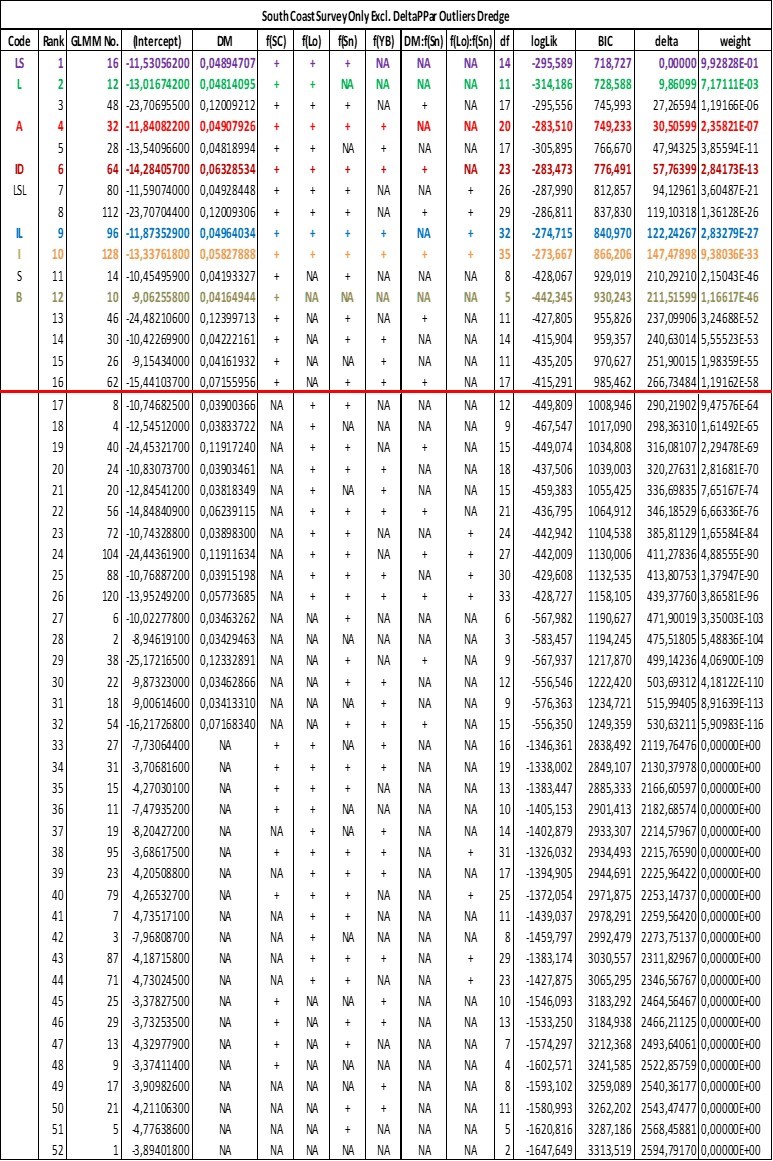


Table 5: Cross Validation summary of several selected GLMM models using both Survey and Observer records pertains to west or South Coast for the fitting process.

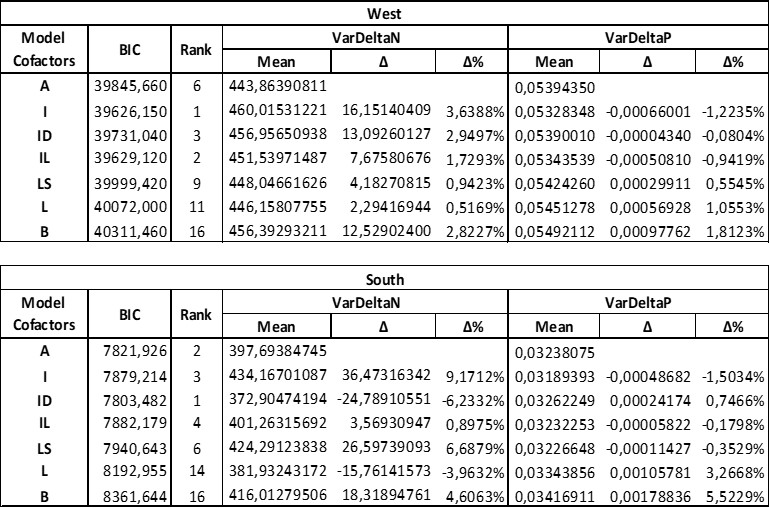


Table 6: Cross Validation summary of several selected GLMM models using only Survey records pertains to either west or South Coast for the fitting process.

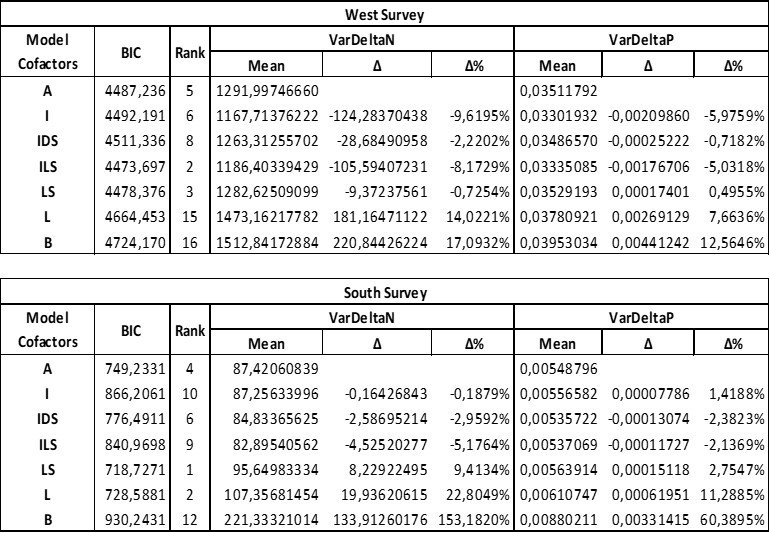


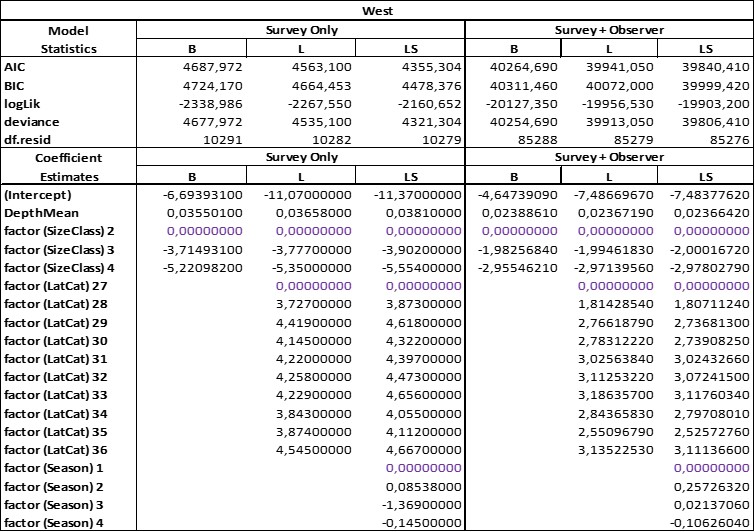
Table 7: Summary table of model statistics and fixed effect coefficient estimates for the B, L and LS GLMM Models using only survey records or survey and observer records that pertains to the West Coast during the model fitting process.

Table 8: Summary table of model statistics and fixed effect coefficient estimates for the B, L and LS GLMM Models using only survey records or survey and observer records that pertains to the South Coast during the model fitting process.

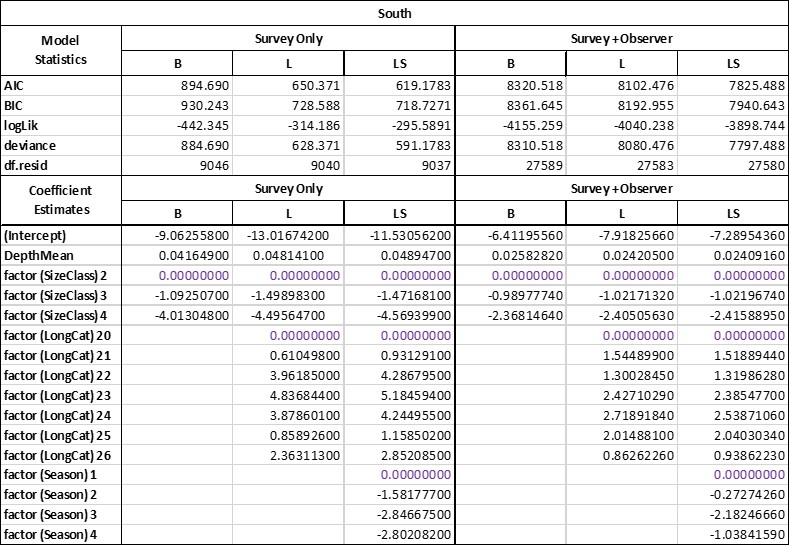


Table 9. Comparison of model parameter estimates obtained either without or with the inclusion of the Trawl ID as a random effect, for Model 6b of Glazer et al, 2018 (FISHERIES/2018/OCT/SWG-DEM57. The parameter values estimated for this model are the shaded values given in Table 2 of FISHERIES/2018/OCT/SWG-DEM57.



Table 10. Output from GLMM and GLM analyses, for Model 6b carried out either with or without the Trawl ID random effect.



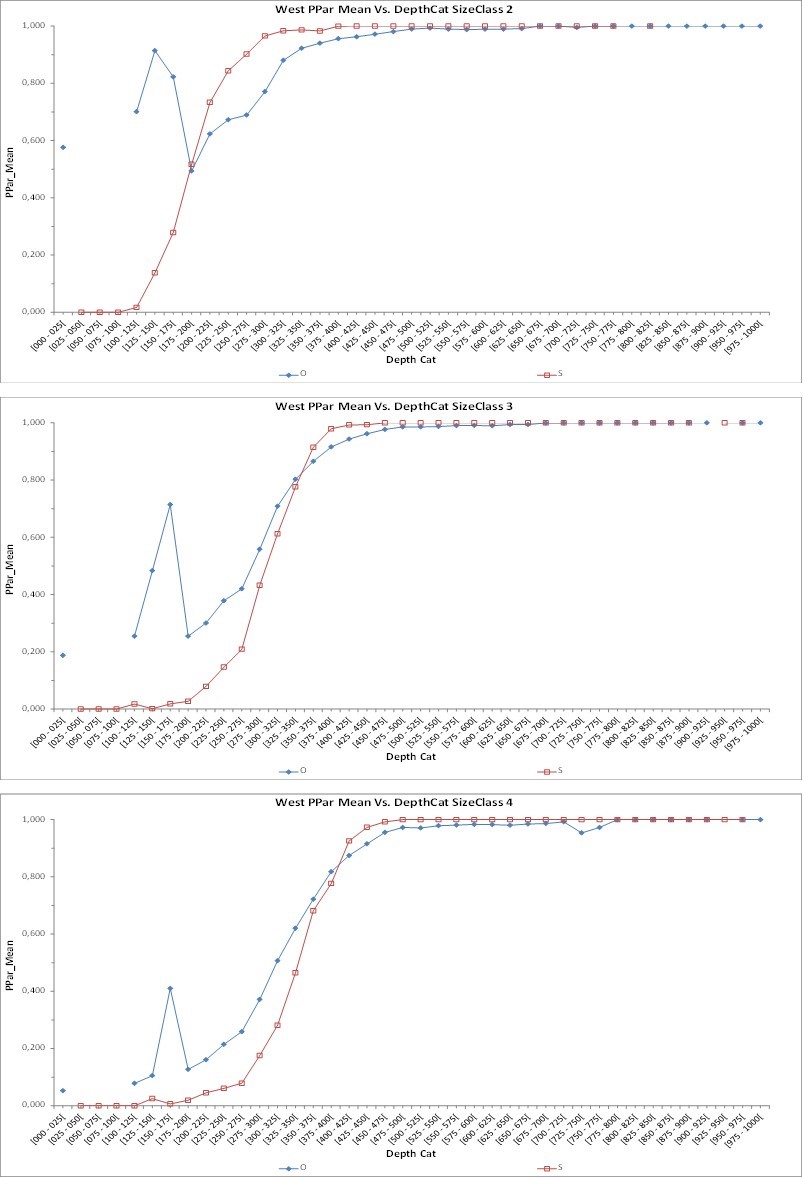


Figure 1: PPar\_Mean versus DepthCat line plots for West Coast survey (red) and observer records (blue) pertaining to the different size classes 2, 3 and 4. These plots are for average values within depth bins “DepthCat” as indicated.

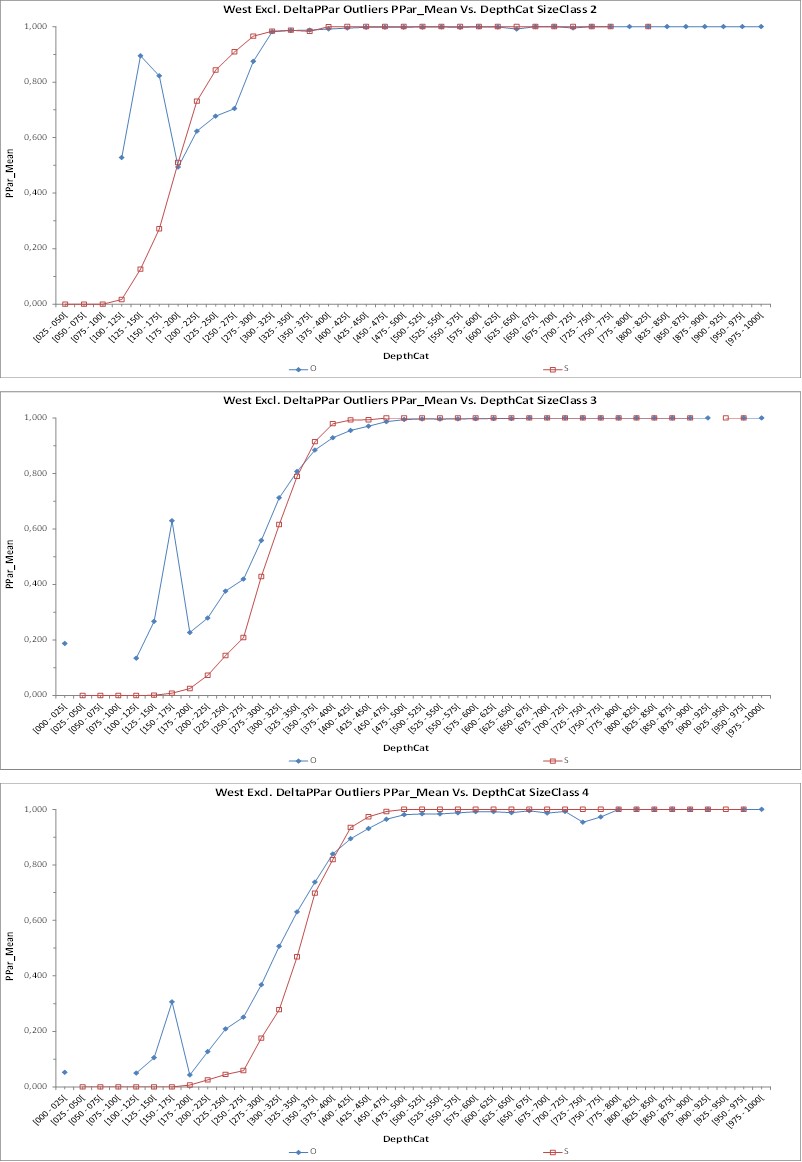


Figure 2: PPar\_Mean versus DepthCat line plots for West Coast survey (red) and observer records (blue) pertaining to the different size classes 2, 3 and 4 where the DeltaPPar outliers were excluded. These plots are for average values within depth bins “DepthCat” as indicated.

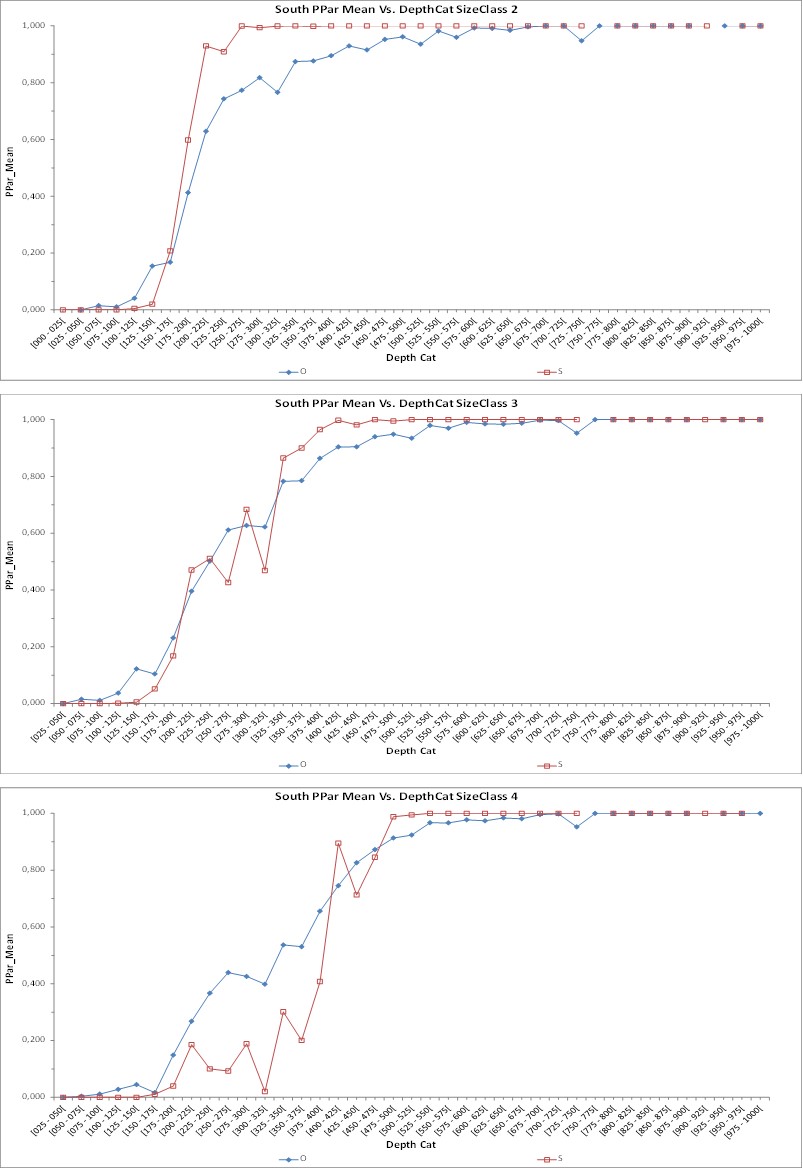


Figure 3: PPar\_Mean versus DepthCat line plots for South Coast survey (red) and observer records (blue) pertaining to the different size classes 2, 3 and 4. These plots are for average values within depth bins “DepthCat” as indicated.

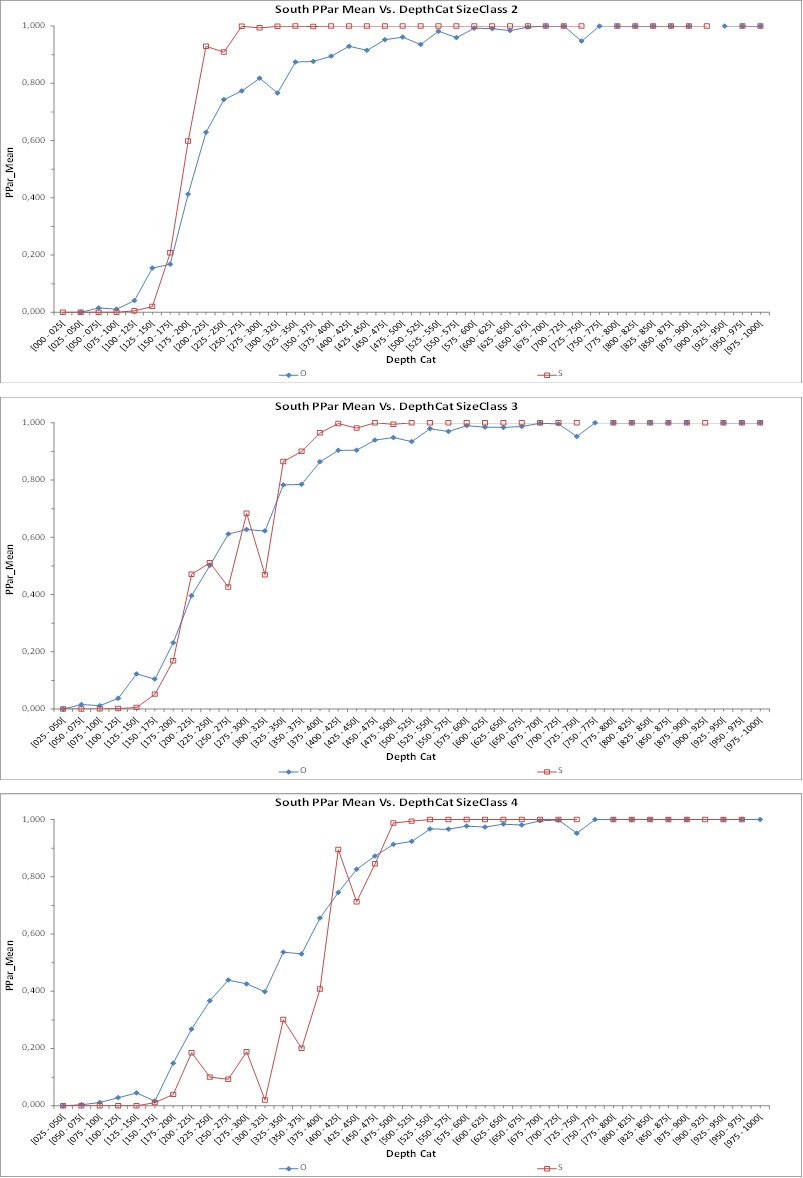


Figure 4: PPar\_Mean versus DepthCat line plots for South Coast survey (red) and observer records (blue) pertaining to the different size classes 2, 3 and 4 where the DeltaPPar outliers were excluded. These plots are for average values within depth bins “DepthCat” as indicated.

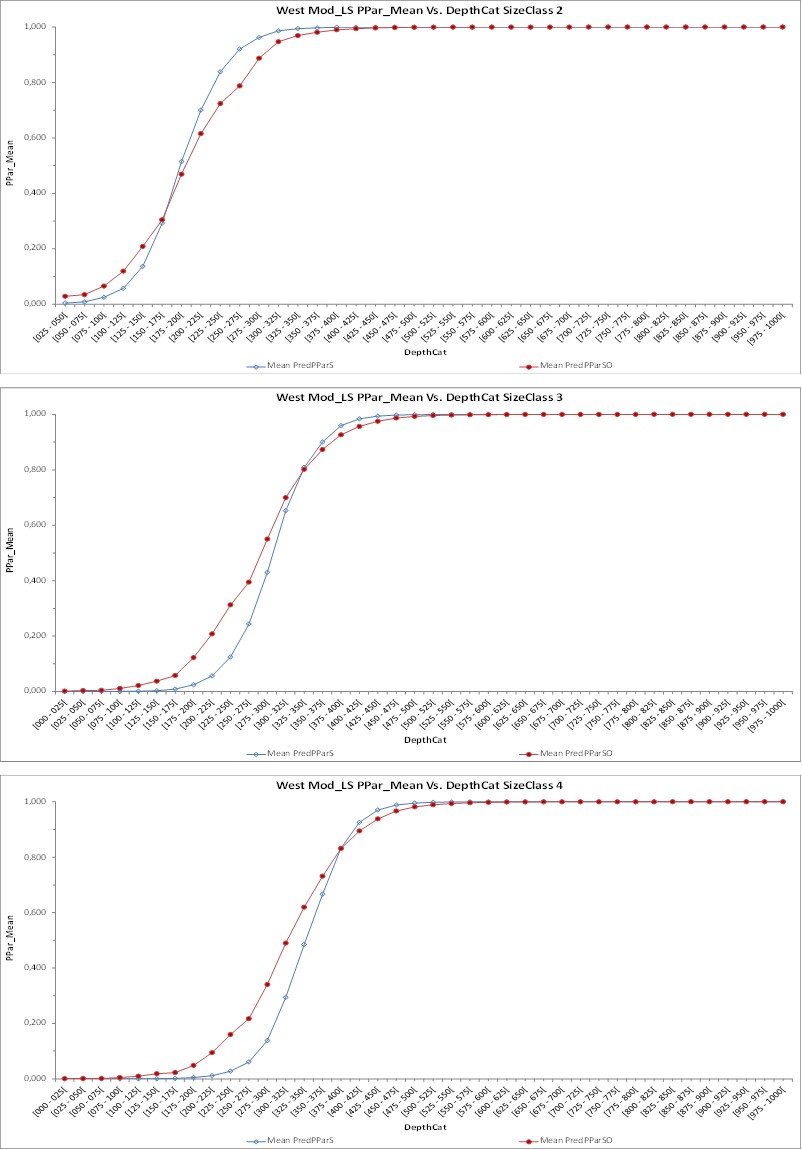


Figure 5: Predicted PPar\_Mean Versus DepthCat line plots for the West Coast survey and observer records based on a LS GLMM model which was either fitted initially with only survey records (Mean PredPParS) or survey and observer records (Mean\_PredPParSO) for the different size classes 2, 3 and 4. These plots are for average values within depth bins “DepthCat” as indicated.

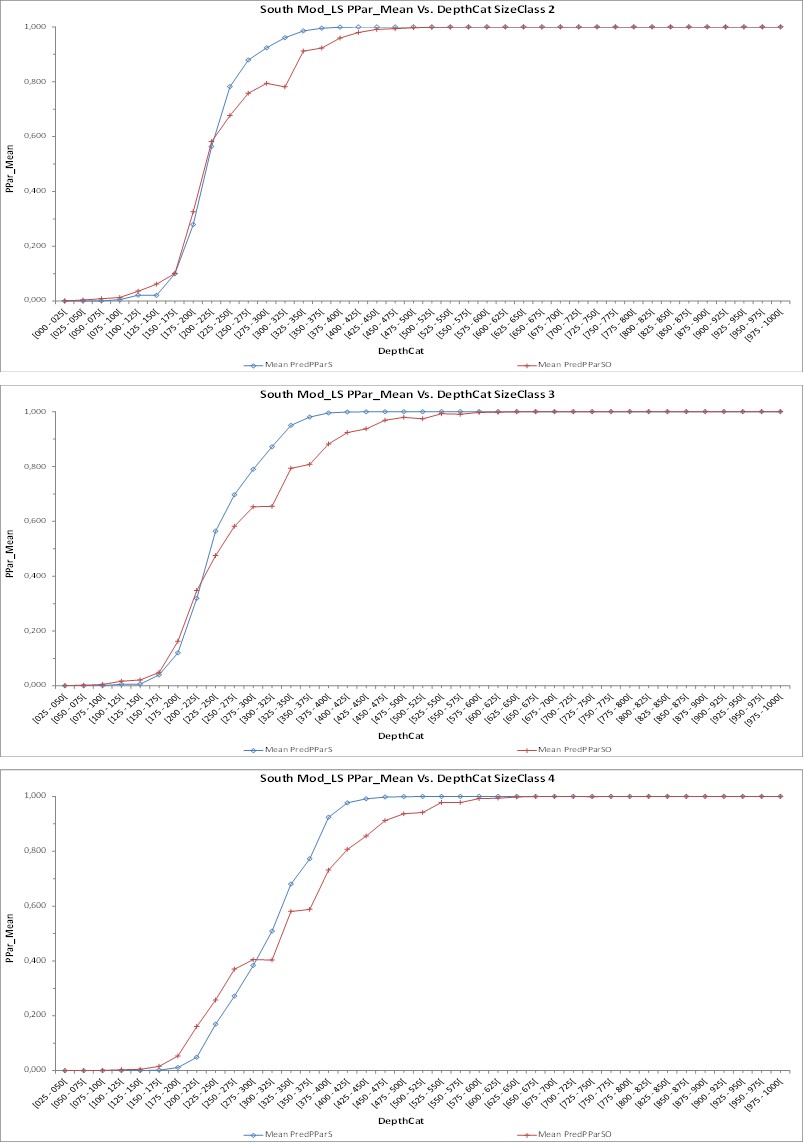
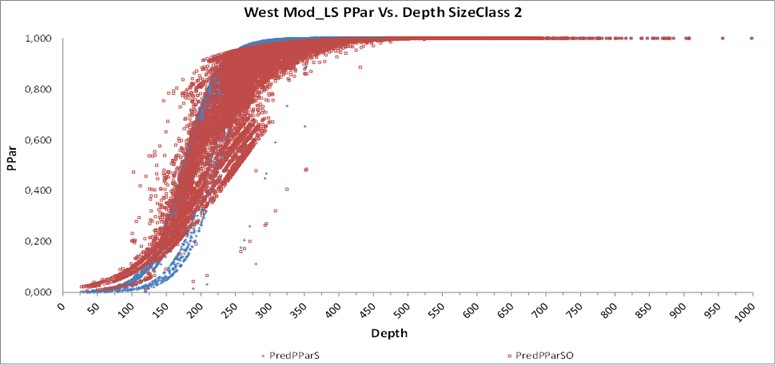
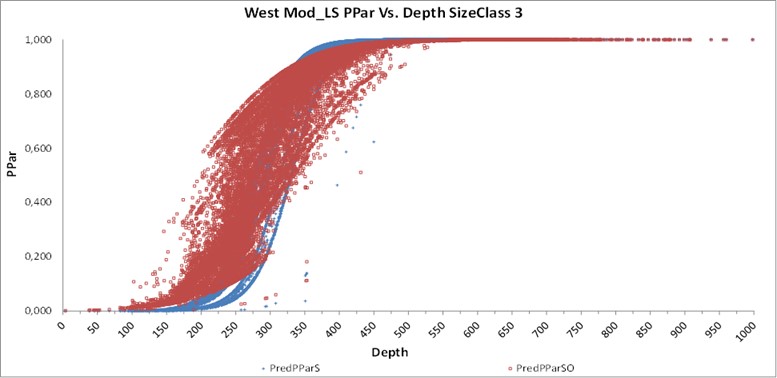


Figure 6: Predicted PPar\_Mean Versus DepthCat line plots for the South Coast survey and observer records based on a LS GLMM model which was either fitted initially with only survey records (Mean PredPParS) or survey and observer records (Mean\_PredPParSO) for the different size classes 2, 3 and 4. These plots are for average values within depth bins “DepthCat” as indicated.





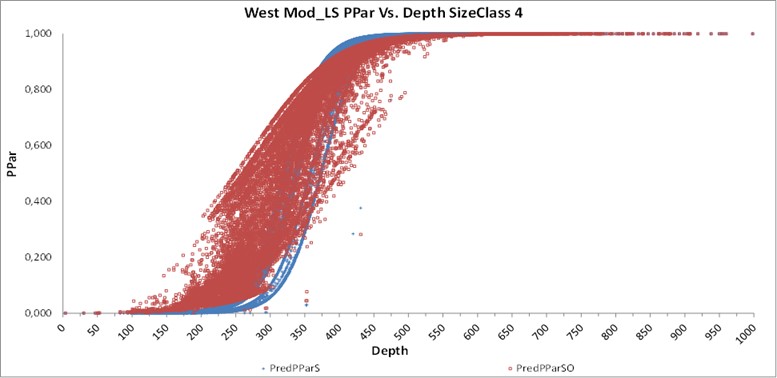
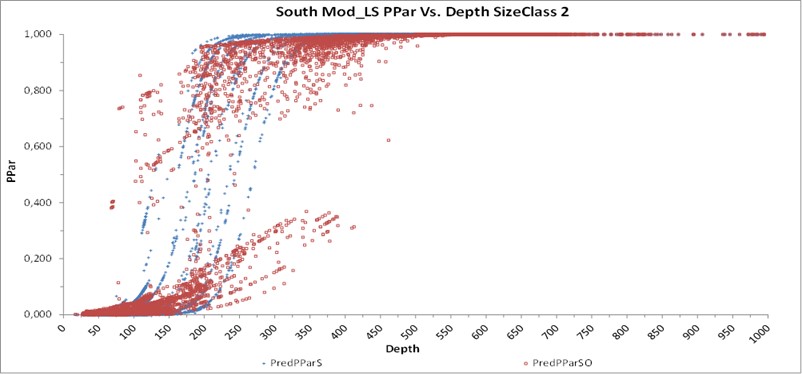
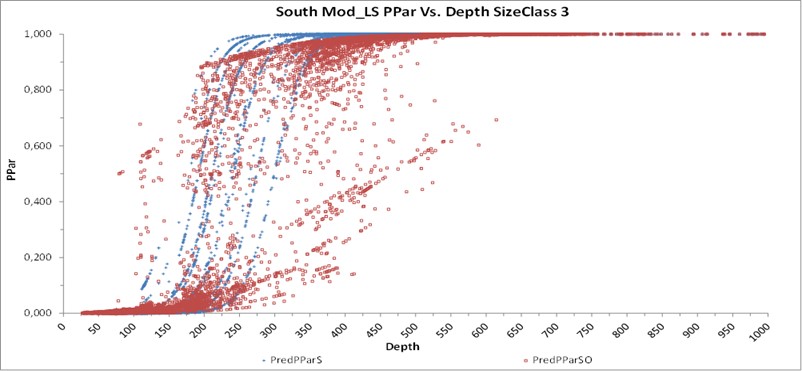


Figure 7: Predicted PPar Versus Depth dot plots for the West Coast survey and observer records based on a LS GLMM model which was either fitted initially with only survey records (Mean PredPParS) or survey and observer records (Mean\_PredPParSO) for the different size classes 2, 3 and 4.





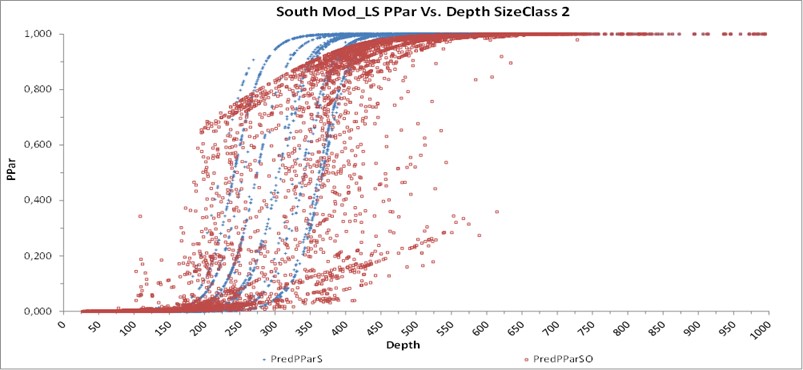


Figure 8: Predicted PPar Versus Depth dot plots for the South Coast survey and observer records based on a LS GLMM model which was either fitted initially with only survey records (Mean PredPParS) or survey and observer records (Mean\_PredPParSO) for the different size classes 2, 3 and 4

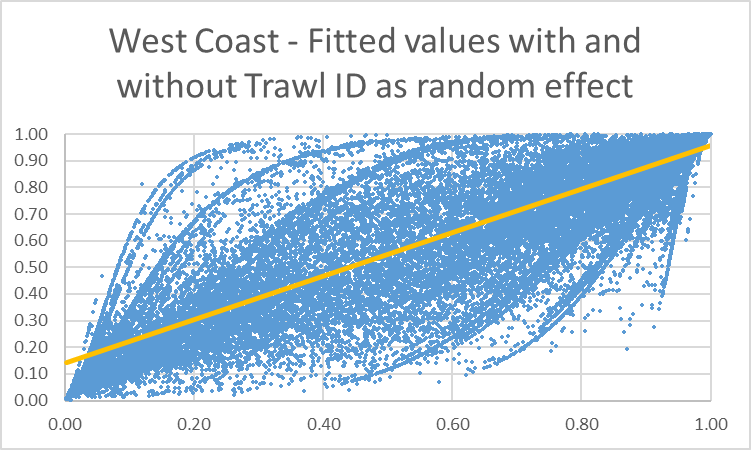


Figure 9. Plot of predicted proportion of M. paradoxus for random effect model (x-axis) versus GLM results without random effect (y-axis), for the West Coast.

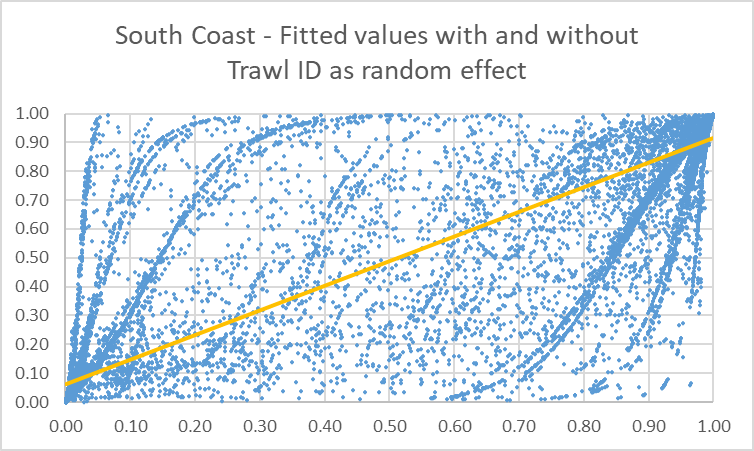
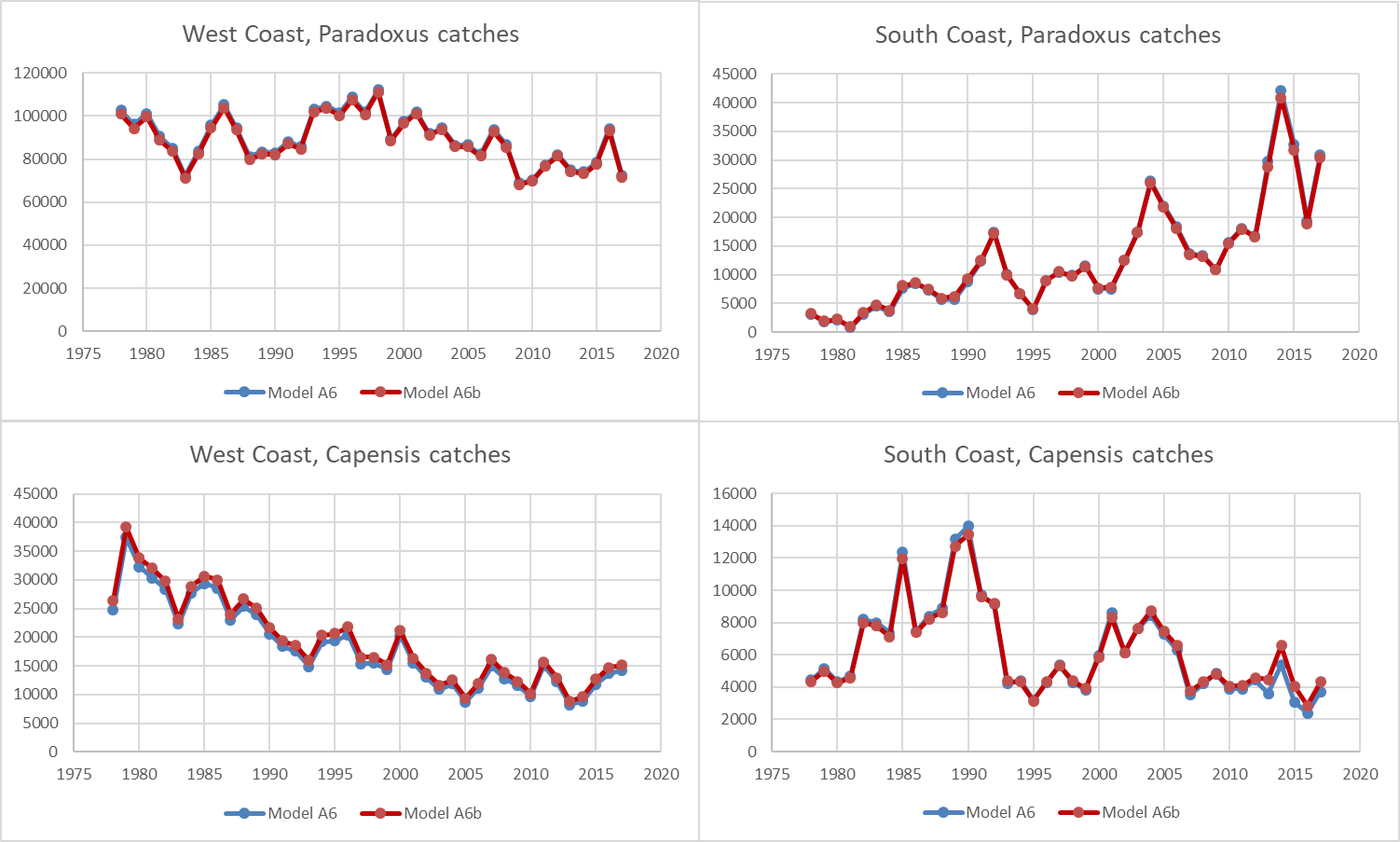


Figure 10. Plot of predicted proportion of M. paradoxus for random effect model (x-axis) versus GLM results without random effect (y-axis), for the South Coast.



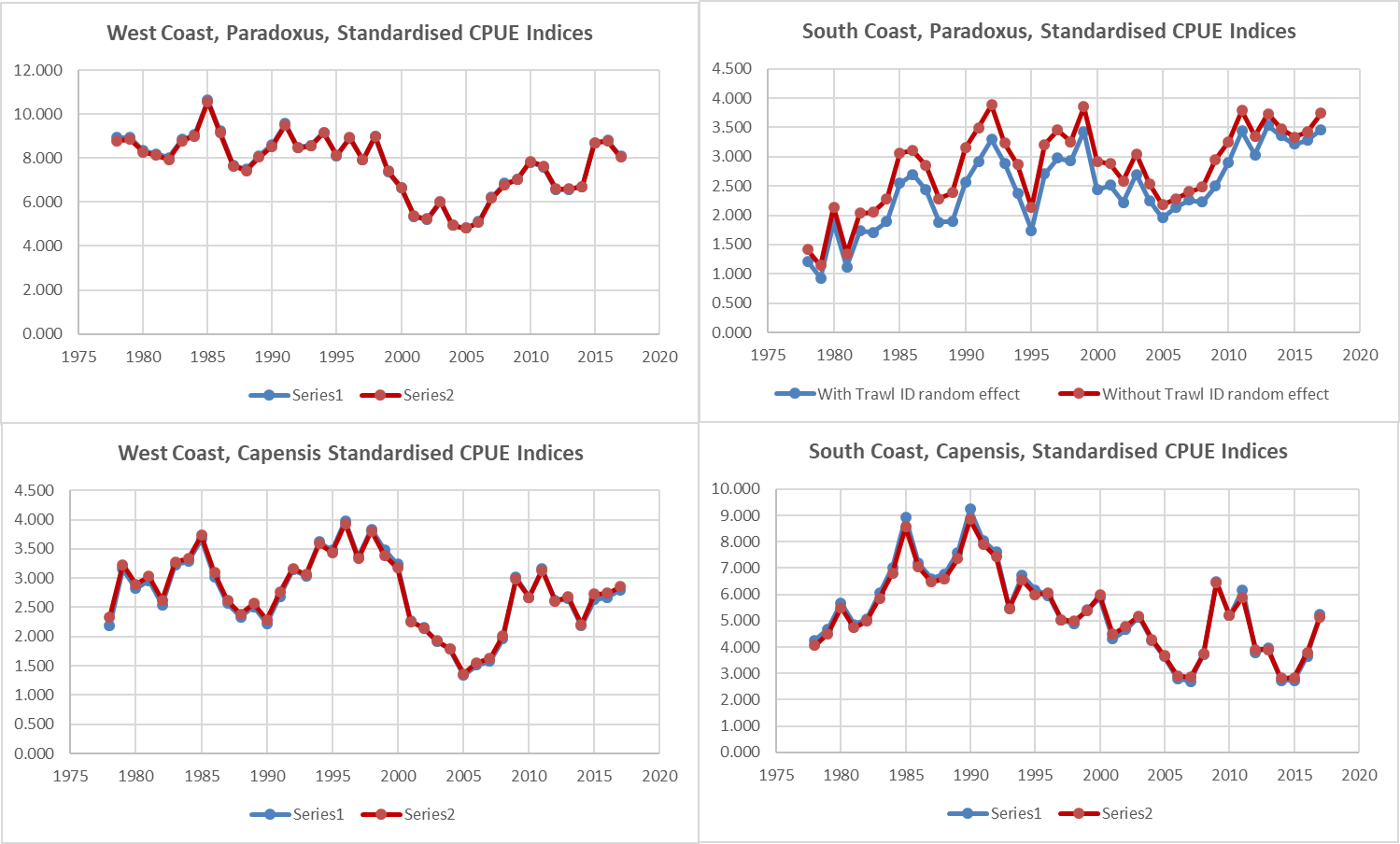


Figure 11. Upper four panels – Catches obtained with or without the use of Trawl ID in fitting the GLMM/GLM model for the proportion of M. paradoxus in catches. Lower panel – GLM standardised CPUE indices.

# Appendix A. The species splitting formula used prior to OMP-18

The existing model is based on a GLM with a binomial distribution and a logit link function. Model effects are additive in logit space, via an equation of the following form for the West Coast:



with 

where: P is the proportion of ***Merluccius paradoxus***;

 is the intercept;

 is the size class specific parameter*;*

 is the covariate parameter for depth*;*

The same model form is used for the South Coast. The model that was previously in use in the management of the South African hake resource is given below as “New model parameters”, noting that the table below (Table 11) includes also results for the following cases;

1. Old model parameters (given in FISHERIES/2013/FEB/SWG-DEM/12)
2. New model parameters (includes the correct stepped West – South Coast split)
3. The new model including block-year factors in 5 year blocks.

The results of the three models described above are given in Table 11.

Table 11. Model parameter estimates for the hake species split model – the “New” values are being used in hake management.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **West Coast** | | |  | **South Coast** | | |
|  | **Old** | **New** | **Block year effect** |  | **Old** | **New** | **Block year effect** |
| µ | -12.851 | -12.978 | -12.846 |  | -23.183 | -22.674 | -23.43 |
| λsmall | 5.788 | 5.928 | 6.209 |  | 10.997 | 8.8 | 8.888 |
| λmedium | 2.049 | 2.137 | 2.329 |  | 7.391 | 5.733 | 5.776 |
| λlarge | 0 | 0 | 0 |  | 0 | 0 | 0 |
| ϒ | 0.037 | 0.037 | 0.039 |  | 0.073 | 0.084 | 0.085 |
| 1983-1988 |  |  | -1.352 |  |  |  | -0.089 |
| 1989-1993 |  |  | -0.936 |  |  |  | 0.127 |
| 1994-1998 |  |  | -0.179 |  |  |  | 0.899 |
| 1999-2003 |  |  | -0.594 |  |  |  | 0.725 |
| 2004-2008 |  |  | -0.673 |  |  |  | 0.323 |
| 2009 + |  |  | 0 |  |  |  | 0 |
| **variance of residual proportions** | **0.047** | **0.05** | **0.048** |  | **0.02** | **0.018** | **0.017** |

# Appendix B. Species splitting model – final version adopted by the DSWG 2018.

A GLMM with a logit link function and a binomial distribution is applied. Both west and south coast data were modelled using the equation:

with

[This model is Model A6b of Glazer et al, 2018 (FISHERIES/2018/OCT/SWG-DEM57. The parameter values estimated for this model are the shaded values given in Table 2 of FISHERIES/2018/OCT/SWG-DEM57. ]

where :

is the proportion of ***M. paradoxus*** by mass for a given trawl

is the model intercept

is the mean depth of the trawl in metres, and is the associated parameter for the covariate

is a categorical variable, being the latitude bin on the West Coast and the longitude bin on the South Coast

is a categorical variable for small, medium or large size classes

is the interaction between size class and depth

is the interaction between size class and position.

The GLMM model was run without any record specific weighting. This means that the dependent value for each record is the observed mass proportion of **M. paradoxus**. The parameter value estimates are shown here inTable 12. These will not be updated during the implementation period of the OMP.

Table 12. Coast-specific parameter values for the species splitting model (see Table 2 of FISHERIES/2018/OCT/SWG-DEM57).

