CSIRO OCEANS AND ATMOSPHERE
www.csiro.au

Improve catch rate standardizations to account for changes in targeting

Malcolm Haddon [a], Rik Buckworth [b], Natalie Dowling [a], George Leigh [c], David C. Smith [a]

a: CSIRO Oceans and Atmosphere, GPO Box 1538, Hobart, TAS 7001, Australia.
b: CSIRO Oceans and Atmosphere, GPO Box 1538, Brisbane, QLD 7001, Australia.
c: QDPI, Brisbane, QLD 7001, Australia.

June 2018 FRDC Project: 2012/201

FRDC
FISHERIES RESEARCH & DEVELOPMENT CORPORATION

National Library of Australia Cataloguing-in-Publication entry

ISBN: 978-1-4863-0834-7

FRDC Report: Improve catch rate standardizations to account for changes in targeting. FRDC Project 2012/201. Fisheries Research Development Corporation and CSIRO Oceans and Atmosphere, Hobart. By Haddon, M., Buckworth, R., Dowling, N., Leigh, G., and D. C. Smith (2017) - ONLINE

Notes: Includes bibliographical references.

Subjects: Fish stock management.

Commercial Fisheries Harvest Strategies.

Author: Malcolm Haddon, Rik Buckworth, Natalie Dowling, George Leigh, David Smith

Title: Improving catch rate standardization for Australian fisheries.

FRDC Project: 2012/201

Year of Publication: 2018

CSIRO. Oceans and Atmosphere.

This publication (and any information sourced from it) should be attributed to:
Haddon, M., Buckworth, R., Dowling, N., Leigh, G. and D.C. Smith (2018) *Improving catch rate standardization account for changes in targeting.* CSIRO. FRDC Final Report 2012/201. Hobart. 162 p.

Researcher Contact Details

Name:	Malcolm Haddon,
Address:	CSIRO Oceans and Atmosphere, Castray Esplanade, Hobart, GPO Box 1538, Tasmania, 7001, Australia
Phone:	03 6232 5097
Fax:	
Email:	Malcolm.Haddon@csiro.au

FRDC Contact Details

Address:	25 Geils Court Deakin ACT 2600
Phone:	02 6285 0400
Fax:	02 6285 0499
Email:	frdc@frdc.com.au
Web:	www.frdc.com.au

In submitting this report, the researcher has agreed to FRDC publishing this material in its edited form.

Contents

1 Acknowledgments

We would like to thank Dr Beth Fulton for her heroic efforts to modify Atlantis to generate simulated shot-by-shot CPUE data. In addition, we would like to thank the members of what is now the South-East Resource Assessment Group, for their discussions of CPUE standardization. Special mention needs to be made of Dr Miriana Sporcic who worked with one of us (MH) on the standardization of the CPUE in the SESSF. Dr Jemery Day also requires special mention for his efforts to re-analyse the fishery independent survey data making the outputs more directly comparable with the CPUE standardizations used in assessments. Many of the ideas in this report stem from standardizing data especially for the SESSF but also for the Northern Prawn Fishery.

2 Executive Summary

2.1 What the report is about

In Australia many stock assessments are dependent upon catch-per-unit-effort (CPUE) to act as an index of relative abundance of fished stocks through time. But CPUE trends can be affected by many factors other than just stock size changes. Around Australia, and internationally, numerous and disparate approaches are used to conduct standardizations of CPUE using statistical methods to account for the effects of these other factors (e.g. the effect of which vessel is fishing, and where and when it is fishing). The objective in all cases is to discover trends in the CPUE that better reflect how the stock's relative abundance is changing through time rather than reflecting changes in the fisher's behaviour. In attempts to improve how such analyses are conducted and reported in Australia, stock assessment scientists from CSIRO and Queensland DPI explored an array of different aspects of CPUE standardizations. The overall aim was to generate a series of recommendations to act as a guide or a set of suggestions when it becomes necessary to use CPUE data in a stock assessment.

2.2 Background

Fishery stock assessments need some way of indexing the relative abundance of a fished stock through time if they are to succeed in summarizing the stock's dynamics and generate appropriate management advice. In Australia, with a few exceptions, the only index of relative abundance available for most fished stocks is fishery dependent CPUE. It has been common practice internationally to standardize CPUE trends with respect to factors unrelated to stock abundance to ensure that any trends observed are not due to those other factors (such as which vessel was fishing, in what area, at what depth, and at what time of year). Despite the process of CPUE standardization being almost obligatory although still not universally adopted, there is no single accepted method for conducting such analyses. This possibly reflects the fact that there are diverse methods of fishing for a diverse array of different species. Not surprisingly, this has also led to a diverse array of different types of fishery dependent data that varies in the detail, frequency of collection, and geographical scale over which it is summarized. It can vary from shot-by-shot information with the latitude and longitude of the start and finish to monthly summary data from 5° × 5° squares. Despite this diversity it is still possible to draw some general conclusions and make general recommendations regarding the uses of CPUE.

2.3 Aims/objectives

There were six objectives to this work:

1. Review the most appropriate catch rate standardisation strategies when targeting is well defined in multi-species fisheries.
2. Compare alternative catch rate standardization strategies in those fisheries where both fishery independent and fishery dependent data are available.
3. After modifying Atlantis SE, simulate shot-by-shot commercial catch rate data and use this in simulation tests for the most robust standardization strategies in mixed fisheries when targeting is unknown and management interventions influence catch rates.
4. Use simulated catch rate data to conduct MSE testing of the influence of potential biased data and standardization strategies on the outcome of stock assessments that rely on catch rate trends and targets. [MODIFIED: simulated catch rate data/trends were replaced with 'single species simulated data'].
5. Use simulated catch rate data to test the potential influence of effort creep (technical improvements in fishing power) on stock assessments. [MODIFIED: simulated catch rate data/trends were replaced with 'single species simulated data'].
6. Based on the results of objectives 1 to 5 write a reference manual on the application of the most robust CPUE standardization strategies for Australian fisheries.

Objective 3 and thus 4 and 5, had to be modified when it became clear that the simulated data possible from Atlantis remained at too coarse a level even after significant modifications to the Atlantis simulation framework. Instead, single species simulated data was developed and used. This still enabled discussions of the issues of bias and standardization strategies, as well as the potential effects of effort-creep (which is where technical changes to fishing improve efficiency but do so invisibly to subsequent analyses).

2.4 Methodology

Separate sections were written dealing with a) potential issues that arise with the CPUE data itself, b) fishery dependent CPUE and fishery independent survey estimates of CPUE, c) alternative methods of statistical standardization, and d) the simulation of shot-by-shot data. In each case methods are described relating to data selection, data analysis, and other numerical methods used to solve the problems being addressed in each section.

2.5 Recommendations

2.5.1 Guidelines

The final objective of this work was to write a reference manual on the application of the most robust CPUE standardization strategies for Australian fisheries. It should be clear that the range of fisheries in Australia (from benthic hand collected fisheries, to trawl fisheries, to pelagic purse-seine and lining fisheries) means that there is no single standard approach to CPUE standardization that will necessarily work well with every fishery. Nevertheless, it remains possible to write out a set of guidelines that will improve the defensibility of any conclusions drawn from CPUE standardizations as well as improve the presentation of results from such analyses to assessment groups and other interested parties. Many of the points made below are included in Chapter 8 (starting on page 145).

2.5.2 Documentation

One pillar of defensibility is complete and explicit documentation of all procedures used in any stock assessment or analysis so that the analysis can be repeated quickly and easily. However, most people interested in the outcome of an analysis focus primarily on the summary or abstract of results and only desire a brief document. Nevertheless, in the interests of openness and defensibility, many of the suggested recommendations (below) for more text, tables, and plots, should be included at least as supplementary materials in appendices. With the growth of electronic documents and reduction in the use of printed documents the size of the final document should not be an impediment to an improvement to how such analyses are presented. If a printed document is required, then the supplementary material need not be printed but should be referred to throughout the primary document.

In the case of CPUE standardization it is necessary to:

1. Have an explicit section in any report on a standardization that focusses solely on the data selection and preparation processes and choices.
2. Describe and explain every choice in any data selection made.
3. Ideally tabulate and plot the distributions of catch, effort, CPUE, depth of fishing, month of fishing, and any other factors/variables included in the analysis to illustrate the quality of the data being used (helps identify whether there are outliers or there is rounding, or whether the data has unexpected properties, or just what those properties are).
4. Be explicit about the statistical models fitted, and how the model parameters (especially the year, or time-step, effects) are derived.
5. Be explicit about the assumptions behind the statistical distributions used in the statistical models.
6. Plot diagnostics relating to the statistical fit of the model to the data.
7. Identify and plot the relative influence of the different factors included in any analysis. Do not rely solely on the variance or deviance accounted for by each factor but also summarize the impact each factor has on the standardized CPUE trend.

2.5.3 Ideal Sensitivity Options

If enough time is available (albeit this is an unlikely scenario):

1. Apply the same statistical model structures but with different underlying statistical distributions to describe the residual structure (e.g. log-normal vs Gamma distributions). This tests for sensitivity to the basic assumptions used.
2. Apply different statistical model structures using the same statistical distribution for the residual errors structure to consider the sensitivity to model structure.
3. Conduct a retrospective analysis through at least the last half of the available years of data to search for consistency and/or for major changes in the factors influencing any trends in CPUE.

2.6 Keywords

Southern and Eastern Scalefish and Shark Fishery, SESSF, catch-rates, CPUE, standardization, fisheries data, catch and effort data, index of relative abundance.

3 Introduction

3.1 Why Use CPUE?

Natural resource management is a cyclical, heuristic, and adaptive process involving the monitoring, assessment, and management of an exploited resource (**Figure 1**). The primary problem in the management of natural fisheries resources is the provision of workable and relatively consistent advice about the state of exploited stocks (Punt *et al.*, 2018), which can then be used as the basis for on-going defensible management actions into the future.

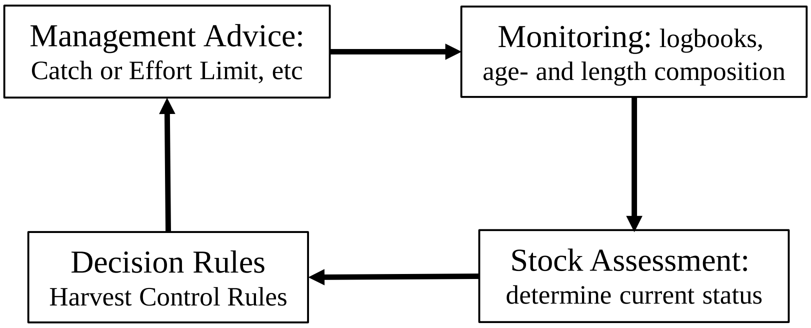

Figure 1. The cyclic process of monitoring, assessment, and management for fished natural resources, which may operate on an annual or longer time scale. The Decisions Rules constitute any form of translation of the stock assessment outputs into management advice. They can be a formal harvest control rule or some other process; they need to lead to repeatable outcomes.

The natural resource management cycle should be reviewed regularly and if a change in any of the component processes is recognized as being required because of changed circumstances within the fishery concerned this would amount to adaptive management.

Fishery stock assessments perform best when they include a valid and informative index of relative abundance through time; these are used to locate the stock dynamics on a real scale of abundance through time. In this way, the current state of depletion relative to hypothetical unfished levels, and many other statistics of management interest in Australia, especially those identified within the Commonwealth Harvest Strategy Policy (HSP), can be determined (DAFF, 2007; Rayns, 2007; Smith et al., 2008; Smith et al., 2014). Australian Commonwealth fisheries generally rely on commercial catch rates (CPUE) as an index of relative abundance and often CPUE remains the main index of relative abundance even after Fishery Independent Surveys have been initiated (for example in the Northern Prawn fishery). Such fishery independent surveys are only really useful for assessment purposes after a sufficiently long time series to exhibit trends has been developed and, importantly, if their inter-annual variation is small enough to allow any trends through time to be clearly identified; for example, a survey optimally designed for a particular target species may not provide a reliable index for a bycatch species taken in the same fishery.

The use of commercial CPUE data is always dependent upon several fundamental assumptions. Chief among these assumptions is that there is some relatively direct relationship between stock size and related CPUE (**Figure 2**) and that this relationship remains the same through time (an assumption of stationarity). Generally, the relationship is assumed to be linear, which means as stock size goes up or down, so does the catch-rate in a directly proportional manner; although it is generally acknowledged that such assumptions are only approximations. The assumption of stationarity is a major assumption which is broken by such things as technological improvements to fishing gear.

This is well recognized in the term “effort creep”, which is short-hand for the influence that improved fishing gear or methods can have on the effectiveness of a single unit of effort. Such ‘effort creep;’ effectively breaks the assumption of stationarity and this can lead to significant and misleading bias in any consideration of CPUE as an index of abundance. This problem will be considered in more detail later.

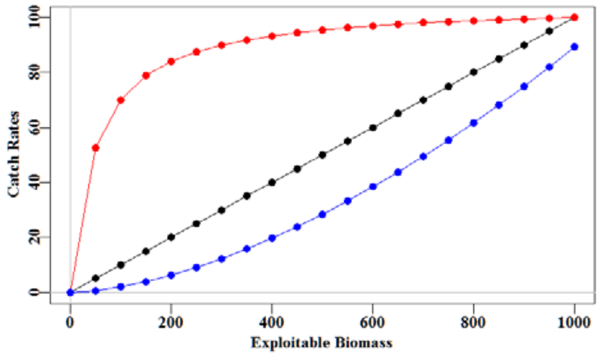

Figure 2. Idealized and deterministic representation of the relationship between exploitable biomass and CPUE. The general assumption is that the linear black line is what happens in nature whereas the other possibilities may be the case, which would distort and bias the interpretation of observed CPUE. It is also possible that with real fisheries the actual relationship might not be static nor a nice smooth curve such as those illustrated.

In addition to the assumptions of linearity and stationarity, there are many factors, other than changes in stock size, which can affect catch rates. For example, if catch rates are typically higher in the winter months but prices change to become better in the summer months there could be a shift of effort into the summer months to maximize profit-per-unit-effort rather than CPUE. Such changes in fishing behaviour would lead to a drop being perceived in the annual nominal CPUE that was actually due to the changed fleet behaviour and not to do with the relative abundance of the stock. In an effort to avoid some of these potentially spurious influences on any average trend, CPUE data are generally standardized statistically (Kimura, 1981). However, the standardizations can obviously only account for those factors for which there is data readily available for inclusion in the analyses.

3.1.1 Development of CPUE Standardizations in the SESSF

The Commonwealth Southern and Eastern Scalefish and Shark Fishery (SESSF) provides an example of the development of the use of CPUE data in Australia. The current fishery derives originally from the South East Trawl (SET) fishery, which became the South East Fishery (SEF) in the mid-1980s. Many changes to the management of the SEF began to be introduced from 1985 onwards. As stated in Tilzey and Klaer (1994, p1):

> *Before 1985, the fishery was virtually unregulated and trawl fleet capacity expanded rapidly during the 1970s and early 1980s. Input management controls based on limiting entry into the fishery were introduced in 1985. A boat replacement (unitisation) policy was introduced in 1986 to prevent further expansion of fleet capacity. However, the subsequent development of the orange roughy fishery and a failure to prevent entry to the SEF resulted in additional increases in fleet capacity and fishing effort.*

An important innovation towards the end of 1985 was the introduction of a new log-book which required trawl and Danish seine fishers to complete a shot-by-shot log-book rather than summary information by trip or even by month as continued to be collected by each State. This meant that the quality and resolution of catch and effort data began to improve greatly in the Commonwealth, although it took a few years before the data in that database became less variable between years than seemed biologically plausible (e.g. **Figure 3**). Such large changes through short periods of time imply there are large process errors in the data as well as the more common measurement errors. These large variations mostly settle down after the introduction of individual transferable quotas from 1992 onwards.

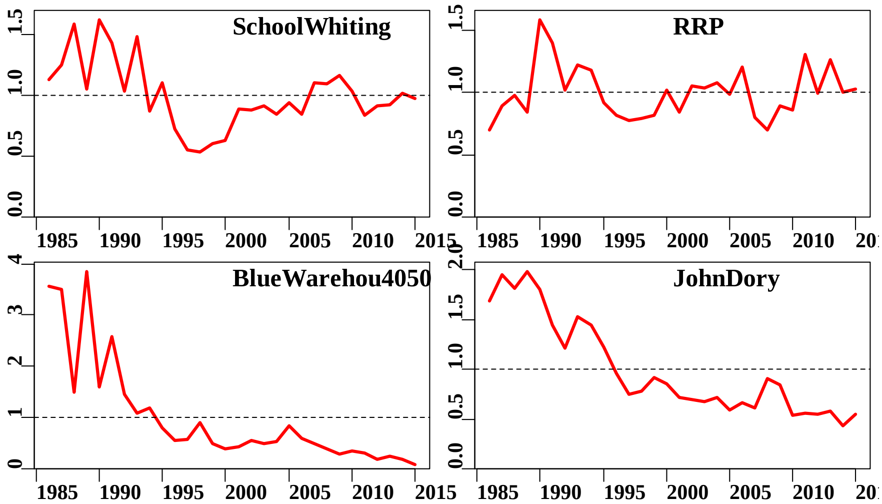

Figure 3. The optimum standardizations (on y-axis) for four fisheries from Sporcic and Haddon (2016, p10) illustrating the greater inter-annual variation in the years immediately following the introduction of the new log-book in Oct 1985. RRP is Royal Red Prawn. Obviously, the x-axis relates to years.

The advent of the high yielding short-lived fishery for orange roughy had long lasting effects upon the SEF. The first large catches off western Tasmania only really began in 1986 and the whole fishery rapidly expanded to a maximum in 1990 after which catch limits by orange roughy zone began to be introduced (**Table 1**).

In 1990, the on-going expansion in fleet capacity and a decline in some major fish stocks (especially Eastern Gemfish, but also Redfish; see Appendix 2 for species names) led to the development of a new fisheries management plan that introduced output management controls on 1 January 1992 in the form of Total Allowable Catches (TACs) and individual transferable quotas (ITQs). These were set for 16 major SEF species groups: Blue-Eye Trevalla, Blue Grenadier, Blue Warehou, Gemfish, Jackass Morwong, John Dory, Mirror Dory, Ocean Perch, Orange Roughy, Pink Ling, Redfish, Royal Red Prawn, Silver Trevally, Spotted Warehou, and Tiger Flathead. Such a major shift in management meant there was a greatly increased need for stock assessment advice although information sufficient for the generation of management advice suitable for output-controlled fisheries was limited to only five of the 16 stocks in 1992.

The use of CPUE data could only begin with the development of a time-series of sufficient length to permit a useful analysis of trends through time. The early data from the catch and effort database was often more variable between years than biologically plausible if the assumption of a relationship between CPUE and stock size was valid (**Figure 3**). The catch rate estimates were initially simple ratio estimates (total catch divided by total effort), then a range of alternative statistical models were examined to determine whether they provided a consistent measure of the catch-rate trends for a few limited species for which more complex stock assessments were being developed. Eventually the need to standardize the CPUE of more and more species developed and a more generally agreed approach based around general linear models (e.g. Haddon, 1998) was settled upon.

Table 1. Orange Roughy catches (t), as reported in the AFMA catch and effort logbook database. Records only start in August 1985. A 700m deepwater closure was introduced in 2007 and Orange Roughy was declared a conservation dependent species effectively shutting the Orange Roughy fishery. After demonstrating a stock rebuild the Eastern zone was re-opened, in a limited manner, in 2015.

Year	Eastern	Southwest	Southeast	Western	Cascade	NERemote	GAB	South Tas Rise
1985	5.5	57.5	0.9	128.6	0.3			
1986	32.8	604.1	26.7	3924.9		0.3		
1987	310.3	320.8	31.8	5118.0	1.8	10.4	406.2	
1988	1948.4	468.9		4722.2		2.7	2820.2	1.7
1989	18345.2	4993.7	2626.0	1365.1	258.5	1.5	3793.2	1.0
1990	16198.5	14898.7	9897.7	801.6	1822.3	215.7	1056.5	35.9
1991	9727.3	3496.3	8025.1	625.4	39.5	437.7	423.1	
1992	7622.7	2412.8	5241.6	1108.2	468.5	131.9	741.8	
1993	1793.8	2484.3	4758.4	964.4	91.8	42.0	647.3	
1994	1481.2	2165.1	2307.8	800.6	478.5	128.4	82.4	
1995	1817.0	1430.5	613.5	962.4	78.3	7.9	345.1	
1996	1818.6	503.1	278.4	1180.3	868.5	54.8	359.1	3.6
1997	1909.8	217.6	232.5	297.0	1092.6	22.4	332.0	1460.3
1998	1858.0	80.5	215.1	316.1	1448.4	33.0	647.9	2878.4
1999	1892.7	69.9	95.0	210.5	1534.9	29.4	819.7	1834.1
2000	1900.0	156.5	130.7	169.3	1536.5	15.3	349.3	791.4
2001	1783.9	142.2	198.9	200.8	1363.0	14.9	374.6	169.4
2002	1521.5	67.2	90.5	255.7	1462.5	38.7	217.6	102.3
2003	747.9	94.2	114.9	217.5	1563.6	66.5	226.4	11.3
2004	719.7	42.1	97.1	283.1	1444.6	40.0	150.1	48.5
2005	713.8	55.9	37.6	264.6	1262.5	16.7	117.1	12.0
2006	577.4	4.3	1.2	139.3	701.7	20.5	215.2	0.2
2007	116.1	4.9	16.9	28.6	204.0	1.7	44.4	

The early assessments in the SEF mostly revolved around a consideration of such CPUE data: “Most current stock assessments in the SEF rely primarily on analysis of catch and effort data (including information on discarded catch from the ISMP) combined with some information on age and length composition of the catch and limited biological information” (Tilzey, 1999, p34). With some exceptions (Orange Roughy and Eastern Gemfish, for which more advanced models had been developed) performance criteria for each fishery were limited:

> *AFMA has set performance criteria based on, among other things, trends in catch per unit effort (CPUE). The catch rate criterion seeks to maintain CPUE above its lowest annual average level from 1986 – 1994. In using this criterion, AFMA recognized that there were a number of factors other than stock abundance that could affect catch rates. …. The AFMA performance criteria do not specify how catch*

per unit effort is to be determined. SEFAG has attempted to standardize catch rates, but no satisfactory method has so far been developed. (Tilzey, 1999, p 34 – 35)

The option of using statistical standardizations in the SESSF was recognized early on (Klaer, 1994) but mostly ratio mean catch rates were used for a number of years after the introduction of quotas in 1992. The introduction of using geometric mean CPUE based on shot-by-shot records rather than ratio means was a first step in the improved statistical treatment of CPUE data. Even in the 1998 Fishery Assessment Report (Tilzey 1999) standardized catch rates were only used for a limited number of fisheries (Haddon 1998; Haddon, 1998b; Haddon, 1999; Haddon and Hodgson, 2000), and then usually to complement the development of age-structured stock assessment models (Punt, 1998; Punt et al, 2001). At that time there were Fishery Assessment Groups for just a few individual species, which eventually became amalgamated into the more general South East Fishery Assessment Group. Up until 2006, standardizations with separate reports were conducted for individual species but as the number of species with more formal stock assessments increased so did the number of CPUE standardizations. Haddon (2007) was the first report which combined eight species across different combinations of zones and fisheries to lead to a total of 14 standardizations selected from 127 statistical models. The species included were: Blue-Eye Trevalla, Blue Grenadier, Blue Warehou, Tiger Flathead, Jackass Morwong, Redfish, Silver Trevally, and Spotted Warehou. Since then the number of species, and stocks has greatly increased (Sporcic and Haddon, 2016). The latest analyses included 23 species spread across 43 different combinations of stocks and fisheries, not including the commercial shark species and some other particular analyses (e.g. Haddon, 2016b). The number of different statistical models is now considerable.

3.1.2 Recent Management and Other Changes

With the advent of the Commonwealth Harvest Strategy Policy in 2007, with its associated structural adjustment or buyback scheme occurring between Nov. 2005 and Nov. 2006, the character of various Commonwealth fisheries has altered remarkably in a number of different ways. On top of these management changes there is also the potential for changes in fishery dynamics due to climate change impacts on such things as sea temperatures and the geographical distribution of species, and the average productivity of species (Pecl et al, 2017). Such changes would have their respective impacts on reported commercial CPUE data. Thus, there is the potential that stock assessments that use these CPUE data as an index of relative abundance will become compromised if these changes have altered the character of the CPUE and such changes are not taken into account.

Other major changes have been seen in the Southern and Eastern Scalefish and Shark Fishery (SESSF) and the Northern Prawn Fishery (NPF), for example, which have both seen a remarkable reduction in the number of active vessels brought about by the structural adjustment in 2006/07. Similarly, the re-organisation of the Queensland State fleet and demersal fisheries has seen large changes in fisher behaviour and the structure of the fleet. In the multi-species, multi-gear SESSF the buyout reduced the trawl fleet by 40% and non-trawl vessels by 16%; although particular fisheries for individual species within the SESSF often saw greater reductions in vessels reporting the capture of those species (Vieira et al., 2010). The structural adjustment was able to increase the combined profitability of the remaining vessels because the available quota was distributed among fewer vessels (Vieira et al., 2010). In addition to changes in profitability the reduction in the various fleets led to relatively large changes in fishing behaviour. These changes were also influenced by some of the most important species caught in the SESSF achieving their target reference point. Many fishers are now claiming that they have now started to avoid catching such economic driver species as Tiger Flathead (*Neoplatycephalus richardsoni*). Flathead are now reported as being relatively simple to catch and fishers also report that they are avoiding catching Flathead as fast as they could if they tried (in other words they avoid them) so as to enable them to catch the quota they hold for other species as well as their Flathead quota. If such changes in targeting behaviour were in fact happening right across the remaining fleet it would clearly bias catch rates downwards. However, without explicit information regarding the changed targeting behaviour, such reductions in catch rate would be interpreted as a reduction in stock size. In fact, in species with fully quantitative assessments (Tier 1) such data would presumably become inconsistent with other data relating to age-structure and length-structure of the fished stocks. So such changes in the quality of the commercial CPUE data can either bias subsequent stock assessments or lead to them becoming less certain due to conflicts among the different data streams. Because of the reliance of Commonwealth stock assessments on commercial catch rate data for an index of relative abundance there is an urgent need to understand the impacts on CPUE of all of the changes imposed on Commonwealth fisheries with the advent of the HSP and structural adjustment. If the time series of CPUE have been disrupted this needs to be demonstrated so that appropriate actions can be implemented in the annual assessments. It is not the case that the catch rates of all species will be affected to the same amount. The research would need to identify those species for which changes to their stock assessment would be required and those for which little or no change was needed.

On top of these management influences there is also the potential for alterations in fishery dynamics due to climate change impacts on such things as sea temperatures and the subsequent geographical distribution of species. These changes would be affecting the assumption of stationarity by altering the relationship between the stock size in a given area and its catch rate. The potential importance of this issue should not be under-estimated as non-stationarity in growth rates has already been demonstrated in a number of species and the assessment of jackass morwong in the SESSF has already been changed to reflect a switch to a less productive state by the east coast stock (Wayte, 2010). As mentioned previously, changes in the technological aids used when catching fish has also had an impact on the assumption of stationarity. ‘Effort creep’, brought about by such things as the advent of GPS, GPS plotters, and colour bottom lock depth sounders has undoubtedly improved the efficiency of fishing vessels. Unfortunately, insufficient information was collected at the time of adopting such technology that accounting for such changes in relative fishing power is difficult or impossible.

3.2 Indices of Relative Abundance

Formal stock assessments require some form of index of relative abundance in order for them to track dynamic changes in the population size of harvested fish populations. Absolute abundance indices are possible (possibly from tagging studies or egg production studies) but these can only be considered absolute estimates if relatively stringent assumptions and conditions are met; invariably great uncertainty remains.

Stock assessments that relate to stock biomass need an index of relative abundance. The more complicated and inclusive ‘Integrated Assessments’ (Maunder and Punt, 2013) can include indices of relative abundance and catches by different fishing methods, age- and length-composition data from different sources, tagging data, and whatever else is available. When there are multiple data streams in such models the question arises about what relative weight to ascribe to each data series (Richards, 1991; Francis, 2011; Punt, 2017). Francis (2011) suggested a guideline which states: "do not let other data stop the model from fitting abundance data well" (page 1124), and this has become a strong influence on stock assessments since. As such, gaining an understanding of whatever index of relative abundance is in use takes on a greater importance.

Two common indices of relative abundance are time series of commercial CPUE data (as used in the stock assessments of Flathead, Pink Ling, etc) and time series of fishery independent survey abundance indices (as used in the eastern Orange Roughy stock assessment (Haddon, 2017).

3.2.1 Fishery Independent Survey Abundance Indices

Fishery independent surveys, whether they be swept area trawl surveys, acoustic surveys, or even standard long-line sets, are all considered to be able to provide the best view of a stock's size available. Of course, with any fishing, and especially where a survey needs to be run within specific dates across specific areas, there are no guarantees that any particular survey will provide usefully accurate estimates of the stock biomass of a species, especially in mixed fisheries. Just as with commercial CPUE data it is not only the within year precision of mean estimates of relative abundance that matter in stock assessments but how consistent the between year estimates are. For example, if there appear to be large inter-annual changes in relative abundance from a survey time-series then a long time-series is needed before any trends in the data could become informative in a stock assessment.

3.2.2 Commercial CPUE

Commercial CPUE have been used in fishery assessments from early in the history of fisheries science (Garstang, 1900; Russell, 1931). CPUE data are used in very many fishery stock assessments in Australia as an index of relative abundance through time. Invariably, the assumption is made that there is a direct relationship between catch rates and the amount of exploitable biomass. However, many factors can influence catch rates, including who was fishing with what vessel and gear, in what depth, in what season, in what area, and whether it was day or night (plus other factors, although information may not be available for all factors of importance).

To use CPUE as an index of relative abundance means that it would be best to remove the effects of variation due to changes in these other factors on the assumption that what remains will provide a better estimate of the dynamics of the underlying stock biomass. This process of adjusting the time series of CPUE for the effects of other influential factors is known as standardization and the accepted way of doing this is to use some statistical modelling procedure that focuses attention onto the annual average catch rates adjusted for the variation in the averages brought about by all the other factors identified. This process is termed statistical standardization.

The primary assumption behind using commercial catch rates in stock assessments is that they reflect the relative abundance of the exploitable biomass through time. The 'through time' phrase is especially important as it implies that any relationship between CPUE and stock abundance remains consistent through time. This is important because in addition to the various factors of location, depth, gear, vessel, etc, there are other factors and events for which there may be no available data. The legitimacy behind using commercial CPUE can be questioned when there are factors significantly influencing catch rates which cannot be included in any standardization. In the Northern Prawn

Fishery (NPF) for example, changes in the fishing gear and how it has been deployed have increased the effective fishing power of individual vessels between 4 and 6 times between 1970 – 2002 (Dichmont et al, 2006). This sort of effect has been termed 'effort creep' as well as 'technological interactions', although this latter term is often restricted to interference between vessels affecting their CPUE. In the SESSF it is likely that the introduction of GPS and GPS Plotters, for example, also led to a form of effort creep but it has not been documented and thus not available for inclusion into CPUE standardizations to date.

In addition to technology improvements, as mentioned earlier, over the last two decades there have been a number of major management interventions in the SESSF including the introduction of the quota management system in 1992, the introduction of the Harvest Strategy Policy (HSP) and associated structural adjustment in 2005 – 2007, and the switch from a calendar year fishing season to one from May to April starting in May 2007. In addition, the combination of quotas that can limit catch and the HSP, is now controlling catches in such a way that many fishers have reported altering their fishing behaviour to try to take into account the availability of quota and their own access to quota needed to land the species taken in the mixed species SESSF. It may be coincidence, but in some species the dates of those major management interventions, 1992 and 2007, correlate strongly with major changes in CPUE (**Figure 4**).

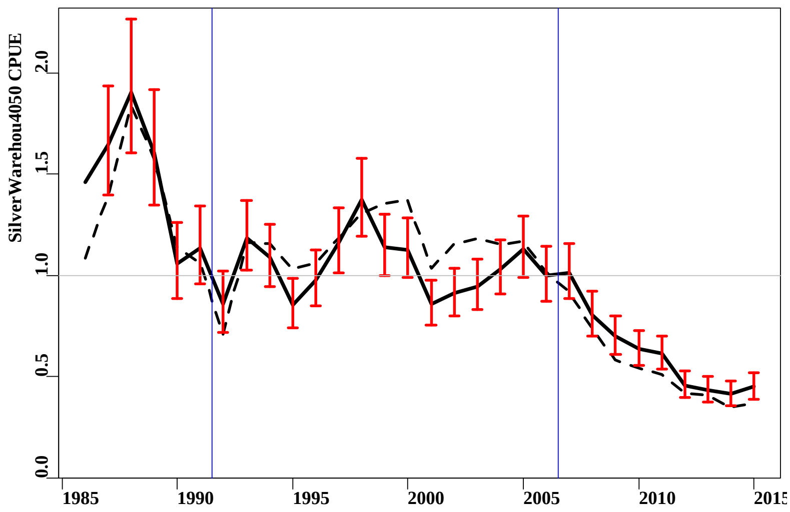

Figure 4. The standardized CPUE for Silver Warehou (*Seriolella punctata*) from SESSF zones 40 and 50. The dashed line is the geometric mean CPUE and the solid line the optimum standardized CPUE with the red 95% confidence intervals around the mean estimates. Both time-series are scaled to a mean of 1.0 to ease visual comparisons. The fine blue lines occurring between 1991 and 1992, relate to the introduction of quota management, and between 2006 and 2007, relate to the introduction of the structural adjustment and the Harvest Strategy Policy. Obviously, the x-axis relates to years.

Some stocks, such as tiger flathead (*Neoplatycephalus richardsoni*), are near or around their target stock size and catch rates are at historically high levels. As a result of this success, some fishers report having to avoid catching species, such as flathead, so as to avoid having to discard and to stay within the bounds of their own quota holdings. Such influences on catch rates tend to bias the catch rates downwards, or at very least add noise to any CPUE signal, which could lead to misinformation passing to any assessment. Currently, there is no way to handle this issue, but care needs to be taken not to provide incorrectly conservative advice or inappropriately high catch targets. Included in the management changes is the on-going introduction of numerous area closures imposed for a range of different reasons, which have different influences on different fisheries depending on their circumstances.

A major on-going issue is whether or not there is a consistent relationship between CPUE, even standardized CPUE, and each stock's relative abundance. Currently there is no way in which to demonstrate this solely on the basis of the data themselves. To make such a validation would require some means of reliably estimating the stock's abundance through time so as to compare with any apparent trends in the commercial CPUE through time. Fishery independent surveys may provide a means of calibrating or confirming any trends seen in commercial CPUE. In the SESSF, however, such surveys have only been running every two years since 2008 (and the first trial year may not have been comparable with following surveys).

3.2.3 Using Indices of Relative Abundance in Stock Assessments

For a stock assessment model to be able to generate valid management advice for a particular fished stock it needs to be fitted to data from that stock. What this means is that a fishery population dynamics model needs to be fitted to data from the fishery with that data consisting of the catches, discards, CPUE or survey indices (or both), and ideally age and length composition data plus any other information available. Both CPUE and survey indices can change dramatically between years whereas, except where some catastrophic event has occurred (Gorfine et al, 2008), or a species is naturally highly volatile, such as Arrow squid or Commercial Scallops, such large and rapid changes in population size are biologically implausible. What the population dynamics model implies in such instances is that the population size trend passed smoothly through the central tendency of the variation unless other more informative data drives the population dynamics in a different direction. Thus, if a particular data set has very large inter-annual variation, whether it be CPUE or a fishery independent index, then the fitting process will not be greatly influenced by that data series.

Whatever the case, the use of CPUE in Australian fishery stock assessments is a potential source of problems and issues for those stock assessments. This present project aims to consider the strengths and weaknesses of CPUE data and make recommendations for improving how it may be used in the future.

3.3 Project Objectives

1. Review the most appropriate catch rate standardisation strategies when targeting is well defined in multi-species fisheries.
2. Compare alternative catch rate standardization strategies in those fisheries where both fishery independent and fishery dependent data are available.
3. After modifying Atlantis SE, simulate shot-by-shot commercial catch rate data and use this in simulation tests for the most robust standardization strategies in mixed fisheries when targeting is unknown and management interventions influence catch rates. [MODIFIED: simulated catch rate data/trends were replaced with 'single species simulated data'].
4. Use simulated catch rate data to conduct MSE testing of the influence of potential biased data and standardization strategies on the outcome of stock assessments that rely on catch rate trends and targets. [MODIFIED: simulated catch rate data/trends were replaced with 'single species simulated data'].

5. Use simulated catch rate data to test the potential influence of effort creep (technical improvements in fishing power) on stock assessments. [MODIFIED: simulated catch rate data/trends were replaced with 'single species simulated data'].
6. Based on the results of objectives 1 to 5 write a reference manual on the application of the most robust CPUE standardization strategies for Australian fisheries.

The third to fifth objectives had to be modified once it had been determined that Atlantis SE was only able to generate pooled mean CPUE estimate for relatively coarse scales within the simulated single species fisheries. Atlantis is designed as an ecosystem and hence multi-species model. The attempts to generate single species data from the Atlantis simulation framework will only ever achieve a limited resolution. An alternative objective adopted was:

3. Simulate shot-by-shot commercial catch rate data and use this in simulation tests for the most robust standardization strategies in mixed fisheries when targeting is unknown and management interventions influence catch rates.

4 Potential Issues when using CPUE Data

4.1 Introduction

Commercial catch rates (CPUE) often constitute the only available index of relative abundance in many fisheries, including in Australia. This makes the important assumption that CPUE is at least proportional to a fished stock's biomass, which implies that if biomass goes up or down the expectation is that so will CPUE, although possibly with a time-lag. In addition to this assumption it has long been recognized that changes in nominal catch rates can be influenced by many factors other than changes in stock size (Kimura, 1981; Ye and Dennis, 2009). Statistical catch rate standardization aims to reduce the effects of other factors, such as vessel, month, and depth of fishing, on any CPUE trends through time so that the standardized index should more closely reflect the actual relative abundance (Maunder and Punt, 2004; Campbell, 2015). In fact, many aspects of the process other than the actual analytical methods used can influence the outcome of CPUE standardization. These other aspects include the quality and amount of fisheries data available, and whether the data used have been filtered or censored in any way. It is very common to read mention of such data selections being applied but rare that such practices are explained, defended, or even described in detail.

4.1.1 The Statistical Methods Used

For a CPUE standardization to act as an index of relative abundance in a stock assessment then generally the objective of the analysis is to provide a detailed description of any trends through time rather than attempting to make predictions of how those trends may develop into the future. This objective influences the methods that might be used for such analyses and how their results are to be interpreted. An array of methods has been used for conducting statistical CPUE standardizations, including Linear Models, Generalized Linear Models, Generalized Additive Models, Generalized Linear Mixed Models, and others (Venables and Dichmont, 2004). Comparisons have been made between the various methods when applied to the same data sets but, so far, no universally optimum analytical strategy has been identified. Part of the reason for this is no doubt the diversity of the quality and quantity of data that are available from different fisheries. This can vary from summary data across different physical and temporal scales (perhaps monthly in five-degree squares) down to shot-by-shot data that includes enormous detail but also potentially a good deal of noise.

4.1.2 The Data Used

At whatever scale of fishing operation, wherever catch and effort values are estimated by the fisher then such CPUE data generally suffers from rounding errors. When fitting statistical models to such observations all the statistical methods used attempt to generate predicted CPUE values to compare with each observation. Data quality issues, especially the rounding of catch and effort values, can make the selection of an appropriate probability density function with which to model the data difficult. Invariably the analysis and data selected in any particular case is a compromise that attempts to discover which data are informative with respect to observed catch rate trends within a particular fishery (where a fishery is for a given species in a defined region using a given gear) and account for any statistical properties of the catch rate data.

How the available data are treated prior to analysis is an aspect of the practical implementation of catch rate standardizations that is rarely discussed, in particular how data to be included in the analyses are selected and what impact such selections can have on the outcomes of the analysis. The primary objective of any initial data selection is usually to focus attention on data that can be considered more representative of a stock's dynamics. If CPUE is related to the stock biomass (will be informative about the biomass levels) then catches in the fishery are expected to affect that biomass and hence the CPUE; increased catches should eventually lead to decreased CPUE and decreased catches to higher CPUE (which is only an approximate description but a reasonable starting point for a developed fishery). Thus, data are selected that are intended to represent where most of the fishing occurs, the core of the fishery, by those particular vessels that are attempting to capture the species concerned. Such criteria would reject data from the periphery of a fishery and data from vessels that were not intending to capture the species. A problem with such criteria being used to guide data selection is that the core of a fishery may change and targeting of a species is generally unknown and generally has to be inferred from the data itself.

Typically, commercial catch and effort data are collected from fishers in some form of log-book (paper or electronic). This data can include many different aspects of the fishing operations and be reported at different geographical and temporal scales. Logbooks will commonly contain data on an array of factors relating to location, date/time, depth, effort, and catch, all in relation to a specific gear and vessel. Fortunate analysts also have data on each vessel's characteristics relating to its relative fishing power and how that may have changed through time (Bishop *et al.*, 2004; Bishop, 2006). Other, less fortunate, analysts may only have each vessel's name and can only assume that the vessel has had stable characteristics and skippers through time (Haddon, 2014), both of which assumptions being highly doubtful!

4.1.3 Identifying the Fishery Core

CPUE data from commercial fisheries can be copious. In the Australian South Eastern Scalefish and Shark Fishery (SESSF; Smith et al., 2008) detailed fisheries data has been collected since 1986 and for the more important target species there can be 100,000s of records across the years. Fitting statistical models to such data has its own array of problems.

Ideally, when fitting statistical standardization models to data, each level of each factor being included will have an equal amount of data available, i.e. the data are balanced across the factors being investigated. In addition, each factor being examined is assumed to be independent (orthogonal to other factors) although they may interact in their effects on the dependent variable (the CPUE). These ideal assumptions are invariably badly compromised by fisheries data. For example, if depth of fishing is included in the analysis then, not unexpectedly, a given species would be expected to exhibit a depth preference, and this can be reflected in the relative numbers of observations made in different depths (**Figure 5**). With factors like Depth it is common to eliminate rarely reported depths as a way of minimizing the number of empty to relatively empty cells in the analysis matrix. Such imbalances across levels of a factor and between factors lower the power of any hypothesis testing but their effect on simply describing the trend in expected mean CPUE is less well understood.

Even with 100,000s of records for a given species not all of them will reflect the core of a fishery where the species can best be captured and possibly targeted and so some data selection is often made. This process usually begins by identifying the fishery to which the analysis is to be applied through selecting records from a fisheries database for a particular species from defined areas taken with a specific fishing gear. The selection of the species may seem obvious although it is especially important in mixed fisheries such as the SESSF.

The areas used need to reflect the distribution of the fishery and the stock, ideally covering an identifiable fishery both in management and biological terms. Even in the absence of evidence-based biological stocks (meaning a mostly isolated reproductive unit) it often remains possible to define evidence-based management units that relate primarily to fishing behaviour but that exhibit a degree of homogeneity of properties sufficient for the region to be analysed separately (Begg and Waldman, 1999; Cope and Punt, 2009). If major management changes have occurred in a fishery this may also suggest that an initial selection of a set of defined blocks of years could be made. These initial stages of data selection merely relate to identifying a particular fishery of interest; it is assumed that there would be a defensible argument available justifying the selection of a given species taken by a given method in particular areas and periods. None of this initial selection should be controversial or in need of much discussion, although, in the interests of repeatability any selection decisions should still be fully documented. That is simple to state but in fact very few documents describing CPUE analyses provide details concerning the criteria used in their data selection. Where data selection becomes more controversial and even less well discussed is in any further data selection following this fishery identification stage.

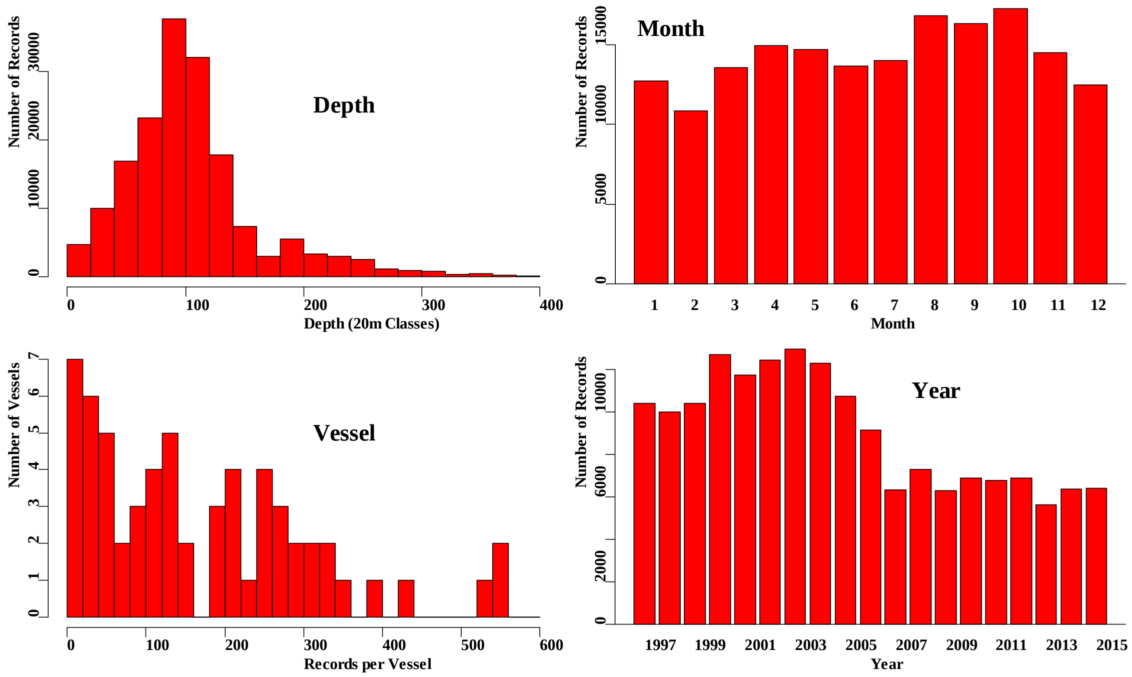

Figure 5. Using Tiger Flathead (*Neoplatycephalus richardsoni*) from SESSF zones 10 and 20 from 1997 – 2015 to exemplify the highly unbalanced distribution of observations of catches by depth, Month, Vessel, and Year. The month factor is the best balanced of the four illustrated. The reduction in the number of records following the structural adjustment at the end of 2006 is clear in the bottom right plot.

4.1.4 Documentation and Defensibility

The standardization of commercial CPUE has become standard practice in almost all fisheries of significant financial value which use CPUE data, although exceptions exist (Linnane et al, 2015; Mayfield et al, 2014). However, many such standardizations are often only reported in the so-called grey-literature and the details of the processes and methods are often very limited even in the better documented ones. It will be argued and recommended here that standardization of available catch and effort data should become routine but also that reporting or documenting the assumptions of the analysis, the details and justifications of any data selections used, and of the specific methods and software used should also become routine.

The data we have from a fishery, in many cases, constitutes the only available evidence about the stock status of the fished species. Any manipulation of such data should therefore be documented in sufficient detail to defend such operations to interested stakeholders. Problems with CPUE data are well known (errors, outliers, non-linearity between CPUE and exploitable biomass, etc) although each data set tends to have its own selection of issues that dominate. Each issue, depending on how it is dealt with, has implications for statistical standardizations. Even if a particular issue is not important for a given data set, in the interests of openness and improved defensibility it would be good practice to document in each analysis the assumptions made and how each known issue has been dealt with.

There are a number of routine operations that are often conducted on CPUE data, but which would be impossible to repeat without documentation. The best form of openness with respect to the analysis of data is for the analyses to be repeatable by others and this provides a clear benchmark for adequate documentation. There is an array of specific issues that should be given attention in any such documentation:

- the identification and removal of outliers and mistakes (data errors),
- data selection aimed at focussing analyses on targeted effort,
- commercial CPUE data while often abundant is often highly unbalanced among the factors that influence its values,
- the factors for which there are data are often correlated rather than independent, and
- commercial CPUE is on estimated catch and effort and these are often rounded numerically leading to unusual (non-parametric) data characteristics.

Each of these aspects to CPUE standardization needs attention and each will be examined in this chapter.

4.1.5 Data Errors

Notoriously, fisheries data from commercial logbooks are often contaminated with errors. Fishers may fill in log books when tired, or in rough seas, and errors can include transcription errors, data entry errors, missing values, and even deliberate errors. Data entry can require the interpretation of hand written logbooks so occasional errors are not surprising, for example, extra digits can accidentally be included e.g. 6000m instead of 600m or 60m as a depth or net length. Deliberate errors can also be recorded by fishers e.g. in the location data where the latitude and longitude of a favourite fishing site will not be revealed and spurious coordinates are reported instead.

Many of the more extreme such errors can be captured by reasonable range checking on data entry, but the practicality of having to enter a large number of records in a short time has sometimes seen such range checking turned off. Under quota systems, logbooks can sometimes become legally binding documents that must be entered as they are written and clarification of obvious errors by contacting individual fishers can be time consuming, especially as fishers can be expected to be at sea for significant amounts of time.

4.1.6 Data Selection Criteria

Given the recognized issues and potential problems with fisheries data it would appear to be a defensible strategy to pre-select data records for inclusion in any analysis with the objective of eliminating potentially spurious or erroneous records and removing empty factor levels from consideration. Effectively, such data selection replaces range checking that might have occurred at the point of data entry with a process of data rejection prior to analysis.

In addition to the removal of outliers, preliminary data selection is also often used when attempting to focus analyses on those records that are taken to be more representative of the fishery, perhaps where the species of interest was more likely to have been targeted or at least expected to be part of the catch in what is a mixed fishery. Fisheries data is often highly unbalanced across the many factors that might influence CPUE, such as year, vessel, depth, month, region, etc (**Figure 5**). For example, a species may very rarely be caught at the extreme deep end of its depth range and inclusion of such uncommon deep records may increase uncertainty more than they improve the standardization. It could thus be argued that it is defensible to select data for analysis from a particular depth range just as data for a particular species from particular areas are selected. This approach is sometimes taken further with the selection of particular vessels based on defined criteria (e.g. present for a minimum number of years in a fishery, or some minimum mean annual catch, etc), all with the justification of focusing on the primary targeted fishery for a species.

While the objectives of such extended data pre-selection can appear clear and reasonable the details and mechanics of the criteria used for making such selections and the implications that these selections have for the analyses are rarely made clear. Nevertheless, the process is common and has been termed data ‘selection’, ‘cleaning’, and even ‘grooming’, although generally very few details are given for how it was done, what criteria were used, or even why it was done. Such a lack of documentation makes such practices less defensible and means that repeating or updating those analyses becomes difficult or impossible.

4.1.7 Section Objectives

There are three objectives to this present section:

- identify characteristics of commercial CPUE data that may complicate statistical standardization analyses and often lead to data selection,
- explore the impact of different types of preliminary data selection on the outcomes of catch rate standardization and, where possible,
- identify strategies for minimizing, or at least identifying any potentially adverse effects, and improving the repeatability of the analyses.

4.2 Methods

4.2.1 Species Data Used

The SESSF is a highly mixed fishery in which specific targeting is relatively limited except for species such as Orange Roughy (*Hoplostethus atlanticus*), which lives in habitats that are separated from most other species (Tilzey and Klaer, 1994). Catch rate data from the SESSF is commonly used in stock assessments every year and the CPUE data are routinely standardized prior to use in these assessments (Haddon, 2014). Three recognized target species will be examined in the following analyses: Tiger Flathead (*Neoplatycephalus richardsoni*), Pink Ling (*Genypterus blacodes*), and Jackass Morwong (*Nemadactylus macropterus*). The first two species remain two of the primary economically valuable species in the SESSF, and Jackass Morwong used to be important, although now its productivity has declined (Wayte, 2013), its relative value to the fishery as a whole has also declined. In the examples considered here the focus will be on trawl fisheries for those three species that operate in the eastern part of the SESSF, which extends over more than 10 degrees of latitude and more than 5 degrees of longitude (**Figure 6**). Log-book data is available at the time of writing from 1986 – 2015/2016.

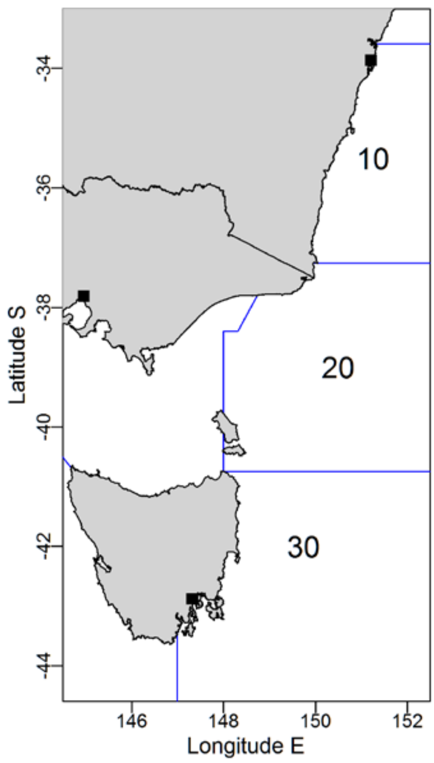

Figure 6. Schematic map of SESSF reporting blocks 10–30, with the fine blue lines representing block boundaries. The locations of Sydney, Melbourne, and Hobart are indicated by black squares from top to bottom. The main fisheries for Tiger Flathead and Jackass Morwong are in zones 10 and 20 while that for the eastern stock of Pink Ling is found in zones 10, 20 and 30.

4.2.2 Other Data Used

To demonstrate that the fuzzy nature of estimated catch and effort data is not found only in the SESSF, data for Tiger prawn catches in the Northern Prawn fishery from 2013 were extracted from the catch and effort data base. The trawling is relatively continuous with the nets being hauled at relatively brief intervals and effort is recorded as total hours fished in the day so CPUE is recorded as catch per shot. The catches reported across the season that were under 100 kg were plotted as frequencies of reporting in 1 kg units to reflect the reporting practices. To provide a further contrasting fishery, commercial catch and effort data from the Tasmanian abalone fishery was also considered where catches are in kgs and effort is in dive hours.

4.2.3 Data Quality

The catches and effort that go into generating CPUE data are generally estimated by the vessel skippers and, as estimates of weight on moving vessels, can be expected to exhibit levels of uncertainty. The detailed distribution of catches and of effort will be examined along with the original ranges of such log-book data to illustrate some of the limitations of commercial CPUE data that lead to a need to exclude outliers. Simple tabulation and plots are sufficient to perform such data checking. Catches, effort, and depths of fishing will be examined in detail for the effect on statistical standardizations of removing different, seemingly extreme, minimum and maximum amounts of effort for each species and fishery.

4.2.4 The Effect of Depth Selection

The selection of particular depth ranges is a common reflection of the fact that species generally have preferred depth ranges within which most catches of those species can be taken (Haddon, 2014). This is well recognized and the assessment groups responsible for reviewing the stock assessments for these species have identified particular depth ranges to be included in the data selection for the three species exemplified here (**Table 2**). In addition to the simple tabulation of the proportion of catches and records with and without the selection, statistical standardizations with and without these depth selections will be conducted and their yearly index estimates compared.

Table 2. Data selection analyses conducted. Years relates to the minimum number of years a vessel reports catches from a fishery. Mean Annual Catch determines the requirement for including a vessel in the analyses. Finally, the minimum catch per record is applied to all data rather than by vessel. In each case these data filters were applied only after the agreed depth range had been selected. A record was included in the available data when a minimum of 1 kg was reported. The agreed depth ranges are those set by the SESSF Assessment Group and Management Advisory Committee as best representing the expected range of the fishery for each given species.

Common Name	Flathead	Pink Ling	Jackass Morwong
Minimum Years	1, 3, 5, 10, 15	1, 3, 5, 10, 15	1, 3, 5, 10, 15
Mean annual catch (t)	0.001, 1, 5, 10, 15	0.001, 1, 5, 10, 15	0.001, 1, 5, 10, 15
Minimum catch (kg) per record	1, 5, 30, 60	1, 5, 30, 60	1, 5, 30, 60
Maximum catch (kg)	1000	750	1000
Zones	10, 20	10, 20, 30	10, 20
Agreed depth range	0 - 400	250 - 600	70 - 300

4.2.5 Selecting Vessels on Years in Fishery and Average Annual Catch

In the SESSF, early standardizations used forms of data selection to focus analytical attention upon the more important vessels in a fishery, for example, by selecting only those vessels that had been in the blue grenadier fishery (*Macruronus novaezelandiae*) for at least two years with an average catch of at least 5 tonnes (Punt *et al*., 2001). This form of selection will be explored further using the three chosen SESSF species and, in each case, the minimum number of years a vessel reported catches from a fishery and the minimum annual catch for a vessel to be included in an analysis range from using all available data up to requiring multiple years or high levels of catch (**Table 2**). Once again standardizations will be conducted on each version of the censored data sets and the outcomes compared to determine the effect of such data selections in terms of the final trend in standardized CPUE and the estimated variation around each trend.

4.2.6 Selecting Records relative to Minimum or Maximum Catches

An alternative selection strategy might be to use only those records where a defined minimum catch was achieved, with the argument that if less was taken in a particular shot then it is unlikely the shot was targeting the species. This data selection strategy will also be examined by comparing standardizations using all available data with those where the data are censored for four different minimum catches per record (**Table 2**).

Removing seemingly large catches is more about attempting to identify and remove outliers, nevertheless, they do exist in the data and comparisons were made by conducting standardizations with and without the data selection to determine the effect of applying such filtering.

4.2.7 The Statistical Models

In each example, trawl catch rates, (kilograms per hour), were natural log-transformed in an attempt to normalize the data and stabilize the variation. A general linear model was used on this transformed data rather than using a generalized linear model with a log-link on the raw data; this has advantages in terms of normalizing the data while stabilizing the variance, which the log-link within the generalized linear model does not always achieve appropriately (Venables and Dichmont, 2004). There is, of course, a good deal of debate over what statistical approach is most effective but that will not be considered here. Instead the effects of using alternative modelling approaches are examined directly in **Chapter 6.**

The statistical models used here were variants on the form: LnCE = Year + Vessel + Month + DepthCategory + Zone + Daynight. The ‘Zone’ factor refers to the SESSF zones (**Figure 6**), of which only zones 10, 20, and 30 down Australia’s east coast were used. The ‘DepthCategory’ factor was a series of usually 20m or 25m depth classes across the depth range selected. It is possible to treat ‘Depth’ as a continuous variable rather than break it into categories and the choice of which approach to use can depend on how many data records are available and how smooth is the relationship between CPUE and depth. Here all analyses treated Depth as a categorical factor for both consistency with usual practice and to avoid an extra complication in the analyses. Finally, there were interaction terms which could sometimes be fitted, such as Month:Zone or Month: DepthCategory. The optimum statistical model, which in the cases considered here generally included one of the interaction terms, was selected on the basis of the minimum Akaike Information Criterion (AIC). Thus, the CPUE, conditioned on positive catches of the species of interest, was statistically modelled with a normal linear model on log-transformed CPUE data:

$$ln\left(CPUE_i\right) = \alpha_0 + \alpha_1 x_{i,1} + \alpha_2 x_{i,2} + \sum_{j=3}^{N} \alpha_j x_{ij} + \varepsilon_i \qquad (1)$$

where $ln(CPUE_i)$ is the natural logarithm of the catch rate for the i-th shot, x_{ij} are the values of the explanatory variables j for the i-th shot and the α_j are the coefficients for the N factors j to be estimated (α_0 is the intercept, α_1 represents the coefficients for the different levels of the first factor, *etc*.). The objective of these analyses is to obtain a set of year effects that are used to represent the relative abundance through time. The expected back-transformed year effect involves a bias-correction to account for the log-normality which then reflects the mean of the distribution rather than the median:

$$CPUE_t = e^{\left(\gamma_t + \sigma_t^2/2\right)} \tag{2}$$

where γ_t is the year coefficient for year t and σ_t is the standard deviation of the log transformed data (obtained from the analysis). To simplify the visual comparison of catch rate changes the year coefficients were all divided by the arithmetic average of the year coefficients:

$$CE_t = \frac{CPUE_t}{\left(\sum CPUE_t\right)/n} \tag{3}$$

where $CPUE_t$ is the yearly coefficients from the standardization, n is the number of years of observations, and CE_t is the final time series of yearly index of relative abundance. If visual comparisons were wanted back on the nominal CPUE scale then the time series from equation (3) can be multiplied by the overall geometric mean catch rate across the nominal CPUE time series, which will rescale the time series to have that overall mean instead of a mean of 1.0; this may make such analyses more meaningful to Industry and Managers inspecting the results even though the multiplication by a constant has no effect on the trend expressed by the time-series.

Even though there are often very large numbers of observations (10s to 100s of thousands) there can also be large numbers of parameters (sometimes hundreds, especially when interaction terms of used). Thus, model selection was on the basis of the minimum Akaike Information Criterion (AIC_c) corrected for small sample sizes:

$$AIC_c = n \times ln\left(\sigma^2\right) + \left(2 \times k\right) \times \frac{n}{n-k-1} \tag{4}$$

and

$$\sigma^2 = \sum \varepsilon^2 / n \tag{5}$$

where n is the number of observation, σ^2 is the estimate of the model variance, as in equation (5), which is included in the number of estimated parameters, k is the number of model parameters estimated, and the ε^2 are the squared residual errors; note in equation (5) the division by n to provide the maximum likelihood estimate of variance (Burnham and Anderson, 2002). When n is large the correction term becomes minor and the equation reduces to the *AIC*.

4.3 Results

4.3.1 Data Quality

The identification of outliers within fisheries datasets is complicated by the fact that such data tends to be intrinsically highly variable. If the reported catches are tabulated in any single year, individual shots can report catches ranging from a single kilogram or less, up to multiple tonnes (see supplementary tables S1 – S3 following this chapter's Discussion). Focussing only on the upper ranges of catch-per-shot the spread of these catches varies through the period 1986 – 2015, which can be visualized in terms of the 90% and 99% quantiles (**Figure 7**), but the variation exhibited differs markedly between the three species. Jackass Morwong, which has declined remarkably during this period (Wayte, 2013), is the only species of the three to have exhibited a continuous directional change.

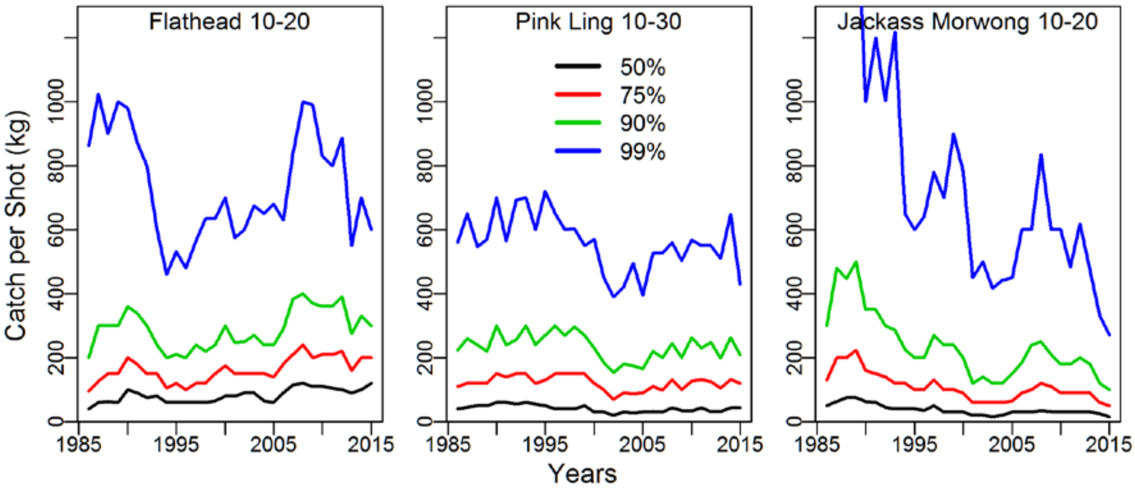

Figure 7. Annual variability in reported catch-per-shot as four different quantiles (50% = median, 75%, 90% and 99%) of the distribution of each year's catches of the three species.

The rounding of estimated catches to the nearest 5 kg, and more often the nearest 10kg, is clearer with Flathead than with Pink Ling or Jackass Morwong. A common belief that a standard fish-bin contains 30kg of fish is reflected by spikes of reports in both species at 30, 60, and 90kg. With Flathead there remain some fishers that consider a fish-bin to contain 32kg, evidenced by smaller spikes at 32, 64, and 96kg. Oddly, the belief with fishers regarding Pink Ling appears to be that a single fish-bin contains either 30kg or 33kg, evidenced by small spikes at 33, 66, and 99kg (**Figure 8**). Why the two species should be either 32kg or 33kg seems more like tradition than fact. Minor catches with Pink Ling and Jackass Morwong, having peaks at 1, 2, 3, 5, 6, 10, and 11kg, are more common than with Flathead, which only has very few reports with 1, 2, 3, or 4 kg.

Similarly, the rounding of the estimated effort is clear when the raw data are plotted at a fine enough scale. Peaks of numbers of records every 15 minutes are apparent from 1 through to 5 or 6 hours (**Figure 8**). The very low numbers implied outside of 1 to 6 hours suggests that omitting such records might have little effect on any standardization. Dominant peaks of reported effort occur at 3, 3.5, and 4 hours for all species, which reflects the mean effort being very similar in all three species being 3.3, 3.5, and 3.4 respectively.

The pattern of dominant catch-per-record in Flathead is different from that in Pink Ling and Jackass Morwong, who are more similar (**Figure 8**). This reflects the lower mean trawl catch rates for Pink Ling and Jackass Morwong rather than where and in what depths they are caught (**Figure 9**).

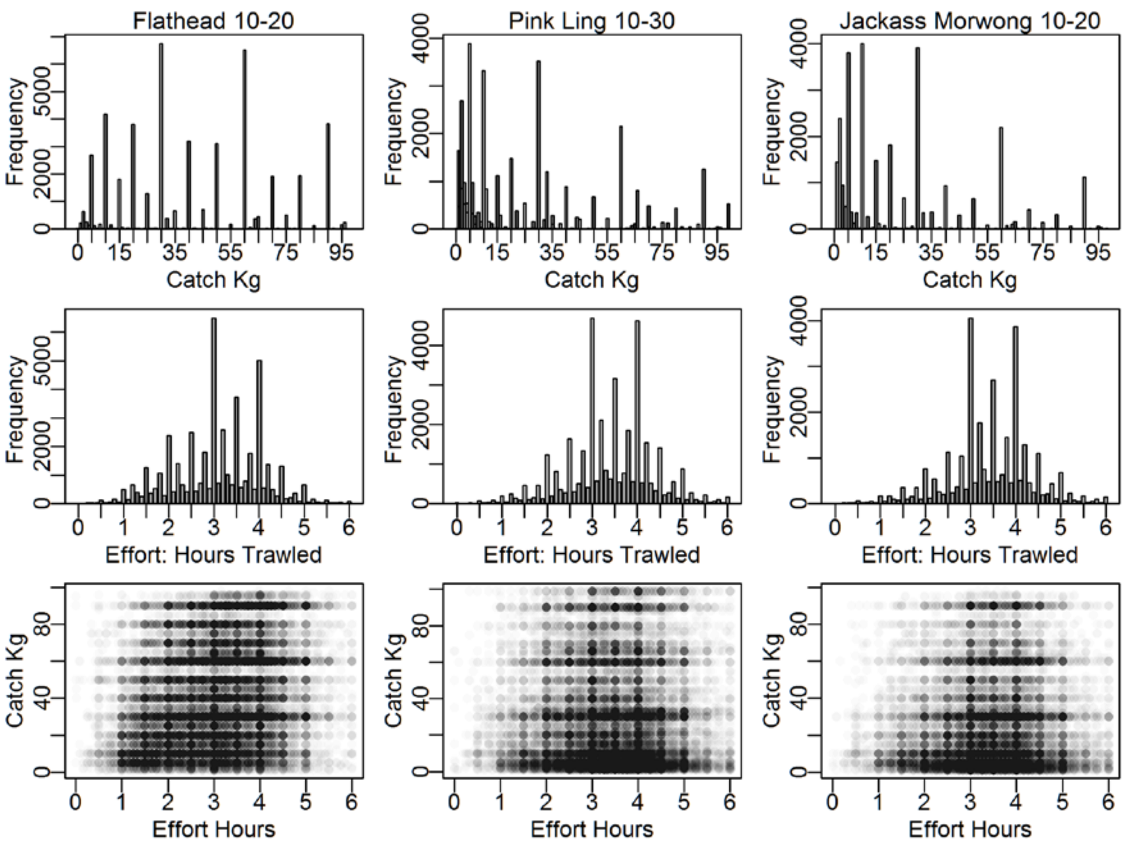

Figure 8. Plots of the catches, effort, and the catch against effort for Flathead, Pink Ling, and Jackass Morwong from the eastern SESSF; the average catch per shot was 127, 70, and 54 kg, the average effort was 3.3, 3.5, and 3.4 hours, and the geometric mean CPUE was 44, 29, and 19 kg/hr, respectively. Each column heading identifies the species and the SESSF zones used. Records were restricted to the years 2000 – 2006, reported catches < 100kg, and <= 6 hours trawling effort. In the bottom plots completely black dots represent 50 or more records with the lightest shade representing a single shot.

Whether any of the larger catch-per-shot values are actually errors or outliers is difficult to determine but such difficulties are less the case with effort data. Across the years 1986 – 2015 reported effort (at least that recorded in the database) for Flathead ranged from -3.65 – 66 hours, for Pink Ling from -33.0 – 25.51 hours, and for Jackass Morwong from -7.5 – 58.67 hours. The negative effort values are clearly errors, but, as in the case of the -33 hours, these can be compounded with other errors such as perhaps accidentally entering two 3's instead of one. The approximate average trawling time tends to range between 3 – 4 hours (**Figure 8**) with reports from fishers saying they would never trawl for more than 10 hours. Of course, other fisheries may exhibit different behaviour.

4.3.2 The Effect of Selecting an Effort Range

An obvious first selection for outliers is to remove any records reporting less than some minimum number of hours and more than some maximum number of hours, which is done in the Great Australian Bight (GAB) fishery but not currently in the SESSF. An examination of the effect upon the standardizations of making a range of data selections based on different combinations of criteria relating to effort was made (**Table 3**). When selections were made with regard to effort, in terms of catches the maximum percent lost from the analyses was only 1.25%, 2.62%, and 1.25% respectively for Flathead,

Pink Ling, and Jackass Morwong. Similarly, for the number of records lost the percentages were 0.697%, 1.66%, and 1.45% respectively.
Even with the strictest selection criteria of >= 1 hour and <= 6 hours for all three species the only differences between standardizations of all the data versus the censored data were extremely minor (see **Figure S1** in the supplementary material). The small proportion of records removed only had minor effects.

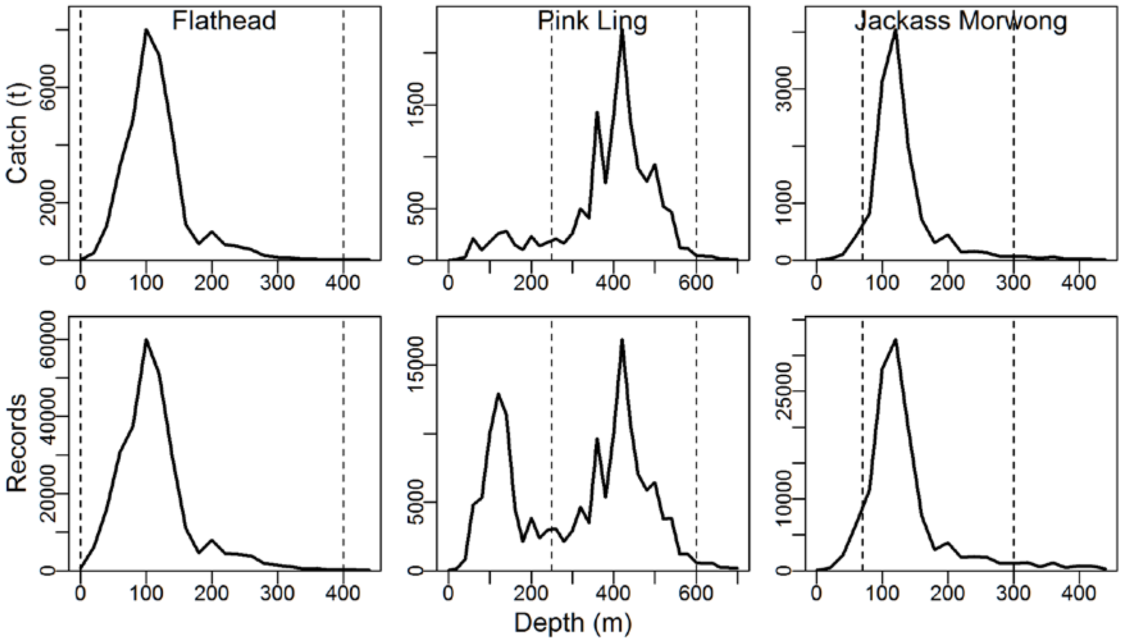

Figure 9. Total catches and total records by depth across the years 1986 – 2015 for Flathead, Pink Ling, and Jackass Morwong. The dashed vertical lines denote the assessment group agreed depth bounds for inclusion in standardizations. Note the y-axis scales differ between plots.

Table 3. The effect of selecting records based on some minimum and possibly maximum amount of effort on the total catch and total records retained for analyses for the three species. The first row of > 0 hours effort relates to the catch and records removed that had 0 or less effort while the second > 0 row relates to the catch and records retained for analysis. The rows below this list the catch and records lost from the analysis by the increasingly stringent selection criteria.

	Flathead		Pink Ling		Jackass Morwong	
Effort Data hrs	Catch	Records	Catch	Records	Catch	Records
All	34288.659	276900	14680.084	164705	12982.832	129278
> 0	456.963	3512	148.941	1703	182.231	1519
> 0	33831.696	273388	14531.143	163002	12800.601	127759
> 0.25	2.684	43	150.007	23	2.019	21
> 0.5	32.387	254	157.311	136	8.119	127
> 0.75	97.899	804	172.182	462	23.166	408
<= 24 & >= 1	150.109	1423	189.736	796	36.077	606
<= 6 & >= 1	423.025	1907	381.225	2706	159.694	1851

When the catch-per-record is plotted against the effort-per-record then the effect of rounding becomes apparent in the fuzzy grid-like outcome (**Figure 8**). These distributions are neither continuous nor discrete and the use of any particular statistical distribution, such as the log-normal, Gamma, Negative Binomial, or Poisson distributions, to act as the residual errors structure when comparing the predicted CPUE for each record with its observed value remains an approximation at best as the statistical model itself can only generate an approximation to the mixture of continuous and discrete distributions of catch and effort.

4.3.3 The Northern Prawn Fishery

The recording of catches in the northern prawn is in fact relatively accurate as they are snap frozen and packed into small cartons quickly after capture (these can vary from 5kg upwards). Nevertheless, there is clear evidence of rounding, as with the SESSF fisheries) when reported catches are plotted in 1kg steps (**Figure 10**).

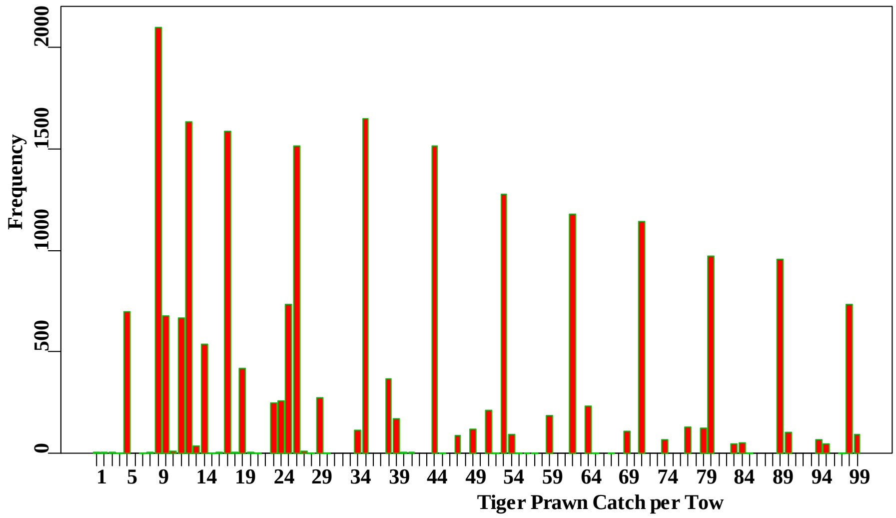

Figure 10. The frequency of catches (kg) from individual trawl hauls in the Northern Prawn fishery from across the whole year of 2013.

Effort is reported as total hours trawled but the fishing operation entails repeatedly hauling the nets and clearing them of prawns so as to maintain quality of the frozen product, thus catch rates are calculated as catch-per-haul or catch-per-record. It is certainly the case that the distribution of CPUE cannot be considered as a smooth or continuous distribution, but neither is it strictly discrete.

4.3.4 The Tasmanian Abalone Fishery

In the Tasmanian abalone fishery the fishing is by hand collection via divers on the shallow rocky sea coast. The required landing dockets are also the source for the catch and effort data. This is an important point because the value of the fishery is so high that the catches are weighed accurately at the point of landing and are thus reliable rather than estimates. Even so they are not smoothly distributed because the abalone divers generally go out with the intent of catching a specific amount of the available quota but the vagaries of good or poor catching mean that their intentions are not always met. Similarly, many divers often go out for a specific period or at least record their effort as some rounded period of time and when combined into a catch rate these again generate the grid like distribution of catch rates seen in other fisheries (**Figure 11**).

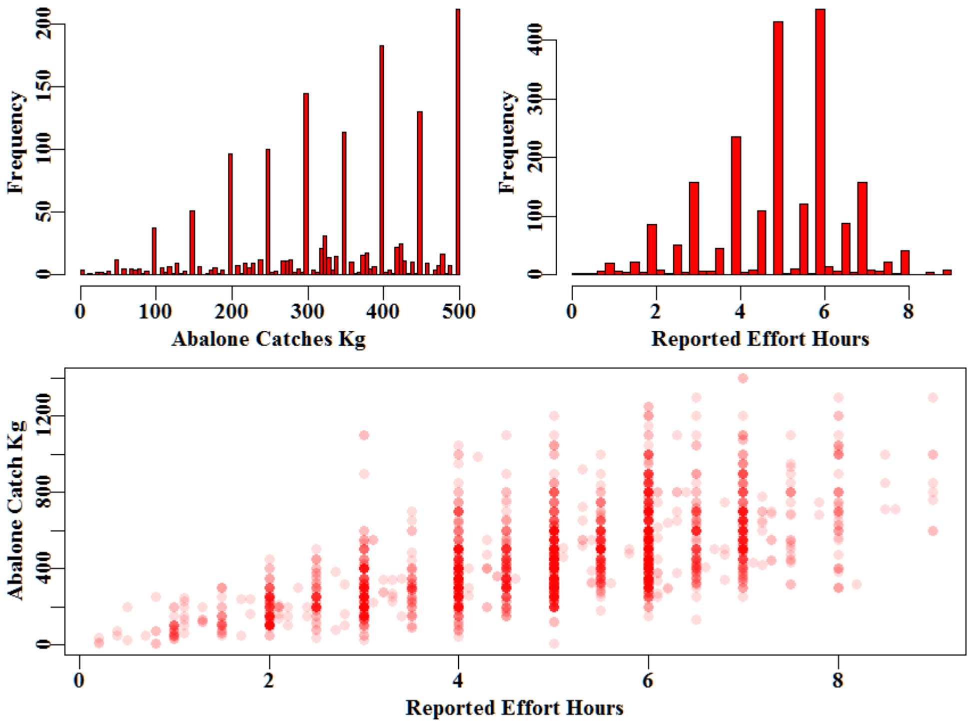

Figure 11. Catch and effort data from the Tasmanian abalone fishery for a single statistical block in the south-west of Tasmania for the years 2012 – 2016. The catches in the top left plot have been truncated at 500 kg to illustrate the granular nature of catches and effort reports were truncated at 9 hours.

4.3.5 Depth as a Factor

In the SESSF, a strongly mixed species fishery, after selecting for species, zones, fishing method, and years, so as to identify the fishery being standardized, it is usual to select for a particular depth range with the intention of focussing any analyses on the main fishery and avoid outliers. For each species the assessment groups have agreed on particular depth ranges (**Figure 9**; **Table 4**). The data for each species already has a number of records that do not have associated depth data, which exemplifies a common problem of missing data when using log-book catch and effort information. Here records with missing information were omitted but the alternative of imputing values also exists (usually, for each record, using the mean estimated for other records within each combination of other factors being used in a standardization). For the three species being used here none has more than 1% of records or catch with missing depth data (**Table 4**). Flathead has the least restrictive of the agreed depth ranges followed by Jackass Morwong and then Pink Ling (**Figure 9**). The Pink Ling example highlights the notion of focussing on the main fishery. In depths shallower than 250m there are very many records which report relatively small amounts of Pink Ling. In this inshore fishery the Pink Ling tend to be smaller, younger, and less common and this is not an area where they are generally targeted by trawlers (**Table 4**; **Figure 9**); in shallower waters they tend to be caught as by-product when targeting other species (such as Flathead and Jackass Morwong). As a result, the RAG has nominated a depth range for the standardization that eliminates more than 13% of the total catches by trawl whereas the agreed depth ranges for Flathead and Jackass Morwong remove less than 1 and 5% respectively.

Table 4. Listings of catch in tonnes and number of records for: all records with depth information, all records within the agreed depth bounds, and those records with missing depth data. Numbers in parentheses are the proportions of the total catch and total records respectively.

Species	Flathead	Pink Ling	Jackass Morwong
Agreed Lower Depth	0	250	70
Agreed Upper Depth	400	600	300
Total Catch	33942	14595	12851
Depth Limited Catch	33860(0.998)	12645(0.866)	12230(0.952)
Catch no depth data	347.1(0.01)	85.4(0.006)	131.9(0.01)
Total Records	274214	163593	128033
Depth Limited Records	273383(0.997)	100735(0.616)	116190(0.908)
Records no depth data	2686(0.01)	1112(0.007)	1245(0.01)

A comparison of standardized CPUE with and without the agreed depth range data restrictions exhibits almost no discernible effects in Flathead and extremely minor effects in Jackass Morwong (see Figure S2 in supplementary material). Even the reduction of 10% of records and 5% of catches, in Jackass Morwong did not translate into any changes that would have an influence on any assessment based upon the standardized CPUE. However, with Pink Ling there were large effects that would have similarly large effects on any subsequent stock assessment. Given the 38% reduction in records and 13.4% reduction in catches (**Table 4**) such large alterations to the CPUE trajectory are not surprising. While the general trend in CPUE is retained the details and gradients of change are altered (**Figure 12**).

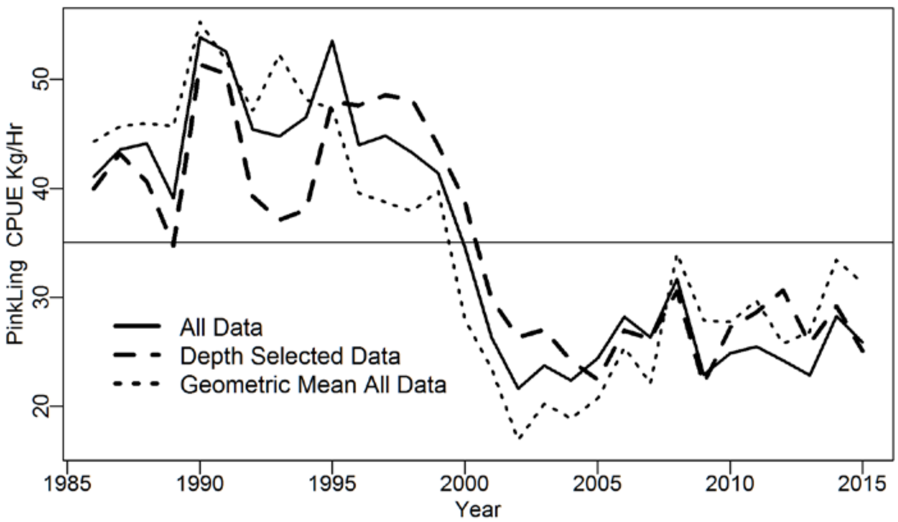

Figure 12. The effect of data selection with respect to depths between 250 – 600 m on Pink Ling. Note the y-axis does not start at 0 so as to emphasize differences between the lines.

4.3.6 Requiring a Minimum Number of Years in the Fishery

The number of years a vessel is required to be in a fishery for it to be included in the analysis had surprisingly little effect upon the final standardization trends (see Supplementary Figures S3 – S9) with Pink Ling exhibiting the largest effects (**Figure 13**), however, these remained essentially trivial relative to the large scale trends and would not be expected to influence a formal stock assessment based upon the degree of change to the overall trend. This was not because there were few vessels present only for a few years. In fact the majority of vessels only report from each fishery for relatively short periods with the median duration in each fishery being 4, 5, and 4 years respectively for Flathead, Pink Ling, and Jackass Morwong (see supplementary Figure S3).

This is a surprising result because the impact of the data selection on the proportion of the total catch, the proportion of the number of records and the proportion of vessels included in the analyses were relatively marked (**Table 5**). For example, with Flathead selecting only vessels with a minimum of 15 years in the fishery reduced the total catch by 27% and the total records by 29% with only 22.5% of the vessels remaining in the analysis. Nevertheless, the trend and even many of the details remained essentially unchanged (supplementary figures S4 – S5). Requiring a minimum number of 3 years made effectively no impact on the standardization and, with Jackass Morwong, when all trends were scaled to a mean of 1.0, only changed the standardizations at the third decimal place at most and sometimes the fourth place; requiring a minimum of five years in the fishery changed that to two and three decimal places.

Plotting the deviations of each trend relative to the standardization that included all vessels (one year) illustrates that sometimes the variations can be up to 10% (**Figure 13**) but even those events have only minor effects upon the trends exhibited by the three species.

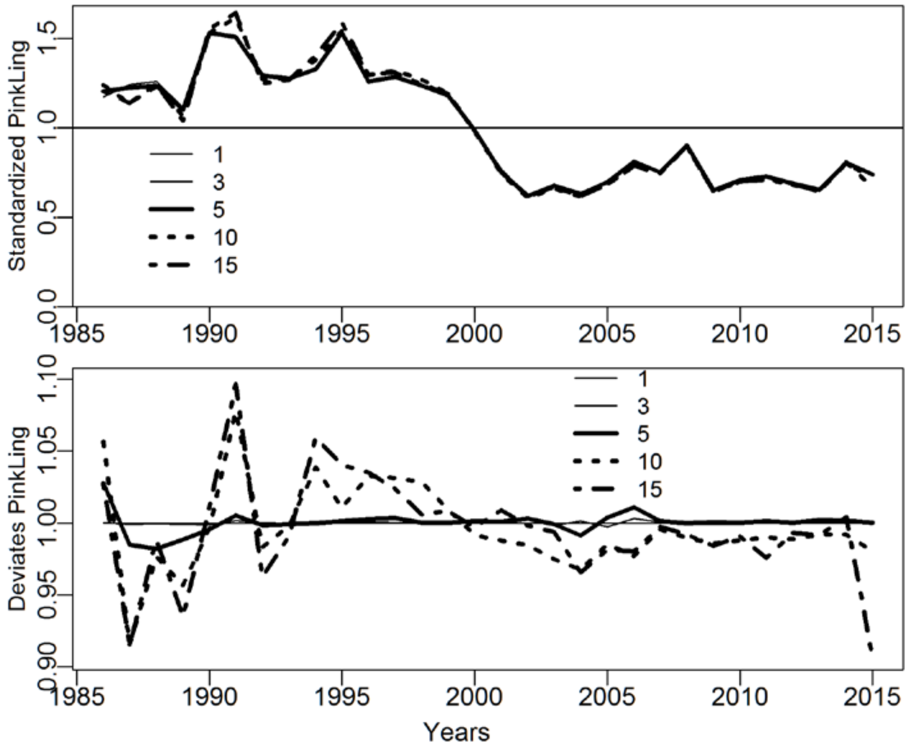

Figure 13. The effect of selecting for a minimum number of years for a vessel to be in the fishery on the CPUE standardizations for Pink Ling with all trend scaled to a mean of 1.0. The deviates, in this case, are ratios relative to the minimum of 1 year (all data) and illustrate the proportional rather than absolute differences. A minimum of 3 years had little effect so the emphasis is on requirements of 5, 10, and 15 years.

Table 5. The effects of requiring a minimum number of years in the fishery from 1 – 15 year on: the number of observations (Nobs), the number of vessels included (Nvess), the total Catch, the proportion of the number of observations (pNobs), the proportion of vessels included (pvess), and the proportion of total catch included (pCatch).

Species	Year	Nobs	Nvess	Catch	pNobs	pvess	pCatch
Flathead	1	271247	187	34289	1.000	1.000	1.000
Flathead	3	267281	118	33801	0.985	0.631	0.986
Flathead	5	256102	92	32442	0.944	0.492	0.946
Flathead	10	225135	58	28099	0.830	0.310	0.819
Flathead	15	198209	42	24738	0.731	0.225	0.721
Pink Ling	1	162110	203	14680	1.000	1.000	1.000
Pink Ling	3	159024	131	14468	0.981	0.645	0.986
Pink Ling	5	154050	103	14051	0.950	0.507	0.957
Pink Ling	10	134069	65	12225	0.827	0.320	0.833
Pink Ling	15	114715	43	10441	0.708	0.212	0.711
Jackass Morwong	1	126721	180	12983	1.000	1.000	1.000
Jackass Morwong	3	124103	117	12737	0.979	0.650	0.981
Jackass Morwong	5	117998	85	11631	0.931	0.472	0.896
Jackass Morwong	10	99714	50	9247	0.787	0.278	0.712
Jackass Morwong	15	87966	37	8180	0.694	0.206	0.630

4.3.7 Requiring a Minimum Average Annual Catch per Vessel

Requiring a minimum average annual catch for a vessel to be included in a standardization analysis is more difficult to scale in an equivalent way for different species. In terms of total catch and total numbers of records, Flathead dominates, followed by Pink Ling, and then Jackass Morwong (**Table 3**). When minimum annual catch-per-vessel values leading to similar percent reductions in catch and numbers of records are compared then, as with the minimum number of years in the fishery, the effect of the minimum annual catch is also minor (see supplementary **figures S10 – S15**; **Table 6**). To lose about 8% of catches required a minimum mean annual catch of about 10t per vessel in Flathead but only about 5 t per vessel for Pink Ling and Jackass Morwong.

With Flathead, selecting for a 5 tonne minimum annual catch per vessel led to maximum deviations from all data trend of only about 3% (**Figure 14**), which visually is barely discernible from the all-data trend. The 10 and 15 tonne annual catch requirements have larger effects, especially with the other two species, which reflects the relatively extreme removals that such a requirement lead to (**Table 6**). Such observations all serve to illustrate that using a data selection approach like this would need to be calibrated for each species to achieve the desired effect.

Table 6. The effects of requiring a minimum average annual catch per vessel from 1kg – 15 tonnes on: the number of observations (Nobs), the number of vessels included (Nvess), the total Catch, the proportion of the number of observations (pNobs), the proportion of vessels included (pvess), and the proportion of total catch included (pCatch).

Species	AvAnnC	Nobs	Nvess	Catch	pNobs	pvess	pCatch
Flathead	0.001	271247	187	34289	1.000	1.000	1.000
Flathead	1	270010	146	34258	0.995	0.781	0.999
Flathead	5	261821	111	33850	0.965	0.594	0.987
Flathead	10	235266	79	31594	0.867	0.422	0.921
Flathead	15	204699	65	29175	0.755	0.348	0.851
Pink Ling	0.001	162110	203	14680	1.000	1.000	1.000
Pink Ling	1	159184	126	14577	0.982	0.621	0.993
Pink Ling	5	136270	74	13274	0.841	0.365	0.904
Pink Ling	10	109095	45	11153	0.673	0.222	0.760
Pink Ling	15	82574	26	8468	0.509	0.128	0.577
Jackass Morwong	0.001	126721	180	12983	1.000	1.000	1.000
Jackass Morwong	1	119896	102	12796	0.946	0.567	0.986
Jackass Morwong	5	99158	65	11906	0.782	0.361	0.917
Jackass Morwong	10	84234	44	10617	0.665	0.244	0.818
Jackass Morwong	15	71722	33	9472	0.566	0.183	0.730

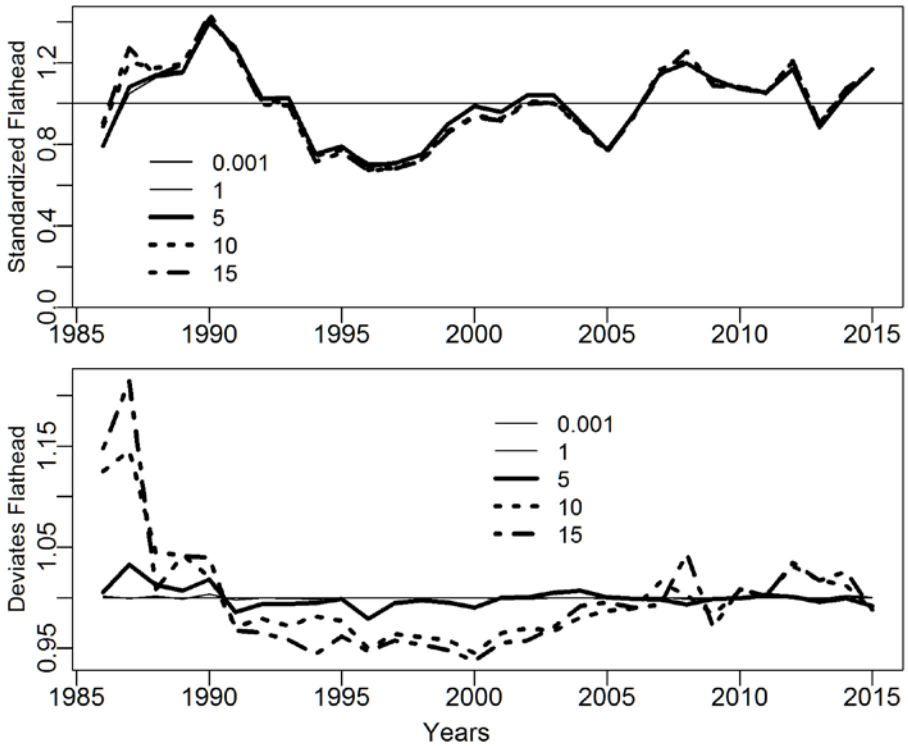

Figure 14. The effect of selecting for a minimum average annual catch per vessel on the CPUE standardizations for Flathead with all trend scaled to a mean of 1.0. The deviates, in this case, are ratios relative to the minimum of 1 kg (all data) and illustrate the proportional rather than absolute differences.

4.3.8 Effect of a Minimum Catch per Record

Setting a selection criterion based on a given minimum catch-per-record can have a relatively large effect on the number of observations while making little impact on the proportion of vessels included or the total catch included in the analysis (**Table 7**). This is a reflection of the fact that the raw CPUE data is strongly positively skewed with most data having lower values. The effect on the number of observations is important because the use of a minimum catch-per-record can have relatively large effects upon the resulting standardization. With each of the three species the effect was to flatten the standardized trend so that it became closer to the overall average (**Figure 15**; also see supplementary **Figures S16 – S21**). Such a flattening would act to obscure any measure of stock depletion if there were a trend in the number of smaller catches (as in Jackass Morwong).

With Jackass Morwong, for example, the effect was progressively greater with each increase in minimum catch-per-record, although selecting on 30 or 60kg per record had about the same effect. Selecting for 30kg per record altered the standardized CPUE trend from a decline from about 2.0 to below 0.5, into a decline from about 1.5 to about 0.75, which would provide a very different interpretation of the stock status. The impact on the other two species is equally large but might not have equivalently misleading consequences.

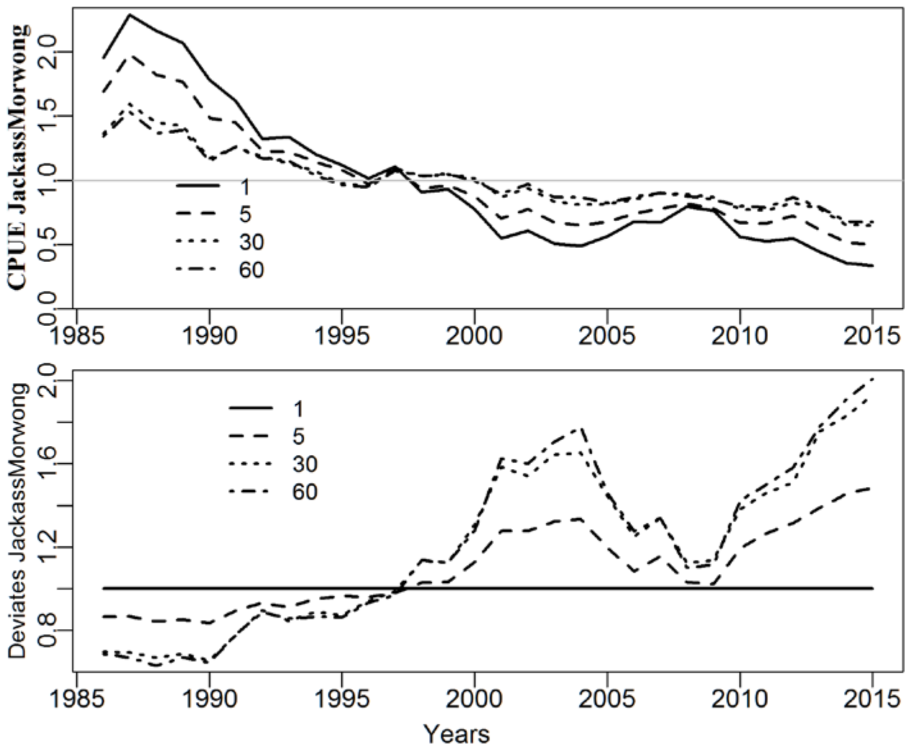

Figure 15. The effect of selecting for a minimum catch-per-record on the CPUE standardizations for Jackass Morwong with all trends scaled to a mean of 1.0. The deviates are ratios relative to the minimum of 1 kg (all data) and illustrate the proportional rather than absolute differences. Catches are kilograms.

Table 7. The effect of altering the minimum catch-per-record (CPR) from 1kg – 60kg on the variance described by the statistical model, the number of observations (Nobs), the number of vessels included (Nvess), the total Catch, the proportion of the number of observations (pNobs), the proportion of vessels included (pvess), and the proportion of total catch included (pCatch).

Species	CPR	adj_r2	Nobs	Nvess	Catch	pNobs	pvess	pCatch
Flathead	1	18.307	270651	187	34288	1.000	1.000	1.000
Flathead	5	16.637	261088	187	34246	0.965	1.000	0.999
Flathead	30	13.498	198901	180	32911	0.735	0.963	0.960
Flathead	60	12.916	147890	171	30344	0.546	0.914	0.885
Pink Ling	1	50.379	159617	203	14678	1.000	1.000	1.000
Pink Ling	5	41.170	137241	202	14597	0.860	0.995	0.994
Pink Ling	30	22.525	88494	183	13707	0.554	0.901	0.934
Pink Ling	60	14.817	63918	172	12552	0.400	0.847	0.855
Jackass Morwong	1	28.061	123980	179	12980	1.000	1.000	1.000
Jackass Morwong	5	23.923	107605	177	12916	0.868	0.989	0.995
Jackass Morwong	30	15.332	65172	165	12076	0.526	0.922	0.930
Jackass Morwong	60	14.452	45242	154	11096	0.365	0.860	0.855

4.3.9 The Effect of a Maximum Catch per Record

The exclusion of what were relatively large catches by species (**Table 2**) led to the removal of about 4.5% of catches but only about 0.4% of records from Flathead and Pink Ling. In Jackass Morwong, even though about 15.4% of catches were removed that still only removed about 0.9% of records. By conducting standardizations on data including the larger catches and contrasting those with data sets without the larger catches (see **Table S4** and **Figures S21 – S24** in the supplementary materials) the only influence these larger shots have is very minor on the general trends and were mostly not visible in plots.

The same patterns were exhibited in the residuals between the trends with and without the larger shots such that in earlier years (before 1992 and the introduction of quota) Flathead and Jackass Morwong show slightly lower trends while after 1992 the trends are slightly higher. With Pink Ling, which is caught in significant amounts in the non-trawl fishery that only began recording detailed log-book data from 1997 onwards, the switch from having deviates below to above the mean occurs in about 1997.

4.4 Discussion

4.4.1 Initial Data Selection

The selection of data for inclusion in any analysis of CPUE from commercial fisheries is a necessary first step to identify a particular fishery in a particular region using a particular method. This is especially the case when dealing with a mixed fishery in which the targeting behaviour is not easily identifiable, but even with a species which is always the explicit target, the area and gear still need to be specified. Any data selection should always be done in a manner where it is possible to defend the selection of given fisheries or parts of fisheries in terms of whether they constitute a coherent whole. For example, in Punt *et al*. (2001) the Australian Blue Grenadier (*Macruronus novaezelandiae*) fishery was split into a spawning fishery and a non-spawning fishery with each defined as operating in particular areas and months within each year and having rather different fleets fishing with rather different selectivity characteristics and CPUE.

Explicit data selection can also be useful for the elimination of outliers and errors, although outliers and errors are not always obvious before conducting a preliminary analysis given the variable nature of fisheries data. Generally, such fundamental data selections, as region, depth, and gear, should not be a problem but designating particular data points as outliers is invariably more difficult to support. The most defensible approach before presenting a set of results involving the removal of putative outliers is to repeat an analysis with and without the selection so that its effects can be made explicit. In the examples considered of selecting only those records with different levels of what might be considered plausible levels of effort the omission of the less plausible records in each case led only to trivial changes in the overall trends of yearly coefficient from the standardizations. The decision about whether to exclude any unusually low or high levels of effort would then become one of what might be considered the most realistic or plausible representation of the fishery concerned (industry input on such matters is valuable). Even where the inclusion of a few records with ridiculous effort values (> 24 hours!) may have very little influence on the outcome of an analysis their removal means that defending the analysis should be simpler and more convincing. But where more prominent effects arise in subsequent trends of CPUE then a stronger case for exclusion might require further independent evidence that the records identified as exceptional really were unusual. In addition, while the CPUE trend derived from a standardization may not be materially changed after a particular data selection, the diagnostics concerning the analysis may be greatly improved.

Potentially less easily defended selections tends to be related more with attempts to focus any analyses on to the main core of a fishery, where the stock dynamics relating to the majority of the stock can be expected to be exhibited. The assumption central to the use of CPUE as an index of relative abundance is that CPUE is directly related to the abundance of the exploitable biomass, with the most common assumption being that CPUE and exploitable biomass are linearly related. Strictly, this would only be the case for a developed fishery in which prior exploitation had depleted the stock down to a more productive state than a naïve unfished stock. In a relatively naïve, lightly fished stock, exceptionally high initial catch rates can be had before the stock is depleted to more highly productive states. MacCall (2009) expressively termed this ‘windfall’ biomass that allows for unsustainable catches to be taken as a stock is fished down, potentially leading to catch rates which would be unrealistically high once sustainable fishing begins.

It is possible to argue that the periphery of a fishery may be more sensitive to exploitation and exhibit more pronounced changes but the periphery is not necessarily where the stock dynamics as a whole is best represented and so, apart from exceptions involving particular circumstances of spatial structuring of the stock, the focus of any analysis should remain in the core of the fishery.

Examples were examined of selecting by acceptable depth range, selecting vessels by how many years they had been reporting from a fishery, selecting vessels by their average annual catch of a species, and finally of selecting records with respect to what were deemed or defined to be the most plausible maximum catch-per-record for targeting to be taking place (selecting for a minimum catch could not be considered an acceptable strategy because of the resulting biases introduced in the outputs). Many other potential factors exist that may influence CPUE and each may have their own criteria for inclusion or exclusion. Whatever the case, each of these criteria can be defensible as being a means of focussing any analysis of the core of the fishery concerned. Here these criteria have been treated independently but in real situations they may be combined. There can be unintended effects when making data selections so even if a plausible defence for making a selection is clear the best argument for the success of a particular data selection strategy is to run analyses with and without the selection in question and explore its actual effects before the resulting time-series is adopted in a stock assessment.

4.4.2 Identifying the Core of a Fishery

The core of a fishery is represented by where the bulk of the available stock occurs and this is assumed to be reflected in where the majority of fishing records derive from (be that in terms of depth, season, location, etc). The intent is to focus on where changes in the bulk of the stock dynamics (its biomass) are most likely to be observed by the fishery through its CPUE. After the basic identification of the fishery of concern, identifying this core of any fishery is generally the primary objective behind most data selection.

If a fishery only occurs away from the major habitats of particular species, perhaps because of unfishable ground, if this is especially marked then the use of CPUE may be compromised as an index of relative abundance for the stock. If this was deemed to be the case then an argument would be required to defend the use of CPUE in an assessment.

4.4.3 The Depth of Fishing

In terms of data selection, the depth of fishing was given particular attention because of its importance in the life of fished species and how that is reflected in its importance when targeting a given species during fishing. The notion of a species having a preferred depth range over which it can best be targeted is simply a reflection of how fish distribute themselves across each continental shelf and slope. Allowing for local variation, as might occur along any coastline, it becomes a sensible strategy for a fisher to exclude certain depth ranges when targeting a certain species (or set of species in a mixed fishery such as the SESSF). This line of argument is the origin of the SESSF assessment group's agreed restrictions on what depths to include in CPUE standardizations used in each assessment. With Flathead and Jackass Morwong the effect of the agreed depth selection, which are often different for different fishing gears, remained minor and made no real changes to the trends. With Pink Ling, however, the selected depth range attempts to split a recognized targeted trawl fishery for Pink Ling from the lesser non-targeted inshore fishery that only catches Pink Ling incidentally to catches of other more targeted species. Additionally, it is reported that the inshore Pink Ling tend to be smaller and presumably younger than those more offshore and deeper. A large impact on the CPUE trend resulted from this depth selection because although about 13% of catches are removed from consideration this also removes about 38% of the records. This is a case where a detailed argument and independent supporting information (e.g. size distribution of inshore fish; different vessels fishing at different times) are provided to defend making such a selection. Even with such a large reduction in records the standardization retains approximately the same shape although the details concerning gradients of change and exactly when changes occurred would affect any stock assessment model that used the outcomes. Once again, without exploring the effects of such data selections the implications would remain unclear, which is the case for both when there is a large effect and when there is in essence no effect.

4.4.4 The Number of Years in the Fishery

When selection criteria were developed for the numbers of years each vessel had been reporting from a fishery, changes to the CPUE trends only began to become visibly noticeable, although still minor, when a minimum requirement of five years in the fishery was applied. For effects to become clearly visible, although still relatively minor, required restricting data to those vessels in the fishery for 10 or even 15 years. Once again the outcome came down to the total number of records remaining in the analysis. While the proportion of vessels remaining after selection dropped down to between 20.6% – 22.5%, the number of records and total catches represented in the data used in the analyses declined far less (losing at most 31% and 37% of records respectively). It would appear that vessels in the fishery for only a relatively short time consequently have fewer records and hence a lower influence. Even so, the trend in CPUE followed by the different vessels in the example fisheries must be similar for so few changes to occur on their removal.

The median number of years for vessels reporting from each fishery was only four or five years in the examples used and other fisheries could, of course, be very different. Whether this would lead to different outcomes could only be determined by exploring the effects of different criteria on different species and fisheries.

4.4.5 The Minimum Annual Catch

Selections based on including only vessels that had more than some defined minimum average annual catch also led to large reductions in the proportion of vessels remaining in the analysis but in the case of Pink Ling could reduce the number of records by 50%. It is not surprising, therefore, that there were clearly visible effects on the CPUE trends when imposing data selection on vessels requiring some of the larger average annual catches (**Figure 14**; see supplementary Figures S10 – S15). It was not predictable how selecting for a minimum average annual catch by vessel would affect the CPUE. With the Flathead and Jackass Morwong data removing those vessels with smaller average annual catches led the CPUE to increase early in the fishery but decrease later. While with Pink Ling increasing the minimum average annual catch per vessel requirement decreased the CPUE in the first 10 years of the fishery but led to increases from 2005 onwards (supplementary Figures S12 – S13). Deciding what constitutes a sensible level of minimum average annual catch is not as straightforward as deciding on some minimum number of years to be in a fishery. The total catch of Flathead is greater than that of Pink Ling, which is greater than that of Jackass Morwong. Their median average annual catch across vessels reflects this at 6.95t, 2.44t, and 1.97t respectively. It is not surprising that requiring a minimum vessel average catch of 10 tonnes has a greater effect in Jackass Morwong and Pink Ling than it does in Flathead. With other factors there would also be a need to calibrate such selection criteria to suit a given species.

4.4.6 Catch per Record

Finally, selections based upon a maximum or minimum catch-per-record are qualitatively different; large catches might be considered as outliers while selecting for some minimum catch might be attempting to focus on shots targeted at some species. Exceptionally large catches do occur but by their nature they are rare, even in much smaller data sets than those used in this study. The inter-annual variation exhibited by the median, and even the 75^{th} quantile, of the catch-per-record for the three species considered was relatively low whereas that for the 99^{th} quantile was much greater (**Figure 7**). Even so, the identification of outlying catches will depend greatly on the fishery concerned. With the three species used here the effect of excluding the larger catches proved to be relatively minor, which is a reflection of the number of records that are excluded. In each case, even where a relatively large proportion of the total catch was excluded (as in Jackass Morwong, which lost up to >15% of all catches) less than 1% of the total records were lost. The resulting effects on the CPUE trends were thus difficult to discern.

The selection on the basis of some minimum catch-per-record differed from the other strategies employed in that it led to consistent and large effects on all species. In each case, the general effect was to flatten any observed trend towards the long term average. Given the positively skewed distribution of catches a strategy of excluding smaller shots, as being unlikely to be targeted at a species, tends to impose hyper-stability on the resulting CPUE time-series and will therefore generate misleading standardizations. The effect was especially marked in Jackass Morwong where there was a long term trend in the number of smaller shots which became hidden when smaller shots were excluded. Even without the presence of underlying trends, strategies involving individual records cannot be recommended given such a fundamental problem.

4.4.7 The Effects of the Example Data sets

Given the large number of observations in the example datasets examined it may not be surprising that the effects of what appear to be relatively severe selection criteria had remarkably little effect on the resulting CPUE trends (except for selecting on a minimum catch-per-record). For species with far fewer records then any selection criteria that eliminates a substantial proportion of those records may well have a large influence on the final time-series of abundance indices. An obvious means of addressing the issue of calibrating any such data selection and defending its effects is to explore the sensitivity of any given data set to such selection strategies and document the findings in detail. This would ensure the repeatability of the analyses and provide a defence capable of withstanding public scrutiny. It is simple to state that this is an ‘obvious’ means of defending any particular selection strategy but, in fact, that this strategy is obvious is negated by the almost total lack of documentation found in the published literature on CPUE standardization but especially the lack in the more informal literature common to stock assessments worldwide. Without such documentation, with detailed specification of what criteria were used and why, then such analyses become almost impossible to reproduce and this adds an unknown degree of extra uncertainty into what are already uncertain biological processes.

Of course, exploring all these options can be time-consuming and there is a universal push to make stock assessments more rapid, cheaper, and more efficient. There is a risk, however, in forcing short-cuts through inadequate time being made available, that inappropriate analyses or data selection practices will go forward without sufficient testing.

4.5 Supplementary Results

Table S1. The number of records and different quantiles for each year's catch-per-shot distributions in kg for Flathead from zones 10 – 20.

Year	records	0%	1%	5%	25%	50%	75%	95%	99%	100%
1986	10546	1	2	5	20	40	96	324	862	3488
1987	8469	1	2	5	25	60	128	450	1024	3500
1988	9568	1	5	10	30	64	150	450	900	9600
1989	9364	1	4	10	30	60	150	500	1000	5000
1990	8402	1	5	10	40	100	200	500	980	2200
1991	8536	1	5	10	35	90	180	480	873	2050
1992	7223	1	5	10	32	75	150	420	800	6000
1993	8982	1	5	10	32	80	150	320	600	2000
1994	10409	1	5	10	30	60	105	270	461	1856
1995	10377	1	4	10	30	60	120	300	532	6000
1996	11150	1	2	10	30	60	100	280	480	1400
1997	10450	1	2	5	30	60	120	313	565	1400
1998	10037	1	2	5	30	60	120	330	636	3000
1999	10446	1	2	10	30	65	150	330	635	2250
2000	13115	1	2	10	30	80	175	400	700	2500
2001	11787	0.5	2	6	30	80	150	320	577	1500
2002	12562	0.5	3	10	35	90	150	330	600	1360
2003	13185	0.5	2	8	30	90	150	360	676	2170
2004	12443	0.5	2	5	30	65	150	330	650	3200
2005	10874	1	3	5	30	60	140	360	680	2500
2006	9284	1	5	10	30	90	180	370	630	2660
2007	6402	1	5	10	50	115	210	510	840	2310
2008	7378	1	5	10	50	120	240	570	1000	3500
2009	6428	0.5	5	10	50	110	200	510	990	4365
2010	6942	1	5	11	50	110	210	500	832	1815
2011	6849	0.3	4	10	50	105	210	496	800	1670
2012	6966	0.6	5	10	50	100	220	525	887	2700
2013	5827	1	5	10	40	90	160	350	550	2100
2014	6430	0.5	5	15	50	100	200	450	700	2255
2015	6469	1	5	20	60	120	200	400	600	1150

Table S2. The number of records and different quantiles for each year's catch-per-shot distributions in kg for Pink Ling from zones 10 – 30.

Year	records	0%	1%	5%	25%	50%	75%	95%	99%	100%
1986	6272	1	2	5	18	40	110	320	560	1330
1987	5501	1	2	5	19	45	120	360	650	1700
1988	5044	1	2	5	20	50	120	310	547	2110
1989	5206	1	2	5	20	50	120	320	571	1500
1990	4163	1	2	5	28	60	150	418	700	9000
1991	4441	1	1	5	25	60	141	320	566	1800
1992	3817	1	2	5	20	55	150	352	694	1500
1993	5173	1	2	5	20	60	150	400	701	3474
1994	6426	1	2	5	20	55	120	330	600	1500
1995	7322	1	2	5	20	50	130	360	720	2000
1996	7308	1	2	4	10	40	150	400	650	2300
1997	8398	1	1	3	15	40	150	360	600	2176
1998	7943	1	1	2	10	40	150	380	605	2000
1999	9062	1	1	2	15	50	150	360	550	1800
2000	8926	1	1	2	10	30	120	300	570	1680
2001	7808	1	1	2	5	30	100	256	450	1100
2002	7330	1	1	1	5	20	70	220	390	1050
2003	7884	0.2	1	1	5	30	90	250	420	1000
2004	6433	1	1	1	5	28	88	264	495	1254
2005	6677	0.3	1	2	6	30	90	231	396	1200
2006	4897	1	1	2	10	33	110	300	528	1500
2007	3525	1	1	2	6	30	99	300	528	1980
2008	4008	1	2	2	11	44	132	330	561	1221
2009	3052	1	1	2	11	35	100	297	505	1320
2010	3176	1	1	2	9	34	127	330	569	6908
2011	3534	1	1	2	8	44	132	318	550	1452
2012	3405	1	1	2	6	33	127	330	550	1690
2013	2652	1	1	2	9	33	106	297	511	900
2014	2652	1	1	2	11	44	132	363	649	1089
2015	2670	1	1	2	11	44	120	280	429	700

Table S3. The number of records and different quantiles for each year's catch-per-shot distributions in kg for Jackass Morwong from zones 10 – 20.

Year	records	0%	1%	5%	25%	50%	75%	95%	99%	100%
1986	5526	1	2	5	20	50	130	500	1429	7000
1987	4569	1	3	5	30	64	200	730	1800	7040
1988	5508	1	4	10	30	76	200	704	1950	7680
1989	4681	1	4	10	30	75	224	800	1700	8960
1990	4690	1	4	10	30	64	160	500	1000	4285
1991	5109	1	2	5	25	60	150	544	1200	7000
1992	3581	1	2	5	20	45	140	480	1005	10000
1993	4258	1	2	5	20	40	120	480	1219	7590
1994	5749	1	2	5	15	40	120	330	650	4500
1995	5239	1	2	5	20	40	100	300	600	5000
1996	6584	1	2	5	15	35	100	300	642	3210
1997	6286	1	1	5	20	50	130	389	780	2400
1998	5026	1	1	3	10	30	100	352	700	3000
1999	4821	1	1	2	10	30	100	340	900	4230
2000	6178	1	1	2	10	30	90	301	785	6000
2001	5255	1	1	2	5	20	60	200	450	1900
2002	6221	1	1	2	5	20	60	210	500	2500
2003	5221	1	1	1	5	16	60	190	418	1800
2004	4828	1	1	2	5	20	60	210	442	1568
2005	5113	1	1	2	10	30	65	220	450	1350
2006	4002	1	1	2	10	30	90	300	600	2640
2007	2724	1	1	2	10	30	100	330	600	2550
2008	3407	1	2	4	12	35	120	350	836	3150
2009	2685	1	1	3	12	32	110	300	600	3750
2010	2734	1	1	2	10	30	90	273	600	2670
2011	2684	1	1	2	10	30	90	270	483	1320
2012	2338	1	1	2	8	30	90	303	619	1600
2013	1513	1	1	2	6	30	90	254	472	1625
2014	1588	1	1	1	5	25	60	180	333	900
2015	1160	1	1	1	3	15	50	150	272	600

Table S4. The effect of selecting for a maximum catch-per-record on catches and numbers of records. The selection entailed retaining only those records with catches less than the maximum listed for each species.

	Flathead	Pink Ling	Jackass Morwong
Maximum catch-per-record	1000	750	1000
Total Catch	34288.658	14680.084	12982.832
Selected Catch	32616.955	14018.257	10982.042
Lost Catch	1671.703	661.827	2000.790
Proportion Kept	0.951	0.955	0.846
Percent Lost	4.875	4.508	15.411
Total Records	276900	164705	129278
Selected Records	275642	164044	128083
Records Lost	1258	661	1195
Proportion Kept	0.995	0.996	0.991
Percent Lost	0.454	0.401	0.924

Table S5. Value ranges found in the within the catch and effort database illustrating the outliers and assumed errors apparent in the data. TW is trawl and DS is Danish seine. The Pink Ling and Jackass Morwong relate to trawl records.

Metric	Flathead TW	Flathead DS	Pink Ling	Jackass Morwong
Effort Minimum hour	-3.65	-14.66	-33	-7.5
Effort Maximum hour	66	23.91	25.51	58.67
Catch Minimum Kg	0.3	0.5	0.2	0.5
Catch Maximum Kg	9600	6620	9000	10000
Vessels	187	54	203	180
Minimum depth m	2	1	4	2
Maximum depth m	2038	1280	6581	1727
Western Longitude	147.997	144.000	147.030	148.000
Eastern Longitude	155.133	150.380	154.283	152.250
Southern Latitude	-40.733	-40.950	-45.650	-40.733
Northern Latitude	-33.600	-37.283	-33.600	-33.600

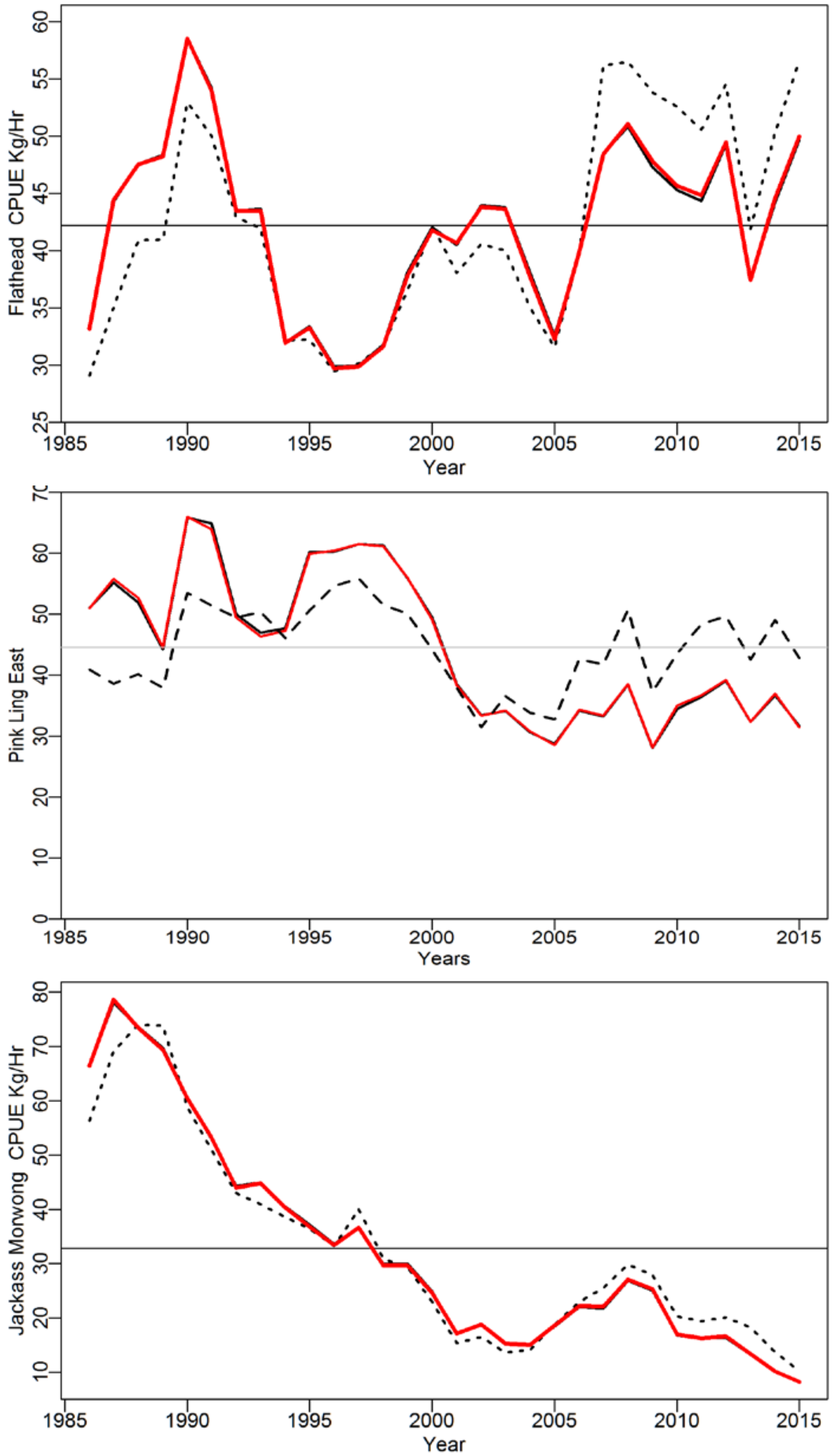

Figure S1. The effect of selecting for a minimum and maximum amount of effort on the CPUE standardizations for Flathead, Pink Ling, and Jackass Morwong. In each case the dotted line is the geometric mean unstandardized CPUE, the red line is the standardized CPUE using only data with >= 1.0 hour and <= 6 hours of trawl effort, and the solid black line (mostly under the red line) is the standardized CPUE for all records with > 0 hours effort. In each case the y-axis does not start at zero in an attempt to make any deviations between the two standardized lines clearer.

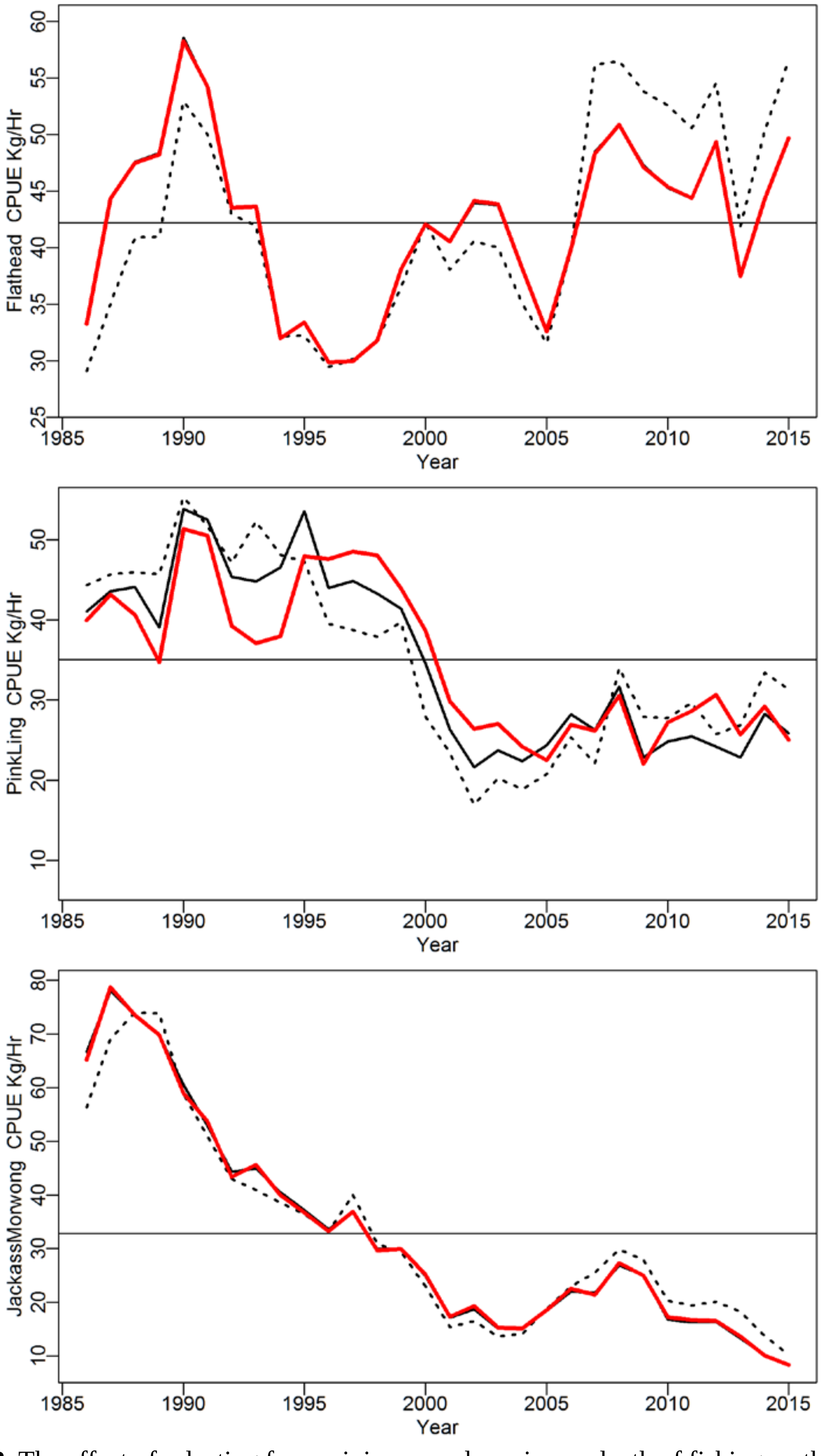

Figure S2. The effect of selecting for a minimum and maximum depth of fishing on the CPUE standardizations for Flathead, Pink Ling, and Jackass Morwong. In each case the dotted line is the geometric mean unstandardized CPUE, the red line is the standardized CPUE using the restricted depth range for each species, and the solid black line (mostly under the red line, except for Pink Ling) is the standardized CPUE for all depths. In each case the y-axis does not start at zero in an attempt to make deviations between the two standardized lines clearer.

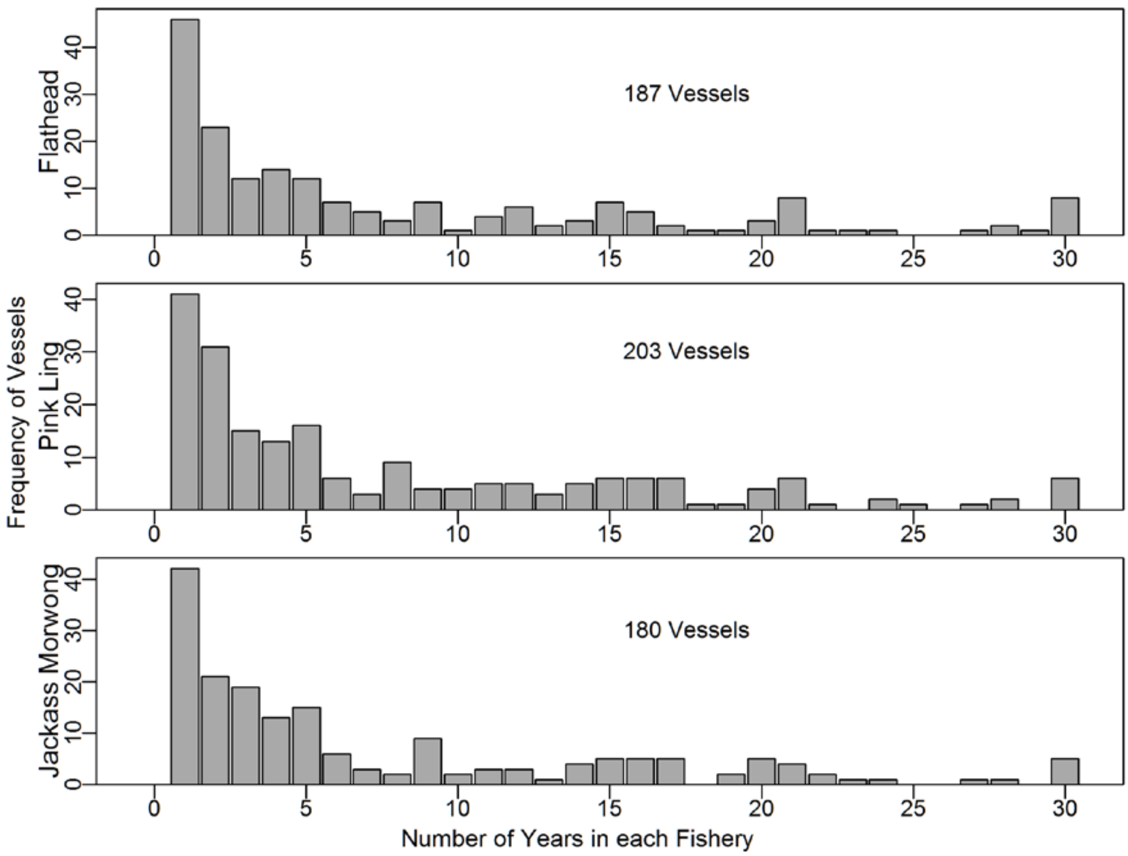

Figure S3. Histograms of the relative frequency of years in which different vessels report from each of the Flathead, Pink Ling, and Jackass Morwong fisheries across the 30 years from 1986 – 2015. The total number of vessels reporting in each fishery across the 30 years is designated in the plot for each species.

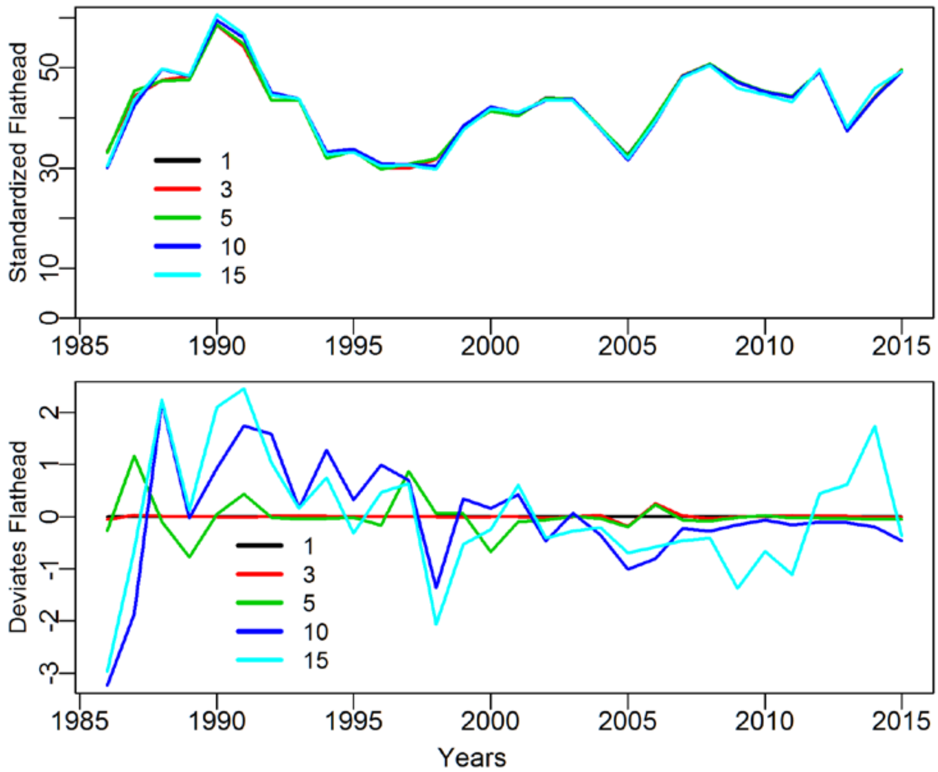

Figure S4. The effect of selecting for a minimum number of years in the fishery on the CPUE standardizations for Flathead. The deviations here are differences between each trend and that for year 1 (all data included).

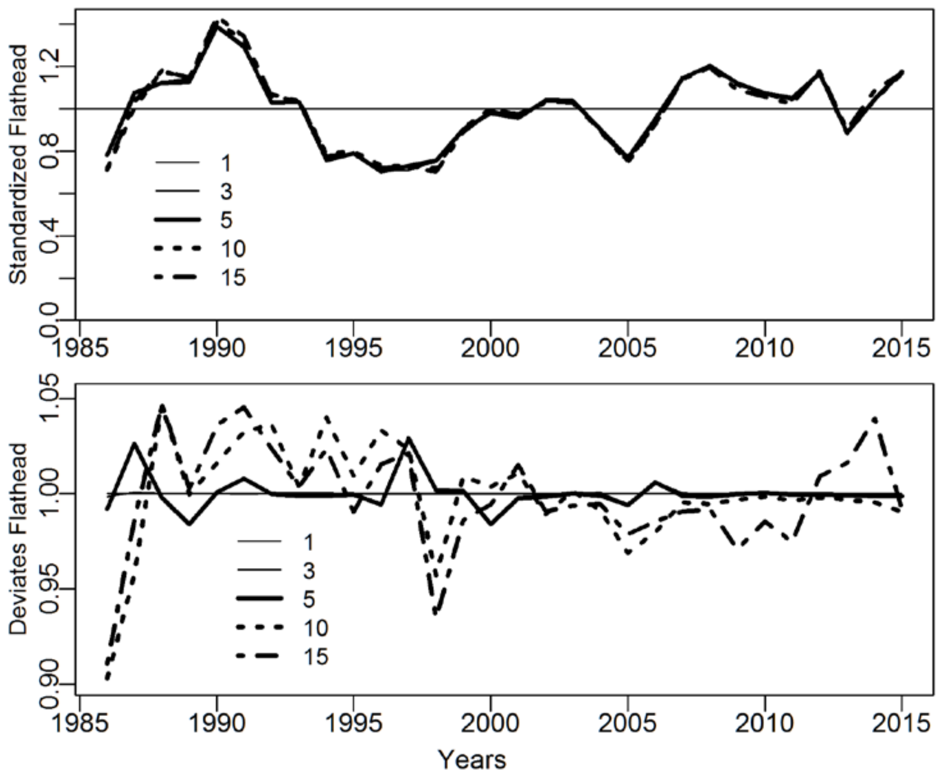

Figure S5. The effect of selecting for a minimum number of years in the fishery on the CPUE standardizations for Flathead with all trend scaled to a mean of 1.0. The deviates, in this case, are ratios relative to the minimum of 1 year (all data) and illustrate the proportional rather than absolute differences.

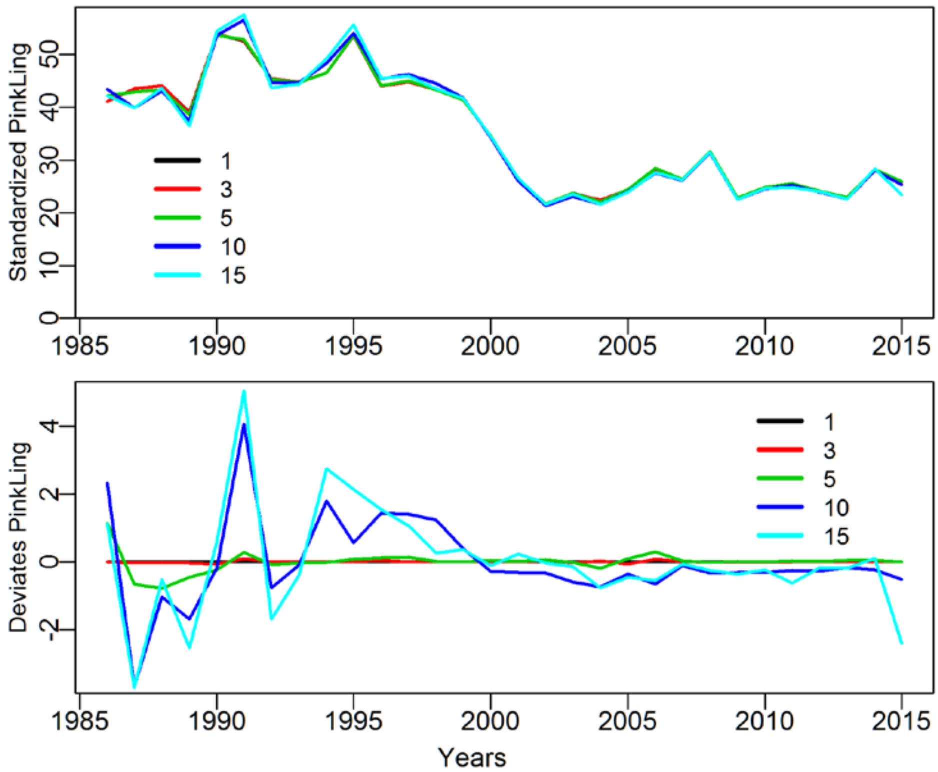

Figure S6. The effect of selecting for a minimum number of years in the fishery on the CPUE standardizations for Pink Ling. The deviations here are differences between each trend and that for year 1 (all data included).

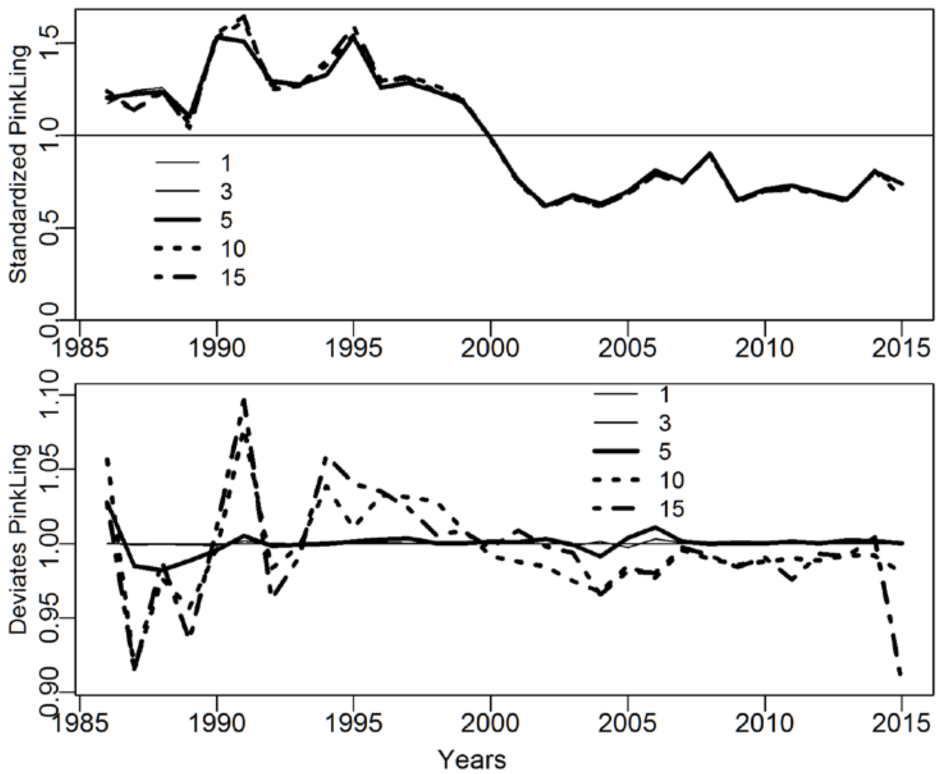

Figure S7. The effect of selecting for a minimum number of years in the fishery on the CPUE standardizations for Pink Ling with all trend scaled to a mean of 1.0. The deviates, in this case, are ratios relative to the minimum of 1 year (all data) and illustrate the proportional rather than absolute differences.

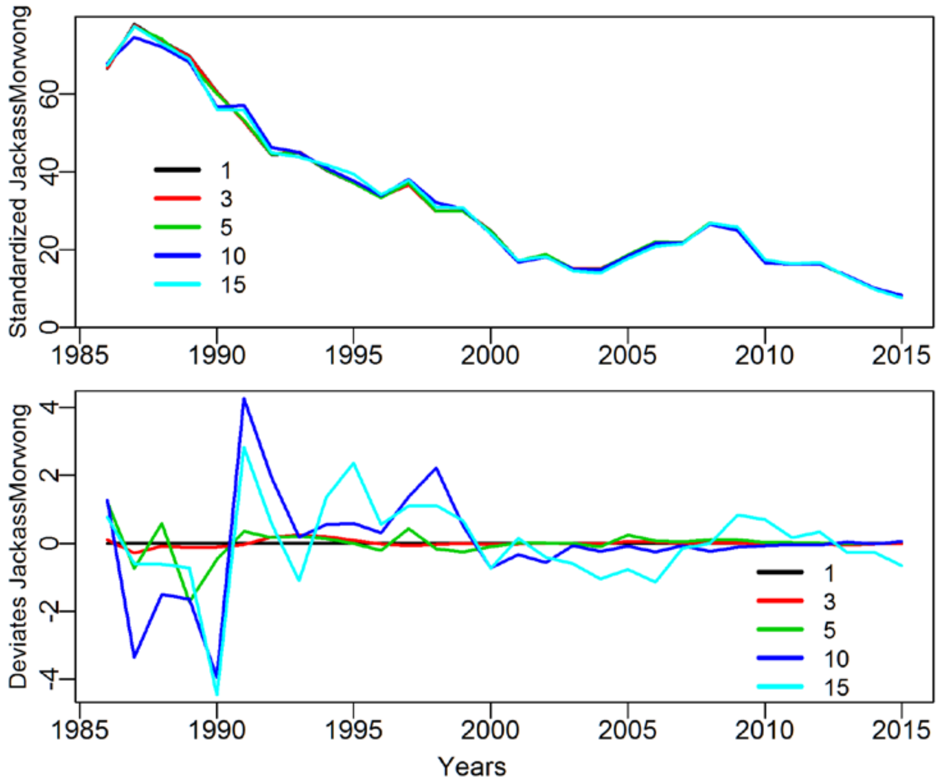

Figure S8. The effect of selecting for a minimum number of years in the fishery on the CPUE standardizations for Jackass Morwong. The deviations here are differences between each trend and that for year 1 (all data included).

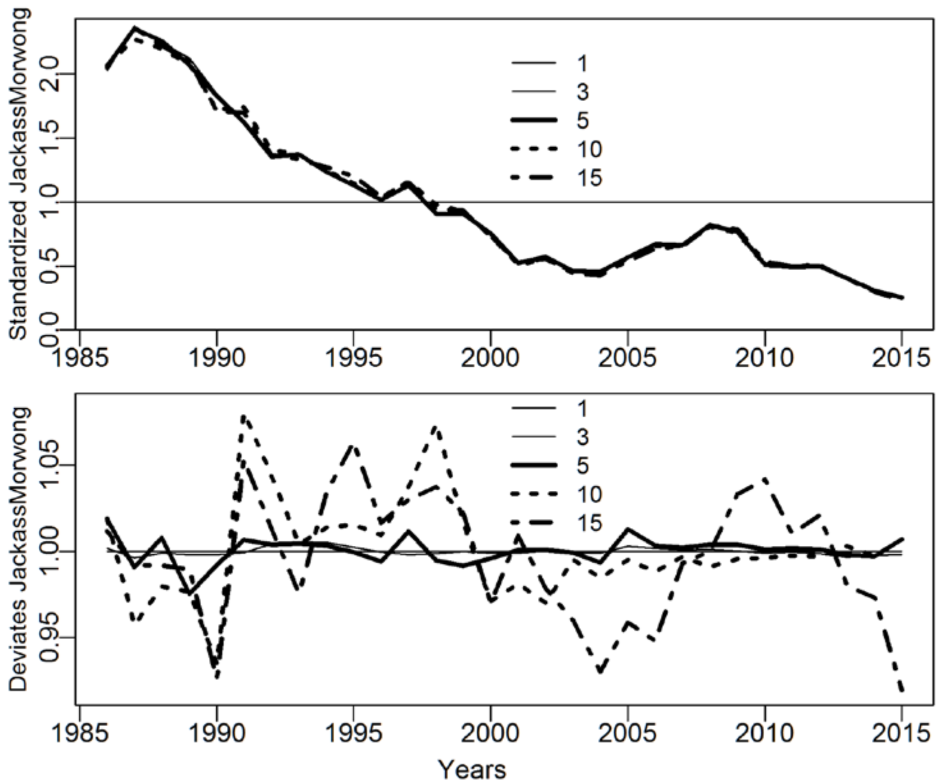

Figure S9. The effect of selecting for a minimum number of years in the fishery on the CPUE standardizations for Jackass Morwong with all trend scaled to a mean of 1.0. The deviates, in this case, are ratios relative to the minimum of 1 year (all data) and illustrate the proportional rather than absolute differences.

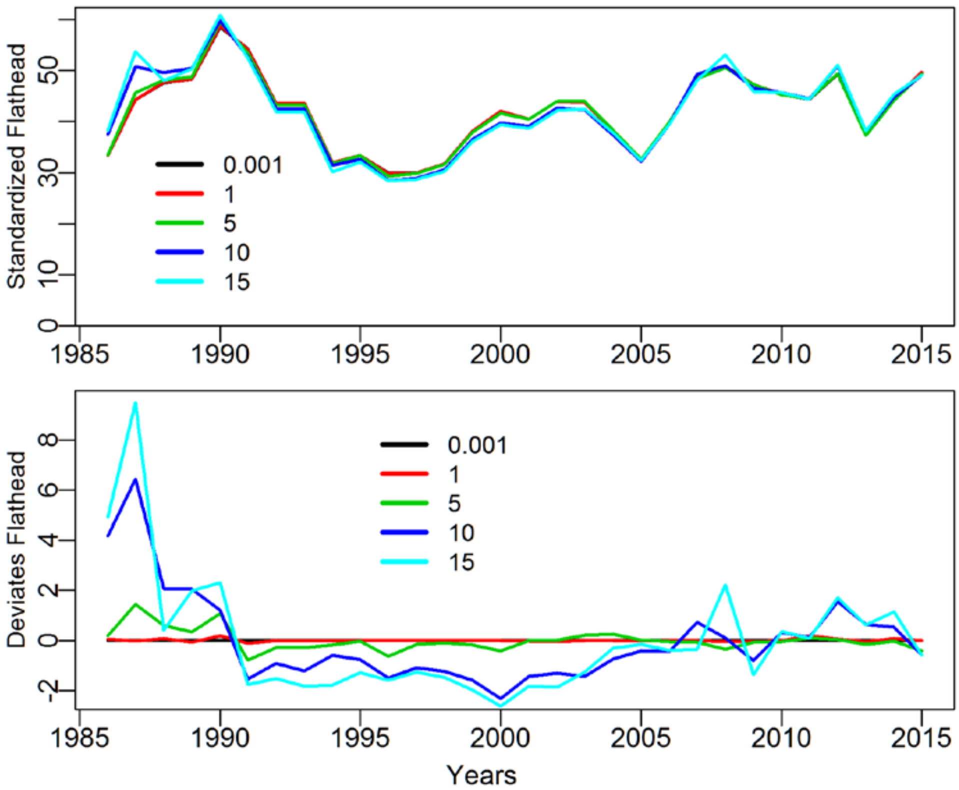

Figure S10. The effect of selecting for a minimum average annual catch per vessel on the CPUE standardizations for Flathead. The average annual catch-per-vessel are in tonnes.

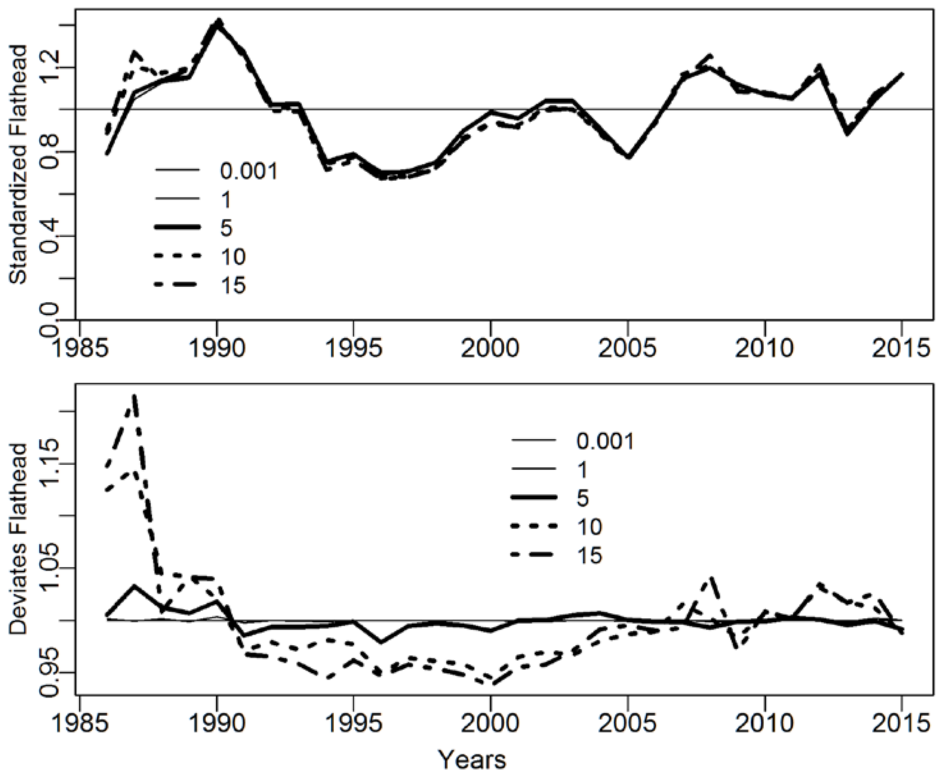

Figure S11. The effect of selecting for a minimum average annual catch per vessel on the CPUE standardizations for Flathead with all trend scaled to a mean of 1.0. The deviates, in this case, are ratios relative to the minimum of 1 kg (all data) and illustrate the proportional rather than absolute differences.

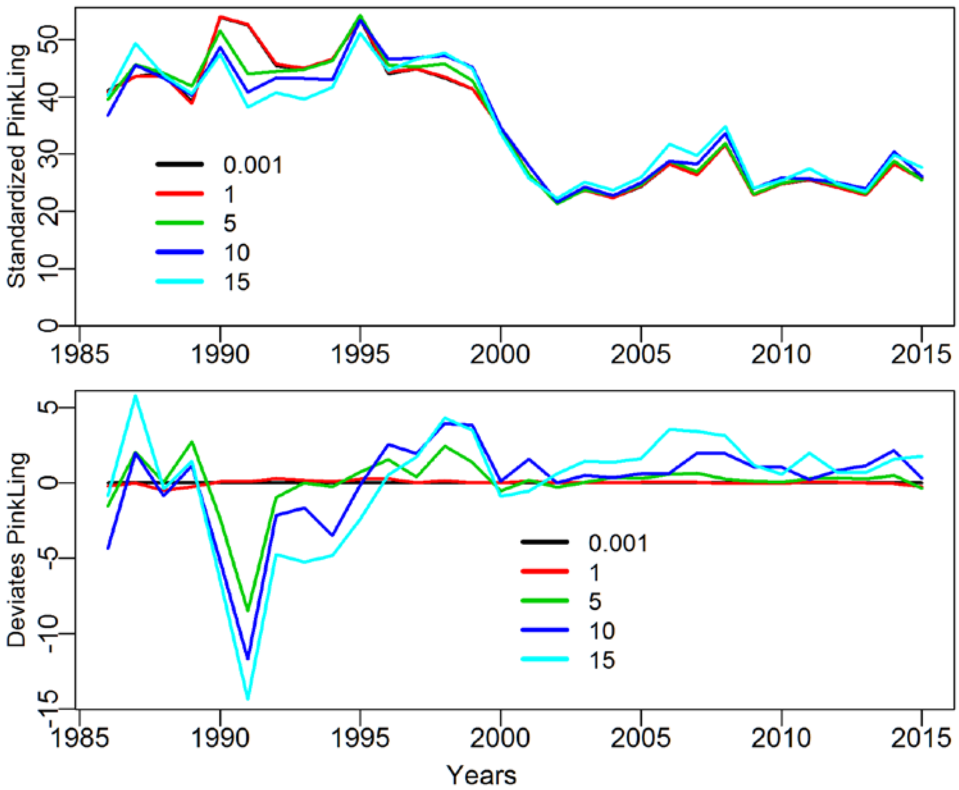

Figure S12. The effect of selecting for a minimum average annual catch per vessel on the CPUE standardizations for Pink Ling. The average annual catch-per-vessel are in tonnes.

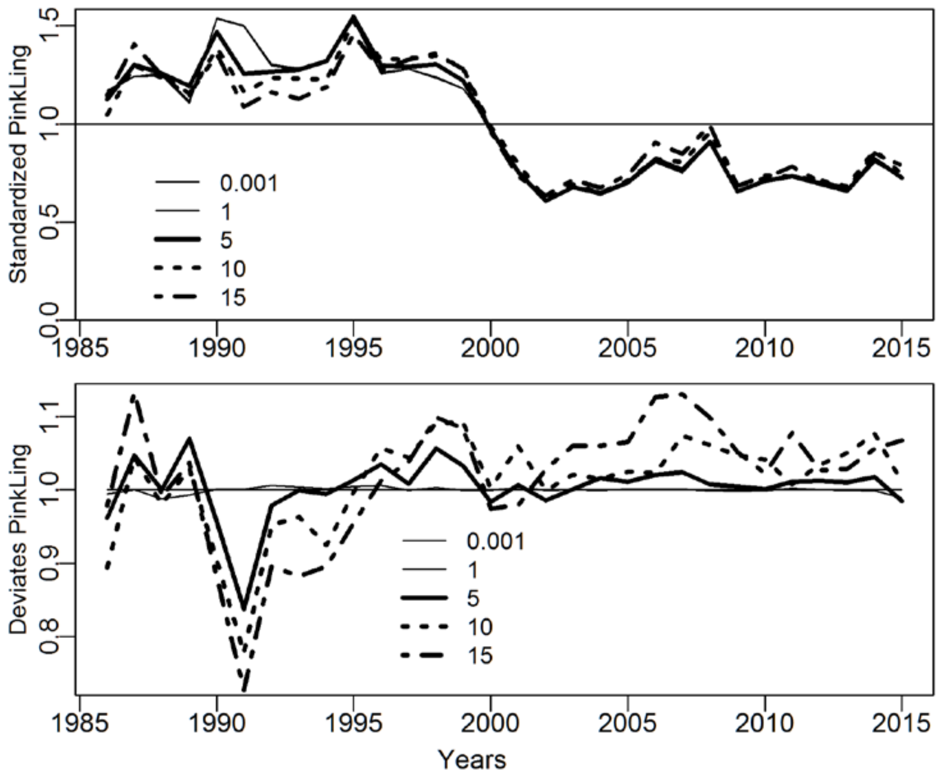

Figure S13. The effect of selecting for a minimum average annual catch per vessel on the CPUE standardizations for Pink Ling with all trend scaled to a mean of 1.0. The deviates, in this case, are ratios relative to the minimum of 1 kg (all data) and illustrate the proportional rather than absolute differences.

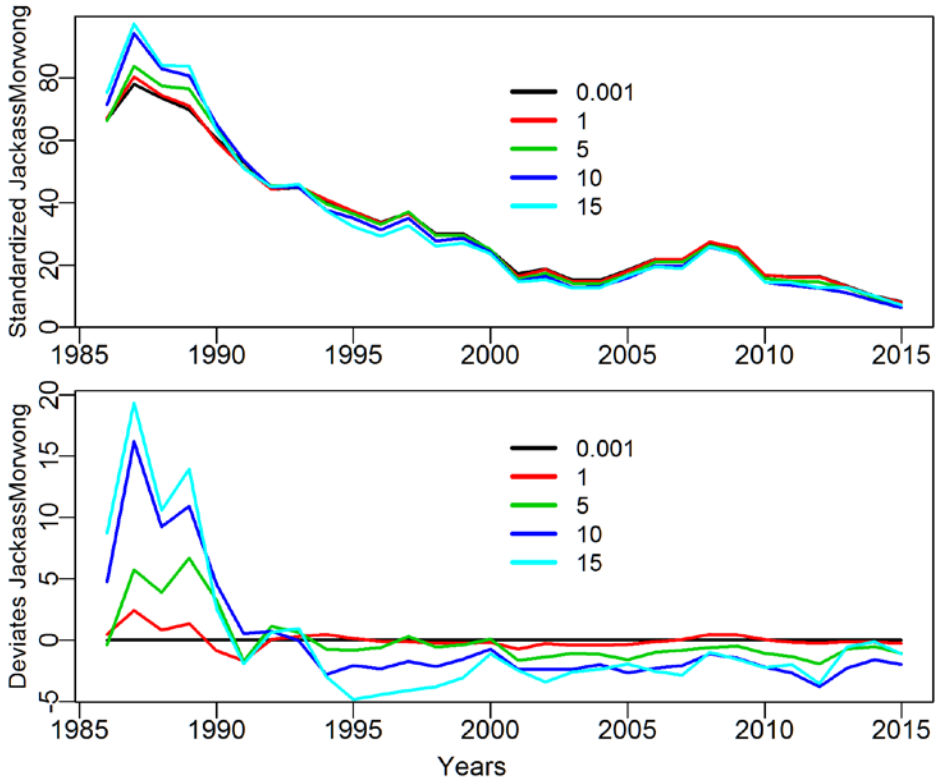

Figure S14. Effect of selecting for a minimum average annual catch per vessel on CPUE standardizations for Jackass Morwong. Average annual catch-per-vessel are in tonnes.

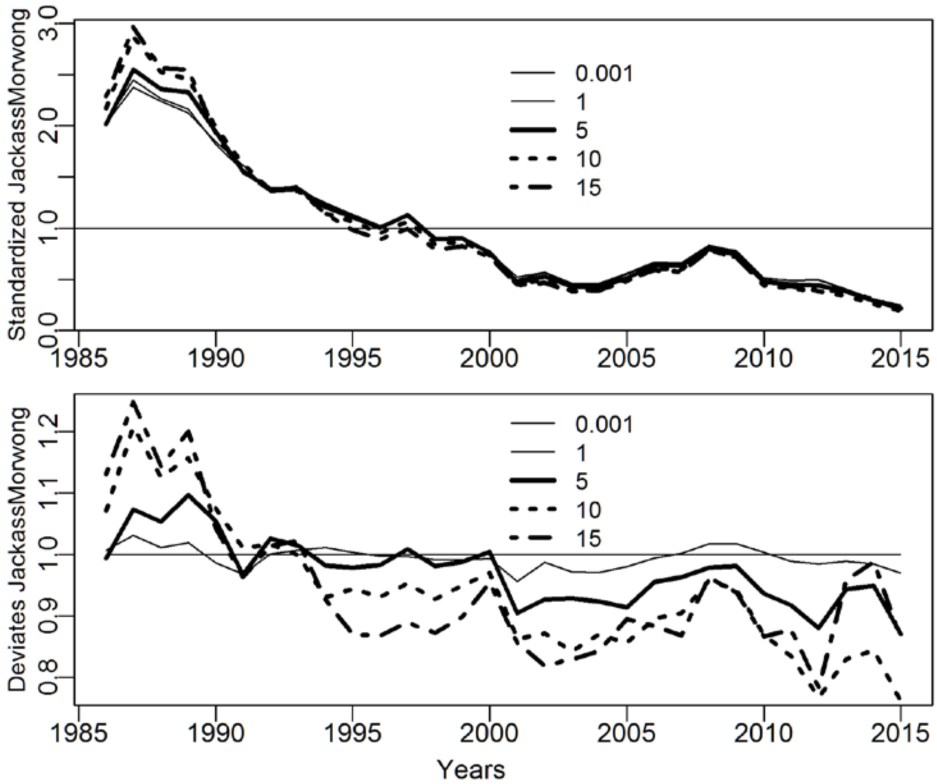

Figure S15. The effect of selecting for a minimum average annual catch per vessel on the CPUE standardizations for Jackass Morwong with all trend scaled to a mean of 1.0. The deviates, in this case, are ratios relative to the minimum of 1 kg (all data) and illustrate the proportional rather than absolute differences.

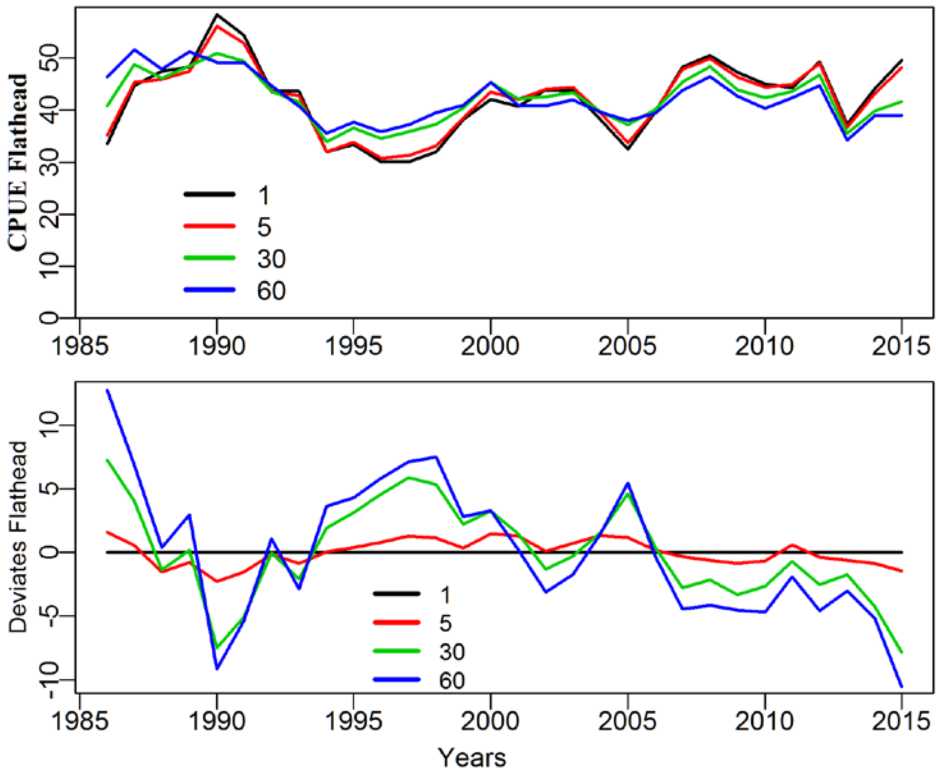

Figure S16. The effect of selecting for a minimum catch-per-record on the CPUE standardizations for Flathead. The deviations here are differences between each trend and that for a catch of 1kg (all data included).

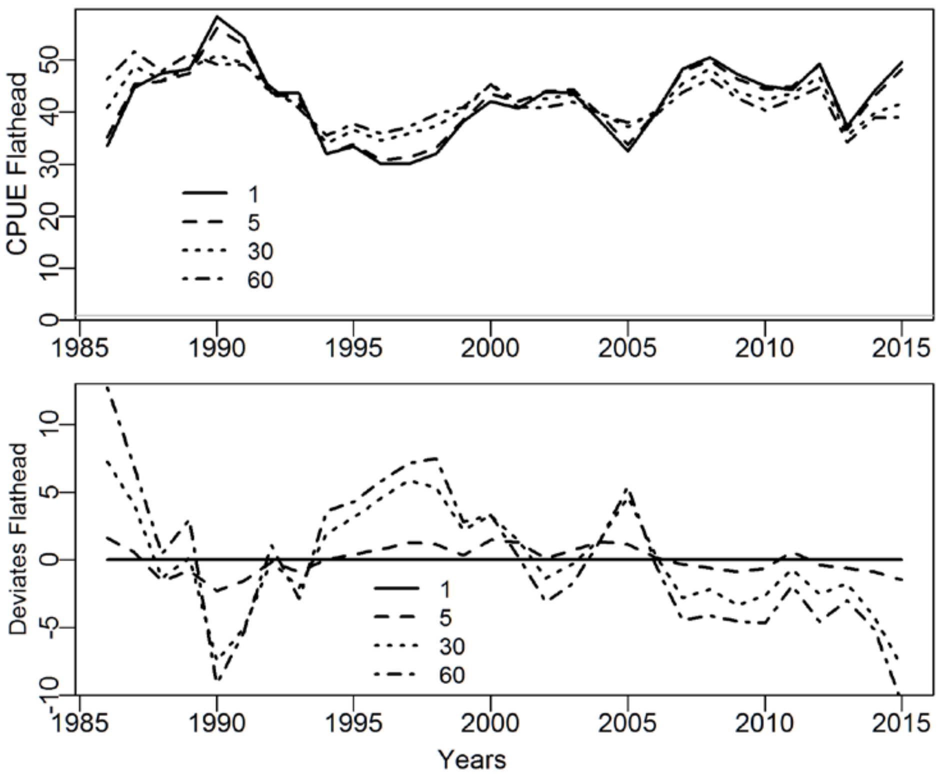

Figure S17. The effect of selecting for a minimum catch-per-record on the CPUE standardizations for Flathead with all trend scaled to a mean of 1.0. The deviates, in this case, are ratios relative to the minimum of 1kg (all data) and illustrate the proportional rather than absolute differences.

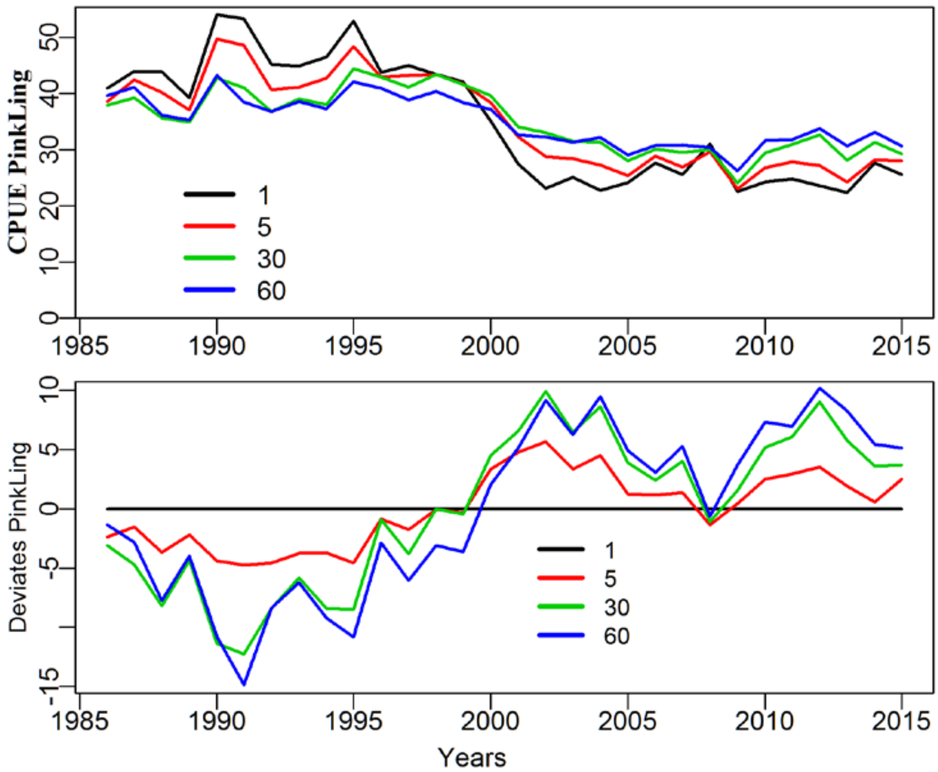

Figure S18. The effect of selecting for a minimum catch-per-record on the CPUE standardizations for Pink Ling. The deviations here are differences between each trend and that for 1 kg (all data included).

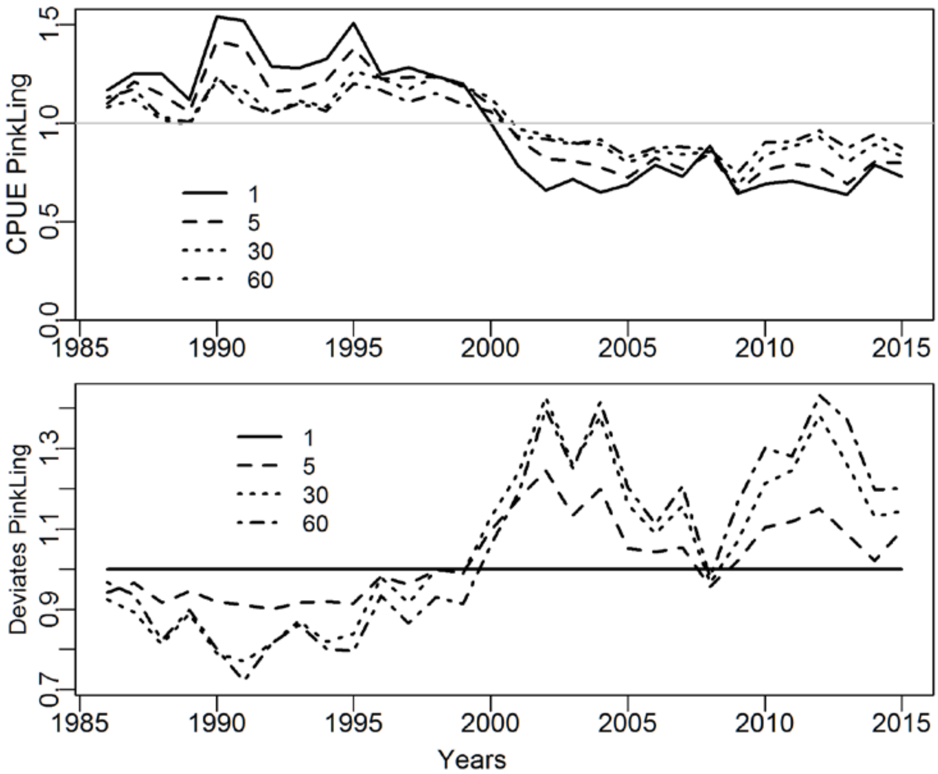

Figure S19. The effect of selecting for a minimum catch-per-record on the CPUE standardizations for Pink Ling with all trend scaled to a mean of 1.0. The deviates, in this case, are ratios relative to the minimum of 1 kg (all data) and illustrate the proportional rather than absolute differences.

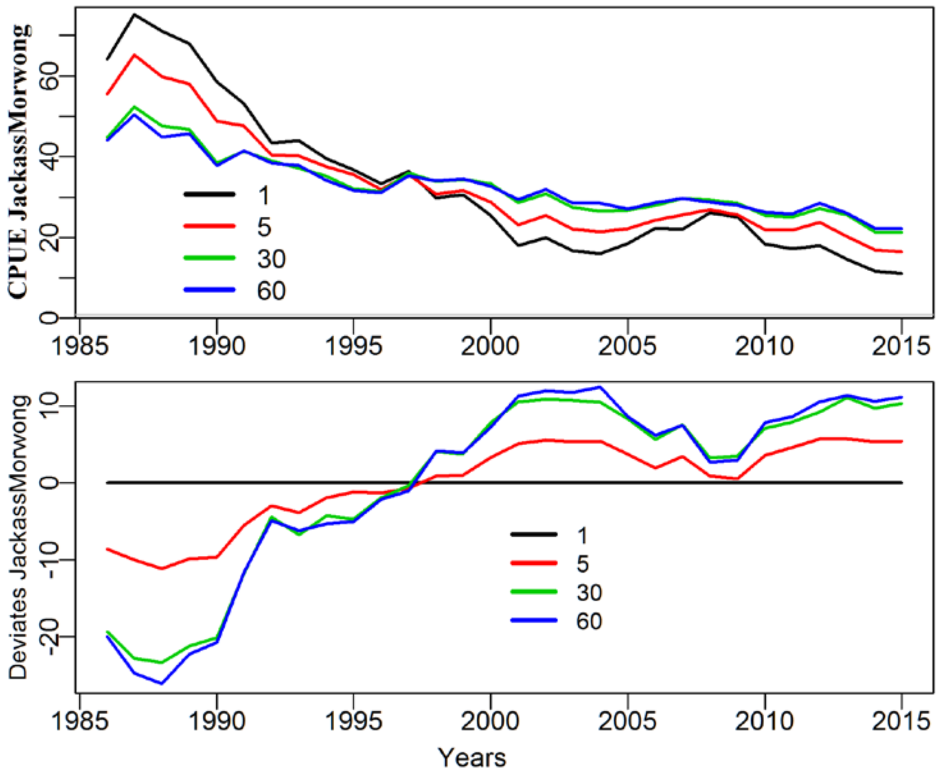

Figure S20. The effect of selecting for a minimum catch-per-record on the CPUE standardizations for Jackass Morwong. The deviations here are differences between each trend and that for 1 kg (all data included). Catches are kilograms.

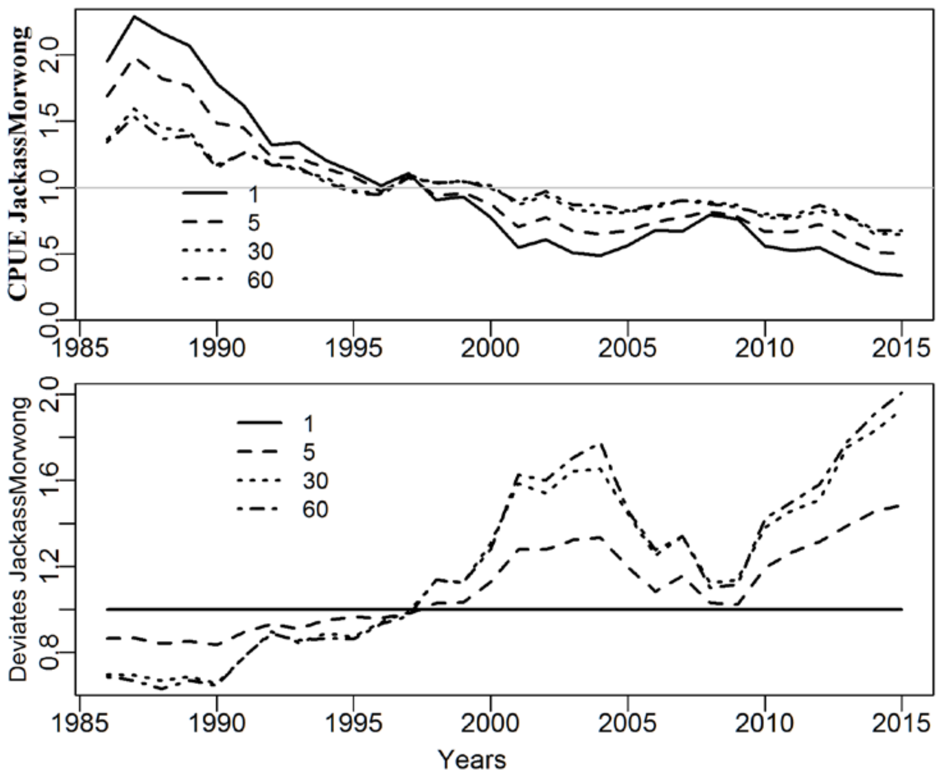

Figure S21. The effect of selecting for a minimum catch-per-record on the CPUE standardizations for Jackass Morwong with all trends scaled to a mean of 1.0. The deviates, in this case, are ratios relative to the minimum of 1 kg (all data) and illustrate the proportional rather than absolute differences. Catches are kilograms.

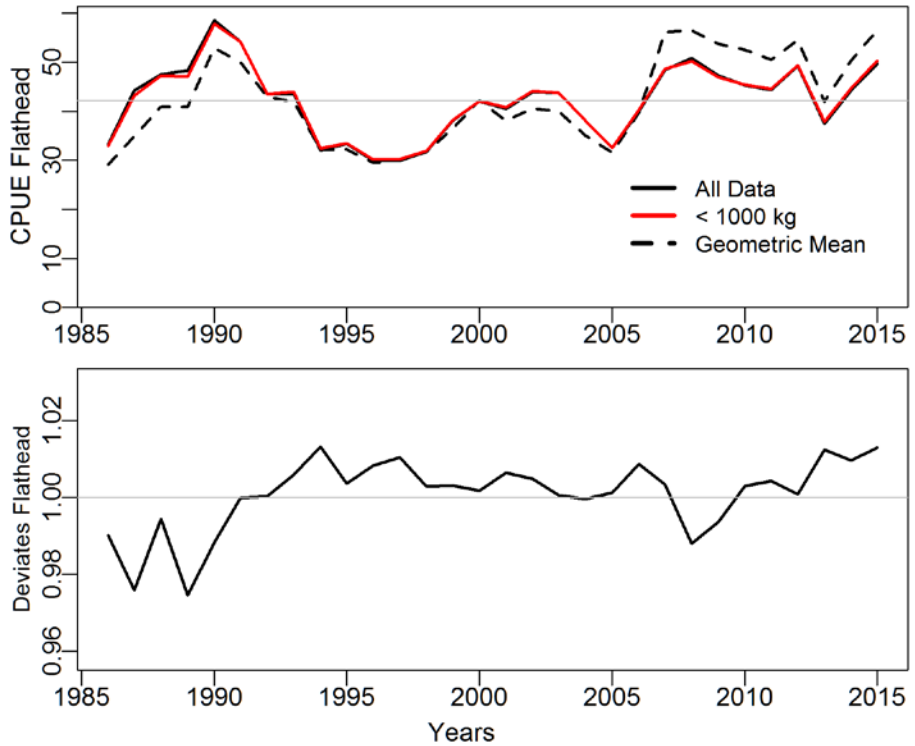

Figure S22. The effect of selecting for a maximum catch-per-record of < 1000kg on the CPUE standardizations for Flathead with all trend scaled to the geometric mean of all the data. The deviates, in this case, are ratios relative to the all data trend line and illustrate the proportional rather than absolute differences.

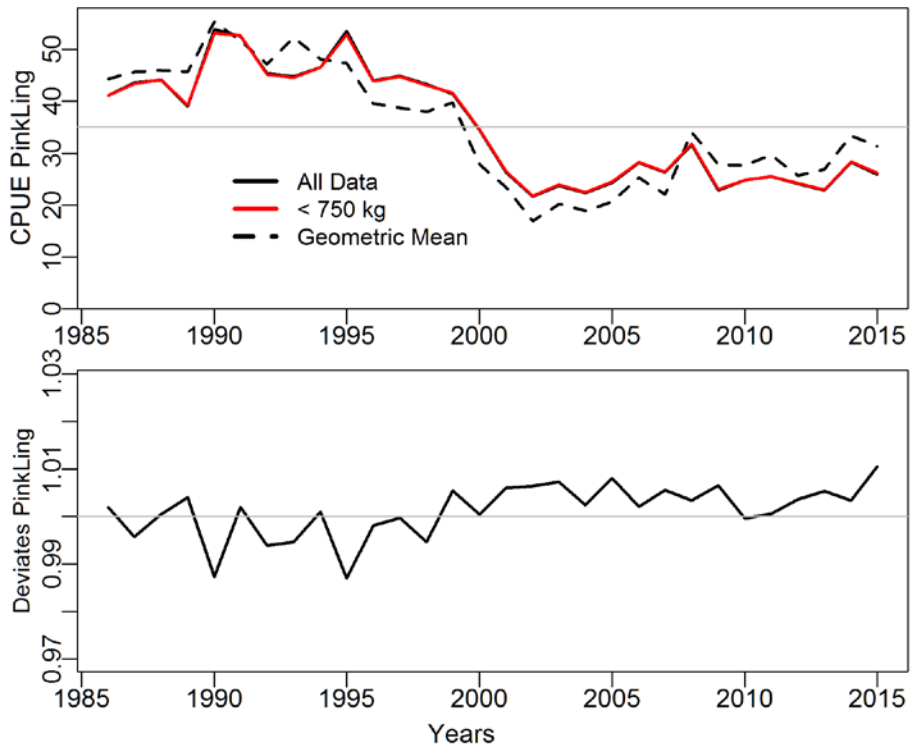

Figure S23. The effect of selecting for a maximum catch-per-record of < 750kg on the CPUE standardizations for Pink Ling with all trend scaled to the geometric mean of all the data. The deviates, in this case, are ratios relative to the all data trend line and illustrate the proportional rather than absolute differences.

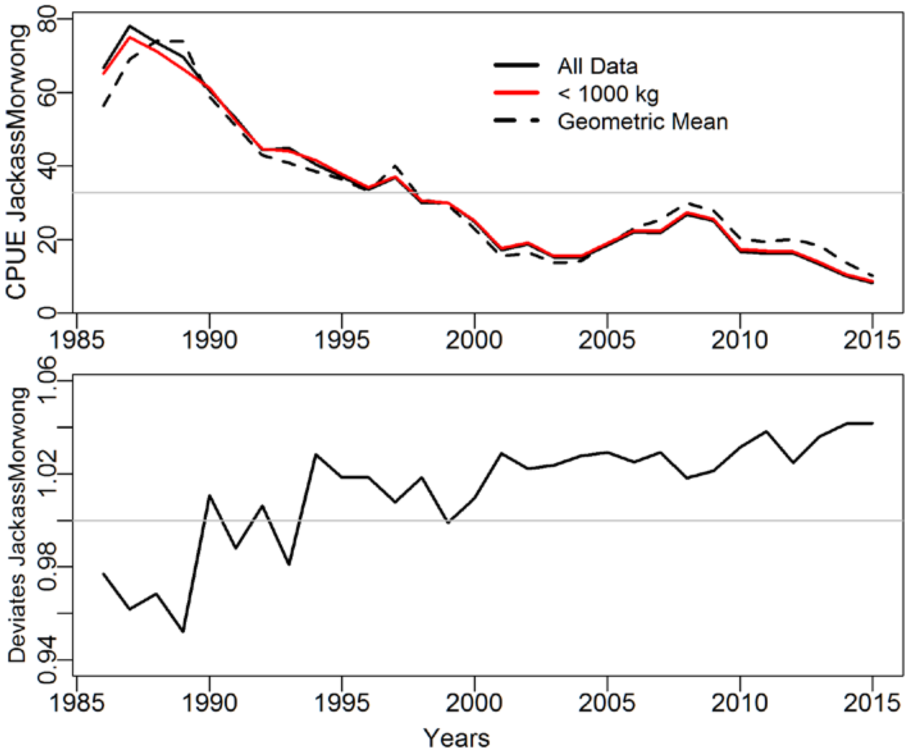

Figure S24. The effect of selecting for a maximum catch-per-record of < 1000kg on the CPUE standardizations for Jackass Morwong with all trend scaled to the geometric mean of all the data. The deviates, in this case, are ratios relative to the all data trend line and illustrate the proportional rather than absolute differences.

4.5.1 Other Data Selection Criteria

The Resource Assessment Group (RAG) tasked with reviewing the assessments for shelf and slope species in the mixed species and mixed gear fishery included within the SESSF agrees on a standard list of data selection criteria and each can be defended (**Table 8**). These are generally of a qualitatively different character to the identification of outliers or errors. Rather, these relate directly to defining exactly what fishery is to be analysed. When the selection criteria for each species or sub-set of each species was being decided the reasons and defences for each section criterion was explicitly included in the annual reports. In one of the latest reports (Sporcic and Haddon, 2016) the selection criteria are included but not the defences. It would be better practice to include the defences in each case explicitly in the methods in each iteration of a document so that it becomes standalone and not dependent on reading an earlier version.

Table 8. The basic constraints over which an initial data selection would be made to a species within the SESSF fishery.

Constraint	Description
Species concerned	A mixed species, mixed fishing gear fishery, so focussing only on records with a particular species is important. Identifying zero shots is especially difficult.
1986 - latest Year	Trawl records first became shot-by-shot from 1986 in the AFMA database, for the Gillnet, Hook, and Trap fishery the equivalent date in 1997. However, some of the earlier records contain what appear to be extreme values and may still be monthly summaries as were collected prior to 1986.
Fishing method	With a mixture of paper logbooks and eLogs there are now multiple codes for the same gear in the database (e.g. TW and TDO for Trawl and Trawl Demersal Otter)
Depths	The depth range selected by the RAG is designed to focus on the major fishing depths and avoid the proliferation of empty cells within categorical variables.
Zones	The SESSF extends over 1000's of km of coastline across which spatial differences occur. The SESSF zones are relatively coarse but analyses are focused within specific areas to avoid too many empty shots.
Fishery	There are numerous fisheries operating around Australia's coastline. Identifying a specific fishery (e.g. SET - South-East Trawl) avoids spurious records from remote fisheries or High Seas fisheries.

5 Fishery Dependent versus Independent Data

5.1 Introduction

The primary objective of most fishery independent surveys of fish stocks is to provide an index (or indices) of relative abundance for at least one target species. In Australia, which could be characterized as having mainly mixed target fisheries (Orange Roughy, *Hoplostethus atlanticus*, is an exception), there are currently Fishery Independent trawl surveys (FIS) in the Great Australian Bight (GAB), in the South East Scalefish and Shark Fishery (SESSF), and in the Northern Prawn fishery (NPF); there are others not considered here, for example, in the Torres Straits, and in various scallop fisheries. Here, comparisons will be made between the standardized commercial CPUE and the FIS outcomes for an array of commercially important species in the GAB and the SESSF. These surveys have recently been reviewed (O'Driscoll and Doonan, 2015a, 2015b) with generally positive outcomes and some constructive suggestions.

The 'Fishery Independent' (FI) aspect of such surveys receives emphasis because this is what sets such surveys apart from Fishery Dependent (FD) data (such as CPUE) and carries with it the implication that data collected in a FI survey is based on a design that aims to obtain a representative sample from the target stock. For this reason, such surveys are usually considered to set a higher standard and provide the best view of how a stock is doing. CPUE data can have many and varied flaws (see **Chapter 4 Potential Issues when using CPUE Data**) whereas FI surveys use standard gear and follow a strict sampling design. Importantly, the use of standard gear and methods (and usually vessels) avoids the problem of 'effort creep' increasing the effectiveness of the survey effort and changing the catchability through time. A survey approach, however, cannot avoid the impact of a directional change in the climate influencing catchability or other outside influences altering productivity. In the Commonwealth Trawl Sector (CTS) and the Great Australian Bight (GAB) fishery the surveys are conducted using methods similar to how the fisheries operate and use commercial vessels (although with standard gear) from those fleets. The survey results are thus directly comparable with the CPUE from the fisheries concerned (the surveys in the NPF, on the other hand, occur at times of each year when commercial fishing is not occurring). Here the outcomes from each of the CTS and GAB surveys to date are directly compared with the standardized CPUE from a selection of the same species with the focus being on the most important species in terms of relative abundance and value. This will achieve the second objective for the project:

5.1.1 Project Objective 2

Compare alternative catch rate standardization strategies in those fisheries where both fishery independent and fishery dependent data are available.

An index of abundance from a FIS can be used as a standardized index of abundance, so a direct comparison with a statistical standardization of commercial CPUE is comparing like with like.

5.2 Methods

5.2.1 South East Australia

In the South East Scalefish and Shark Fishery (SESSF) Australian region a fishery-independent trawl survey has been repeated during the winters of 2008, 2010, 2012, 2014, and 2016 (Knuckey et al. 2017). Originally this survey occurred in both summer and winter but an examination of the results indicated that estimates from the winter surveys provided generally more precise estimates for a wider range of species. To reduce survey costs it was then agreed to carry out only winter surveys despite the analytical design requiring both. Therefore, for this report we have included only the five existing winter survey results for comparison with matched CPUE time series.

To display the survey estimates as well as an indication of the uncertainty associated with each estimate, error ranges were constructed for the survey estimates by assuming a log-normal distribution, and indicative 95% confidence intervals around annual mean values as +/- 1.96 × standard deviation:

$$\begin{aligned} L95 &= \overline{x}_s \times e^{-1.96 \times StErr_s} \\ U95 &= \overline{x}_s \times e^{1.96 \times StErr_s} \end{aligned} \tag{6}$$

where $\overline{x}_s$ is the mean abundance estimate for species *s*, and $StErr_s = \overline{x}_s \times cv_s$, where cv_s is the coefficient of variation for the species *s*. Species were chosen for examination based on whether good quality survey results were available (at least one year with a CV estimate < 0.3), and also whether a standardised CPUE was also available for that species from across at least the same area as covered by the trawl survey (although the survey sometimes covered more area than the fishery and CPUE; e.g. Flathead is primarily caught in SESSF zones 10, 20, and 30, but the survey analysis combined 10 – 50 and some way into the GAB).

Since May 2008 the fishing year in the SESSF runs from May through to April each year, however, before that the fishing year ran from January through to December. To reflect the majority of commercial catch and effort data, standardised abundance indices continue to relate to calendar year for all major quota species in SE Australia; these indices are provided recently by Sporcic and Haddon (2016). To compare standardised CPUE with FIS survey estimates, it was important to choose those indices that best matched in spatial coverage of the trawl survey. As the SE FIS covered both east and west of Tasmania for most species, the standardised CPUE covering both areas was chosen for comparison in most cases (**Table 9**). The areas (or SESSF zones) chosen in the CPUE analyses reflect the stock assessments applied to each species. Thus, for example, in Flathead separate analyses are conducted for Zones 10 – 20, and for zone 30 (eastern Tasmania) because the character of the fisheries in those different zones are very different (e.g. the fish are generally larger next to Tasmania; zone 30). Leaving out such diversity would weaken the stock assessment's capacity to draw valid conclusions.

5.2.2 Comparison between Indices

For all comparisons among different abundance indices, each index was normalised by dividing by the arithmetic average of each series across the years which were shared between all time-series (this entailed using the years 2008, 2010, 2012, and 2014 because the CPUE standardization for 2016 was not then available); this leads to the average for each series across the years being compared being one, which places emphasis on any trends in each time-series rather than their absolute values. Although some series may provide absolute estimates with some level of meaning, for comparative purposes here, all were treated as relative annual indices only.

The scaling of each time series of CPUE, I_y, to produce yearly parameters, P_y, involves:

$$P_y = \frac{I_y}{\bar{I}} \quad \text{where} \quad \bar{I} = \sum_{y=1}^{N} I_y / N \tag{7}$$

where, in the GAB, y is the set of fishing years 2004/2005 – 2008/2009, 2010/2011, and 2014/2015 (there was no survey in 2009/2010), and N = 7, the number of years (or N = 4 and years are 2008, 2010, 2012, and 2014 for the SESSF/CTS). For plotting purposes in the trawl survey series the interpolated values inserted for any missing years is the average of the values immediately before and after, so as to provide a guidance line between the available data, but all comparisons are only made relative to the years in which each time series is represented with real data.

A key property of interest to any stock assessment is not the within-year precision but the between-years variation and consistency. For comparison between each time series, *S*, the sum of the absolute difference (*SAD*) of each value from the expected mean of 1.0 across the *N* years (*y*) 2004/2005 – 2008/2009, 2010/2011, and 2014/2015 for the GAB and for the years 2008, 2010, 2012, and 2014 for the SESSF species considered.

$$SAD_S = \sum_{y=1}^{N} abs\left(1 - I_{S,y}\right) \tag{8}$$

For each fishery these *SAD*s were calculated for the geometric mean, the standardized CPUE index, and the trawl survey index. In addition, the number of the CPUE points that overlapped with the 95% confidence intervals around the survey points was tabulated and the congruence between the two general trends assessed visually. In the SESSF the last FIS index was made in the 2016 calendar year but the standardized CPUE was not then available for comparison. In those cases, to score whether overlap would occur, the line from 2014 and 2015 was extended linearly and if it would then overlap it was scored appropriately.

The two main species from the GAB fishery and 16 species from the SESSF fishery had their standardized CPUE compared with the mean abundance indices from their respective FIS (**Table 9**).

Table 9. Species chosen for comparison between 16 species for the SESSF FIS (Knuckey et al., 2017) and two species for the GAB FIS (Knuckey et al., 2015) data and the standardized commercial CPUE (Sporcic and Haddon, 2016) based on spatial area coverage.

Fishery	Species		Standardisation
GAB	Deepwater Flathead	*Platycephalus conatus*	GAB zones
GAB	Bight Redfish	*Centroberyx gerrardi*	GAB zones
SESSF	Tiger Flathead	*Neoplatycephalus richardsoni*	Zones 10-20 (main fishery)
SESSF	Pink Ling	*Genypterus blacoides*	Zones 10-30 (main fishery)
SESSF	Silver Warehou	*Seriolella punctata*	Zones 10-50
SESSF	John Dory	*Zeus faber*	Zones 10-20 (main fishery)
SESSF	Ocean Perch	*Helicolenus percoides*	Zones 10-20
SESSF	Jackass Morwong	*Nemadactylus macropterus*	Zones 10-50
SESSF	Mirror Dory	*Zenopsis nebulosus*	Zones 10-50
SESSF	Blue Grenadier	*Macruronus novaezelandiae*	Zones 10-50 (Non-Spawning)
SESSF	Redfish	*Centroberyx affinis*	Zone 10 (main fishery)
SESSF	Western Gemfish	*Rexea solandri*	Zones 50-40 (north of -42° S)
SESSF	Blue Warehou	*Seriolella brama*	Zones 10-50
SESSF	Royal Red Prawn	*Haliporoides sibogae*	Zone 10
SESSF	Ribaldo	*Mora moro*	Zones 10-50
SESSF	Blue-Eye Trevalla	*Hyperoglyphe antarctica*	Zones 20-50
SESSF	Eastern Gemfish	*Rexea solandri*	Zones 10-30 (Non-Spawning)
SESSF	Silver Trevally	*Pseudocaranx dentax*	Zones 10-20

5.3 Results

5.3.1 Great Australian Bight

The CPUE standardization for Deepwater Flathead does not change the inter-annual trends in the nominal CPUE in any major way, although for Bight Redfish, since 2010/2011, while the nominal trend has been increasing slightly the standardized trend has been declining slightly, so the two trends are beginning to diverge (**Figure 16**).

The trawl survey indices for Deepwater Flathead are flatter than those for the commercial CPUE as evidenced by the lower SAD value exhibited by the survey trend relative to the CPUE trends (**Table 10**). Even so, the general trends in the two main time series (standardized CPUE and trawl survey) for Deepwater Flathead confirm each other, with the survey exhibiting about the same degree of variation across all values as does the CPUE (**Figure 16**).

The opposite is the case for Bight Redfish where the inter-annual differences across the survey trend are much larger than those exhibited by either CPUE series (**Figure 16**).

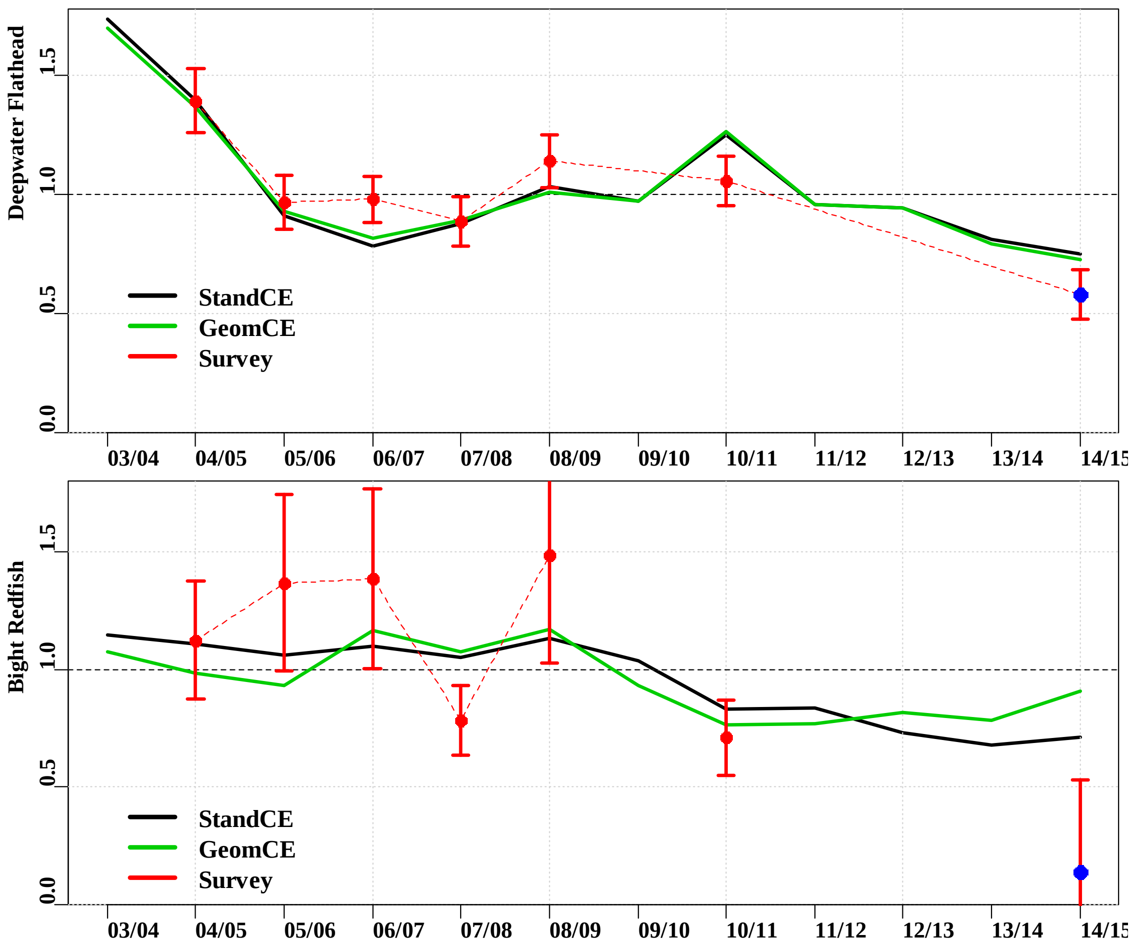

Figure 16. A comparison of the indices from the standardized commercial CPUE and the trawl survey indices for Deepwater Flathead (*Platycephalus conatus*) and Bight Redfish (*Centroberyx gerrardi*) from the GAB. The red lines represent ± 1.96 × StDev in each year for the FIS mean estimates (see **Table 15** and **Table 16**). GeomCE is the scaled geometric mean CPUE; each time series has been scaled to have a mean of 1.0 across years 2004/2005 – 2008/2009, 2010/2011, and 2014/2015. The horizontal black dotted line at 1.0 is the reference against which the sum of absolute differences were calculated.

The differences in the Bight Redfish's time series have been exacerbated by the 2015 survey, the efficacy of which may have been adversely affected by a coincident seismic survey. Over the general period of the survey the commercial CPUE also exhibited a remarkable reduction (**Table 11**; **Figure 17**). To balance the very low latest value to generate an average of 1.0 for the series, the earlier estimates are thus elevated slightly higher above the average of 1.0. The standardized Bight Redfish time-series suggests a recent minor downturn but not as great as suggested by the upper 95% confidence interval of the latest survey. Even before the 2015 survey, however, the indices for Bight Redfish were already more variable inter-annually than the CPUE time series (**Figure 16**; **Table 10**, **Table 15**, and **Table 16**), which is why they succeed in overlapping the CPUE series as often as they do. Even though the number of shared points for Deepwater Flathead and Bight Redfish are the same this is only because variation around the Bight Redfish estimates is so large.

Table 10. The sum of absolute difference (SAD) scores for each time series considered; see equation (2). For the trends to be similar it is necessary, but not sufficient, for their SAD scores to be similar. The lowest SAD score is bolded in each case implying that series deviates least from the expected average of 1.0; the geometric mean is least in 6 instances, the standardized CPUE in 9 instances and the Survey in 2 instances. 'Mostly Similar' means that the mean estimates match up in places but important differences remain. 'Similar but Noisy' means the general trends are similar but the survey results are so variable the actual trend is poorly defined. Shared implies the number of shared years between the CPUE series and the survey years.

Species	Geomean	CPUE	Survey	Trend	Shared
Deepwater Flathead	1.271	1.362	**1.181**	Similar	4/7
Bight Redfish	**0.828**	0.909	2.730	Dissimilar	4/7
Tiger Flathead	0.308	**0.239**	0.810	Some matches	4/5
Pink Ling	0.253	**0.154**	0.239	Similar	4/5
Silver Warehou	**0.787**	0.804	2.363	Dissimilar	2/5
John Dory	1.005	0.997	**0.914**	Similar but Noisy	5/5
Ocean Perch	0.224	**0.187**	2.738	Dissimilar	2/5
Jackass Morwong	**0.998**	1.023	1.517	Similar	3/5
Mirror Dory	**0.505**	0.816	2.782	Dissimilar	2/5
Blue Grenadier	0.952	**0.812**	1.721	Similar but Noisy	4/5
Redfish	**0.769**	0.882	1.937	Similar but Noisy	3/5
Western Gemfish	0.714	**0.438**	1.955	Similar but Noisy	4/5
Blue Warehou	1.224	**1.110**	2.988	Dissimilar	2/5
Royal Red Prawn	0.510	**0.502**	4.244	Dissimilar	2/5
Ribaldo	0.601	**0.418**	1.430	Similar but Noisy	5/5
Blue-Eye Trevalla	1.143	0.973	2.036	Some matches	4/5
Eastern Gemfish	1.008	**0.953**	2.315	Dissimilar	4/5
Silver Trevally	**0.492**	0.739	3.290	Dissimilar	3/5

The outcome of the comparison in the GAB between the FIS and the commercial CPUE is thus mixed with the survey closely following the Deepwater Flathead CPUE except in the last year, which appears to have been exceptional for external reasons.

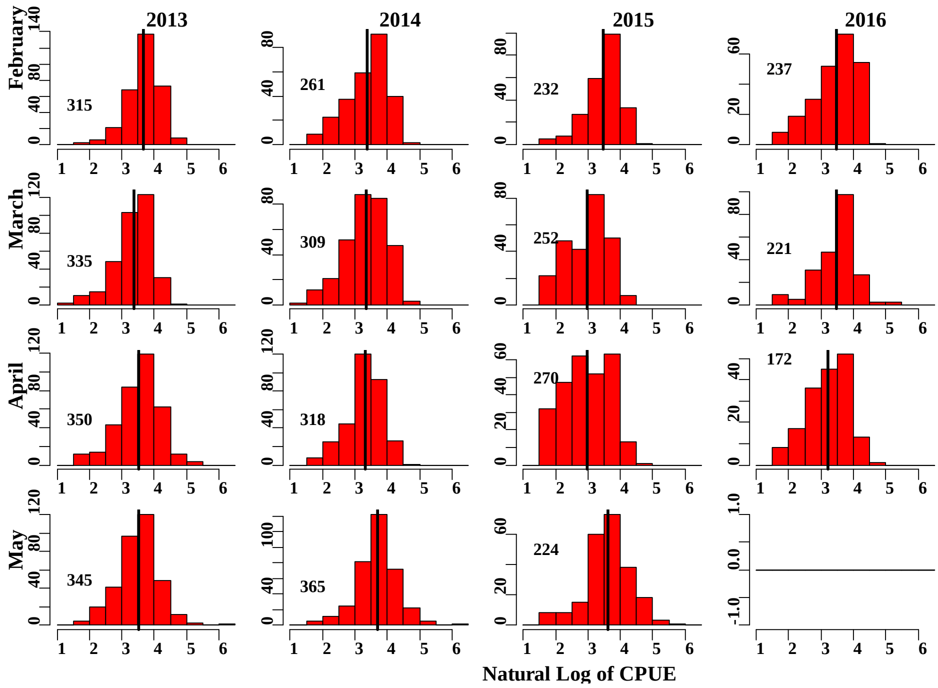

Figure 17. The distribution of log-transformed CPUE for Deepwater Flathead in the years 2013 – 2016 in the months of February to May. The most recent GAB FIS occurred in Mar – Apr 2015 which exhibited the lowest mean catch rates. While this is only an association with the seismic survey, which occurred during this period it is suggestive of a link (see **Table 11**).

Table 11. The geometric mean CPUE for Deepwater Flathead in the GAB across the years 2012 – 2016 in the months from Feb – May. The upper set of numbers are log-transformed mean CPUE, while the lower rows are the back-transformed nominal CPUE (with no bias-correction). The most recent GAB FIS occurred in Mar – Apr 2015 which exhibited the lowest mean catch rates and was associated in time with a seismic survey. The survey periods are highlighted.

Month	2013	2014	2015	2016
2	3.677	3.387	3.470	3.460
3	3.359	3.354	2.983	3.485
4	3.531	3.310	2.981	3.214
5	3.509	3.694	3.622	
2	39.54	29.57	32.15	31.82
3	28.77	28.62	19.75	32.62
4	34.17	27.39	19.71	24.87
5	33.40	40.22	37.42	

Even with this most recent exceptional year the between-year variation is of the same order as that exhibited by the CPUE. The Bight Redfish series, however, might be considered to have three exceptional years (07/08, 10/11, and 14/15), which, if omitted would lead to the series lining up more effectively. This gives the appearance of good years and bad years to survey Bight Redfish, which may imply there is some other factor involved which is currently not understood. Whatever the case the between year variation in the FIS is much greater than in the commercial CPUE.

5.3.2 SESSF Mixed Species

Trends in relative abundance across all zones for Jackass Morwong, Flathead and Pink Ling exhibit similarities between the SESSF FIS and the commercial CPUE (**Figure 18**). However, in all sixteen species considered in the SESSF, the FIS exhibited greater inter-annual variation in annual mean estimates than the commercial CPUE (**Figure 18**, **Figure 19** and **Figure 20**). In seven out of the 16 species considered the trawl survey index exhibited a significantly different trend in relative abundance index to that exhibited by the standardized or geometric mean CPUE (Silver Warehou, Ocean Perch, Mirror Dory, Blue Warehou, Royal Red Prawn, Eastern Gemfish, and Silver Trevally; **Table 10**). Others, such as John Dory, and Blue-Eye Trevalla, have some similarities between the three time-series but the FIS estimates are so noisy inter-annually that it is difficult to draw a firm conclusion. In 12 out of 16 species the SAD score for the trawl survey was two or three times larger than for either of the two CPUE series (**Table 10**).

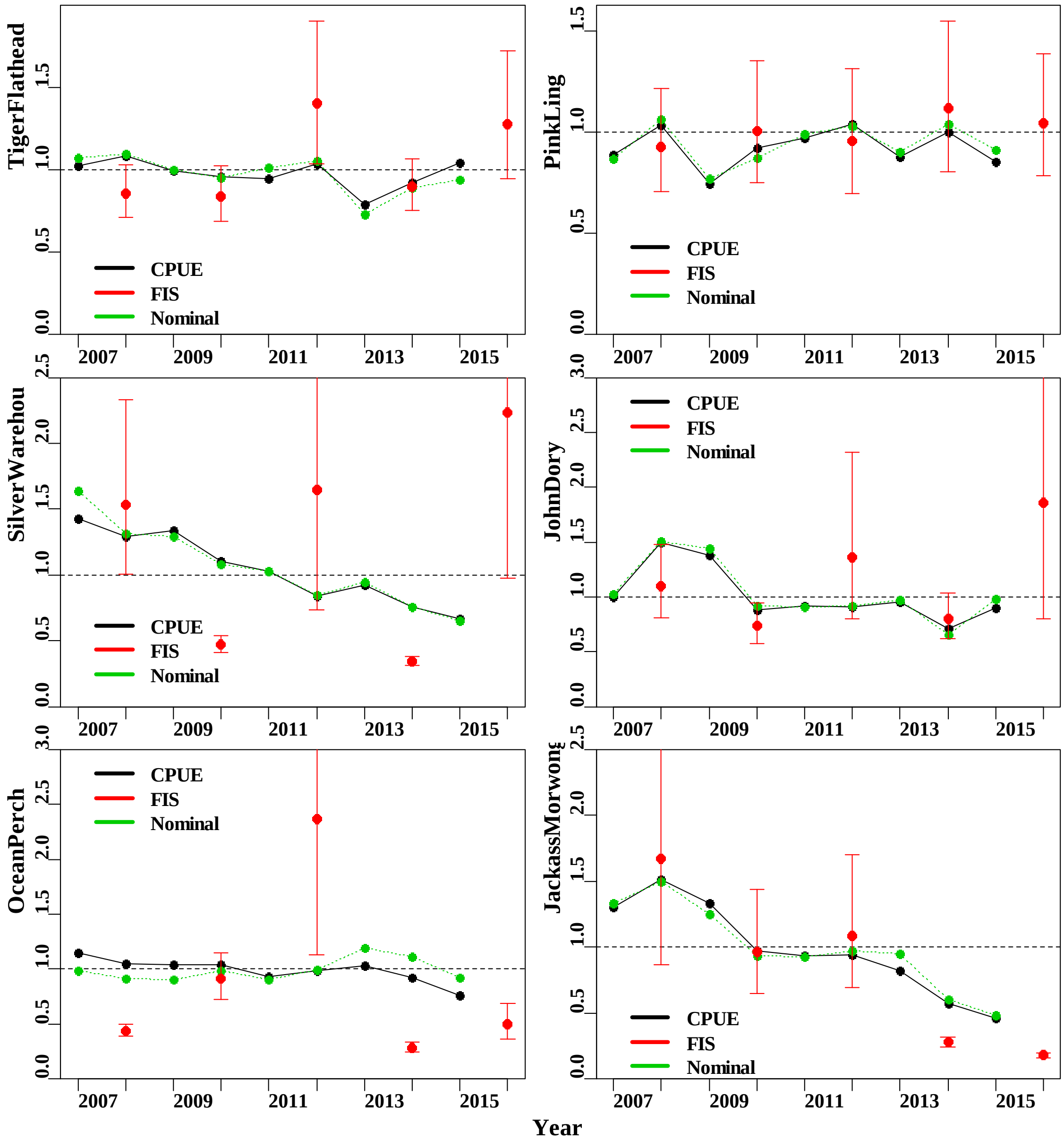

Figure 18. A comparison between the commercial CPUE (both the nominal geometric and optimal standardized means) and the FIS mean estimates for Tiger Flathead, Pink Ling, Silver Warehou, John Dory, Ocean Perch, and Jackass Morwong. The points indicate available mean estimates while the lines are only to aid visual comparison. The vertical red lines on the FIS estimates are approximate 95% confidence intervals derived from the CVs of individual surveys.

For all species considered the geometric mean CPUE over the period from 2007 - 2015 was similar to the standardized CPUE, in some cases very similar; there were differences seen with Blue Warehou and slight differences elsewhere, but in all cases the general trend was essentially the same (**Figure 18** to **Figure 20**; **Table 12** to **Table 16**).

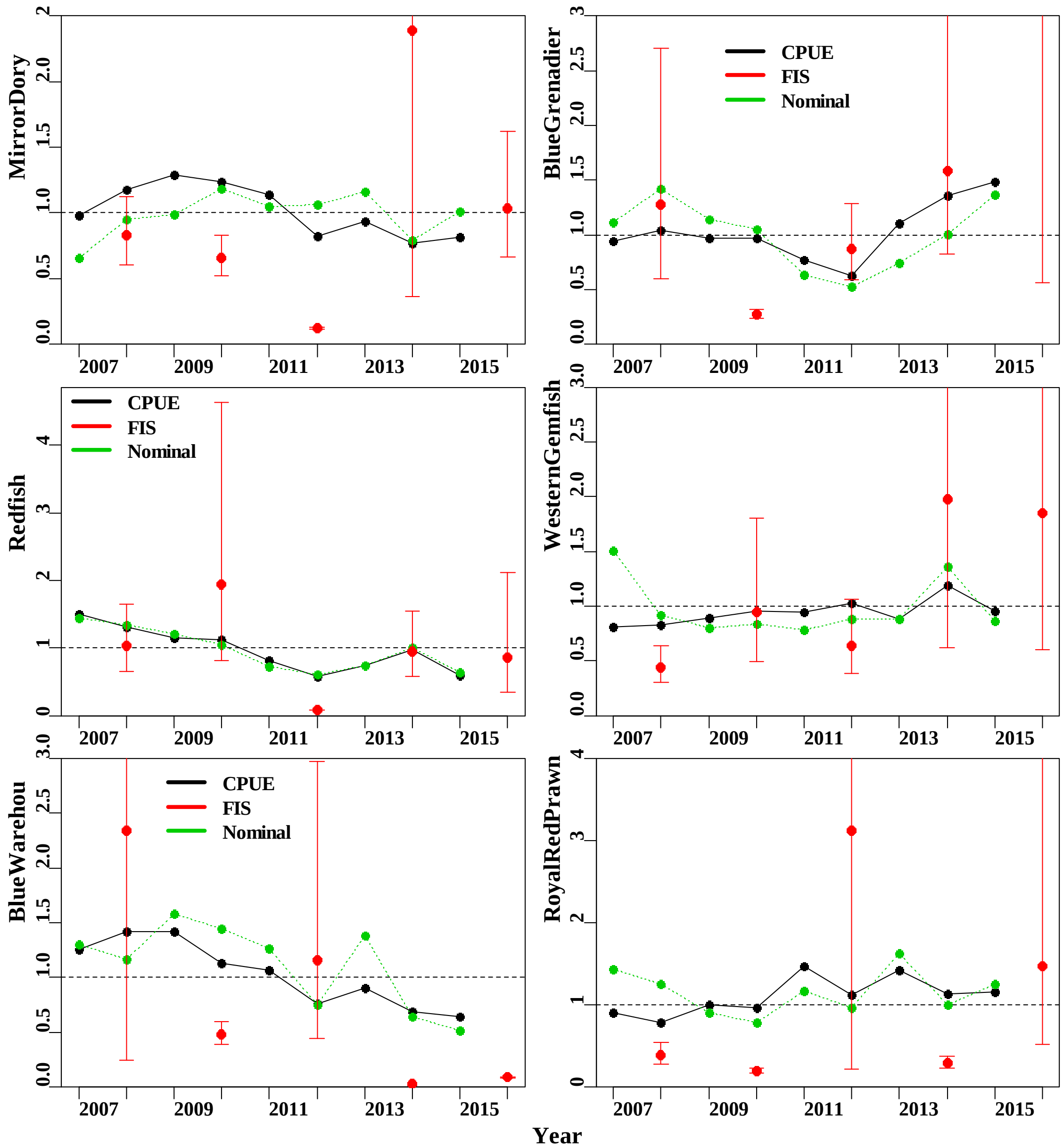

Figure 19. A comparison between the commercial CPUE (both the nominal geometric and optimal standardized means) and the FIS mean estimates for Mirror Dory, Blue Grenadier, Redfish, Western Gemfish, Blue Warehou, and Royal Red Prawn. The points indicate available mean estimates while the lines are only to aid visual comparison. The vertical red lines on the FIS estimates are approximate 95% confidence intervals derived from the CVs of individual surveys.

There was a close correlation between trends of trawl survey indices and CPUE indices for Jackass Morwong (**Figure 18**), and a similar trend in the shared years was also apparent for Pink Ling. The CPUE trends for Flathead were essentially flat in the overlap years whereas the trawl surveys in 2012 and 2016 exhibited relatively large rises, although with increased uncertainly.

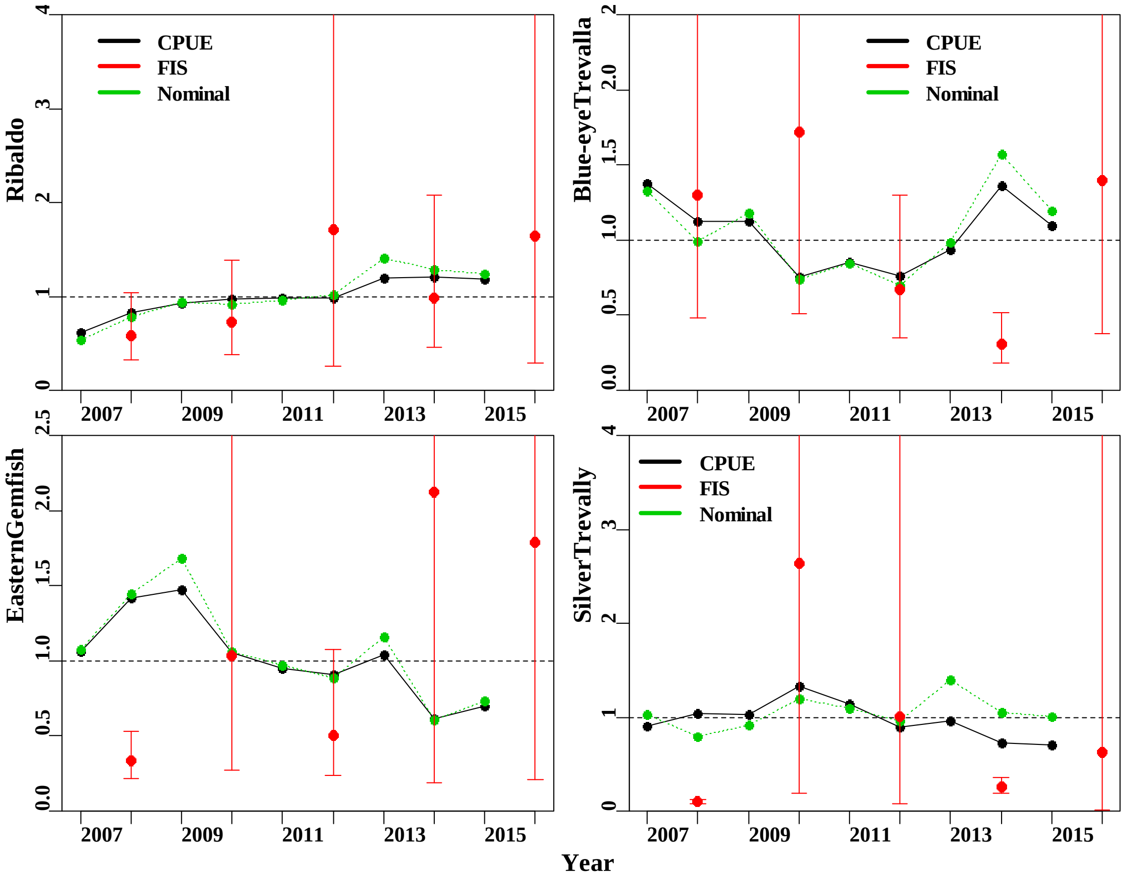

Figure 20. A comparison between the commercial CPUE (both the nominal geometric and optimal standardized means) and the FIS mean estimates for Ribaldo, Blue-Eye Trevalla, Eastern Gemfish, and Silver Trevally. The points indicate available mean estimates while the lines are only to aid visual comparison. The vertical red lines on the FIS estimates are approximate 95% confidence intervals derived from the CVs of individual surveys.

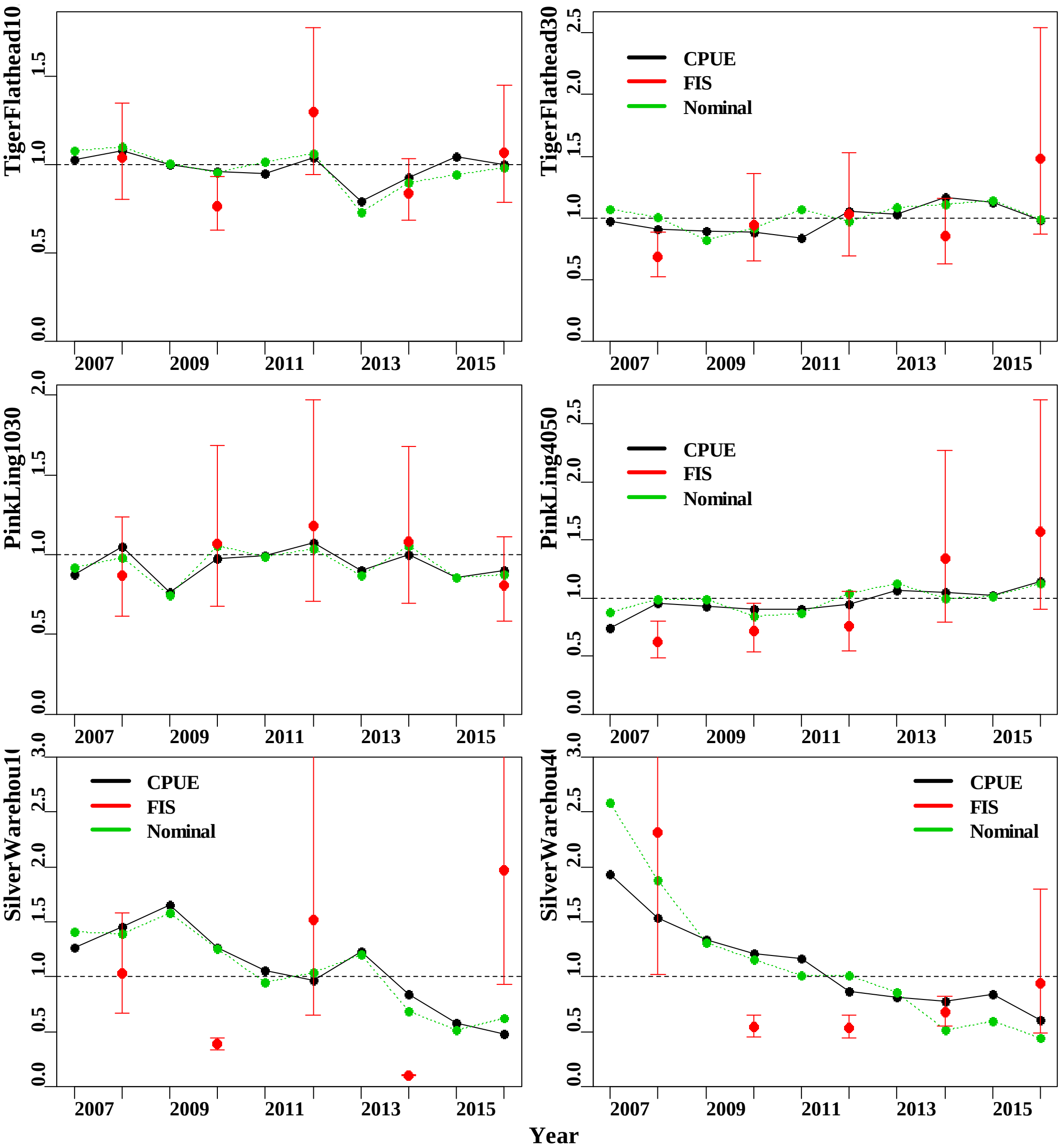

Figure 21. A comparison between the commercial CPUE (both the nominal geometric and optimal standardized CPUE) and the FIS mean estimates for Tiger Flathead1020 and Tiger Flathead30, Pink Ling1030 and 4050, and Silver Warehou1030 and 4050. This differs from **Figure 18** by the zones used in analyses i.e., the FIS indices and logbook CPUE pertain to the east or west zones.

Splitting the survey data into the different eastern and western zones for tiger flathead, pink ling, and silver warehou (**Figure 21**) improved the coincidence between the standardized CPUE series for flathead. For pink ling the eastern zones (10 – 30) exhibited strong similarity with the FIS while in the west (40 – 50) the slight upward trend in the CPUE was more exaggerated in the FIS. The eastern silver warehou FIS results continued to exhibit very large inter-annual variations, and the trend in the FIS on the west differed from the trend in the CPUE.

Despite the continued differences the approach of redesigning the analysis for the FIS may be able to improve the connection between the various trends apparent in the data.

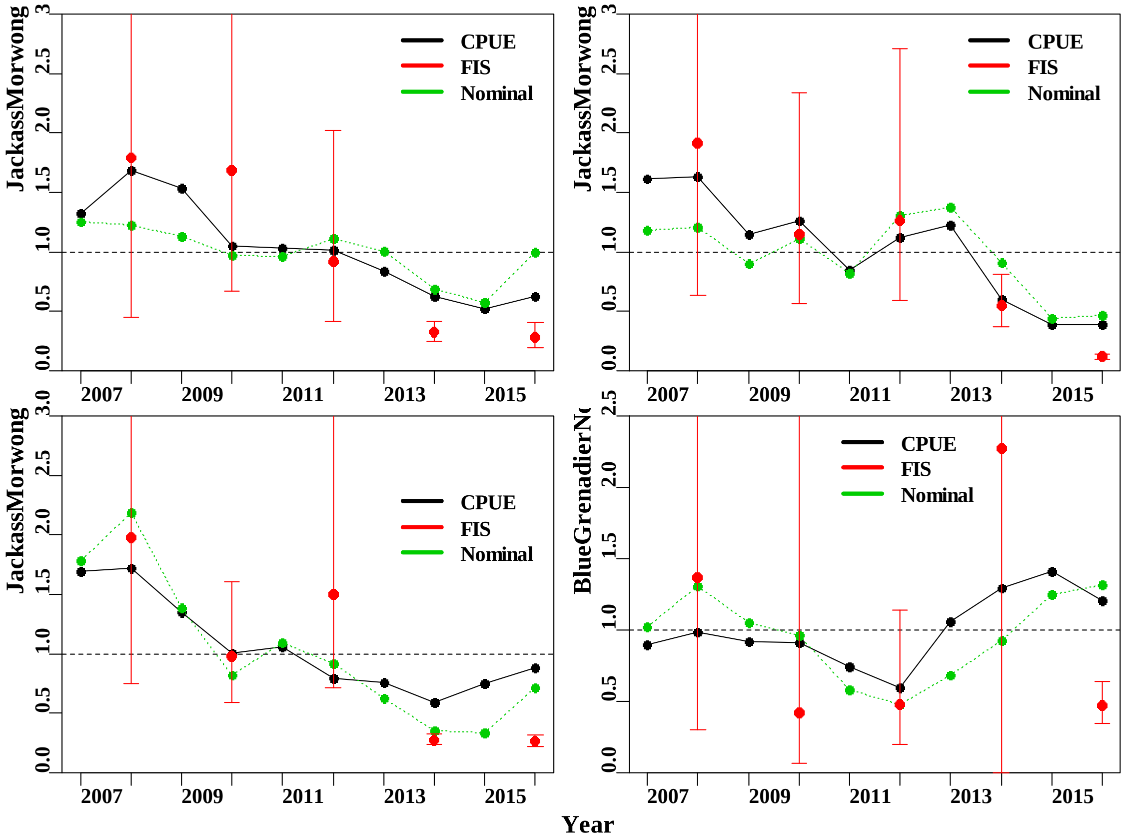

Figure 22. A comparison between the commercial CPUE (both the nominal geometric and optimal standardized means) and the FIS mean estimates for JackassMorwong1020 and JackassMorwong30, and finally JackassMorwong4050 (**Figure 18**). In addition, the Blue Grenadier non-spawning fishery was compared with that from the FIS (**Figure 19**).

When the data for Jackass Morwong is subdivided into the same zones as are used in the stock assessments the FIS remains noisy but consistently following the downward trend in the CPUE data in each region (**Figure 22**). Focussing only on the Blue Grenadier non-spawning fishery, however, led to the FIS results becoming much more variable and less informative as an index of relative abundance since 2007.

5.4 Discussion

Relative abundance indices derived from fishery-dependent catch per unit effort (CPUE) data can be influenced by error or bias from four major sources: (1) measurement error due to the fishing fleet changing its spatial behaviour and fishing by different amounts in important strata between years, (2) susceptibility changes of the species to the fishing gear among strata (3) availability changes of the species among strata, and (4) improvements, or otherwise, in the gear used when fishing leading to changes in behaviour or the effectiveness of fishing.

The strata may include time periods (season, month), depth, fishing fleet and area (zone or block). Sampling and sub-sampling effects due to fishing behaviour are those that are intended to be accounted for in CPUE standardisation procedures, although it is well recognised that many such effects may not be well modelled, and some not measured (e.g. changing skippers or fishing gear through time when using ‘Vessel’ as a factor). In terms of changing susceptibility of the fish to the gear it can be argued for the SESSF that the introduction of quotas artificially changed the targeting behaviour of fishing vessels. Similar arguments can be made with respect to the structural adjustment made to Commonwealth fisheries between Nov 2005 – Nov 2006. In recent years, for example, it has been proposed that fishers are more likely to target “mixed bags” of species to better balance individual quota holdings. Availability changes due to, for example, species moving in and out of an area of the fishery among years is not specifically dealt with and can only be examined as a process error within a stock assessment framework. The issue of availability or susceptibility would only be exacerbated by spatial closures, of which there are now a large and complex array in the SESSF. In short and in general, CPUE as an index of relative abundance does not have a particularly robust reputation.

Fishery independent surveys, on the other hand, are designed particularly to minimise errors from sources (1; sampling errors) and (2; susceptibility) above – i.e. a stratified design is employed to standardise and minimise measurement errors among years and the fishing procedure (vessels, gear, operators, gear operation procedures) is also standardised. It is not possible to account for availability changes in a fishery independent survey and closures can affect the ability to sample in some regions although special licences to survey in closures are possible in principle.

For stock assessments globally, because a FIS is supposed to minimize a number of sources of bias or error that may remain in standardised fishery-dependent CPUE data, an abundance index derived from a FIS is generally considered to set a better standard to one obtained from CPUE. Exceptions to this generality would be:

1. when the measurement error of the FIS is known to be very large. Knuckey et al. (2015, 2017) use a rule of thumb to reject estimates where the CV > 0.3, although that measure of variation relates only to the within year variation,
2. the number of available index points is very small (the SESSF now have five while in the GAB there are seven), or
3. the implied process error of FIS index is very large (e.g. the FIS index changes greatly from one year to the next compared to an assessed biomass that can only change gradually due to the species biology).

5.4.1 Great Australian Bight

In the GAB there have been six comparable points between the CPUE and the FIS with the trends for Deepwater Flathead exhibiting relatively close agreement. Stock assessments have therefore shown that whether the FIS series is included or not in an assessment that also has standardised CPUE shows relatively small differences. However, the seventh FIS in the GAB, that ran during March/April 2015, deviated from the general CPUE trends quite markedly, even for Deepwater Flathead. It can be argued that the FIS in the GAB has provided a validation that trends in the CPUE appear to provide an adequate representation of relative abundance. There have not been strong reasons proposed by fishers in the GAB trawl fishery for why fishery CPUE should not generally reflect abundance, or at least available biomass. It has been a disappointment to industry and the RAG that the inclusion of otherwise of the survey results into the GAB stock assessments has had little influence on their outcomes. It has also been argued however, that if good reasons become apparent in future for fishery CPUE to not reflect abundance in ways not accounted for in the standardisation, that the FIS results would then become more influential in stock assessment outcomes.

The 2014/2015 survey in the GAB appears to have been negatively influenced by a coincident seismic survey (further comparisons between commercial CPUE and coincident seismic surveys should be analysed to determine whether this is a singular event or a repeatable effect). The recent stock assessment of Deepwater Flathead (Haddon, 2016) included a sensitivity that excluded all data from the FIS to see the influence this data had on the assessment outcome. The data from the FIS included the index of relative abundance but also the size-composition of the catch, and the conditional age-at-length data. Despite the most recent index of relative abundance being markedly lower, removing all the FIS data altered the predicted depletion level from the final base-case level of 45% up to 51%. This increase was a reflection of the length and age-composition data rather than the FIS index of relative abundance. The recent assessment was the second to include FIS length and age composition data rather than just the index of relative abundance (the first was Bight Redfish in 2015; Haddon 2015), which is simply a reflection of the limited number of years of observation in the SESSF making inclusion of the FIS data problematical. Sufficient data is needed to permit the estimation of a new selectivity for the FIS data. It is not sufficient to only discuss the effect of the FIS index of relative abundance as the selectivity of the FIS in both the GAB and the SESSF leads to smaller fish being taken. Thus, the length and age-composition data can also be informative about the state of the stock and thus needs to be included in any assessment.

5.4.2 South East Australia

The time series of the South East FIS results is still short, with only 5 available points. Only in 2016 (when four points were available) did Tier 1 stock assessments begin to include the full range of data from the FIS. For example, the Flathead stock assessment (Day, 2016) included both the index of relative abundance from the FIS as well as length-composition data (ageing data is still to be generated). An important complication, however, is that the SESSF FIS is designed with the complete coastline included (Peel et al., 2012). What this means is that observations from all around the coastline are included in the mean estimates, which fails to recognize that some species, such as Flathead, are much more limited in their distribution. The recent assessment (Day, 2016) tried separating the Flathead FIS analysis into the same SESSF zones as have previously been used in the assessment. These complications precluded any simple clarification about the effect of including the FIS data. Future assessments should be able to improve on these analyses as agreement is reached in the RAGs about how best to proceed.

The measurement error distributions for many of the FIS points across species are rather wide when compared to, for example, Deepwater Flathead in the GAB. Of all of the species results, the best alignment of standardised CPUE and FIS was for Jackass Morwong and Pink Ling. In general, the trend from standardised CPUE is less variable than that from the FIS, which suggests that the SE FIS probably also includes a comparatively large process error component for many of the species. This will only become better characterised if the number of sample points increases.

In the RAGs there has been discussion of perhaps using fishery indicators (e.g. FIS) as a method for tactically setting TACs using a simple harvest control rule, at least during periods between more formal stock assessments. But, given what appears to be relatively large measurement and process errors of the current SE FIS, its usefulness for abundance indices that might be used for tactical TAC setting purposes is yet to be demonstrated and currently would be so variable that it may not be practical.

Most FIS data only become useful for stock assessment purposes after a time-series has become established so that the longer-term trends become apparent. Hence with only five observations in the SESSF and seven observations in the GAB (with different inter-annual spacing) these time-series are only now becoming long enough to become potentially influential in stock assessments. The inter-annual variation between the mean abundance indices in the SESSF is currently too great in most species for the time-series to have much value or validity in describing longer term trends. Exceptions might be Flathead, Pink Ling, and Jackass Morwong, with the first two being major financial drivers for the SESSF. More years of data will be required before a conclusion concerning the full benefit of operating the FIS becomes more apparent. Currently, for some key stocks (e.g. Gummy shark – not shown, Blue-Eye Trevalla, Silver Warehou, Mirror Dory), its inter-annual variability is so great that it is not sufficiently informative about trends to be used to replace the CPUE as an index of relative abundance into the future.

Table 12. The trawl survey indices (FIS), the geometric mean CPUE (Geom) and the standardized CPUE (CPUE) for four species. CV is the FIS Coefficient of Variation, and a prefix of 'rs' indicates rescaled to a mean of 1.0 over the years for which they are compared (2008, 2010, 2012, and 2014). Empty cells imply no available data. Data derived from Sporcic and Haddon (2016) and Knuckey et al, (2017)

Species	Year	FIS	CV	Geom	CPUE	rsFIS	FStErr	rsGeom	rsCPUE
Ocean Perch	2007			9.918	1.078				1.143
Ocean Perch	2008	6.90	0.14	9.192	0.991	0.437	0.14	0.061	1.051
Ocean Perch	2009			9.036	0.982				1.042
Ocean Perch	2010	14.34	0.13	9.865	0.984	0.909	0.13	0.118	1.043
Ocean Perch	2011			9.100	0.875				0.927
Ocean Perch	2012	37.38	0.16	9.967	0.932	2.369	0.16	0.379	0.988
Ocean Perch	2013			12.012	0.969				1.027
Ocean Perch	2014	4.49	0.30	11.174	0.866	0.285	0.30	0.085	0.918
Ocean Perch	2015			9.300	0.715				0.758
Ocean Perch	2016	7.82	0.33			0.496	0.33	0.164	
Blue Grenadier	2007			86.472	0.792				0.946
Blue Grenadier	2008	15.83	0.30	110.980	0.874	1.276	0.30	0.383	1.045
Blue Grenadier	2009			89.099	0.812				0.971
Blue Grenadier	2010	3.38	0.28	81.869	0.810	0.273	0.28	0.076	0.968
Blue Grenadier	2011			49.221	0.647				0.774
Blue Grenadier	2012	10.75	0.23	40.803	0.523	0.867	0.23	0.199	0.626
Blue Grenadier	2013			58.218	0.926				1.107
Blue Grenadier	2014	19.65	0.21	77.969	1.138	1.584	0.21	0.333	1.361
Blue Grenadier	2015			106.437	1.238				1.480
Blue Grenadier	2016	58.20	0.23			4.693	0.23	1.079	
Jackass Morwong	2007			12.250	0.688				1.299
Jackass Morwong	2008	41.51	0.20	13.789	0.800	1.671	0.20	0.334	1.511
Jackass Morwong	2009			11.469	0.705				1.331
Jackass Morwong	2010	23.97	0.21	8.553	0.516	0.965	0.21	0.203	0.975
Jackass Morwong	2011			8.541	0.495				0.935
Jackass Morwong	2012	27.00	0.21	8.943	0.497	1.087	0.21	0.228	0.938
Jackass Morwong	2013			8.713	0.434				0.820
Jackass Morwong	2014	6.87	0.24	5.507	0.305	0.277	0.24	0.066	0.576
Jackass Morwong	2015			4.408	0.242				0.457
Jackass Morwong	2016	4.41	0.33			0.178	0.33	0.059	
Mirror Dory	2007			26.000	0.946				0.978
Mirror Dory	2008	36.56	0.19	37.524	1.135	0.827	0.19	0.157	1.174
Mirror Dory	2009			38.878	1.246				1.289
Mirror Dory	2010	29.21	0.18	46.711	1.194	0.660	0.18	0.119	1.234
Mirror Dory	2011			41.228	1.103				1.141
Mirror Dory	2012	5.39	0.24	41.967	0.796	0.122	0.24	0.029	0.823
Mirror Dory	2013			45.703	0.906				0.937
Mirror Dory	2014	105.77	0.40	31.256	0.744	2.391	0.40	0.956	0.769
Mirror Dory	2015			39.918	0.787				0.813
Mirror Dory	2016	45.81	0.22			1.036	0.22	0.228	

Table 13. The trawl survey indices (FIS), the geometric mean CPUE (Geom) and the standardized CPUE (CPUE) for four species. CV is the FIS Coefficient of Variation, and a prefix of 'rs' indicates rescaled to a mean of 1.0 over the years for which they are compared (2008, 2010, 2012, and 2014). Empty cells imply no available data. Data derived from Sporcic and Haddon (2016) and Knuckey et al, (2017)

Species	Year	FIS	CV	Geom	CPUE	rsFIS	FStErr	rsGeom	rsCPUE
Redfish	2007			10.772	0.529				1.506
Redfish	2008	14.37	0.23	10.006	0.463	1.034	0.23	0.238	1.316
Redfish	2009			9.019	0.407				1.157
Redfish	2010	26.89	0.23	7.824	0.395	1.935	0.23	0.445	1.125
Redfish	2011			5.479	0.288				0.820
Redfish	2012	1.14	0.31	4.607	0.203	0.082	0.31	0.025	0.578
Redfish	2013			5.558	0.262				0.746
Redfish	2014	13.20	0.26	7.497	0.345	0.950	0.26	0.247	0.981
Redfish	2015			4.807	0.210				0.598
Redfish	2016	12.02	0.53			0.865	0.53	0.458	
Western Gemfish	2007			11.017	0.637				0.812
Western Gemfish	2008	1.26	0.44	6.736	0.646	0.437	0.44	0.192	0.824
Western Gemfish	2009			5.884	0.700				0.893
Western Gemfish	2010	2.72	0.35	6.126	0.751	0.944	0.35	0.330	0.957
Western Gemfish	2011			5.705	0.742				0.946
Western Gemfish	2012	1.85	0.40	6.483	0.805	0.642	0.40	0.257	1.027
Western Gemfish	2013			6.481	0.695				0.886
Western Gemfish	2014	5.70	0.30	9.935	0.935	1.977	0.30	0.593	1.192
Western Gemfish	2015			6.345	0.746				0.951
Western Gemfish	2016	5.32	0.31			1.846	0.31	0.572	
Blue Warehou	2007			5.668	0.260				1.255
Blue Warehou	2008	38.10	0.49	5.090	0.294	2.342	0.49	1.148	1.422
Blue Warehou	2009			6.912	0.294				1.421
Blue Warehou	2010	7.84	0.23	6.306	0.234	0.482	0.23	0.111	1.133
Blue Warehou	2011			5.525	0.221				1.066
Blue Warehou	2012	18.74	0.42	3.266	0.158	1.152	0.42	0.484	0.763
Blue Warehou	2013			6.028	0.187				0.905
Blue Warehou	2014	0.39	0.47	2.791	0.141	0.024	0.47	0.011	0.682
Blue Warehou	2015			2.260	0.133				0.641
Blue Warehou	2016	1.42	0.37			0.087	0.37	0.032	
Royal Red Prawn	2007			252.814	0.800				0.900
Royal Red Prawn	2008	0.12	0.44	221.099	0.695	0.390	0.44	0.172	0.783
Royal Red Prawn	2009			158.960	0.890				1.002
Royal Red Prawn	2010	0.06	0.35	138.310	0.858	0.195	0.35	0.068	0.966
Royal Red Prawn	2011			206.357	1.308				1.473
Royal Red Prawn	2012	0.96	0.44	169.276	0.995	3.122	0.44	1.374	1.120
Royal Red Prawn	2013			286.917	1.264				1.423
Royal Red Prawn	2014	0.09	0.44	176.369	1.005	0.293	0.44	0.129	1.131
Royal Red Prawn	2015			219.912	1.030				1.159
Royal Red Prawn	2016	0.45	0.36			1.463	0.36	0.527	

Table 14. The trawl survey indices (FIS), the geometric mean CPUE (Geom) and the standardized CPUE (CPUE) for four species. CV is the FIS Coefficient of Variation, and a prefix of 'rs' indicates rescaled to a mean of 1.0 over the years for which they are compared (2008, 2010, 2012, and 2014). Empty cells imply no available data. Data derived from Sporcic and Haddon (2016) and Knuckey et al, (2017)

Species	Year	FIS	CV	Geom	CPUE	rsFIS	FStErr	rsGeom	rsCPUE
Ribaldo	2007			3.249	0.439				0.609
Ribaldo	2008	2.62	0.52	4.733	0.598	0.578	0.52	0.301	0.830
Ribaldo	2009			5.698	0.669				0.927
Ribaldo	2010	3.28	0.46	5.596	0.701	0.724	0.46	0.333	0.972
Ribaldo	2011			5.829	0.706				0.979
Ribaldo	2012	7.77	0.57	6.163	0.713	1.715	0.57	0.978	0.989
Ribaldo	2013			8.581	0.861				1.194
Ribaldo	2014	4.45	0.39	7.816	0.872	0.982	0.39	0.383	1.209
Ribaldo	2015			7.538	0.857				1.189
Ribaldo	2016	7.42	0.54			1.638	0.54	0.885	
Blueeye Trevalla	2007			1.627	1.222				1.374
Blueeye Trevalla	2008	1.26	0.39	1.215	1.001	1.302	0.39	0.508	1.126
Blueeye Trevalla	2009			1.448	1.003				1.128
Blueeye Trevalla	2010	1.66	0.36	0.912	0.668	1.716	0.36	0.618	0.751
Blueeye Trevalla	2011			1.041	0.755				0.850
Blueeye Trevalla	2012	0.65	0.50	0.856	0.678	0.672	0.50	0.336	0.762
Blueeye Trevalla	2013			1.207	0.832				0.936
Blueeye Trevalla	2014	0.30	0.85	1.931	1.210	0.310	0.85	0.264	1.361
Blueeye Trevalla	2015			1.464	0.973				1.094
Blueeye Trevalla	2016	1.35	0.48			1.395	0.48	0.670	
Eastern Gemfish	2007			4.243	0.650				1.065
Eastern Gemfish	2008	0.30	0.69	5.707	0.866	0.337	0.69	0.233	1.419
Eastern Gemfish	2009			6.645	0.898				1.472
Eastern Gemfish	2010	0.92	0.66	4.193	0.646	1.034	0.66	0.682	1.058
Eastern Gemfish	2011			3.840	0.581				0.951
Eastern Gemfish	2012	0.45	0.76	3.511	0.556	0.506	0.76	0.384	0.910
Eastern Gemfish	2013			4.597	0.637				1.043
Eastern Gemfish	2014	1.89	0.58	2.404	0.374	2.124	0.58	1.232	0.613
Eastern Gemfish	2015			2.888	0.427				0.700
Eastern Gemfish	2016	1.59	0.61			1.787	0.61	1.090	
Silver Trevally	2007			11.809	0.769				0.901
Silver Trevally	2008	0.24	1.09	9.108	0.885	0.097	1.09	0.106	1.036
Silver Trevally	2009			10.519	0.882				1.033
Silver Trevally	2010	6.53	0.51	13.777	1.139	2.636	0.51	1.344	1.334
Silver Trevally	2011			12.567	0.975				1.142
Silver Trevally	2012	2.50	1.30	11.092	0.768	1.009	1.30	1.312	0.899
Silver Trevally	2013			16.102	0.823				0.963
Silver Trevally	2014	0.64	0.62	12.088	0.624	0.258	0.62	0.160	0.731
Silver Trevally	2015			11.620	0.598				0.700
Silver Trevally	2016	1.55	3.40			0.626	3.40	2.127	

Table 15. The trawl survey indices (FIS), the geometric mean CPUE (Geom) and the standardized CPUE (CPUE) for Deepwater Flathead (*Platycephalus conatus*) from the GAB. Empty cells imply no available data. Data derived from Sporcic and Haddon (2016) and Knuckey et al, (2015)

	Standardized Commercial CPUE							Trawl Survey Indices		
Fish Year	N	Catch	Geomean	CPUE	StDev	Index	StDev	Year	Abund	CV
1987/1988	453	76.84	27.6907	0.4627	0.0479			1988		
1988/1989	815	314.074	56.0806	0.9285	0.0502			1989		
1989/1990	1126	397.497	53.0361	0.9525	0.0507			1990		
1990/1991	1501	423.226	49.0776	1.0288	0.0497			1991		
1991/1992	1781	611.214	54.5388	0.9416	0.0481			1992		
1992/1993	984	509.217	76.9248	1.1969	0.0500			1993		
1993/1994	900	585.645	91.4997	1.5357	0.0504			1994		
1994/1995	1745	1258.893	106.3058	1.9451	0.0478			1995		
1995/1996	1862	1559.439	125.2137	1.8880	0.0477			1996		
1996/1997	2784	1466.636	79.3934	1.2512	0.0469			1997		
1997/1998	2908	1012.471	50.9703	0.8871	0.0467			1998		
1998/1999	2558	682.171	34.6696	0.6603	0.0471			1999		
1999/2000	2102	545.837	39.1315	0.8030	0.0482			2000		
2000/2001	2413	775.52	43.0405	0.8683	0.0477			2001		
2001/2002	2448	912.971	51.5431	1.0343	0.0477			2002		
2002/2003	3144	1632.131	73.4099	1.5005	0.0472			2003		
2003/2004	4536	2188.227	68.4174	1.4128	0.0470			2004		
2004/2005	5551	2100.187	55.0520	1.1384	0.0467	1.3930	0.0696	2005	12152	0.05
2005/2006	5349	1358.407	37.5227	0.7409	0.0468	0.9646	0.0579	2006	8415	0.06
2006/2007	4254	969.179	32.9286	0.6358	0.0467	0.9789	0.0489	2007	8540	0.05
2007/2008	4003	971.174	35.9047	0.7155	0.0472	0.8855	0.0531	2008	7725	0.06
2008/2009	3118	775.737	40.6974	0.8421	0.0475	1.1397	0.0570	2009	9942	0.05
2009/2010	3205	829.729	39.1349	0.7922	0.0474			2010		
2010/2011	2805	930.288	50.8864	1.0177	0.0477	1.0577	0.0529	2011	9227	0.05
2011/2012	3270	788.742	38.5448	0.7800	0.0475			2012		
2012/2013	3611	876.182	37.9414	0.7666	0.0473			2013		
2013/2014	3304	672.62	31.9933	0.6620	0.0474			2014		
2014/2015	2572	484.746	29.3345	0.6114	0.0480	0.5806	0.0523	2015	5065	0.09

Table 16. The trawl survey indices (FIS), the geometric mean CPUE (Geom) and the standardized CPUE (CPUE) for Bight Redfish (*Centroberyx gerrardi*) from the GAB. Empty cells imply no available data. Data derived from Sporcic and Haddon (2016) and Knuckey et al, (2015)

	Standardized Commercial CPUE					Trawl Survey Indices				
Fish Year	N	Catch	Geomean	CPUE	StDev	Index	StDev	Year	Abund	CV
1987/1988	184	32.753	29.2533	2.3613	0.0952			1988		
1988/1989	492	85.88	32.9965	2.2877	0.1015			1989		
1989/1990	827	171.577	31.8857	1.5603	0.0996			1990		
1990/1991	1023	250.2255	36.6457	1.3858	0.0980			1991		
1991/1992	1101	240.443	27.4447	1.2674	0.0962			1992		
1992/1993	718	120.188	18.3377	0.9355	0.0985			1993		
1993/1994	695	107.418	16.2182	0.9267	0.0990			1994		
1994/1995	1282	159.907	11.9237	0.6480	0.0946			1995		
1995/1996	1395	175.277	11.8016	0.7837	0.0947			1996		
1996/1997	2036	329.777	15.3383	0.9467	0.0930			1997		
1997/1998	1930	365.931	16.0229	0.9572	0.0933			1998		
1998/1999	1812	440.296	20.2349	1.0999	0.0933			1999		
1999/2000	1478	324.421	17.1853	0.9572	0.0955			2000		
2000/2001	1697	387.531	15.6494	0.8446	0.0947			2001		
2001/2002	1637	225.642	10.8567	0.6621	0.0949			2002		
2002/2003	2118	364.3121	13.4661	0.7043	0.0937			2003		
2003/2004	3154	841.725	20.1099	0.9685	0.0933			2004		
2004/2005	3808	758.0925	18.3742	0.9395	0.0929	1.1252	0.1279	2005	20887	0.13
2005/2006	3553	722.8482	17.4248	0.8977	0.0930	1.3672	0.1913	2006	25380	0.16
2006/2007	3293	873.7396	21.7750	0.9310	0.0927	1.3852	0.1938	2007	25713	0.16
2007/2008	2743	683.535	20.0988	0.8911	0.0935	0.7860	0.0756	2008	14591	0.11
2008/2009	2443	648.786	21.9054	0.9565	0.0941	1.4873	0.2341	2009	27610	0.18
2009/2010	2298	445.717	17.3788	0.8760	0.0941			2010		
2010/2011	1851	277.889	14.2664	0.7049	0.0948	0.7105	0.0808	2011	13189	0.13
2011/2012	2188	322.865	14.4195	0.7075	0.0945			2012		
2012/2013	1873	255.705	15.2641	0.6208	0.0950			2013		
2013/2014	1494	187.558	14.6071	0.5750	0.0959			2014		
2014/2015	1396	233.371	16.9298	0.6031	0.0966	0.1386	0.2000	2015	2573	0.28

6 Alternative Methods of Statistical Standardization

6.1 Introduction

There are already a number of reviews and summaries in the published literature concerning alternative approaches that can be used to conduct statistical standardizations of catch-rate data (Bishop et al., 2004; Maunder and Punt, 2004; Venables and Dichmont, 2004; Campbell, 2015; **Table 17**).

Table 17. A list of the methods found in a rapid review of standardization analyses of CPUE data found in the formal published and so-called 'grey' literature. The LM and GLM still appear to be the most common although the reason for using a particular method in any particular fishery is rarely documented. This is not an exhaustive list.

Acronym	Comments
LM	Linear Models – essentially linear regression (or if only categorical factors then ≡ anova) – only normal residual errors (confusingly general linear models, which also require normal response variables, can be referred to as GLM), and can include non-linear terms as polynomials (see Neter *et al*., 1996)
GLM	Generalized Linear Models – Need not assume normally distributed response variables, can use Normal, Gamma, Poisson, Binomial, Exponential, Negative Binomial, etc; uses a link function to relate the mean of the response variable (CPUE) to the equation used as the linear predictor of CPUE; can include non-linear terms as polynomials (McCullagh and Nelder, 1989)
GLMM	the same as GLM except the linear predictor can include fixed and random effects (factors). There remains a good deal of debate over how best to interpret the meaning of the terms 'fixed' and 'random' when applied to model effects. This on-going (never-ending) debate appears to stem from there being a Bayesian and a Frequentist difference of interpretation. In practice, if one can present a defensible and reasonable argument for using a random effect then one should be used, if not then don't (McCulloch and Searle, 2001).
GAM	Generalized Additive Models – no need to assume linear or any particular functional form relating the response variable to the predictor. GAMs extend GLMs by allowing the inclusion of non-parametric smoothers in addition to the usual parametric equations (Hastie and Tibshirani, 1990)
GAMM	Once again akin to the relationship between GLM and GLMM, GAMMs extend GLMMs through allowing the inclusion of random effects factors into GAMs (Chen, 2000).
GEE	Generalized Estimating Equations are, as their name suggests, very general in that they provide a means of modelling the catch rate trends and the variation around that trend, with any correlation structure between variables, separately (Hardin and Hilbe, 2003). These have not yet been used extensively in a CPUE standardization sense.

Bishop et al (2004) compared four different modelling approaches with the same data set from the Northern Prawn Fishery (NPF) in the Gulf of Carpentaria; the methodologies compared were linear models, generalized linear models, mixed-models (those that include both fixed and random effects), and generalized estimating equations. This NPF data set is exceptional in having a large amount of information about the fleet and changes in fishing gear through time, which was obtained in attempts to manage high levels of increases in fishing power (effort creep) through time. The summary conclusions were that the more complex models and the inclusion of either the correlation structure between variates or the use of random effects models all yielded similar results to those from the simpler models. Such detailed catch-per-unit-effort data is exceptional in Australia, while in other data sets the only accurately rendered information about fishing gear through the years of a fishery is often limited to the vessel doing the fishing. In addition, the NPF CPUE data is a very large data set made up of many 100,000s of records through the years of the fishery (about 1970 to the present). In the SESSF, for example, Flathead taken by trawl in SESSF zones 10 and 20 have just over 270,000 records but many have 10's of thousands rather than 100's of thousands. Even so, with so many data points in each year the apparent statistical precision can be unrealistically high. This is apparent when fitting a stock assessment model to such CPUE data. It is necessary to adjust the implied variation about each data point (by lowering the effective sample size) to match that predicted by the assessment model. This is primarily necessary to allow for inter-annual variation in CPUE and if it is not done the model can be forced to fit to noise rather than a signal about the stock size through time.

It is still possible to use an array of alternative methods (**Table 17**) with less comprehensive data although some of the issues with data quality illustrated in the earlier section (**Chapter 4. Potential Issues when using CPUE Data**), such as the rounding of catch and of effort details, could compromise methods that might be sensitive to non-standard forms of variation. Of potentially more importance, the less data one has across the years of a fishery the greater the uncertainty around any year's mean estimate, irrespective of what method is used to make it, so that identifying change through time (generally the intent of the study) can become more difficult. With Blue-Eye taken by auto-line, for example, if one includes classical 95% confidence intervals and compare those with a similar analysis conducted on Flathead over the same period, the effect of the much larger number of records each year for Flathead is marked (**Table 18**; **Figure 23**)

Table 18. A contrast between the number of CPUE records and total catch (Catch) by method for Flathead by trawl and Blue-Eye by Auto-Line across the years 2002 - 2015. Flathead has between 12.4 – 54.8 times as many records as Blue-Eye but only about 3.5 – 11 times as much catch. A large amount of Flathead is also taken by Danish Seine.

	Flathead Trawl		Blue-Eye Auto-Line	
Year	Records	Catch (t)	Records	Catch (t)
2002	12377	1449.200	226	131.366
2003	12885	1587.068	432	156.966
2004	12236	1338.284	956	227.589
2005	10613	1144.064	859	237.854
2006	9069	1138.895	604	237.218
2007	6294	1068.883	459	308.245
2008	7213	1309.623	414	205.017
2009	6222	1037.656	465	279.887
2010	6694	1086.889	416	202.140
2011	6618	1070.668	384	151.689
2012	6798	1149.380	365	158.120
2013	5597	683.369	301	156.342
2014	6343	943.858	225	176.813
2015	6368	984.696	224	155.946

As previously noted the documentation regarding how data are selected for inclusion in a CPUE standardization is usually poor. Generally, the reasons for selecting a particular statistical technique with which to conduct a standardization are equally poorly documented or defended. As a result, it is possible to get the impression, when reading examples from the formal and the grey literature, that the methods selected are more related to what has generally been used before in a particular fishery or relates to the expertise or previous experience of the analysts involved.

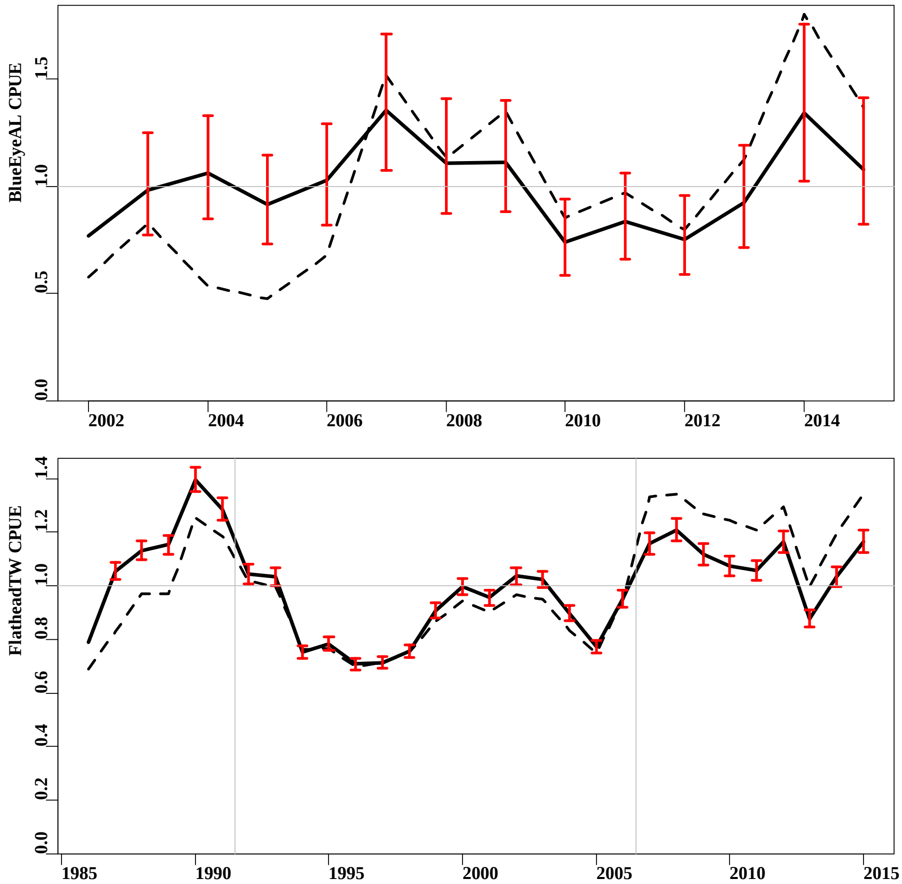

Figure 23. A comparison of the 95% confidence intervals of Blue-Eye taken by auto-line (top plot) and Flathead taken by trawl (bottom plot) where Blue-Eye has between 220 and 900 records per years while Flathead has between 6200 and 12800 (see **Table 18**), and yet both have many factors included in their standardizations. The dashed lines are the geometric mean CPUE, the black lines are the optimum standardized CPUE, and the red bars are the log-normal 95% confidence bounds around the mean estimates. The horizontal grey line = 1.0, the overall average of all trends, and the vertical grey lines at 1991.5 and 2006.5 relate to the introduction of ITQs and the Harvest Strategy Policy and associated structural adjustment.

A direct comparison of different methods using identical data and structurally identical statistical models (as far as possible) would enable an examination of the idea that some methods are preferable to others. Additionally, it would generate specific examples to assist with the production of guidelines for standardizations in a later chapter. Such a comparison would include a consideration of classical and other diagnostics for quantifying and include presenting the relative quality of model fit to the available data. Here we will examine a trawl fishery for flathead (*Neoplatycephalus richardsoni*) that has over 270,000 records over 30 years from 1986 – 2015, as well as an auto-line fishery for

Blue-Eye Trevalla (*Hyperoglyphe antarctica*) that has only about a 20^{th} the number of records (6330 over 14 years, 2002 – 2015). These species and fisheries were selected as being markedly different although both constitute significant economic drivers in the SESSF. Flathead is the primary shelf trawl fishery and the Blue-Eye fishery is an iconic deeper water slope fishery taken mainly by line methods.

The objectives of this section are therefore to:

1) Compare an array of alternative statistical methods using the same data and statistical models.
2) Illustrate different diagnostic plots and tables and how they are used to compare statistical models of CPUE.

6.2 Methods

Here comparisons were made between Linear Models (LM), Generalized Linear Models (GLM), and Generalized Additive Models (GAM), with two variants of the LM, three variants of the GLM, and four variants of the GAM. With Flathead (*Neoplatycephalus richardsoni*) the analyses were conducted on catches taken by trawl in depths between 0 – 400 m from SESSF zones 10 and 20 across the years 1986 – 2015. For Blue-Eye Trevalla (*Hyperoglyphe antarctica*) the analyses were conducted on catches taken by auto-line in depths between 200 – 600 m from SESSF zones 20, 30, 40, and 50 (see **Figure 24**) across the years 2002 – 2015. The initial analysis using the linear model (or the GLM with the identity link function) is identical to that analysis used in the stock assessment in the SESSF. Hence its use as a standard reference provides a direct commentary on the current CPUE assessment.

6.2.1 Statistical Models Used

After preliminary analyses of available data, different optimal models were used for Flathead and Blue-Eye (**Table 19**).

Table 19. The optimum statistical models compared within each of Flathead and Blue-Eye. All rows, except for *CELog*, and *CEGam*, through log-transforming the CPUE or catch and effort, are effectively assuming normally distributed residuals and are using the equivalent of the identity link between the mean estimates and data.

	Blue-Eye Trevalla	
Class	**Model Used**	**Name**
LM	LnCE ~ Year+Vessel+Month+Zone+DepCat+Month:Zone	LnCE
LM	LnC ~ Year+Vessel+Month+Zone+DepCat+Month:Zone+LnE	LnC
GLM	LnCE ~ Year+Vessel+Month+Zone+DepCat+Month:Zone	glmLnCE
GLM	CE ~ Year+Vessel+Month+Zone+DepCat+Month:Zone	CELog
GLM	CE ~ Year+Vessel+Month+Zone+DepCat+Month:Zone	CEGam
GAM	LnCE ~ s(Long)+Year+Vessel+Month+Zone+DepCat+Month:Zone	gamLon
GAM	LnCE ~ s(Lat) +Year+Vessel+Month+Zone+DepCat+Month:Zone	gamLat
GAM	LnCE ~ s(Long, Lat) + Year + Vessel + Month + DepCat	gamLL
GAM	LnCE ~ s(Long, Lat) + Year + Vessel + Month + DepCat	gamTrim
	Flathead	
LM	LnCE ~ Year+Vessel+DepCat+Zone+Month+DayNight+Month:Zone	LnCE
LM	LnC ~ Year+Vessel+DepCat+Zone+Month+DayNight+Month:Zone+LnE	LnC
GLM	LnCE ~ Year+Vessel+DepCat+Zone+Month+DayNight+Month:Zone	glmLnCE
GLM	CE ~ Year+Vessel+DepCat+Zone+Month+DayNight+Month:Zone	CELog
GLM	CE ~ Year+Vessel+DepCat+Zone+Month+DayNight+Month:Zone	CEGam
GAM	LnCE ~ s(Long)+Year+Vessel+DepCat+Zone+Month+DayNight+Month:Zone	gamLon
GAM	LnCE ~ s(Lat) +Year+Vessel+DepCat+Zone+Month+DayNight+Month:Zone	gamLat
GAM	LnCE ~ s(Long, Lat) +Year + Vessel + DepCat + Month	gamLL
GAM	LnCE ~ s(Long, Lat) +Year + Vessel + DepCat + Month	gamTrim

6.2.2 Back-Transforming the Year Parameters

Where the residual error structure was assumed to be distributed following a Gamma distribution, obtaining the annual parameters entails a back transformation of the log-space parameters using the exponential function to put them back into the linear space. Thus, if the model used was:

glm(usemod, family=Gamma(link="log"), data=sps2)

and 'usemod' was CE ~ Year+Vessel+Month+Zone+DepCat+Month:Zone (as it was for Blue-Eye), and sps2 was the selected fisheries data used in the analysis. The model output would include an array of year parameters Y_i. The back transformed year effects would be:

$$CPUE_t = e^{(Y_i)} \tag{9}$$

The Gamma function requires all values to be positive (> 0) and so it is not possible to use a log-transformation on CPUE with Gamma distributed residual errors (unless they all happened to be positive, which would mean all catch rates would need to be greater than 1.0 kg/hr, which is not the case in Flathead and many other species.

However, if the model used lognormal residual errors with, for example a model such as:

glm(usemod, family=gaussian(link="identity"), data=sps2)

with 'usemod' as LnCE ~ Year+Vessel+Month+Zone+DepCat+Month:Zone (as it was for Blue-Eye), then, for the log-normal models the expected back-transformed year effect involves a bias-correction to account for the log-normality:

$$CPUE_t = e^{\left(Y_t + \sigma_t^2/2\right)} \tag{10}$$

where Yt is the modelled Year coefficient for year t and σ_t is the standard deviation of the log transformed data (obtained from the analysis). This returns the mean rather than the median of the nominal scale standardized CPUE. Generally, this has little effect but sometimes (when sample sizes are relatively small) it can be very influential.

With the GAM analyses the parameters relating to the different factors outside of the smoother terms, are back-transformed as for the log-normal based analysis; that is they include a bias-correction to account for the log-normality; equation (10).

6.2.3 Data Selection Criteria

Data selections were made on years, fishery, method, depths, and SESSF zone (**Figure 24**; **Table 20**).

6.2.4 Comparison of Standardization Outputs

Not all standardizations will be directly comparable statistically because they have different assumptions about the distribution of CPUE and the residuals used when estimating the optimal year effects. In addition, there will be very different numbers of parameters when comparing methods like the GAMs (with the local empirical smoothers). Nevertheless, within standardization type it is possible to make some comparisons of such things as the Akaike's Information Criterion. Less formally, however, graphical comparisons will be used to illustrate the differences between outcomes with both simple plots of the different trends, and plots of the differences between the reference standardization and the alternatives.

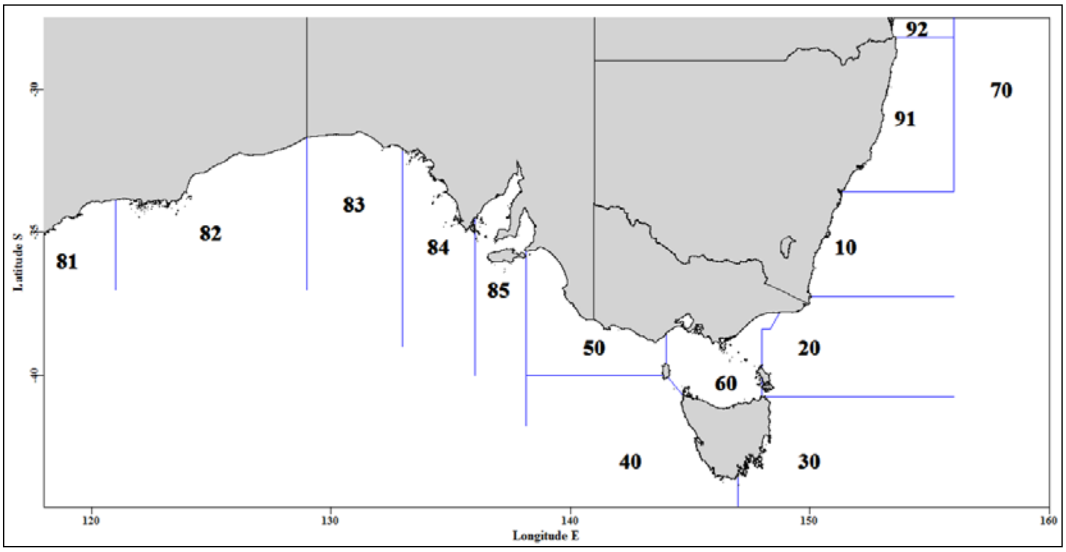

Figure 24. A schematic diagram depicting the statistical reporting zones in the SESSF, as used in this document. The GAB fishery is to the west of zone 50. The main SESSF trawl zones are zones 10 – 50. Each zone extends out to the boundary of the EEZ, except for zones 50 and 60, and for zones 92 and 91, which are bounded by zone 70.

Table 20. The data selection criteria used for the two species considered.

Species	Criteria	Values
Flathead	Years	1986 – 2015
	Fishery	SET
	Method	Demersal trawl (TW and TDO)
	Depth	0 – 400
	Zone	10 and 20 (see **Figure 24**)
	Comment	One vessel that only fished in one year, that had catch rates over 10 times that of anyone else in any years, was removed as having made some fundamental error in how they reported their effort.
Blue-Eye	Years	2002 – 2015
	Fishery	GHT, SEN
	Method	Auto-Line
	Depth	200 – 600
	Zone	20, 30, 40, and 50 (see **Figure 24**)
	Years in Fishery	> 1
	Single Line Drops	Single Line Drops were identified as an anomaly in the auto-line data by Haddon (2016b); as with the now standard analysis these were removed from consideration.
	Comment	The auto-line fishery has been operational since 1997 but catches were low and the number of vessels very limited until 2002 (Haddon, 2016b). The requirement to have > 1 year in the fishery was used to remove some exceptional outliers (see text).

6.3 Results

6.3.1 The Standardization of Flathead

To prepare a CPUE standardization of Flathead (*Neoplatycephalus richardsoni*) taken by trawl in the SESSF, for inclusion in the Integrated Stock Assessment, the usual procedure is to treat all included factors as categorical and fit a linear model to the log-transformed CPUE data (Haddon, 2014). In effect, this is a highly unbalanced ANOVA where the primary interest is in the year terms. Here, using the CPUE data from 1986 – 2015 the optimal model was (see Equation (11); **Table 19**; **Figure 25**):

LnCE ~ Year + Vessel + DepCat + Zone + Month + DayNight + Month:Zone (11)

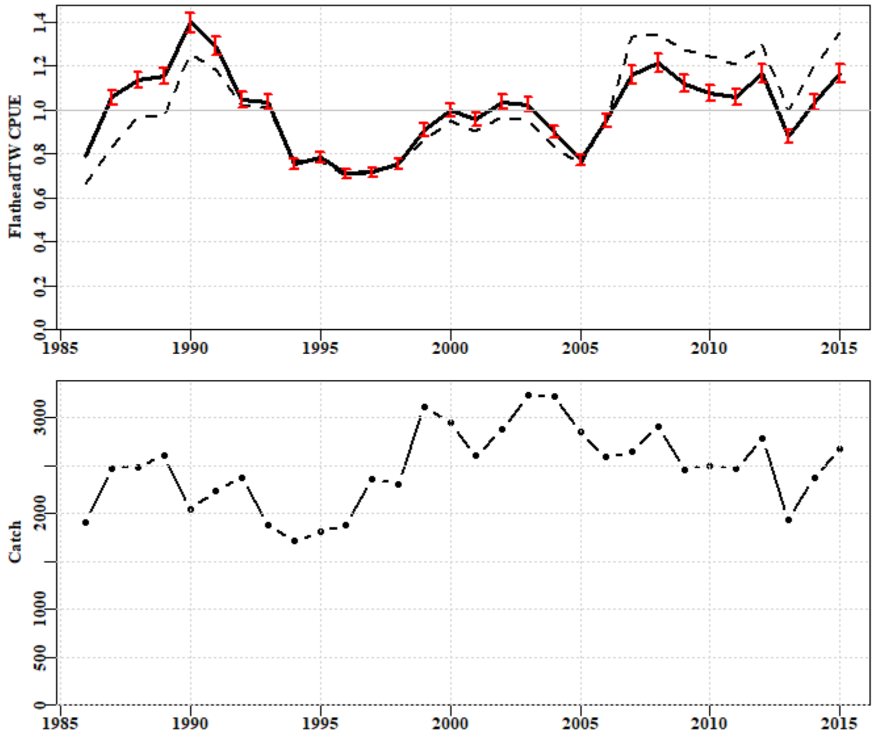

Figure 25. The reference standardization for Flathead from zones 10 and 20. The geometric mean CPUE (dashed line) and the optimal standardized CPUE (solid line; from equation (11)), with red coloured 95% confidence intervals around the mean estimates. The standardized curve deviates from the standardized curve in the years prior to the introduction of quota (1992) and in the years from 2007 onwards following the completion of the structural adjustment at the end of 2006. The lower plot is total reported catches for flathead by all methods and areas.

While there is no guarantee that this is the best representation of the 'true' underlying CPUE (**Figure 25**; **Table 21**) it can validly be used as a reference against which to compare all the other analyses. It is possible to apply standard diagnostic tests and plots to the quality of the model fit to the available data, although with approximately 270,000 data points the sheer weight of data operates to counter-balance the presence of an occasional outlying point (**Figure 26**). The Q-Q plot that compares the distribution of residuals to a normal distribution (**Figure 26**) exhibits an apparent shift away from normality when negative residuals extend beyond about -2 and positive residuals above about 3.5. However, out of 270,410 observations there were only 8913 from -2 downwards and only 101 from 3.5 up, which are equivalent to about 3.29% and 0.037% of all observations respectively (if the lower bound is taken as -1.75 the proportion of low valued non-normal data increases to about 4.9%.

In the case of Flathead, there is great interest in whether the large changes identified in the CPUE trend following the introduction of quota in 1992 and the structural adjustment at the end of 2006 constitute breaks in the time-series. However, here we are concerned with whether the application of different statistical methods of standardization lead to different outcomes of sufficient scale to affect a stock assessment.

Using statistical techniques to compare such different approaches as linear models, generalized linear models, and generalized additive models is not always possible given their different assumptions and the different numbers of parameters (especially with the GAMs) and different likelihood structures. Nevertheless, graphical comparisons are always possible and statistical comparisons, at least within methods, remain valid (**Figure 26**).

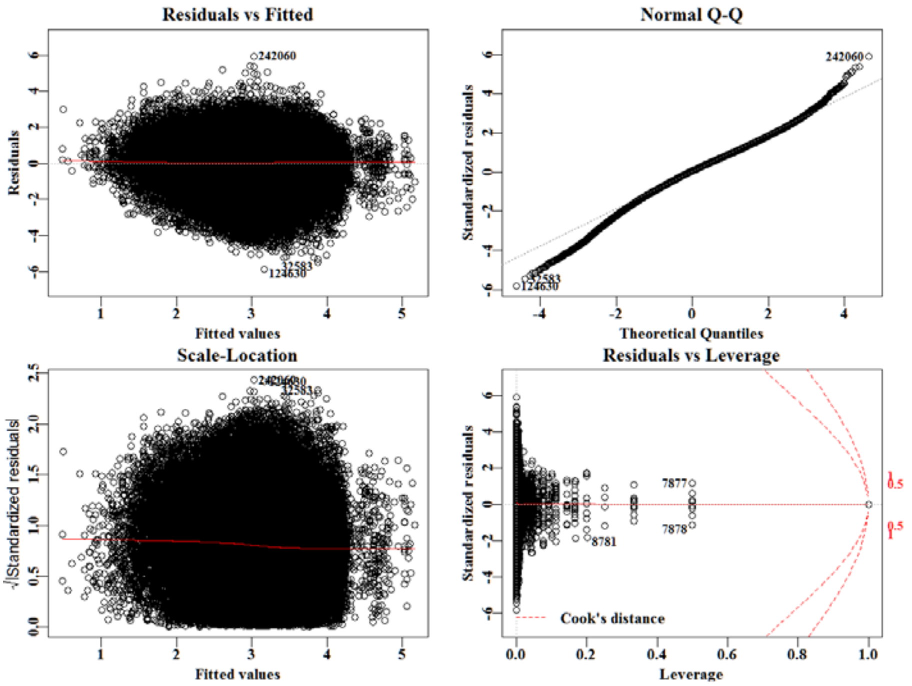

Figure 26. Classical diagnostic plots for a linear model, although such standard tests were not designed to be used with the approximately 270,000 data points available for Flathead taken by trawl in SESSF zones 10 and 20 between 1986 – 2015.

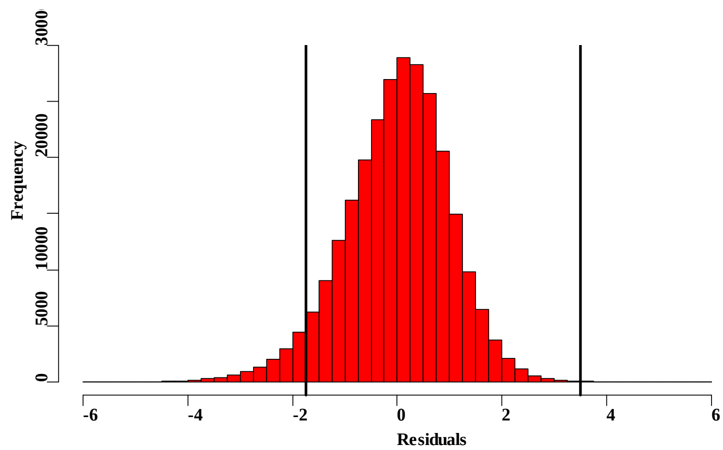

Figure 27. The distribution of all residuals from the LM on LnCE. It has a longer left-hand tail than a properly normal distribution would expect, which is why the QQplot in **Figure 26**, deviates from the line denoting a normal distribution at about -1.75. Above +3.5 there were only about 0.037% of records although below -1.75 there were almost 5% of records. Again, the enormous number of observations dominates the analysis.

6.3.2 Comparing the CPUE Trends Predicted by the Different Models

The optimum models from each standardization procedure generate yearly indices that would be used as indices of relative abundance in suitable stock assessment models (**Table 21**). In the case of Flathead (*Neoplatycephalus richardsoni*) the assessment would use Stock Synthesis 3 (Methot and Wetzel, 2013) while with Blue-Eye Trevalla (*Hyperoglyphe antarctica*) the assessment method is an empirical harvest control rule designated the Tier 4 analyses in the SESSF (Wayte, 2009, Little et al, 2011a).

Table 21. Year parameters from nine standardization procedures applied to Flathead CPUE taken by trawl between 1986 – 2015 in SESSF zones 10 and 20. The geometric mean CPUE (Geom) is unstandardized. LnCE is the simple linear model (with identical results under LnCE under GLM). LnC standardizes the log of catches using log(effort) as an offset. CELog uses normal residuals with a log link-function. CEGam uses Gamma distributed residuals with a log link. In the GAMs 'Long' only smooths on Longitude, 'Lat' smooths on Latitude, 'LongLat' smooths on both, and LLtrim repeats the LongLat analysis after removing data reported from peripheral locations as a start on reducing edge effects on the GAM smoother.

	Linear Models			Generalized Linear Models			General Additive Models			
Year	Geom	LnCE	LnC	LnCE	CELog	CEGam	Long	Lat	LongLat	LLtrim
1986	0.6894	0.7890	0.7646	0.7890	1.0415	0.8875	0.8243	0.8472	0.8330	0.8276
1987	0.8315	1.0549	1.0289	1.0549	1.3962	1.1907	1.0790	1.0899	1.0725	1.0699
1988	0.9694	1.1327	1.1100	1.1327	1.2367	1.2011	1.1535	1.1510	1.1329	1.1305
1989	0.9716	1.1537	1.1302	1.1537	1.3702	1.2415	1.1268	1.1377	1.1172	1.1130
1990	1.2559	1.3954	1.3892	1.3954	1.3402	1.4087	1.3320	1.3204	1.3174	1.3227
1991	1.1841	1.2867	1.2764	1.2867	1.3916	1.4048	1.2273	1.1964	1.2074	1.2082
1992	1.0168	1.0435	1.0375	1.0435	1.0937	1.1532	1.0189	1.0156	1.0062	1.0063
1993	0.9963	1.0347	1.0306	1.0347	0.9030	1.0458	1.0397	1.0368	1.0443	1.0425
1994	0.7633	0.7537	0.7492	0.7537	0.6008	0.7573	0.7545	0.7535	0.7579	0.7563
1995	0.7655	0.7840	0.7834	0.7840	0.8431	0.8187	0.7656	0.7716	0.7836	0.7828
1996	0.7004	0.7079	0.7063	0.7079	0.6690	0.7301	0.7020	0.7077	0.7178	0.7156
1997	0.7164	0.7143	0.7136	0.7143	0.5896	0.7351	0.6975	0.7074	0.7166	0.7178
1998	0.7526	0.7550	0.7494	0.7550	0.7991	0.7784	0.7366	0.7321	0.7460	0.7448
1999	0.8657	0.9067	0.9019	0.9067	0.8702	0.9072	0.8782	0.8797	0.8928	0.8908
2000	0.9453	0.9959	0.9887	0.9959	1.0199	1.0461	0.9949	0.9833	0.9996	0.9969
2001	0.9027	0.9559	0.9517	0.9559	0.9200	0.9326	0.9401	0.9360	0.9490	0.9518
2002	0.9687	1.0358	1.0321	1.0358	0.9728	0.9720	1.0232	1.0099	1.0335	1.0313
2003	0.9515	1.0235	1.0242	1.0235	1.0205	1.0082	1.0180	1.0082	1.0177	1.0169
2004	0.8326	0.8976	0.8998	0.8976	0.9516	0.9092	0.8882	0.8799	0.8759	0.8759
2005	0.7504	0.7720	0.7801	0.7720	0.8657	0.8109	0.7587	0.7541	0.7520	0.7507
2006	0.9447	0.9512	0.9631	0.9512	0.9180	0.9165	0.9497	0.9431	0.9318	0.9327
2007	1.3309	1.1570	1.1699	1.1570	1.0997	1.0926	1.1748	1.1674	1.1683	1.1696
2008	1.3410	1.2094	1.2220	1.2094	1.1956	1.1815	1.2536	1.2533	1.2617	1.2637
2009	1.2669	1.1180	1.1354	1.1180	1.0720	1.0675	1.1697	1.1648	1.1666	1.1672
2010	1.2449	1.0740	1.0888	1.0740	1.0214	0.9945	1.1184	1.1217	1.1084	1.1104
2011	1.2087	1.0563	1.0682	1.0563	1.0323	1.0187	1.0745	1.0735	1.0735	1.0750
2012	1.2965	1.1639	1.1762	1.1639	1.1351	1.0974	1.1839	1.1994	1.1878	1.1902
2013	0.9949	0.8775	0.8915	0.8775	0.7741	0.7510	0.8914	0.9055	0.9012	0.9058
2014	1.1955	1.0336	1.0484	1.0336	0.9285	0.9308	1.0466	1.0686	1.0495	1.0528
2015	1.3460	1.1661	1.1887	1.1661	0.9280	1.0103	1.1785	1.1843	1.1779	1.1804

6.3.3 Comparing LMs with GLMs: Flathead

As they should be, because they are mathematically equivalent, the year parameters from the linear model of the log-transformed CPUE data and from the generalized linear model of the log-transformed CPUE using the identity link function are identical (*LM:LnCE* vs *GLM:LnCE*; **Table 21**), and thus are not compared in any plots. The GLM analysis of untransformed CPUE data using a log link function, *CELog*, does, however, differ markedly in places from the analyses of log-transformed data, *LnCE*. The other GLM model explored, which used Gamma residual errors, *CEGam*, also differs from the LM but not so markedly (**Figure 28**).

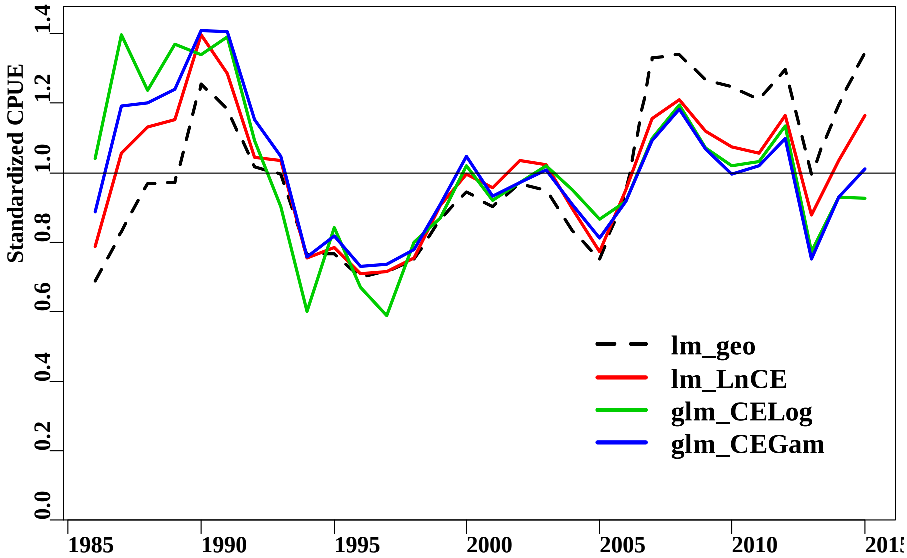

Figure 28. The geometric mean (black dashed), the reference standard LM on LnCE (red), relative to the two GLMs, the first being with untransformed CE using a log-link (green) and the second using Gamma residual errors (blue). Obviously, the x-axis relates to years.

The general trend of all three analyses (*LnCE*, *CELog*, and *CE_Gam*) is similar in that prior to 1992 they are all above the unstandardized annual mean CPUE, between 1992 – 2006 they are closer to the unstandardized trend, and after 2006 they are all below the geometric mean. Among the three trends the *CELog* analysis is more variable and exhibits a few larger deviations in 1994, 1997 and 2005. The *CELog* analysis exhibits more inter-annual variation than any of the other trend, including the geometric mean CPUE (**Figure 28**).

A clearer perception of the differences between the trends can be obtained by a consideration of how much they deviate from the reference CPUE trend (that is, *LnCE*), by subtracting the year parameters (in each case scaled to an overall mean of 1.0) from those generated in the *LnCE* analysis how the different trajectories differ becomes more apparent (**Figure 29**). While it is possible to use a GLM to analyse the untransformed data merely by using a log-link function this leads to a more variable, less stable results, which is expressed through the more jagged appearance of the *CELog* line. This reflects the fact that the log transformation normalizes the data and stabilizes the variance of the CPUE whereas the distribution of the catch and the effort may not provide a reasonable approximation to log-normality (this may especially be a problem if there is marked rounding of catch and or effort values as illustrated in Chapter 4).

When the 95% confidence intervals around the mean estimates are compared to determine any overlap the *CELog* analysis (using untransformed CPUE with a log-link) only overlaps with the reference *LnCE* trend in 18 years out of 29 (the first years cannot be compared, although they are clearly different in this case (**Figure 37**). The *CEgam* analysis, on the other hand, overlaps in 22 years out of 29.

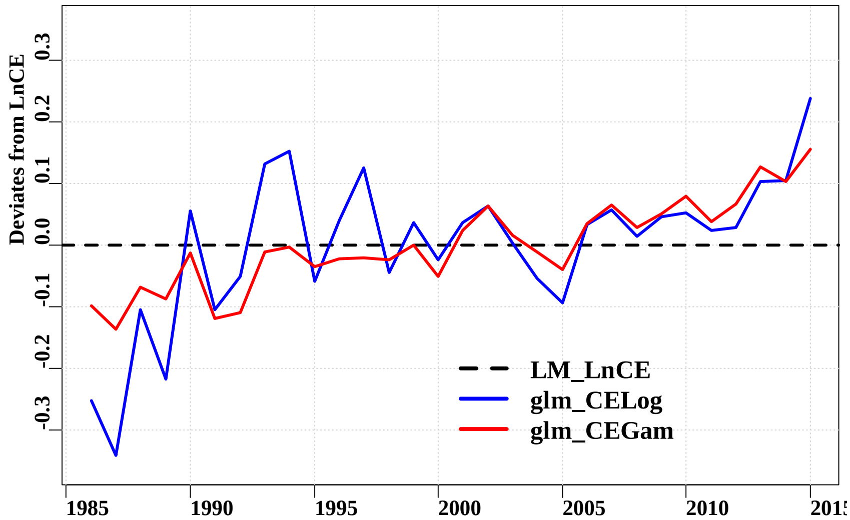

Figure 29. The differences between the reference linear model on the log-transformed data (*LnCE*; the flat dashed line along zero), and the two GLM models, one in untransformed data using a log-link (blue), and the other using the Gamma distribution instead of log-normal (red). Obviously, the x-axis relates to years.

6.3.4 Comparing LMs with GAMs: Flathead

Much better agreement between standardized trends occurs when the Linear Model using log-transformed CPUE data (*LnCE*) is compared with the analyses using GAMs. The GAM which put a smoothing function onto Latitude overlapped the reference analysis in 28 out of 29 years (and the year without overlap was just barely different), with the remaining three GAMs overlapping in all 29 years (**Figure 30**, **Figure 31**, **Figure 32**, and **Figure 37**; **Table 21**) .

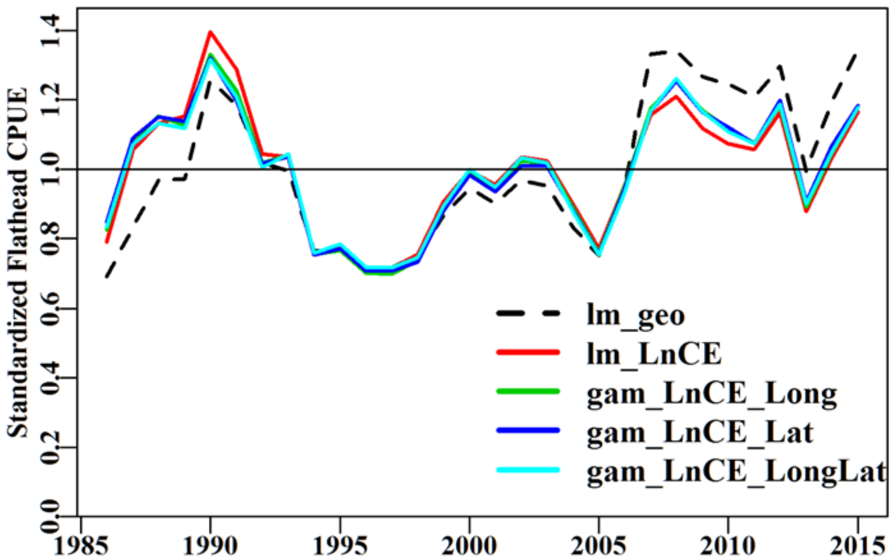

Figure 30. A comparison of the year indices from the usual standardization, the geometric mean, and the two GAMs involving longitude and latitude, and the GAM using both Latitude and Longitude. The GAMs are almost exactly on top of each other with slight deviations at the second and third decimal place being apparent in **Table 21**. Obviously, the x-axis relates to years.

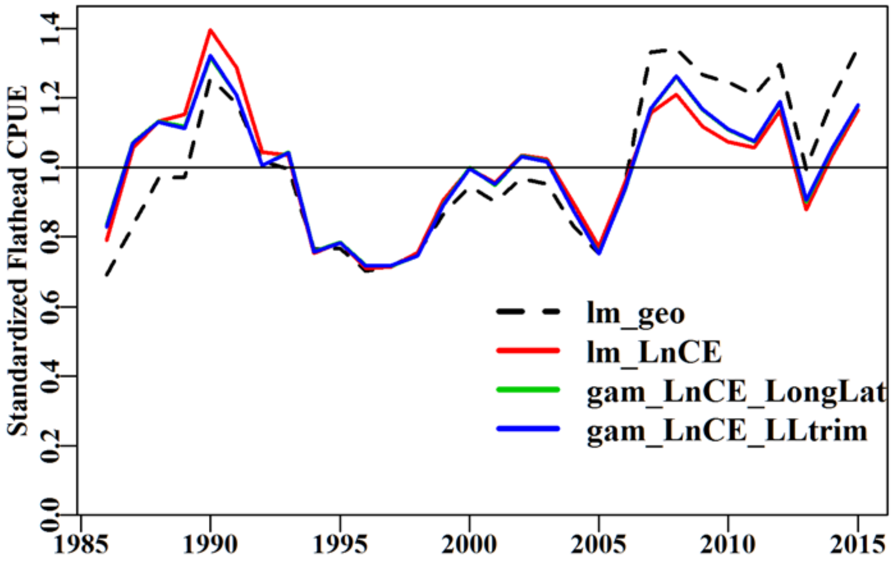

Figure 31. A comparison of the year indices from the usual standardization, the geometric mean, and the two GAMs involving both longitude and latitude together, one with the data trimmed to reduce edge effects. The two GAMs are almost exactly on top of each other with very slight deviations at the second and third decimal place being apparent in **Table 21**. Obviously, the x-axis relates to years.

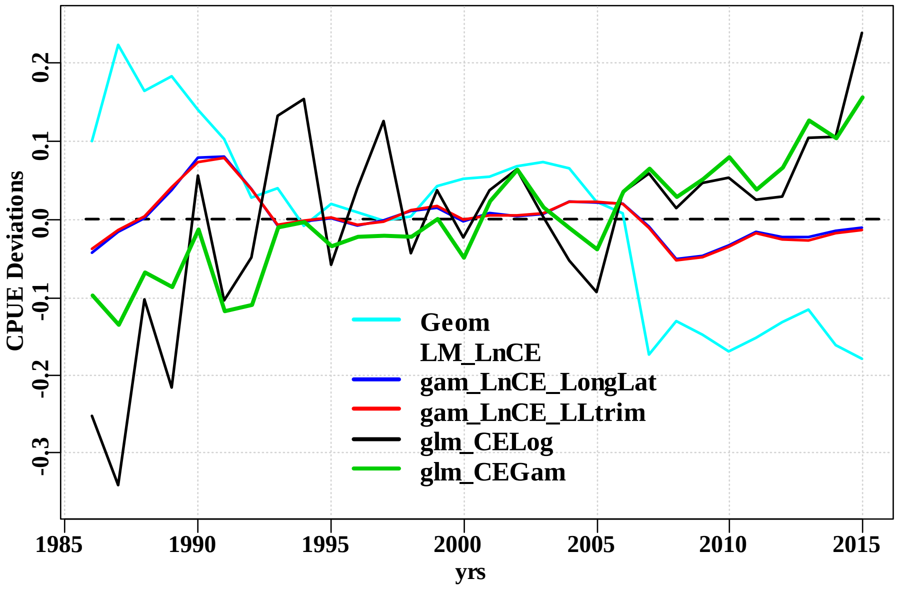

Figure 32. A comparison of the deviations of each standardization using a GAM relative to the reference analysis of *LnCE*. The pale blue line is the unstandardized geometric mean CPUE. The slight differences between the *gamLL* and the *gamLLTrim* are more apparent in this format. Obviously, the x-axis relates to years.

The GAMs operate by fitting a locally smooth relationship between the effect of the variable being smoothed and the CPUE (or, in this case log-transformed CPUE). Plotting these smooth curves provides information on the relative expected effect on predicted mean CPUE (**Figure 33**and **Figure 34**). When working with location data that has this resolution then it would be most appropriate to use both Longitude and Latitude even though their influence on CPUE is, in this case, correlated.

The contour plot for the GAM using all available data has some implausible features off the coast in areas of sparse, possibly erroneous data points. Removing many of those outlying points has improved the patterns apparent in the contour plots so that they are far fewer implausible features (**Figure 34**). Even so, the effect on the year parameters, which are the point of the analysis, remains very minor (**Figure 30** and **Figure 31**).

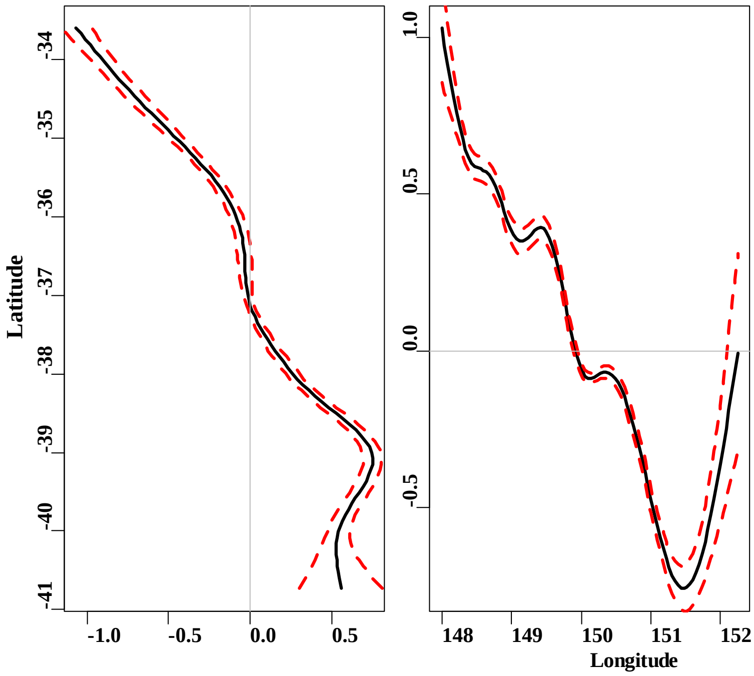

Figure 33. The smoothing trends from the GAMs using Latitude and Longitude separately. A very strong correlation is apparent between the two.

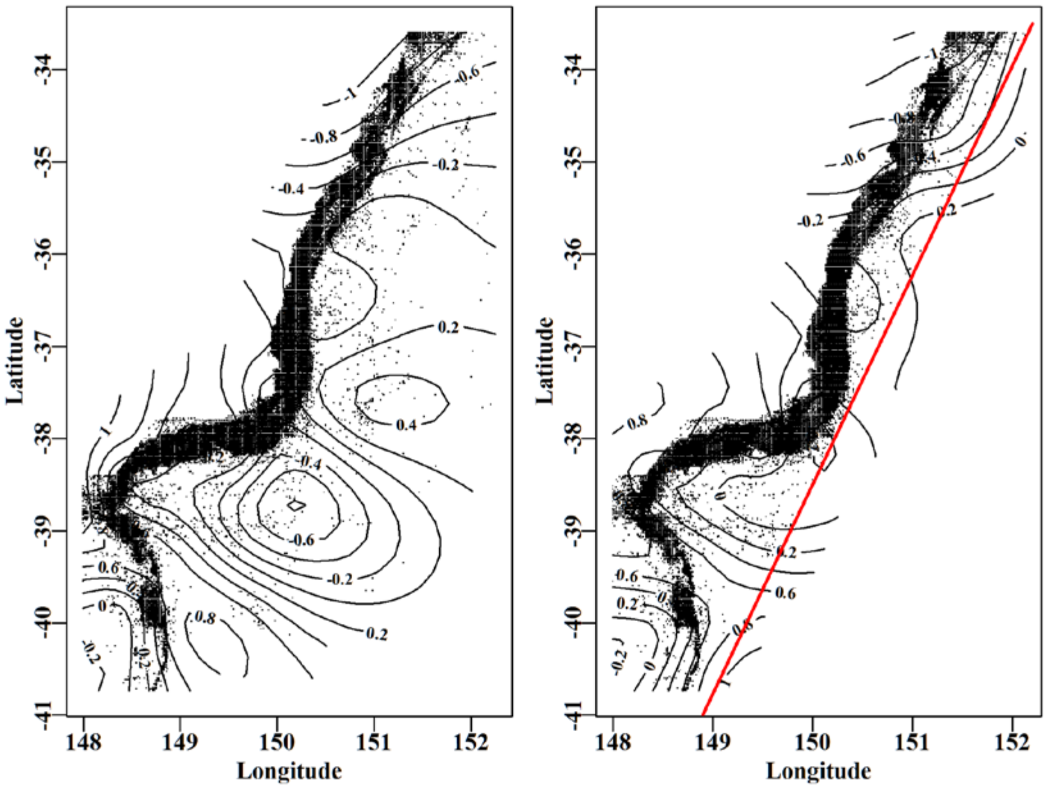

Figure 34. On the left, the GAM smoother from all longitude and latitude data for Flathead taken by trawl in zones 10 – 20 from 1986 – 2015, the horseshoe of points south of -38° reflecting the 'Horseshoe' fishing region to the east of Bass Strait. The dark mass of fine points the 270,000+ reported locations for the start of each trawl and the contours relate to the smoothing parameters with positive values around the Horseshoe with negative values into NSW and down along the Tasmanian east coast. The plot on the right is a repeat except all data points to the west of the red line have been deleted as being implausibly deep or too far offshore.

6.3.5 Standardizing CPUE or Catch: Flathead

Structurally the LM and GLM models compared are very similar although how they handled the CPUE's variance structure certainly differed. Most approaches used a log-transformation in an attempt at normalizing the distribution of the data, or of the mean estimates for each factor. However, the *LnC* model, standardized the log of catches instead of catch over effort, and included the log of effort as an offset on the right hand side of the model's equation. Mathematically this may appear to be the same, but in reality, the form of variation for catch and effort separately tends to be very different to that of CPUE (**Figure 35**). These differences are primarily a reflection of the rounding of the continuous values of catch and effort discussed in sections 4.3.1, 4.3.3, 4.3.4 in the Results section of Chapter 4. Making a ratio of even rounded data can improve the ability of a standard statistical distributions to describe the data. Even with the rather different variance structures between the *LnCE* and *LnC* approaches the predicted mean trajectories are very similar (**Figure 36**).

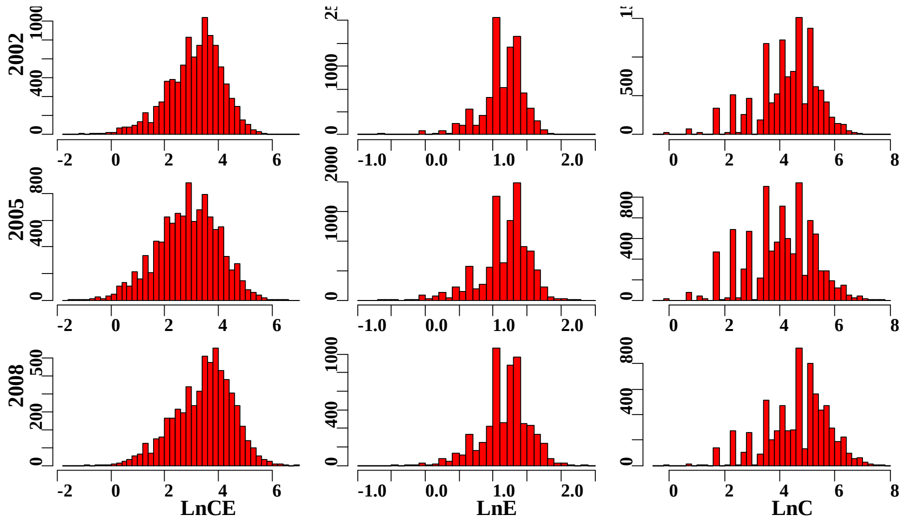

Figure 35. A comparison of the observed distributions of the natural log of catch-rates, the natural log of effort (total hooks), and the natural log of catches from Blue-Eye Trevalla taken by the auto-line fishing method in the SESSF.

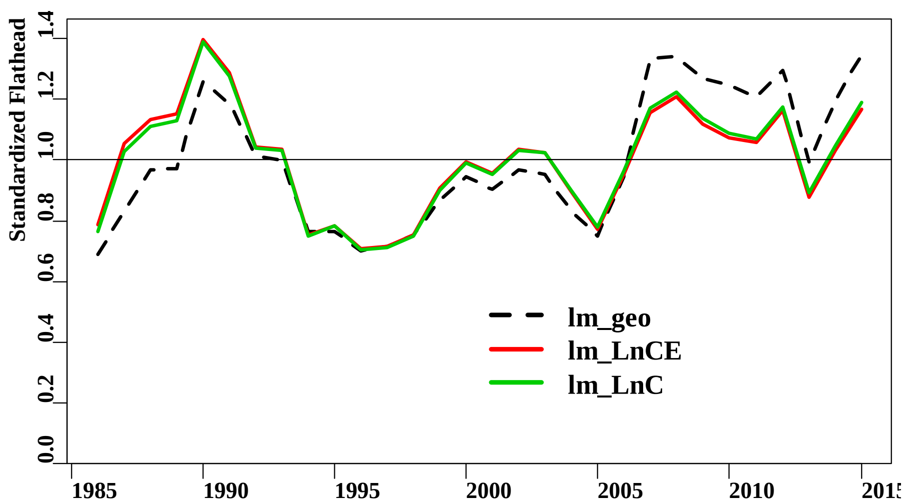

Figure 36. A comparison of the year parameter trajectories for the *LnCE* reference against the unstandardized geometric mean (*lm_geo*) and the standardized Catch trajectory (*LnC*). Only very slight differences are apparent before 1990 and after 2006.

As a further confirmation of the effective equivalence of the *LnCE* and *LnC* lines they overlap relatively tightly over all 29 comparable years (**Figure 37**).

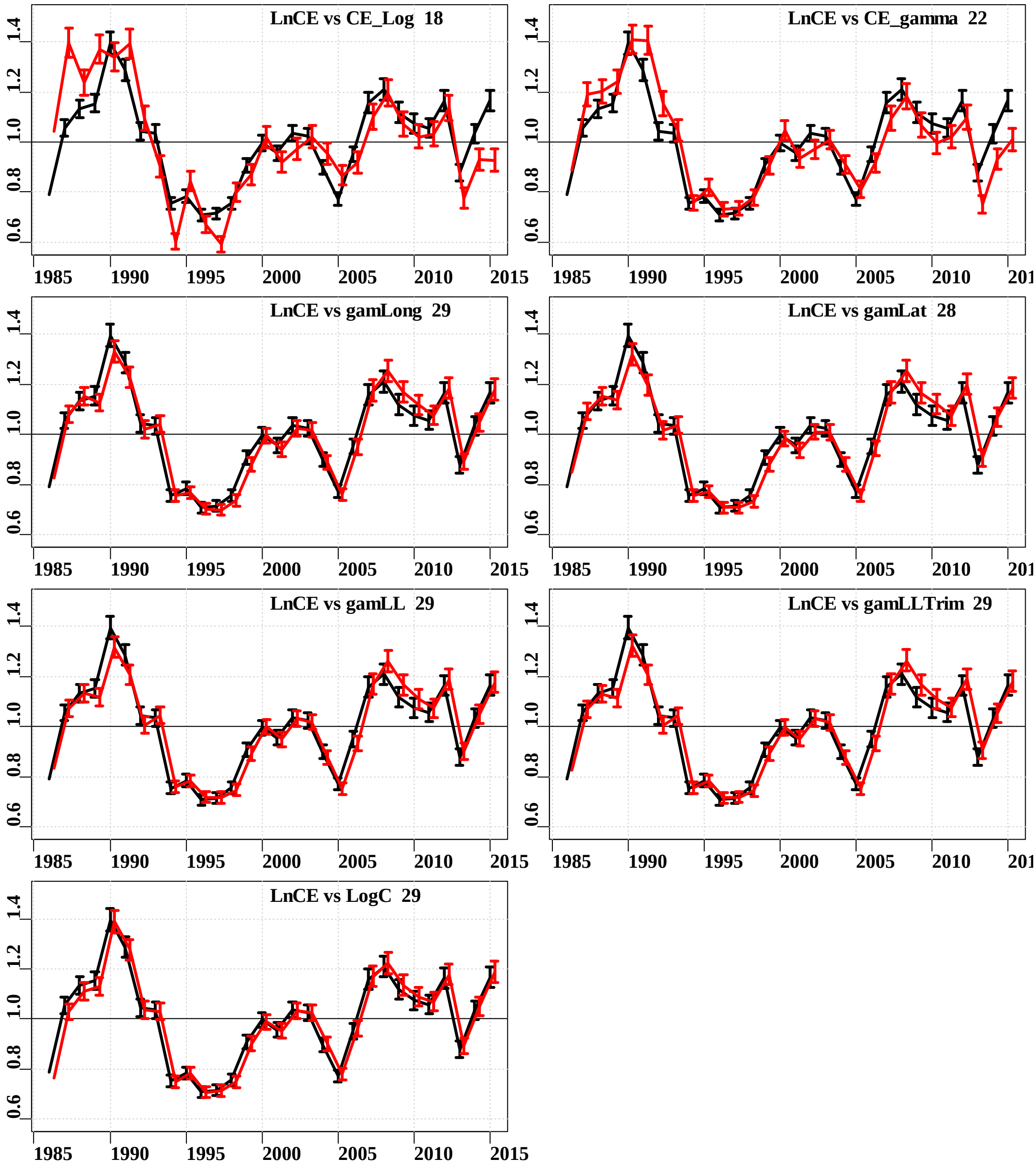

Figure 37. A comparison of each alternative analytical strategy with the reference analysis based upon the simple linear model of the log-transformed CPUE data (*LnCE*). In each case the year parameters are plotted alongside each other with *LnCE* trajectory always depicted in black and the comparison in red. The comparison trend is offset by 0.3 of a year to make comparison of the 95% confidence intervals easier. The number at the end of each plot's name is the number of years out of 29 where there is overlap of the confidence intervals. The y-axis, in each case is the standardized CPUE and, obviously, the x-axis relates to years.

6.3.6 The Standardization of Blue-Eye

Blue-Eye Trevalla (*Hyperoglyphe antarctica*) has always been fished primarily using line-catching methods (drop-line and now auto-line). Trawling has also invariably landed Blue-Eye when taken but reported catches never attained the degree of the line methods (**Table 22**). As shown by Haddon (2016b), the records for auto-line only begin to become representative of the wider area of the fishery from 2002 onwards. For this reason, all analyses will be restricted to that period.

Table 22. The number of records and catches per year for auto-line, drop-line, and trawl vessels reporting catches of Blue-Eye Trevalla from 1997 – 2015. Data filters were to restrict the fisheries included to SET, GAB, SEN, GHT, SSF, SSG, and SSH. Methods were limited to AL, DL, TW, and TDO. Finally, only CAAB code = 37445001 that identifies *Hyperoglyphe antarctica* were included. This table is a direct copy of Table 1 from Haddon (2016b). The grey cells indicate the years used.

Year	AL Catch	AL Record	DL Catch	DL Record	TW Catch	TW Record
1997	0.267	3	271.942	575	104.567	1500
1998	27.253	50	343.505	738	82.074	1398
1999	61.590	77	377.032	971	100.329	1712
2000	90.932	93	384.409	1075	95.042	1893
2001	47.884	76	335.873	797	90.218	1809
2002	134.067	234	223.074	619	67.998	1548
2003	219.676	487	221.649	587	28.918	1210
2004	329.608	1338	158.491	515	48.767	1558
2005	301.303	1142	93.779	363	42.969	1169
2006	354.582	1087	114.639	327	66.105	924
2007	455.097	667	46.011	127	38.321	834
2008	281.384	612	15.549	76	36.046	806
2009	325.893	578	30.158	105	39.386	618
2010	236.620	488	42.023	225	43.480	647
2011	267.318	562	59.381	230	23.268	624
2012	217.816	465	34.107	119	10.792	424
2013	190.515	360	7.762	47	22.893	358
2014	227.041	305	10.242	68	29.381	340
2015	198.232	282	46.711	92	25.128	301

The current approach to the stock assessment of Blue-Eye is to use a SESSF Tier 4 harvest strategy, which uses an empirical harvest control rule, based around catches and standardized CPUE (Little et al, 2011; Haddon, 2016c). The standardized catch rate is therefore an important component. Once again the usual procedure is to treat all included factors as categorical and fit a linear model to the log-transformed CPUE data (Haddon, 2014). Here, the CPUE data (as kg-per-hook) from auto-line vessels was used by selecting years, 2002 – 2015, depths between 200 – 600 m, but restricting the analysis to SESSF zones 20 – 50, where the fishery was relatively well developed early on. Development of the fishery into the GAB came a few years later. Using these data the optimal model was (see equation (12); **Figure 38**):

$$LnCE \sim Year + Vessel + Month + Zone + DepCat + Month{:}Zone \qquad (12)$$

With only 8308 records (relative to 270,000+ records for Flathead only in zones 10 and 20) the confidence limits around the mean CPUE estimates are much wider than those apparent in Flathead (compare **Figure 25** with **Figure 38**). For Blue-Eye the standardization has the effect of flattening the geometric mean CPUE with the outcome having the appearance of a noisy but relatively flat line.

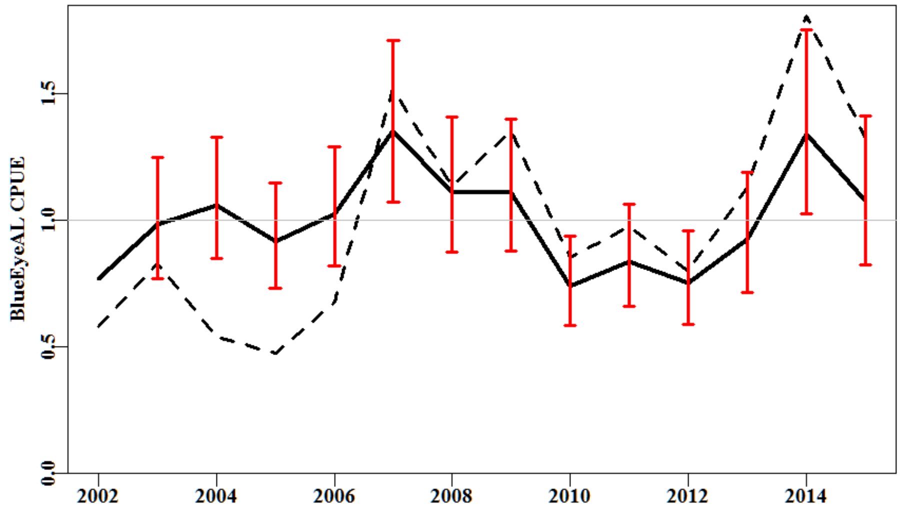

Figure 38. The reference standardization for Blue-Eye Trevalla caught by the auto-line method from the years 2002 – 2015 for SESSF zones 20, 30, 40 and 50. The dashed black line is the un-standardized geometric mean CPUE while the solid black line is the optimum standardized model with red coloured 95% confidence intervals around the mean estimates (see **Table 26**).

The diagnostic plot for the *LnCE* model exhibits standardized residuals that deviate away from that expected from a fully normal distribution (**Figure 39**, **Figure 40**).

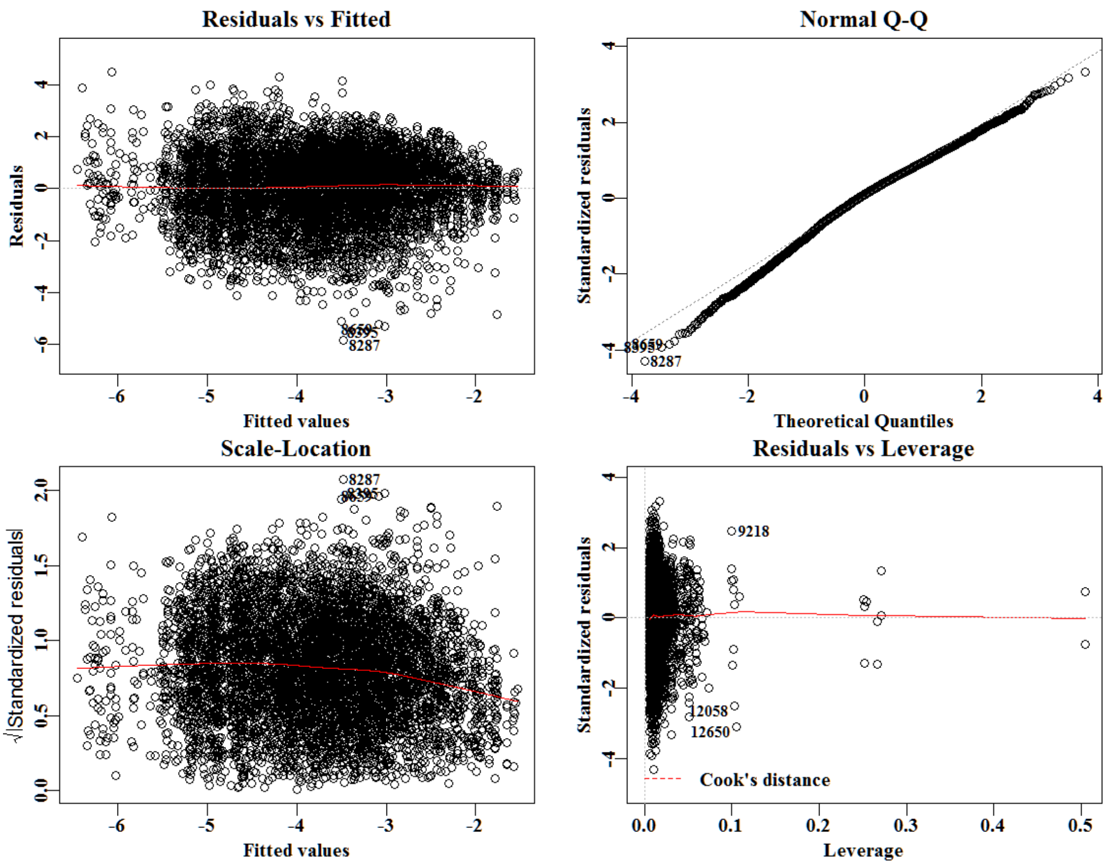

Figure 39. Classical diagnostic plots for a linear model, here the model is LnCE ~ Year + Vessel + Month + Zone + DepCat + Month:Zone applied to Blue-Eye taken by auto-line from zones 20, 30, 40, and 50 over the years 2002 – 2015.

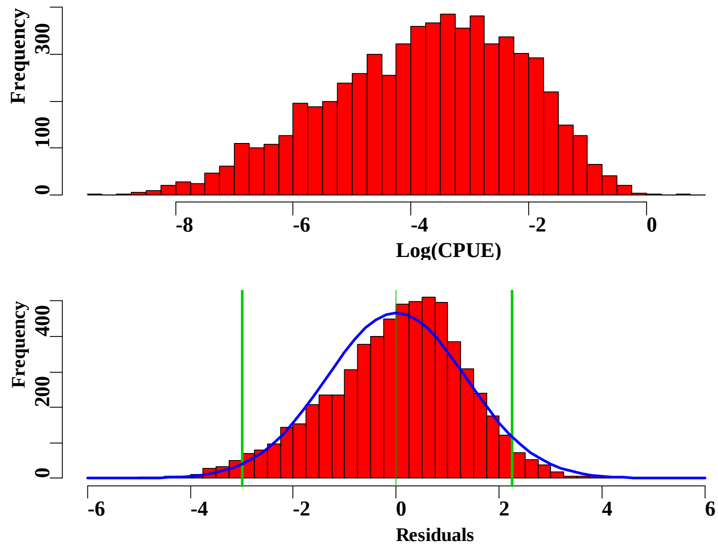

Figure 40. The distribution of log(CPUE) and of the residuals from the *LnCE* model with a fitted normal distribution (blue line) having a mean = 0.0, and a standard deviation of 1.35. The green lines are the 95% quantiles of the residuals and the thin green line denotes zero.

The qqplot in **Figure 39** appears to suggest a normal distribution extending from the right-hand side of the histogram down to residual values around 1.5 – 1.75, after which the model fit deviates from normal. The histogram of Log(CPUE) and of the simple model residuals (**Figure 40**) certainly indicates the data used are not completely normally distributed. Industry members regularly suggest that they are either targeting Blue-Eye or Pink Ling and it may be that the lower CPUE values (left-hand of the Log(CPUE) values) are from the mixed catches where Pink Ling might dominate the catches. There is evidence that the lowest CPUE values are associated with those shots having the least proportion of Blue-Eye (**Figure 41**), indicating that further exploration of the interaction between the two targeted fisheries is needed (**Figure 42**).

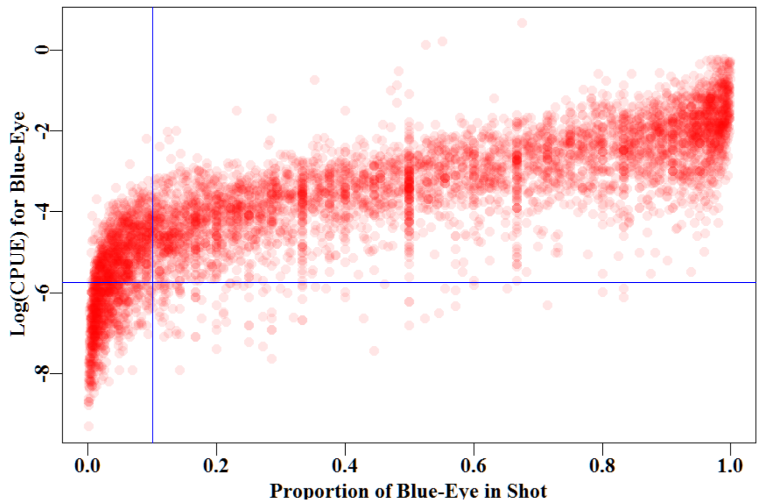

Figure 41. The distribution of natural-log of Blue-Eye CPUE as kg/hook, from auto-line in zones 20 – 50 from 2002 – 2015, as influenced by the proportion of Blue-Eye to Pink Ling in each shot. The colour intensity ranges from 1 observation to 10+ for the most intense red.

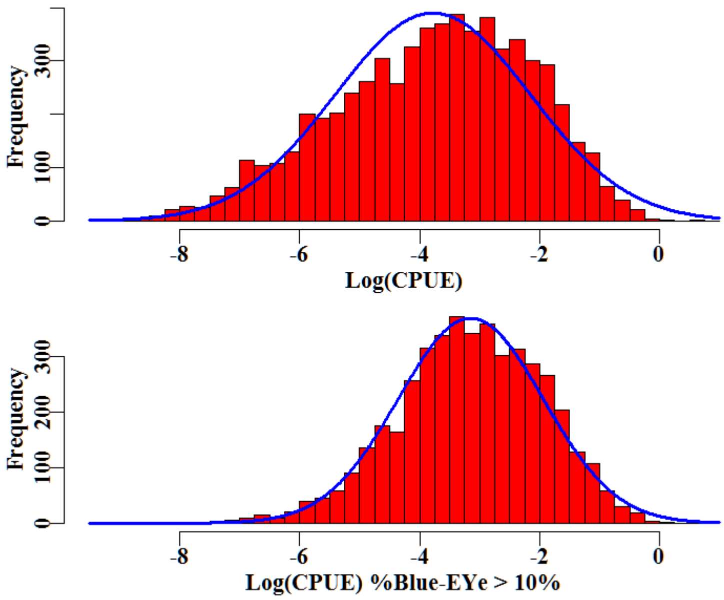

Figure 42. The distribution of Log(CPUE) for Blue-Eye including all data, compared with the same plot except only using those shots where the proportion of Blue-Eye is greater than 10%. The data selection on Blue-Eye proportion increases the naïve geometric mean CPUE, from the fitted normal distributions, from 0.023kg/hook to 0.043kg/hook, an 86% increase and the normality of the log-transformed data improves greatly despite the simplicity of the data selection, which loses some higher catch rates as well as the lower levels (**Figure 41**).

A sensitivity on the reference standardization (*LnCE*) was conducted where the only records used were those where the proportion of Blue-Eye in the combined Blue-Eye and Pink Ling catch was > 10% (**Figure 42**, **Figure 43**). This is an overly crude selection criterion (especially as the average proportion of Blue-Eye in the first few years is lower than in the later years) but suffices to illustrate that the two trajectories effectively overlap; even in 2006 their confidence intervals around their respective means would strongly overlap. Even the reduction from ~6300 records to ~4400 records after the data selection was not enough to significantly alter the standardized trajectory.

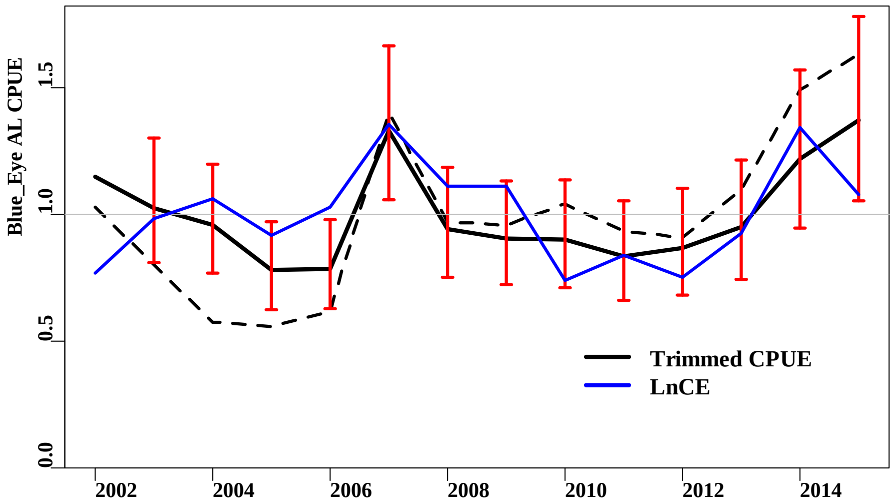

Figure 43. A comparison of the trajectory (black line, red 95% confidence intervals) obtained by selecting only those records containing >10% Blue-Eye (relative to Pink Ling), with the reference CPUE trajectory (*LnCE*; blue line).

Most of the effect of trimming those records containing less than 10% Blue-Eye (when considering the total catch of Blue-Eye and Pink Ling) occur in the years from 2002 – 2005. This reflects a large change in the average proportion of Blue-Eye in shots containing both over the same period (**Figure 44**). This was also a period of large changes in the Pink Ling fishery with the significant catches by an array of methods over the period 1997 – 2003 (**Table 25**); including significant catches using fish traps. How these catches interact with Blue-Eye catches by auto-line is open to further exploration.

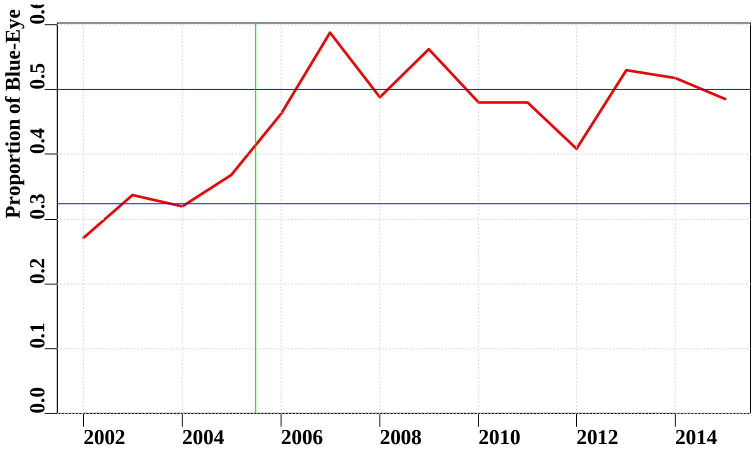

Figure 44. The proportion of Blue-Eye by weight in individual auto-line shots when the total catch of Blue-Eye and Pink Ling are considered. The mean proportion between 2002 – 2005 was 32.4% while the mean proportion from 2006 – 2015 was 50%.

One aspect of trimming data in this way, by selecting only those records with greater than a given proportion of the catches being the species of interest is that if there is a strong trend in the other species, this might influence the outcome of the standardization of the CPUE for the species of interest. Thus, it is a sensible strategy at least to plot how the proportion of records eliminated changes through time (**Figure 45**). The correlation between the standardization and the proportion of records eliminated through time was nowhere near significant and any potential influence of trends in that proportion through on the standardization would appear to be negligible.

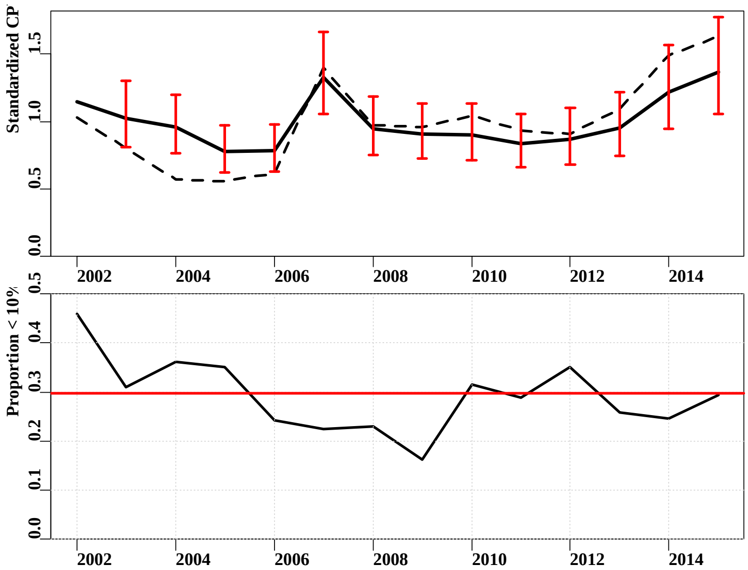

Figure 45. The standardization of the trimmed Blue-Eye by auto-line CPUE data (see **Figure 43**) compared with the proportion of records which contain < 10% of Blue-Eye by combined weight of Blue-Eye and Pink Ling. The correlation between the standardization line and the proportion of records = -0.0198, which is not significant.

6.3.7 Including the Ratio of Blue-Eye:Pink Ling

A second sensitivity was conducted whereby the proportion of Blue-Eye relative to Pink Ling in each catch was categorized into 10 levels (0 = 0 – <10%, 1 = 10 – <20%, 2 = 20 – <30% …., 10 = 100%) and that factor included in the standardization either using all records or after excluding records with <= 10% Blue-Eye (pBlue = 0). The outcome was a greatly improved model fit with much more of the variation in the data accounted for and much tighter residuals (**Table 23**).

Table 23. Statistics from four statistical models for Blue-Eye auto-line CPUE standardization With pBlue means the proportion of Blue-Eye was included. All means that records with < 10% Blue-Eye were included. df = degrees of freedom.

	Residual Standard Error	Degrees of Freedom	Multiple R-squared	Adjusted R-squared	F-statistic	F df
All No pBlue	1.337	8090	0.3053	0.2952	30.13	118
All with pBlue	0.8389	8080	0.7270	0.7227	168.1	128
No pBlue	1.051	6268	0.3072	0.2941	23.55	118
With pBlue	0.7856	6259	0.6134	0.6055	78.18	127

Despite the greatly improved statistical model fit from including the proportion of Blue-Eye in each record, the overall trend in CPUE through time was not greatly changed (**Figure 46**) except the first two years of the series when the proportions of Blue-Eye in the catches were relatively lower and, given the highest coefficients are for the top five proportions (**Table 24**) the implied CPUE is increased.

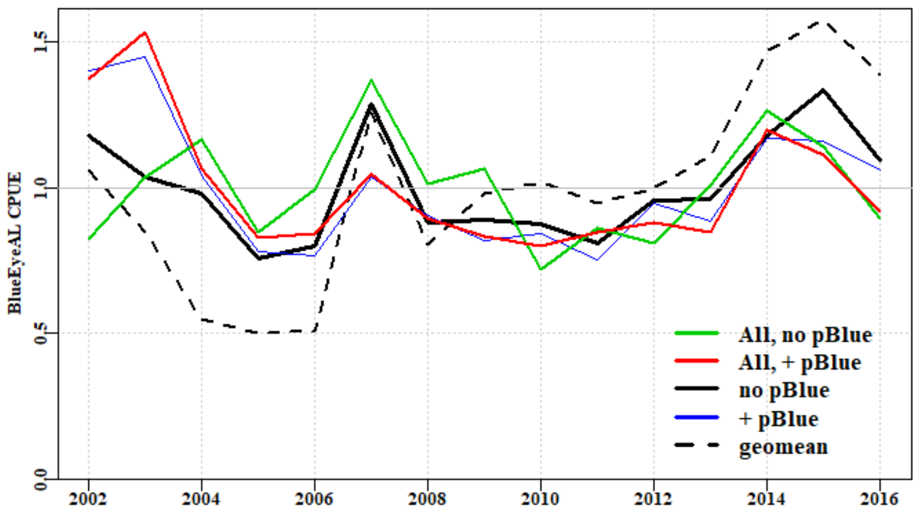

Figure 46. The auto-line CPUE standardizations. All includes all auto-line records between 2002 – 2016, lines without ‘All’ imply that records with <= 10% Blue-Eye were excluded. +pBlue means that the proportion of Blue-Eye is included as a factor (see **Table 23**).

The trends including the effects of pBlue are similar and somewhat flatter, although this is influenced by the higher values in the first two years. In those years the proportion of shots with proportions 90% - <100% was reduced.

Table 24. The standardized coefficients for the pBlue factor. The lowest value occurs when pBlue = 0, where the coefficient is only about that for pBlue = 1, divided by 3.

pBlue	1	2	3	4	5	6	7	8	9	10
Coefficient	0.225	0.325	0.487	0.637	0.705	1.007	1.319	1.682	2.655	0.958

Table 25. The reported catch of Pink Ling (*Genypterus blacodes*) taken in the GHT fishery within the SESSF. AL – auto-line, BL – bottom-line, DL – drop-line, FP – fish trap, GN – gillnet, RR – rod and reel, and TL – trot-line. Detailed data is unavailable prior to 1997 when a new log-book and database was introduced for the GHT fishery, or the South-East Non-trawl fishery (SEN) as it then was.

Year	Unknown	AL	BL	DL	FP	GN	RR	TL
1997	0.030	95.910	9.165	10.309	79.146	48.803		0.529
1998		78.967	5.165	11.718	69.956	21.411		
1999		105.611	2.264	16.909	115.622	25.622		
2000		54.720	25.025	13.749	108.525	27.714		
2001		176.418	38.408	6.945	75.045	20.872		0.104
2002		379.349	3.967	10.928	64.560	14.502		0.275
2003		382.861	0.097	7.382	16.660	23.759		0.216
2004	0.012	730.746	1.174	10.897		0.695		0.195
2005	0.015	524.530	0.441	0.988		0.459		
2006		419.957		1.625		0.397		
2007		294.700	3.610	0.215		0.394		
2008		365.753	4.212	0.466		0.276		
2009		253.500	2.605	0.331		0.088		
2010		318.296	0.584	0.685		0.088		
2011		373.726	1.474	0.788		0.281		
2012		363.861	1.990	0.472		0.257		
2013		242.633	2.350	0.169		0.079		
2014		274.408	5.428	0.338		0.051	0.050	
2015		223.613	2.449	0.099		0.040	0.002	3.789

Table 26. The optimum model predicted year parameters for each model compared for Blue-Eye. All columns are scaled to a mean of 1.0. Geom is the unstandardized geometric mean CPUE, the other column names are listed in **Table 19**.

	Linear			Generalized Linear			Generalized Additive			
Year	**Geom**	**LnCE**	**LnC**	**glmLnCE**	**CELog**	**CEGam**	**gamLon**	**gamLat**	**gamLL**	**gamTrim**
2002	0.5807	0.7678	0.9385	0.7678	0.8735	0.8838	0.7585	0.8451	0.9016	0.9018
2003	0.8283	0.9817	1.2945	0.9817	0.7664	0.8807	1.0145	1.1099	1.2213	1.2404
2004	0.5387	1.0614	1.2052	1.0614	0.6938	0.9398	1.0464	1.0046	1.1178	1.1236
2005	0.4769	0.9165	1.0932	0.9165	0.5968	0.9745	0.9289	0.8984	0.8702	0.8708
2006	0.6755	1.0275	1.2148	1.0275	0.7395	0.8673	1.0224	0.9522	0.9510	0.9517
2007	1.5221	1.3541	1.3264	1.3541	1.2753	1.3119	1.3662	1.3481	1.3428	1.3392
2008	1.1371	1.1092	1.0901	1.1092	0.9834	0.9547	1.0583	1.0600	1.0234	1.0189
2009	1.3548	1.1108	1.1196	1.1108	0.9156	0.9992	1.0420	0.9762	0.9670	0.9603
2010	0.8536	0.7403	0.6919	0.7403	1.0220	0.8176	0.7212	0.7129	0.6976	0.6954
2011	0.9739	0.8371	0.7203	0.8371	0.9470	0.8658	0.8526	0.8654	0.8939	0.8929
2012	0.8005	0.7511	0.6825	0.7511	1.0373	0.9129	0.7130	0.7826	0.7599	0.7647
2013	1.1292	0.9235	0.7602	0.9235	1.1694	1.0151	0.9183	0.9668	0.9212	0.9172
2014	1.8064	1.3409	1.0320	1.3409	1.5446	1.3013	1.4477	1.3608	1.3221	1.3106
2015	1.3224	1.0781	0.8308	1.0781	1.4355	1.2755	1.1100	1.1169	1.0103	1.0124

6.3.8 Comparing LM on LnCE and LnC

Some analysts prefer to standardize the log transformed catch, putting log transformed effort as an offset within the body of the independent side of the model (e.g. Bishop et al, 2004); i.e.

$$LnC \sim Year + Vessel + Month + Zone + DepCat + Month{:}Zone + LnE \qquad (13)$$

As long as the *LnE* is treated as an offset (Venables and Ripley, 2002) the outcome of this analysis is identical to the linear model on *LnCE* using a log-normal residual error structure. This makes sense as the offset *LnE* term relates to the amount of effort used when calculating catch-per-unit-effort. However, if *LnE* is treated as just an ordinary variable then the outcome is very different (**Figure 47**).

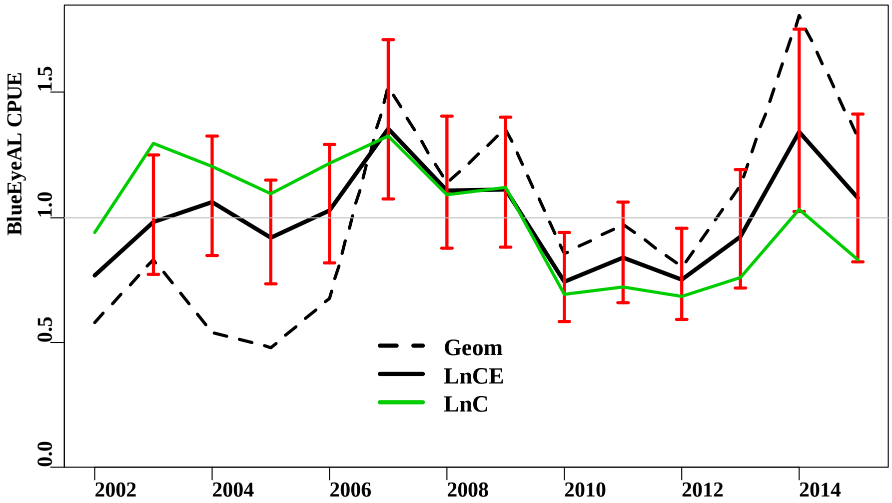

Figure 47. A comparison of the reference standardization *LnCE* with the *LnC*.

The *LnC* standardization (without offset) predicts a mean estimate for each year which departs further from the geometric mean CPUE than the reference standardization (*LnCE*; **Figure 47**), although the two series almost overlap from 2007 – 2010. An examination of the standard diagnostic plots, however, indicates some major differences (compare **Figure 39** with **Figure 48**), primarily because of the use of catch vs cpue.

The rounding of estimated catches of Blue-Eye in the catch and effort logbook database is clearly apparent in the plot of the residuals against the fitted values and direct plots of the log(catch) and log(effort) (**Figure 48** and **Figure 49**). While the qqplot appears to match that from the *LnCE* analysis how data being distributed in such a fashion might affect a standardization whose residuals are based on some smooth probability distribution is unknown.

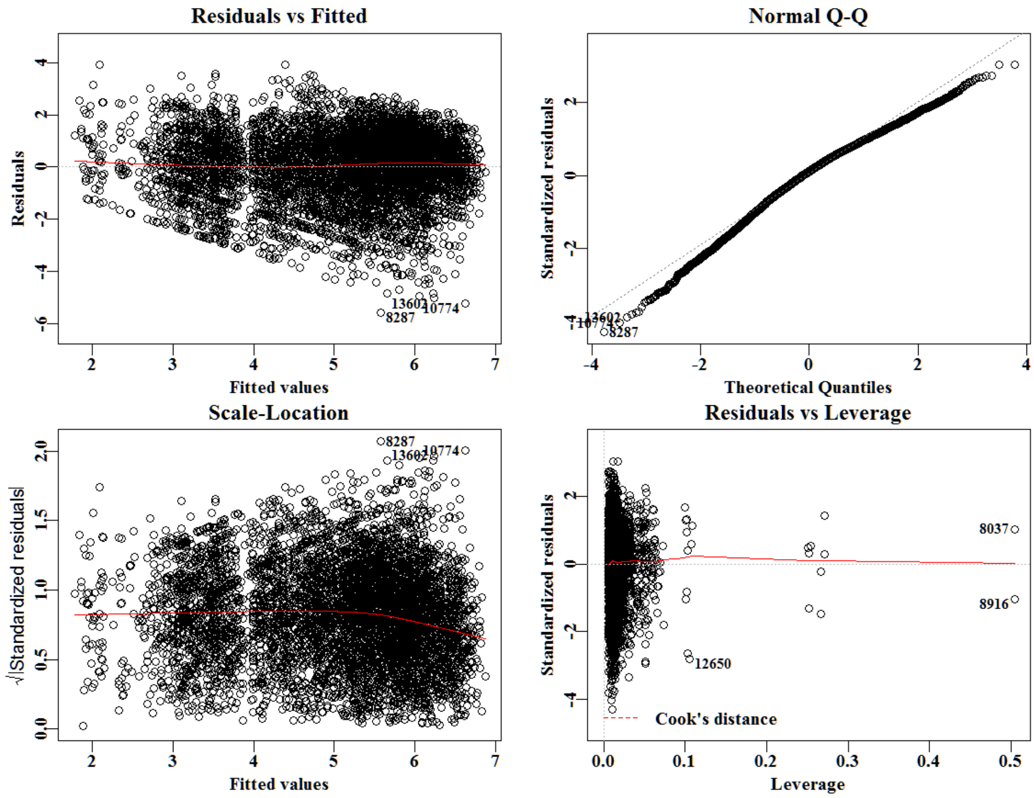

Figure 48. Diagnostic plots for the standardization based on log transformed catches (*LnC*). The rounding of both catches and of the number of hooks leads to the common occurrence of identifiable runs of values through the residuals in the left hand plots.

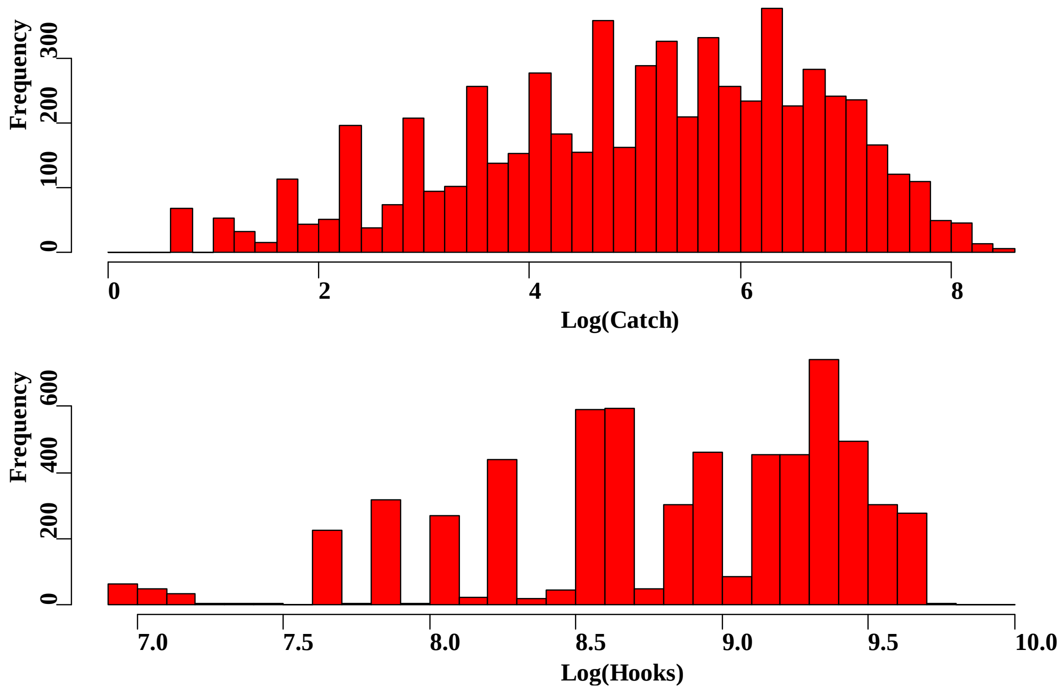

Figure 49. The distributions of reported catches and reported numbers of hooks from the autoline fishery for Blue-Eye in zones 20, 30, 40, and 50, from the years 2002 – 2015. The rounding of values is apparent by the peaks at particular values. This is even more apparent when looking at the untransformed data.

6.3.9 Comparing LMs with GLMs: Blue-Eye

Once again the GLM on LnCE with an identity link is, naturally, identical to the simpler linear model (**Table 26**). The comparison of the reference standardization with the GLM of untransformed CPUE using a log-link (*CELog*) and with the GLM of CPUE with Gamma residual errors (*CEGam*) for Flathead found some differences with the *CELog* scenario. With Blue-Eye there were also differences between the reference standardization (*LnCE*) and the *CELog* but, once again, only relatively minor differences with the *CEGam* analysis (**Figure 50**). The *CELog* analysis more closely follows the unstandardized geometric mean CPUE rather than flattening the 2002 – 2015 trajectory.

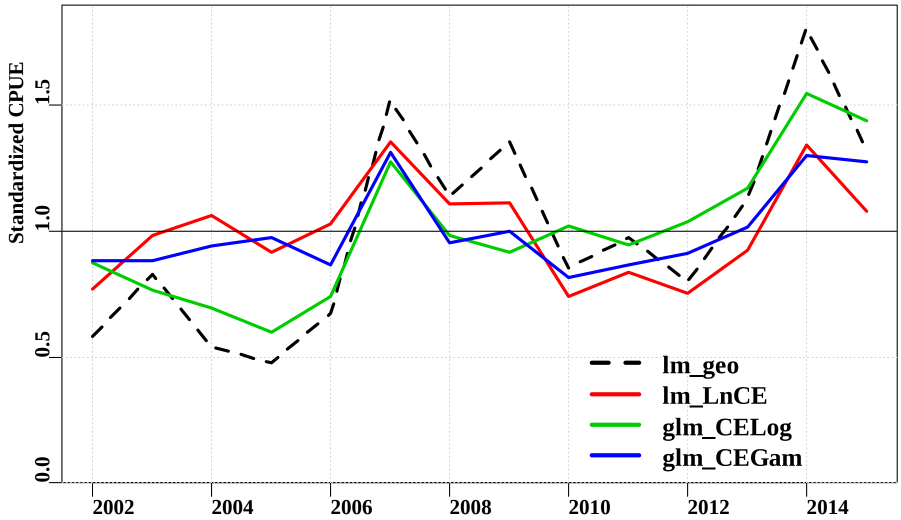

Figure 50. A comparison of the standardized CPUE for the reference *LnCE* standardization with the *CELog* and the *CEGam* scenarios.

The variability of CPUE invariably increases with the mean CPUE (the variances are heterogeneous) and the Linear Model on the log-transformed CPUE (*LnCE*) has advantages over the use of a GLM on the untransformed CPUE using a log-link (*CELog*) in that it normalizes the data while stabilizing the variance. The *CELog* analysis assumes the variance of the original data is stable across different levels of CPUE (Venables & Dichmont, 2004). Using the *LnCE* approach also has the benefit that it is less sensitive to outlying data points (**Figure 51**).

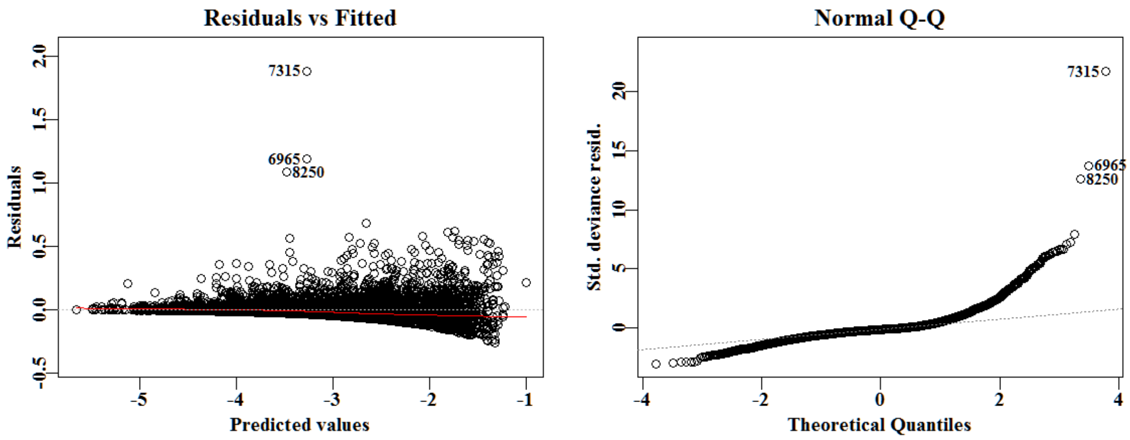

Figure 51. Two diagnostic plots for the *CELog* analysis. The presence of outliers is important and all records with CPUE > 0.5 kg/hook are influential, which is far less the case with the *LnCE* analysis.

If some of the diagnostic plots for the *CELog* analysis are examined some points stand out as outliers and each of these has very high CPUE values relative to the rest (**Figure 52**). There are only three records with CE > 0.8 and 13 records between 0.7 and 0.8.

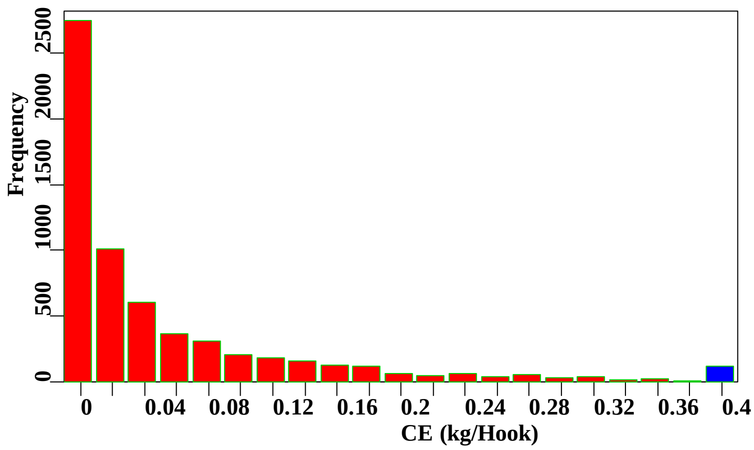

Figure 52. The distribution of CPUE data for Blue-Eye taken by auto-line.

6.3.10 Comparing LMs with GAMs

From 2004 – 2015 the three GAMs (*gamLat*, *gamLong*, and *gamLatLon*) follow approximately the same trajectory, which is generally similar to the *LnCE* analysis overlapping in places (**Figure 53**). The GAM that placed a smoothing function on Longitude (*gamLong*) most closely follows *LnCE*, although with deviations, all trajectories, however, remain close to the reference standardization (**Figure 53**).

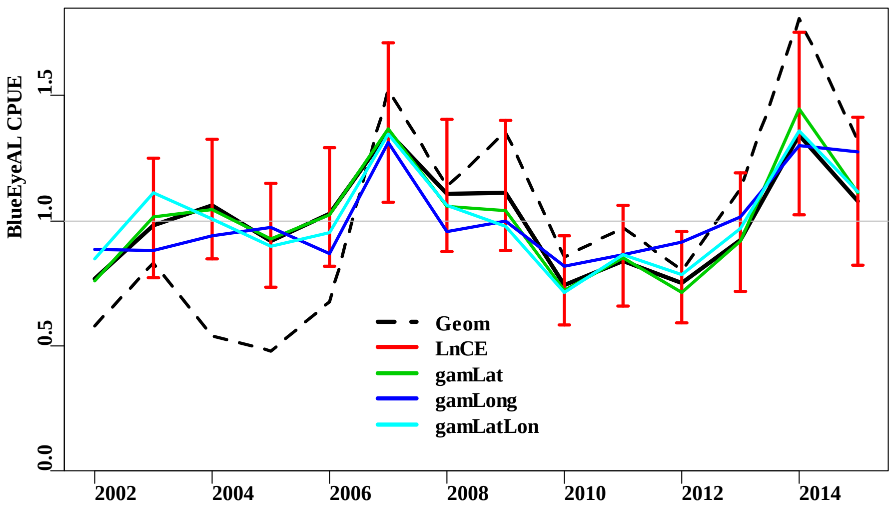

Figure 53. A comparison of the reference standardization with the three main GAM models.

In each case the smoother had a highly significant influence on the standardization (**Table 27**, **Table 28**; **Figure 54**). The large changes in the smoothing term with longitude and with latitude illustrate that in the *gamLong* and *gamLat* standardizations these smoothers added to the spatial information over and above the inclusion of the ‘Zone’ factor. Using the Akaike Information Criteria (AIC) to aid in model selection, the *gam-*

Lat standardization provides a better description of the data than the *gamLong*. However, there is an almost equal improvement (reduction) in the AIC when both longitude and latitude are combined in the non-parametric smoothing term (**Table 28**).

Table 27. The parametric terms from the GAMs. 'df' is degrees of freedom, 'F' is the F-statistic, and 'p' is the probability. When both longitude and latitude are included in the model the 'Zone' factor, and its interaction, was omitted as redundant.

		gamLong		*gamLat*		*gamLatLon*		*gamLLtrim*	
	df	F	p	F	p	F	p	F	p
Year	13	8.33	< 2e-16	7.28	2.36E-14	8.76	<2e-16	8.77	<2e-16
Vessel	11	62.75	< 2e-16	81.37	< 2e-16	64.55	<2e-16	65.50	<2e-16
Month	11	19.75	< 2e-16	12.77	< 2e-16	37.90	<2e-16	37.63	<2e-16
DepCat	15	1.56	0.0756	1.28	0.208	1.23	0.238	1.26	0.218
Zone	3	10.48	7.16E-07	8.78	8.29E-06				
Month:Zone	33	5.09	< 2e-16	4.39	9.22E-16				

Table 28. The non-parametric terms from the GAMs, with 'edf' being the effective degrees of freedom, 'Ref.df' being the reference degrees of freedom, dfAIC being the effective degrees of freedom for the AIC statistic, and 'AIC' being the Akaike's Information Criteria (smaller is better).

Model	edf	Ref.df	F	p	dfAIC	AIC
gamLong	8.89	9.00	49.32	< 2e-16	96.89	21449.2
gamLat	8.74	8.98	108.89	< 2e-16	96.74	20962.1
gamLatLon	27.32	28.79	63.15	< 2e-16	79.32	20574.3
gamLLtrim	26.89	28.68	63.73	< 2e-16	78.89	20529.2

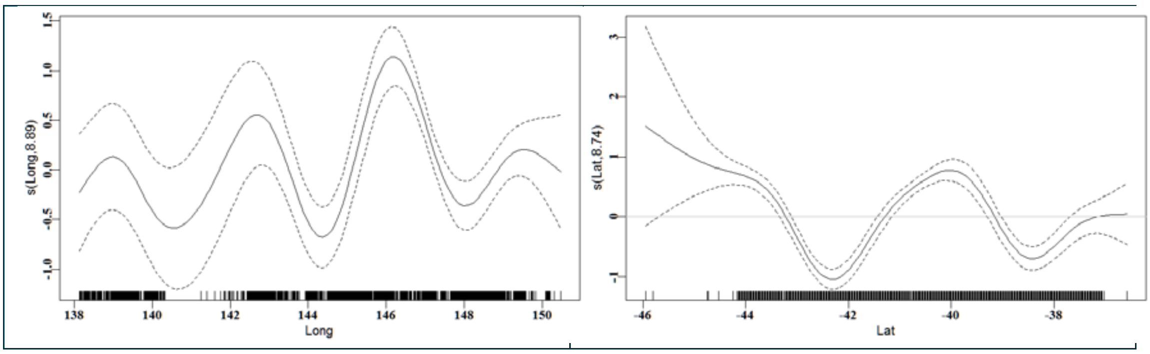

Figure 54. The smoothers from *gamLong* (left-hand) and *gamLat* (right-hand) exhibiting marked contrast in influence across the range of the fishery. The 'rug' along the bottom of each graph illustrates the density of data points.

When a smoothing term was placed onto both longitude and latitude at the same time the smoother produces a two-dimensional plot of the relative influence of the smoother on the log transformed CPUE (**Figure 55**). While the details of the surface can be affected by even a few outlying points, which the log-book location data invariably has, these outliers tend to have little influence on the final CPUE trend (**Figure 56**).

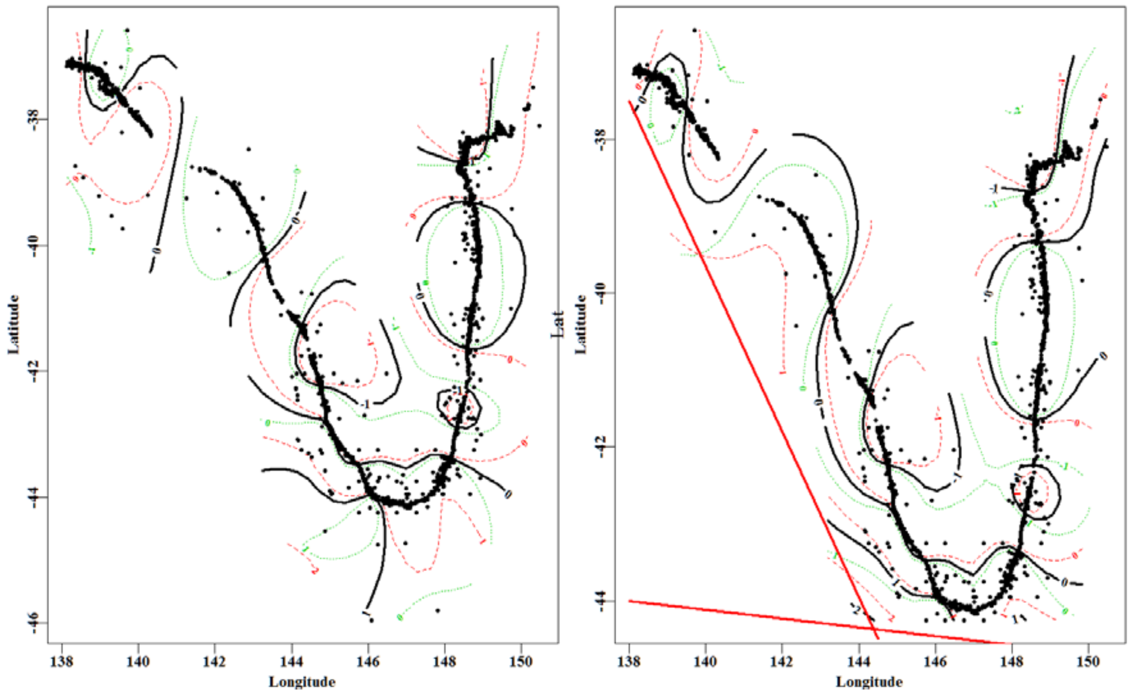

Figure 55. Two versions of the gamLatLon GAM standardization. The left-hand side is the analysis using all data, some of which is obviously out over the abyssal plain and some has been reported to have been caught on land. The right-hand plot illustrates the effect of removing the few points that have been reported to the right and below the red lines (however, the few points on land remain). The removal of just a few points helps linearize the trend along the coastline.

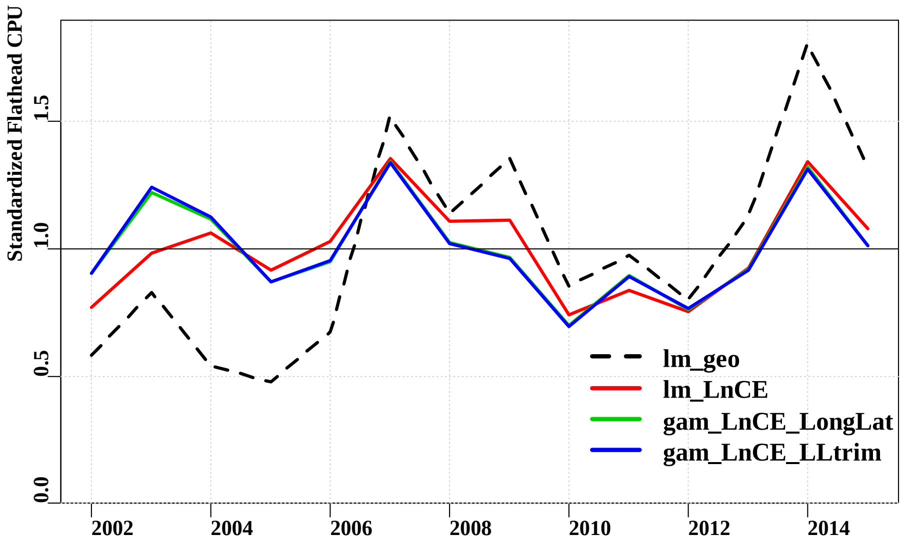

Figure 56. A comparison of the *gamLonLat* with the *gamLLtrim* illustrating that while the appearance of the smoother has altered (**Figure 55**) the effect on the standardization trend in Blue-Eye by auto-line is minimal with just slight differences apparent in 2003.

It remains difficult to determine which of the approaches could be characterized as the 'best'. A further consideration that could be used in such a decision is the spread and distribution of the residuals after an optimal statistical model has been fitted (**Figure 57**). The distributions of residuals from the standardizations that rely on untransformed CPUE (*CELog* and *CEGam*) have a wider spread than those that use log-transformed CPUE and, at the same time, are skewed more to the left away from the zero point.

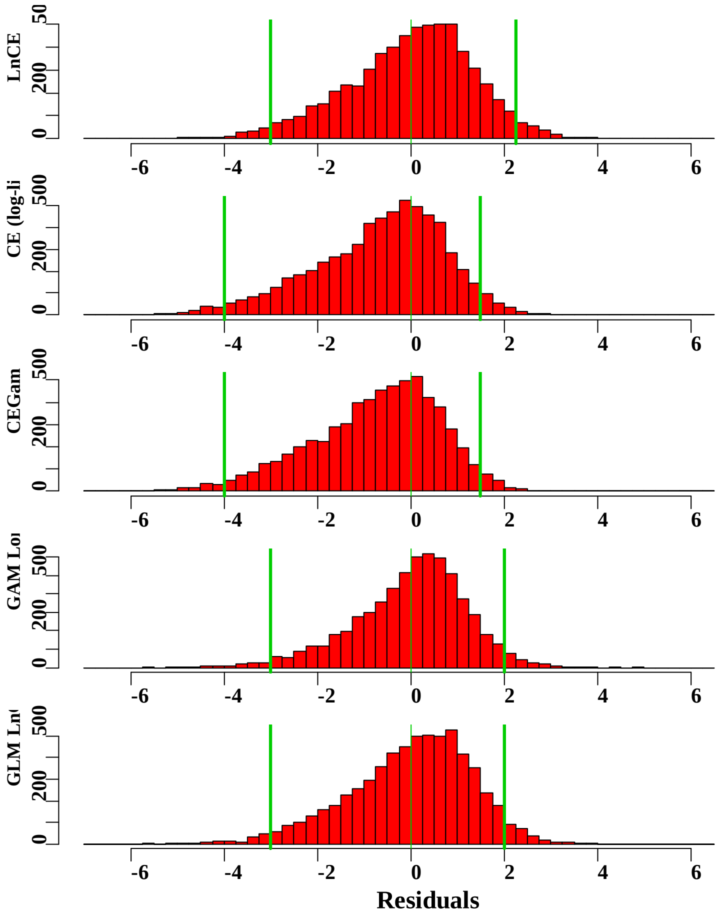

Figure 57. A comparison of the distribution of the residuals from the reference *LnCE*, the *CELog*, the *CEGam*, the *gamLonLat*, and the *LnC* standardizations. The thick vertical lines are approximations to the 95% quantiles from the distributions as described by each histogram. The fine central green line merely denotes zero. The approximate 95% quantiles for *LnCE* was -3 – 2.25, for *CELog* and *CEGam* was -4 _ 1.5, and for *gamLonLat* and *LnC* was -3 – 2.0.

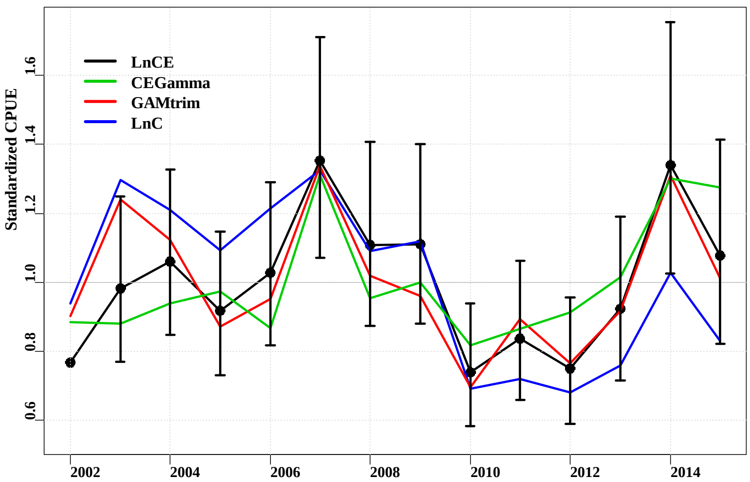

Figure 58. The reference standardization relative to the *CEGam*, the *gamLLtrim*, and the *LnC* models. The vertical black lines are the 95% confidence intervals around the *LnCE* model. Only the *LnC* model falls outside the 95% intervals, and that only in 2003. The *gamLLtrim* follows the reference model most closely, except in the first two years and in 2008 and 2009.

6.4 Discussion

6.4.1 Introduction

The number of CPUE standardization approaches considered in this chapter was only a sub-set of those available but encompassed those that get used most often. The two main species examples considered, Flathead (*Neoplatycephalus richardsoni*) and Blue-Eye (*Hyperoglyphe antarcticus*), cover the spectrum of extremely data-rich, with 100,000s of records and relatively data-limited with only 8308 records across 2002 – 2015. When the number of records is very large (270,000+ for Flathead) every factor used becomes statistically significant and the apparent precision of the mean estimated CPUE each year is very tight. However, the statistical models being used assume that each year's sample is independent of every other year, which, when working on the same population through time, is obviously invalid. The analyses involving such large numbers of records become very robust and stable but there can be inter-annual variation much greater than is biologically plausible if the CPUE is considered to linearly reflect relative abundance. For example, with Flathead from 1986 – 1990 the standardized CPUE, based on the simplest linear model changes from about 0.8 – 1.4 (about a 60% increase) and was followed by a decline from 1990 – 1994 down to 0.75. Such large ups may possibly be due to above average recruitment occurring multiple times, but the following decline, which occurred in a period of relatively low catches (**Figure 25**; page 87), does not seem plausible given the biology of the species and the very high recruitment suggested by the large rise. This suggests there are other inter-annual factors at play of which we remain ignorant. The true variability of CPUE needs to take account of this inter-annual variation as well as the estimation uncertainty that derives from the sample size.

The confidence intervals around the mean CPUE estimates become intuitively more plausible when considering the data-limited case of Blue-Eye Trevalla. Examination of **Figure 38** and **Figure 43** (pages 97 and 99) suggest that despite the apparent ups and downs of the Blue-Eye CPUE trajectory in each case, the only years that differ significantly from the long term mean of 1.0 are 2007 in both cases, and 2014 and 2015 in each year respectively.

An important conclusion drawn for all CPUE standardizations is that the use and display of diagnostic plots and statistics should become routine. This may appear to be an obvious statement but it needs to be made as the dearth of such diagnostics in the grey literature should emphasize that such documentation is not yet as routine as it should be.

Prior to standardization the characterization of the available data is a necessity so as to clarify the coherency of the data through time and qualitatively identify any particular changes in the data quality. The interactions between Blue-Eye and Pink Ling provide many examples where understanding the character of the original data helps in the interpretation and in making decisions on how best to analyse each fishery.

6.4.2 The Assumptions of LM and GLM

The application of generalized linear models and related methods to the standardization of fisheries catch-effort data breaks a number of the basic assumptions of such linear models. Venables and Ripley (2002) list the following assumptions for GLM:

- There is a response y observed independently at fixed values of stimulus variables $x_1, \ldots, x_p$,

- the stimulus variables may only influence the distribution of y through a single linear function called the *linear predictor* $\eta = \beta_1 x_1 + \ldots + \beta_p x_p$

where the predictor variables are expected to be mutually independent or orthogonal to each other.

In real world fisheries one would never expect the stimulus or explanatory variables to be at fixed or controlled values and in fisheries it is also rare for those explanatory variables to be fully independent of each other. Fortunately, these classical statistical methods are not always affected unduly by deviations from these assumptions, although some of the deviations in fisheries are very marked. The potential for unknown and unintended bias should not be ignored. This in itself should be encouragement to attempt such analyses using multiple analytical strategies to determine the sensitivity of each particular data set to alternative statistical treatments. The breaking of underlying assumptions is another line of argument that confirms statistical CPUE standardization have many issues of which analysts need to be aware and wary.

6.4.3 The Choice of Statistical Approach

When designing a model to conduct a stock assessment of a fished species the decisions made about the model form and its components have important implications from what it is possible to conclude from the analysis. In addition, there is the contribution made to the overall uncertainty of any conclusions that derives from the model structure. This 'model uncertainty' is often a major contributor to the total uncertainty as different model structures can lead to very different conclusions concerning such fundamental outcomes as what constitutes a sustainable catch.

Similar things about structural decisions can be stated about the statistical models used in CPUE standardization. Even though the primary aim may be to describe the trends in CPUE there is an upper limit to the complexity of a statistical model as the inclusion of numerous factors can lead to later, less influential factors attempting to account for spurious patterns in the remaining noise. So, some attempt to constrain the number of factors from inclusion is a reasonable strategy to adopt. Decisions on the number of parameters (with associated variables or factors) can be made statistically (e.g. each factor needs to account for some, usually arbitrary, level of variation in the data) or by using some other rule of thumb, such as the influence on the mean trend of CPUE through time needs to indicate more than just noise.

Even in the relatively data-limited case of Blue-Eye the uncertainty in the analysis meant that in almost all cases, even where two different approaches (e.g. log-normal vs Gamma) differed, they did not differ significantly in many years. Often the difference between the unstandardized CPUE and the standardized trends was greater than the difference between the standardized trends by different approaches. Once again before final decisions are made it is a reasonable strategy to trial different approaches and if large differences in trend are found discover why the difference arises under the different circumstances of analysis rather than simply selecting a single approach.

Given the structural issues arising from, for example, the rounding of the components of CPUE, it is not reasonable to make hard and fast decisions regarding the selection of the probability distributions of the residuals between the observed CPUE (or catch) and that predicted by some approximate statistical model of the process giving rise to those CPUE (or catches). While the residuals can be better behaved statistically than the catches or effort values, meaning a probability density function can provide a plausible description of them, the statistical models that give rise to the predicted values (from which the residuals are derived) will be at best an approximation. If one statistical distribution does a consistently better job of describing the residuals in the CPUE standardization for a fishery then it is obviously a reasonable strategy to adopt that distribution rather than sticking to another out of tradition. But the optimum statistical distribution can vary between species, and even within a species between years as new data are included. It is difficult to know which approach is best until more than one approach is tried. Empirically, when a CPUE standardization is first being developed for a fishery, it is best to trial alternative approaches and adopt that which generates the most stable outcome through time (in a retrospective analysis). Once established, however, in an ideal world it would be best to test alternatives every few years to ensure nothing has changed.

6.4.4 Using LnC ~ … LnE

It is clear that if an analyst wishes to work with *LnC* rather than *LnCE* then the effort term used needs to be treated as an offset, otherwise some serious bias can be introduced and the response becomes unreliable, especially in cases where there are large trends or changes in effort through time.

As long as an analyst is consistent through time with how they conduct their analyses, then using either catch or catch-effort as the dependent variable is suitable (provides identical results; at least when using a log-normal residual error structure). Treating effort as an offset is necessary to scale the predicted values relative to the effort expended and some find it more natural to treat it separately than to include effort directly in a calculation of CPUE. Given the weakness of the assumptions when applying GLM to fisheries CPUE data this issue of separating catch and effort has complications of its own. Relative to CPUE, the underlying distributions of catch and of effort are generally more severely affected by the rounding of estimates of catch and effort (see **Figure 35**). It is possible that data-sets that exhibit more extreme degrees of aggregation in the separated data than was seen in the example data-sets used here, may constitute such a fracture of the statistical assumptions that the outputs become unstable. Rather than simply adopting one method over another it is best to compare both when first exploring a new of unknown data-set.

An examination of the classical diagnostic plots from a standardization based on the log transformed catches (**Figure 48**, on page 104) indicates that there is a failure to correct for the expanding variation in the residuals as catches increases. One of the assumptions of the analysis is that the variances are constant, which is certainly less compromised when using log-transformed CPUE (**Figure 39**, on page 97). Similar things can be said about using the GLM and Gamma errors with a log-link (**Figure 51**, on page 105).

Generally, making a selection of approaches entails balancing any trade-offs between which assumptions are being stretched by which approach.

6.4.5 The Influence of Mixed Fisheries

The auto-line fishery for Blue-Eye Trevalla and Pink Ling provides a relatively simple example of a mixed fishery that has two economically important species known to be targeted at different times (though when each is targeted is not reported with sufficient accuracy to usefully separate the two). How best to approach the analysis for each of the species is a problem that has yet to be solved. It remains a serious problem as many fisheries, especially in Australia, are mixed fisheries where the specific target for any particular shot is not necessarily known. Indeed, in such circumstances fishers often report having an expectation of catching an array of species rather than one single target. What is presented here is a potential direction for further work to provide one approach at clarifying this issue.

The procedure adopted for standardizing Blue-Eye Trevalla is conditioned on positive catches of Blue-Eye, meaning that any records containing Blue-Eye taken by a particular fishing gear are identified and included (as long as they meet other selection criteria relating to fishing zone and depth of fishing. This means that included in those shots that may have been directly targeting Blue-Eye there will be shots that were in fact targeting Pink Ling and only caught Blue-Eye as a by-product. This is very plain when considering the catch-rates of Blue-Eye relative to the proportion of Blue-Eye relative to Pink Ling in a shot (**Figure 41**, on page 98).

A simple option is to remove those records containing the lowest proportions of Blue-Eye, which directly improves the distribution of the log(CPUE) for the Blue-Eye (**Figure 42**, on page 99). An additional action would be to categorize the proportion of Blue-Eye in each shot and included those categories as a factor in the standardizations.

The act of removing records containing low proportions of Blue-Eye has the risk that if there is a strong correlation beyond between the proportion of Blue-Eye and the amount of Blue-Eye caught then if there is a trend in the proportion of records containing less than the selected minimum proportion of Blue-Eye in each shot then such removals could mask an influential trend. Fortunately, with Blue-Eye taken by auto-line, no such clear trend is present in the available data (**Figure 45**, on page 100), but it is important to check for such trends if such a strategy of removing records is pursued.

Removing the records with low proportions of Blue-Eye but not including the proportion of Blue-Eye as a factor in the standardization altered the trend relative to the trend that included all data. The two trends that included the proportion of Blue-Eye as a factor in the standardization were more similar to each other than the two standardizations that did not include pBlue as a factor.

Further exploration of the strategy of including the proportion of each species into standardizations should be made across other species. Such a strategy may lose effectiveness if the number of mixed species begins to increase. But if it is pursued it is essential to include the diagnostic plots of whether trends exist in the proportions of the mixed species changes through time in a trend like manner.

6.4.6 The treatment of Spatial Information

Generally, an important factor relating to CPUE is the geographical distribution of the species concerned. Within a species' distribution range it is expected that CPUE will vary with location (all other things being equal). Some differences in trends were produced when comparing the use of simple, relatively large geographical zones as a proxy for spatial differences, and using precise longitude and latitude values for the centre of each shot (**Figure 56**, on page 108). The mapped surface fitted to the longitude and latitude data can be greatly influenced by outlying, clearly incorrect data points, and while removing those points from consideration improved the surface representation it did almost nothing to the implied CPUE trend. While the effect of using a GAM instead of the LnCE classical standardization was only minor with the Flathead example (**Figure 31**, on page 92), the effect was much greater with the Blue-Eye example (**Figure 56**, on page 108). Before deciding it would be worthwhile comparing the two approaches (and adding more surface fits to the GAM model, perhaps on depth of each shot) to determine the overall effect.

Generally, if enough precise location data is available then a statistical model that uses it should be more informative. However, this would be partly determined by just how precise such data was and what proportion of erroneous data was present. Typically, with such data it is not uncommon to find some records on land, or some in port, or in other implausible locations. An alternative to these extremes of crude geographical zones and the use of longitude and latitude data is to use smaller geographical areas, as perhaps 1 × 1 degree blocks, or possibly smaller. As with all such analyses before settling on a standard approach it is necessary to explore the alternatives to determine the trade-offs between accounting for variation in the available data and including numerous possibly meaningless parameters.

6.5 Appendix: R code used to Conduct Standardizations

6.5.1 Standard Data-Base Extract

Table 29. Variables extracted from the catch and effort database within CSIRO prior to any analyses, plus derived variables denoted by a *.

Variable	**Comment**
Year	calendar year of fishing
Month	month of fishing
Day	day of fishing
Vessel	a unique identifier for each vessel through time
catch_kg	reported catch in kg (does not include discards)
Long	reported longitude of start of shot
Lat	reported latitude of start of shot
LongE	reported longitude of end of shot
LatE	reported latitude of end of shot
Depth	average depth of fishing
DayNight	day, night, mixed, unknown
Zone	SESSF zones
SEF_ZONES	old SESSF zones
Effort	trawl effort (hours); other methods (auto-line) require other fields
Method	fishing method
CE*	the catch/effort (where catch and effort > 0)
LnCE*	the natural log of CE
DepCat*	the depth grouped into 20 metre classes
count*	the number of years in the fishery for each vessel
avC*	the average annual catch for each vessel in the fishery

6.5.2 Software Used

In all analyses the statistical and programming language R was used. Here all functions and routines are listed as used in the nine Flathead examples. The functions form the basis of an r4sessf R package developed to facilitate running CPUE standardizations. Some of these functions are now also included in the ***cede*** R package produced to complement the primary objectives in FRDC project 2017/102.

```
#  Get and select Data --------------------------------------------------------------
load(paste0(datadir,"spsF.RData"))

# spsF is the tiger flathead dataset for the trawl method only


pickrec <- which((spsF$Year > 1985) & (spsF$Year < 2016) &
                 (spsF$DepCat >= 0) & (spsF$DepCat <= 400) &
                 (spsF$Zone %in% c(10,20)))
sps1 <- droplevels(spsF[pickrec,])
dim(sps1)
pickV <- which(sps1$Vessel == 529)   # a major outlier
if (length(pickV) > 0) sps1 <- droplevels(sps1[-pickV,])
pickD <- which(is.na(sps1$CE))       # remove records with no CE
if (length(pickD) > 0) sps1 <- droplevels(sps1[-pickD,])
dim(sps1)

# utility functions
```

```
makecategorical <- function (labelModel, indat) {
    Interact <- grep(":", labelModel)
    nInteract <- length(Interact)
    numvars <- length(labelModel) - nInteract
    for (fac in 1:numvars) {
        if (length(indat[, labelModel[fac]]) > 0) {
            indat[, labelModel[fac]] <- factor(indat[, labelModel[fac]])
        }
        else {
            warning(paste0("Factor name ", labelModel[fac], "does not appear in data.frame"))
        }
    }
    return(indat)
}

makemodels <- function (labelModel, dependent = "LnCE") {
    numvars <- length(labelModel)
    interterms <- grep(":", labelModel)
    ninter <- length(interterms)
    mods <- vector("list", (numvars + 1))
    form <- paste0(dependent, " ~ ", labelModel[1])
    mods[[1]] <- assign(paste0("ff", 1), as.formula(form))
    if (numvars > 1) {
        if (ninter > 0) {
            for (i in 2:(numvars - ninter)) {
                form <- paste0(form, " + ", labelModel[i])
                mods[[i]] <- assign(paste0("ff", i), as.formula(form))
            }
            for (i in interterms) {
                interform <- paste0(form, " + ", labelModel[i])
                mods[[i]] <- assign(paste0("ff", i), as.formula(interform))
            }
        }
        else {
            for (i in 2:numvars) {
                form <- paste0(form, " + ", labelModel[i])
                mods[[i]] <- assign(paste0("ff", i), as.formula(form))
            }
        }
    }
    mods[[(numvars + 1)]] <- labelModel
    return(mods)
}

standLM <- function (inmods, indat, inlab = "", console = TRUE) {
    NModels <- length(inmods)
    labelM <- inmods[[NModels]]
    NModels <- NModels - 1
    ans <- vector("list", NModels)
    names(ans) <- labelM
    Yearnames <- levels(as.factor(indat[, labelM[1]]))
    Nyrs <- length(Yearnames)
    rows <- c("AIC", "AICc", "RSS", "MSS", "Nobs", "Npars", "adj_r2", "%Change")
    WhichM <- matrix(nrow = length(rows), ncol = NModels, dimnames = list(rows,labelM))
    rows <- Yearnames
    Results <- matrix(nrow = Nyrs, ncol = NModels, dimnames = list(rows,labelM))
    ResStErr <- matrix(nrow = Nyrs, ncol = NModels, dimnames = list(rows,labelM))
    modellist <- vector("list", NModels)
    names(modellist) <- inlab
    geomod <- inmods[[1]]
    model <- lm(geomod, data = indat)
    totalssq <- sum(anova(model)[2])
    for (index in 1:NModels) {
        if (console)
            cat(as.character(inmods[[index]]), "\n")
        model <- lm(inmods[[index]], data = indat)
        modellist[[index]] <- model
        modelsum <- summary(model)
        mat <- modelsum$coefficients
        ans <- getlmfact(mat, labelM[1])
        Results[, index] <- scaleCE(ans[, "Coeff"])
        ResStErr[, index] <- ans[, "SE"]
        anv <- anova(model)
        RSS <- tail(anv$"Sum Sq", 1)
        WhichM["RSS", index] <- RSS
        WhichM["MSS", index] <- totalssq - RSS
```

```
        df <- unlist(anv[1])
        ndf <- length(df)
        nobs <- sum(df) + 1
        npars <- sum(df[1:(ndf - 1)]) + 1
        WhichM["AIC", index] <- nobs * log(RSS/nobs) + (2 * npars)
        WhichM["AICc", index] <- nobs * log(RSS/nobs) + ((2 * npars)*(nobs/(nobs - npars - 1)))
        WhichM["Nobs", index] <- nobs
        WhichM["Npars", index] <- npars
        WhichM["adj_r2", index] <- 100 * modelsum$adj.r.squared
    }
    for (index in 2:NModels) {
        WhichM["%Change", index] <- WhichM["adj_r2", index] - WhichM["adj_r2", index - 1]
    }
    WhichM["%Change", 1] <- 0
    pickinter <- grep(":", labelM)
    if (length(pickinter) > 0) {
        lastsimple <- pickinter[1] - 1
        WhichM["%Change", pickinter] <- WhichM["adj_r2",pickinter]-WhichM["adj_r2", lastsimple]
    }
    optimum <- which.max(WhichM["adj_r2", ])
    msg <- paste("Optimum model ", inmods[optimum], sep = "")
    if (console) print(msg, quote = F)
    if (NModels >= 3) {
        count <- 0
        for (i in 3:NModels) {
            if (WhichM["%Change", i] > WhichM["%Change", (i - 1)])
                count <- count + 1
        }
    }
    out <- list(Results, ResStErr, WhichM, optimum, Nyrs, inlab,
        summary(modellist[[optimum]]), modellist[[optimum]],modellist)
    names(out) <- c("Results", "StErr", "WhichM", "Optimum", "Nyrs", "Label", "Parameters",
                    "optModel", "Models")
    class(out) <- "CEout"
    return(out)
}
```

The main code; most plotting and tabulating details omitted for clarity and brevity

```
cat("\n\n Model 1: Linear Model based on LnCE and normal random errors \n\n")

labelM <- c("Year","Vessel","DepCat","Zone","Month","DayNight","Month:Zone")

sps2 <- makecategorical(labelM[1:6],sps1) # convert variables to factors
mods <- makemodels(labelM)

out1 <- standLM(mods,sps2,"FlatheadLM")
cat("Optimum Model: ",mods[[out1$Optimum]],"\n\n")
cat("Summary of Optimum Model  \n")
summary(out1)
cat("\n\n Anova of Optimum Model  \n\n")
anova(out1$optModel)
model1 <- out1$optModel
lmgeo <- out1$Results[,1]             # lm standardized geometric mean

answer1 <- getfact(out1,"Year")
cat("Year parameter Estimates from Model 1  \n")
answer1

# 2. glm on LnCE -------------------------------------------------------------
# source("C:/A_CSIRO/Rcode/CPUEExplore/utils/extra_utils.R")

cat("\n\n Model 1: Linear Model based on LnCE and normal random errors  \n\n")

usemod <-  as.formula(LnCE ~ Year + Vessel + DepCat + Zone + Month + DayNight + Month:Zone)
model2 <- glm(usemod,family=gaussian(link="identity"),data=sps2)
#anova(model2)
summary(model2)

answer2 <- getfact(model2,"Year")
answer2
back2 <- cbind(lmgeo,answer1[,"Scaled"],answer2[,"Scaled"])
colnames(back2) <- c("lm_geo","lm_LnCE","glm_LnCE")
back2                  # should be identical
```

```
# 3. glm on CE log-link -----------------------------------------------------
cat("\n\n Model 3:GLM based on CE and a log link \n\n")

usemod <- as.formula(CE ~ Year + Vessel + DepCat + Zone + Month + DayNight + Month:Zone)
model3 <- glm(usemod,family=gaussian(link="log"),data=sps2)
answer3 <- getfact(model3,"Year")
back3 <- cbind(lmgeo,answer1[,"Scaled"],answer2[,"Scaled"],answer3[,"Scaled"])
colnames(back3) <- c("lm_geo","lm_LnCE","glm_LnCE","glm_CE")
back3   #

# 4. glm on CE gamma errors ---------------------------------------------------------
cat("\n\n Model 4: GLM based on CE and Gamma errors with a log link \n\n")

usemod <- as.formula(CE ~ Year + Vessel + DepCat + Zone + Month + DayNight + Month:Zone)

model4 <- glm(usemod,family=Gamma(link="log"),data=sps2)
m4 <- summary(model4)$coefficients
m4yr <- m4[1:30,]
m4yr[1,] <- c(0,0,0,0)
m4pars <- exp(m4yr[,"Estimate"])  # no bias correction necessary
answer4 <- cbind(m4pars,m4pars/mean(m4pars))
colnames(answer4) <- c("m4pars","Scaled")
back4 <- cbind(lmgeo,answer1[,"Scaled"],answer2[,"Scaled"],answer3[,"Scaled"],
               answer4[,"Scaled"])
colnames(back4) <- c("lm_geo","lm_LnCE","glm_LnCE","glm_CE_Log","glm_CE_Gamma")
back4

AIC(model1, model2, model3, model4)

# 5. gam LnCE wth Long ----------------------------------------------------

cat("\n\n Model 5: GAM on Long, based on LnCE and normal random errors \n\n")

library(mgcv)
library(nlme)
library(gamm4)

pick <- which(sps2$Long > 152.25)
if (length(pick) > 0) sps2 <- sps2[-pick,]

model5 <- gam(LnCE ~ s(Long) + Year + Vessel + DepCat + Zone + Month + DayNight + Month:Zone,
              data = sps2)

answer5 <- getfact(model5,"Year")
back5 <- cbind(lmgeo,answer1[,"Scaled"],answer2[,"Scaled"],answer3[,"Scaled"],
               answer4[,"Scaled"],answer5[,"Scaled"])
colnames(back5) <- c("lm_geo","lm_LnCE","glm_LnCE","glm_CE_Log","glm_CE_Gamma","gam_LnCE_Long")
back5


# 6. gam LnCE with Lat ----------------------------------------------------
cat("\n\n Model 6: GAM on Lat, based on LnCE and normal random errors \n\n")

model6 <- gam(LnCE ~ s(Lat) + Year + Vessel + DepCat + Zone + Month +
                 DayNight + Month:Zone, data = sps2)
print(Sys.time()-startime)
answer6 <- getfact(model6,"Year")
back6 <- cbind(lmgeo,answer1[,"Scaled"],answer2[,"Scaled"],answer3[,"Scaled"],
                        answer4[,"Scaled"],answer5[,"Scaled"],answer6[,"Scaled"])
colnames(back6) <- c("lm_geo","lm_LnCE","glm_LnCE","glm_CE_Log","glm_CE_Gamma",
                    "gam_LnCE_Long","gam_LnCE_Lat")
back6

if (saveall) dev.off()

# 7. gam LnCE LoLa -------------------------------------------------
cat("\n\n Model 7: GAM on Longand Lat, based on LnCE and normal random errors \n\n")

model7 <- gam(LnCE ~ s(Long,Lat) + Vessel + DepCat + Month + Year, data = sps2)
print(Sys.time()-startime)

answer7 <- getfact(model7,"Year")
back7 <- cbind(lmgeo,answer1[,"Scaled"],answer2[,"Scaled"],answer3[,"Scaled"],
               answer4[,"Scaled"],answer5[,"Scaled"],answer6[,"Scaled"],answer7[,"Scaled"])
colnames(back7) <- c("lm_geo","lm_LnCE","glm_LnCE","glm_CE_Log","glm_CE_Gamma",
                    "gam_LnCE_Long","gam_LnCE_Lat","gam_LnCE_LongLat")
back7
#anova(model7)
```

```
# 8. gam LnCE LLtrim -----------------------------------------------------
 cat("\n\n Model 8: GAM on Long and Lat with trimmed data \n\n")

labelM <- c("Year","Vessel","DepCat","Zone","Month","DayNight")
sps3 <- makecategorical(labelM[1:6],sps1) # convert variables to factors

ypts <- c(-41,-33.5); xpts <- c(148.9,152.20)
mod8 <- lm(ypts ~ xpts)
coef(mod8)
sps3$LR <- NA
sps3$LR <- coef(mod8)[2] * sps3$Long  + coef(mod8)[1]
pick <- which(sps3$LR > sps3$Lat)
if (length(pick) > 0) sps3 <- sps3[-pick,]

model8 <- gam(LnCE ~ s(Long,Lat) + Vessel + DepCat + Zone + Month + Year, data = sps3)
print(Sys.time()-startime)

answer8 <- getfact(model8,"Year")
back8 <- cbind(lmgeo,answer1[,"Scaled"],answer2[,"Scaled"],answer3[,"Scaled"],
               answer4[,"Scaled"],answer5[,"Scaled"],answer6[,"Scaled"],
               answer7[,"Scaled"],answer8[,"Scaled"])
colnames(back8) <- c("lm_geo","lm_LnCE","glm_LnCE","glm_CE_Log","glm_CE_Gamma",
                     "gam_LnCE_Long","gam_LnCE_Lat","gam_LnCE_LongLat","gam_LnCE_LLtrim")
back8
```

7 Simulate Shot-by-Shot CPUE Data

7.1 Introduction

If it were possible to simulate CPUE data then before any further analyses were applied it would be possible to set or at least to know what the correct or actual mean values ought to be obtained from estimates derived from subsequent analyses. It should then be possible to determine, within the limits of precision for a particular analytical strategy, whether a given analytical strategy gives rise to a consistent bias, or other distortion, and whether different strategies give significantly different answers. So, the ability to simulate CPUE data would be of value for exploring the relative effectiveness of different methodologies.

The original intention within this project was to attempt to use the Atlantis ecosystem simulation framework (Fulton et al, 2005; Fulton, 2010) to generate multi-species CPUE at a shot-by-shot level. The Atlantis code was extensively modified to provide more details of simulated output data in terms of location, time, and vessels, but, in the end, it became clear that the CPUE data that it was possible to generate, without needing to alter the Atlantis software far more than was plausible in the time available, was still too coarse in terms of being aggregated over vessels and regions. To convert that into more typical shot-by-shot data would have entailed conditioning a separate model of single species data onto the coarser scale output from the modified Atlantis and then back-translate that into shot-by-shot data. The initial trials failed to produce plausible data and already slowed the operational running time of the Atlantis model significantly. The inevitable progress of time (for both people and this project) meant that the backup/fall-back position of simulating single species CPUE data was picked up and attempts to modify the Atlantis software further along these lines were stopped.

7.2 Simulating Single Species CPUE Data

7.2.1 Simulate the Stock and Fishery

The simulation of realistic single species CPUE data can be approached in more than one way. It might be possible, for example, to simulate the population dynamics of a fished stock including the variation in abundance through years combined with variation in how the fish are distributed geographically within years. In addition, one would need to simulate the dynamics of fishing (the number of vessels, where they fished, and with what relative efficiency) and how that changes through time. One difficulty with this approach relates to determining whether the simulated data is realistic or not. This whole-of-system approach to simulating CPUE data relies on making numerous assumptions regarding the movement of fish and how that might relate to spatial structuring within a given fishery, but also assumptions relating to how the vessels relate to the fishery. These both remain active areas of research with, currently, no single or clear answer. Using such an approach it would be possible to compare the outcomes from alternative statistical approaches used to analyse CPUE data and which was able to recover the true underlying relative abundance through time. However, it would not then be possible to conclude that the outcomes represented what would occur with real CPUE data.

Unfortunately, there would be no way to determine whether the CPUE data generated by such a system was realistic; meaning whether or not it represented the real world in a representative manner. The usual argument made when applying management strategy evaluation, which is, at heart, a simulation framework that attempts to simulate a complete system, is that it enables a comparison of methods and enables those methods or strategies that would fail to be rejected. Even if the simulation cannot be guaranteed to represent the system being simulated perfectly, the argument is made that if some approaches fail in a simulation while others succeed, then those that succeed should be more likely to succeed than those that fail. For example, if a given approach led to biased or otherwise distorted outcomes, while others did not, then it would be valid to be warier of the approach that led to biased outcomes than the one that did not. Hence pursuing this strategy of simulating CPUE should be worthwhile. Nevertheless, here, given the limitations of CPUE data as detailed in earlier chapters, and the very many unknowns for which information would be required or need to be assumed to generate a plausible approximation to a real fishery an alternative simpler approach was used to simulate realistic CPUE data.

7.2.2 The Single Species Simulation Used

The alternative approach used here essentially uses the inverse of a statistical standardization to condition a simulation to be like a given fishery. A standardization of CPUE data from a given fishery estimates parameters for each level of each factor within the standardization equation used. Those parameters effectively condition that statistical model onto that fishery with its history and scale. In other words, it is possible to use the statistical model, with all its parameters to predict new values of CPUE for each record, and multiple options are available. It is possible to replicate the original values if the parameters are left unaltered, or, by altering some or all of the yearly parameter values it is possible to change the apparent trends in the predicted CPUE data through time.

The trends and patterns of the yearly CPUE parameters can thus be left the same as obtained from the conditioning data or altered to suit the requirements of particular tests. For example, if testing how great a change in a CPUE trend is required to be effectively detected then the expected mean CPUE trend through time can be adjusted up or down by the necessary amount (the year parameters are adjusted to the desired levels), then the parameter set can be used to simulate shot-by-shot CPUE data by using the parameters and statistical model to generate predicted CPUE values (the fitted values) for each original record. Thus, simulated data can be obtained by populating a new set of records each with their own plausible combination of Year, Vessel, Depth, Month, and other factors included in the standardization.

One limitation is that situations where there are significant year × 'factor' interactions cannot be simulated easily, at least not without including such an interaction with their more complex correlations. Here Year × 'factor' interactions are not considered further but it is likely to be an avenue for productive future work, especially with regard to spatial heterogeneity of biological properties across stocks.

Using a particular standardization to condition a statistical model will automatically mean that the assumptions of the standardization used will be propagated through to the simulated data. If, for example, log-normally distributed errors are included in the simulation then it would not be surprising if an approach using log-normal errors recovered the underlying trends more accurately more often than, say, using Gamma distributed errors. What this means is that such simulated data is unsuitable for answering some classes of questions, especially those relating to identifying whether there is an optimum underlying residual error structure for standardizing CPUE data. Nevertheless, cross-comparisons where data generated using log-normal and using Gamma distributed errors can each be fitted using both approaches to determine whether each approach can recover the trend generated by the other approach.

Even within a single approach, some questions should be directly answerable, such as, for example, how small a change in the annual mean CPUE is detectable given the variation inherent in the available data? Also, does it matter which residual error structure is used when conducting a standardization? Thus, the exploration of the limits and limitations of classical methods remains open to the use of such simulated CPUE data.

In practice, using a standardization to condition a statistical model and using that to generate predicted values for each record in a simulated data set has its own issues. Simply generating the predicted CPUE values for each record from a standardization would not be sufficient as these would be, as expected, less variable than the observed data to which the statistical model was fitted (**Figure 59**). This is simply a reflection that in each record the predicted log(CPUE) or the fitted values, will be mean estimates for each combination of level and factors expressed in each record.

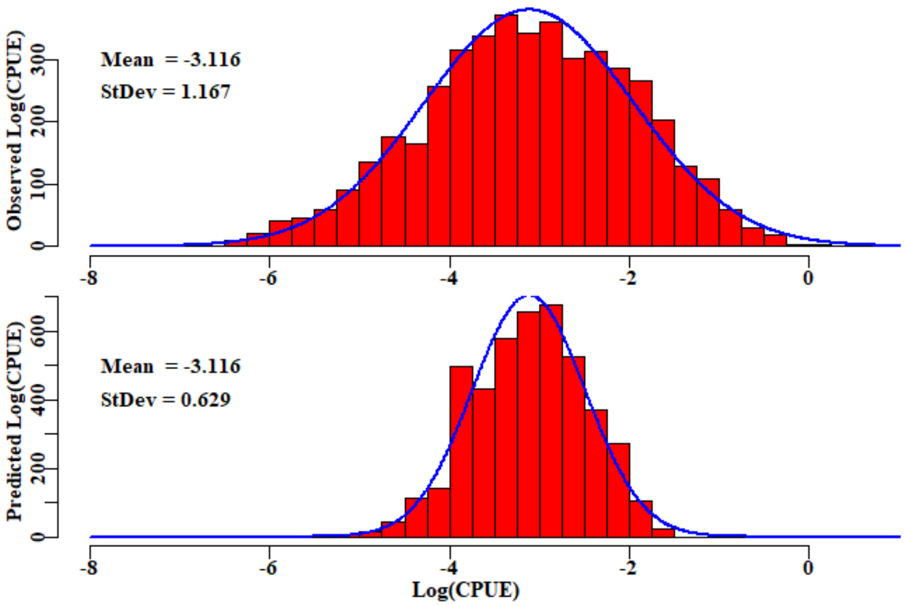

Figure 59. A comparison of the observed log-transformed CPUE for Blue-Eye taken by autoline (using only those records with > 10% Blue-Eye by weight relative to the combined weight of Blue-Eye and Pink Ling - see Chapter 6) and the fitted values taken from the optimal standardization from among those considered. The blue curves are normal distributions fitted to the histogram counts. The mean of both distributions was -3.116 but the standard deviation of the upper plot was 1.1671 and of the bottom plot was only 0.629. The total number of records was 4,448.

7.2.3 Objectives

In this chapter we will be using simulated CPUE data in attempts to achieve three objectives:

1. Develop a methodology for reliably simulating single species CPUE data.
2. Use such simulated data to test whether classical standardization methodologies can estimate true annual mean estimates with acceptable precision and without appreciable bias.
3. Use such simulated data to test whether alternative analytical strategies generate significantly different standardized outcomes.

7.3 Methods

7.3.1 The Standardization used to Condition the Model

Assuming only records with a complete set of fields will be used in a CPUE standardization, then each record will have an entry for the catch taken by a given amount of effort by a particular vessel, in a particular location, depth, month, and year (and any other factors included). One way of representing such CPUE data, in terms of catch, *C*, catchability, *q*, effort, *E*, and biomass, B_t, derived from a particular year (t), vessel (V), depth (d) category, month (m), and area (A), (and any other factors) would be:

$$\frac{C}{E}_{V,A,d,m,t} = \alpha q_V q_A q_d q_m q_t B_t \times e^{N(0,\sigma)} \tag{14}$$

And in a log-space:

$$ln(\frac{C}{E})_{V,A,d,m,t} = ln\alpha B_t + lnq_V + \ln q_A + \ln q_d + \ln q_m + \ln q_t + N(0,\sigma) \tag{15}$$

For example, a useful model for the Blue-Eye auto-line CPUE data has the form:

LnCE ~ Year + Vessel + Month + Zone + DepCat + Month:Zone (16)

with the final Month:Zone term being an interaction term which, if significant, would suggest that the effect of each Zone on CPUE altered through the seasons. As each new factor is included in the standardization the amount of the total variation accounted for increases although the changes to the final trend become less and less, with most of the final trend becoming established with only the first three factors: Year + Vessel + Month (**Figure 60**). The Vessel factor obviously has a major effect, although the clear switch coincides with the end of the Commonwealth Structural Adjustment (Vieira et al, 2010). Then Month of fishing also has some influence on the trend with its effect changing two years after the structural adjustment. The remaining factors appear to only have minor and possibly random influences on the trend through time. While the final annual trend may be captured by the first three factors (**Figure 60** and **Figure 61**) the character of the variation of the fitted values (the predicted Log(CPUE) for each record) only stabilizes when more factors are included in the standardization (**Figure 62**), although the residual distribution stabilizes relatively early in the fitting of sequential factors.

The quality of the model fit using log-normal random residual errors is reasonable. The diagnostic qqplot indicates that the residuals are approximately normally distributed (fall on the expected transformed line) at least down to residuals of about -1.75. However, a histogram of the residuals indicates that only about 5% of observations have residuals less than -1.75 (**Figure 63**).

The form of the equation matters to its interpretation. Equations (14) and (15) imply that the observed CPUE for a particular vessel, month, zone, and depth, in a given year will be related to the stock wide exploitable biomass (B_t) but modified by the particular catchability modifier for each particular vessel, month, zone, and depth (and any other factors included along with interaction terms; any factor not included, even potentially influential ones, will contribute to the error or random noise component). In each case the modifiers, the *q* values, are distributions of proportions with an expectation of 1.0. At heart, the standardization process involves estimating each of these many different *q* values.

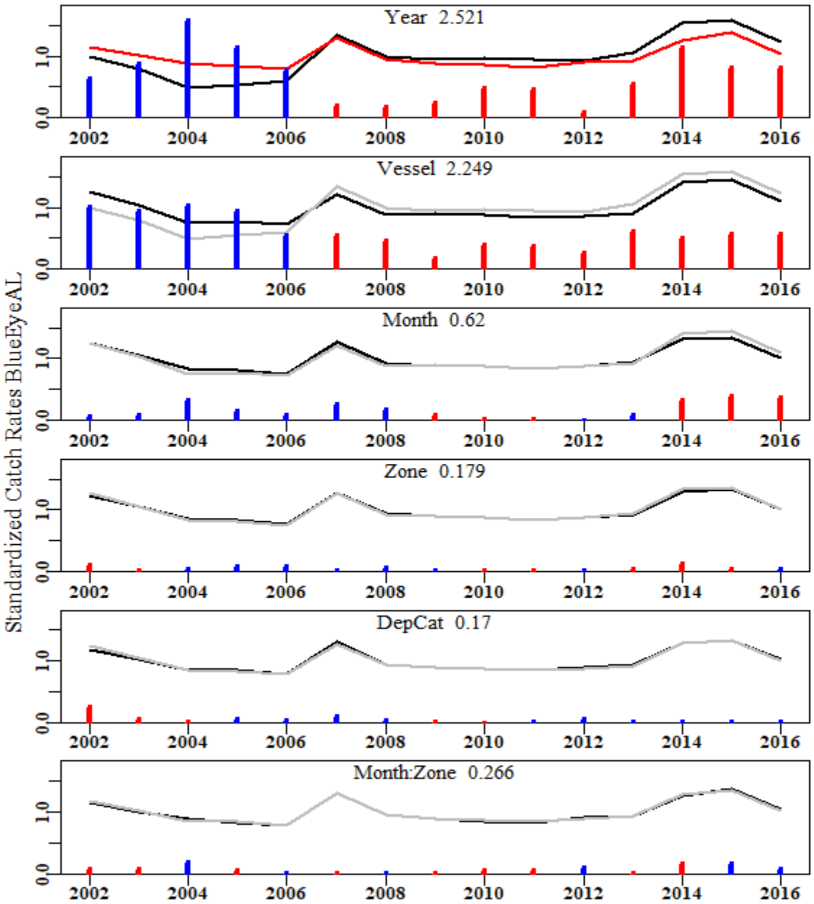

Figure 60. The top plot illustrates the final model (black) relative to the geometric mean (grey = Year alone) and the number in each case represents the sum of the squared differences between the two lines. Hence the Vessel factor accounts for much of the final difference between the un-standardized (grey) and standardized lines (black). The Month plot illustrates the difference between the LnCE ~ Year + Vessel model (grey) and the Year + Vessel + Month model (black) (and so on down the plot). The vertical lines illustrate the scale of difference by year with blue indicating the predicted line is above the previous line and red that it is below.

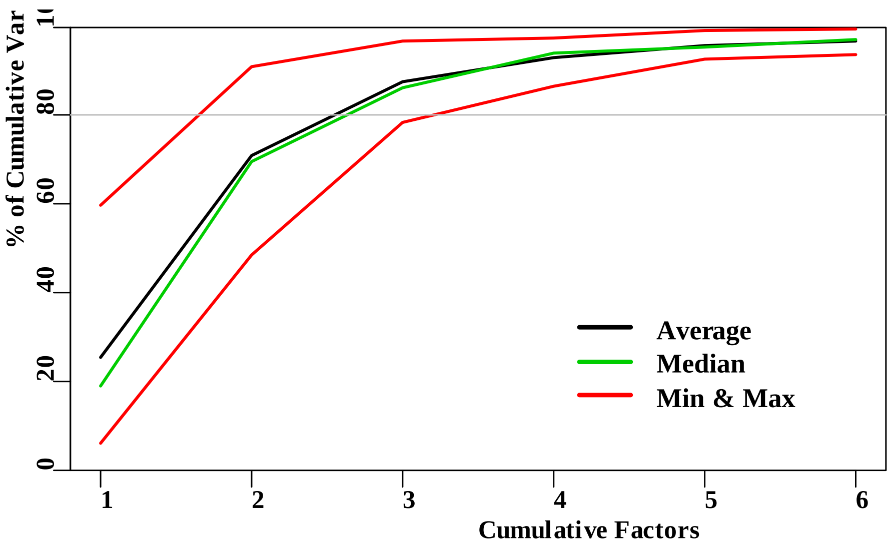

Figure 61. The % cumulative variance as factors are added to an array of standardizations for those fisheries significantly pursued in more than one fishing zone (see Sporcic and Haddon, 2016); that is, not School Whiting, Royal Red Prawns, or Redfish.

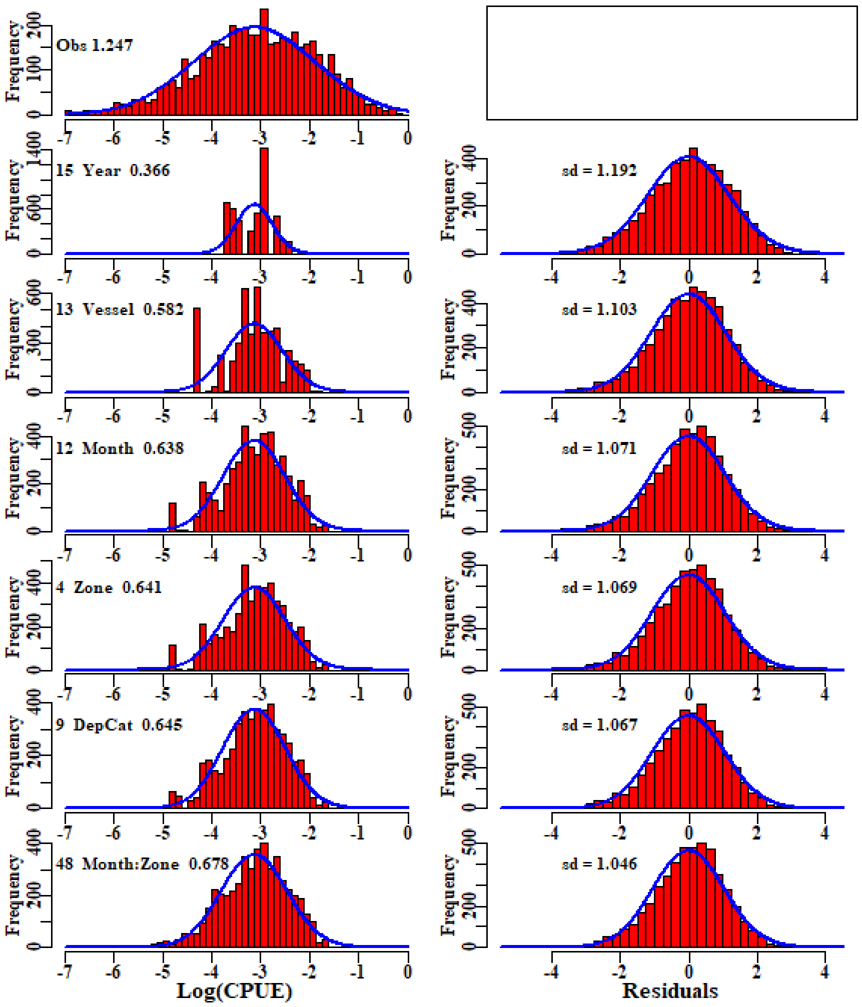

Figure 62. The effect on variation in the predicted Log(CPUE) and related residual values as more factors are added to the standardization. The top plot illustrates the distribution of the observations used. The model sequentially increases in complexity from 'Year' to 'Month:Zone'; see equation (16). Blue lines are normal distributions fitted to histogram counts. The observations have a mean of -3.146 and all the fitted models have means of -3.159. Their labels are the number of parameters, the factor name, and the standard deviation. As model complexity increases so does the variation accounted for, while the standard deviation of both the fitted values and residuals decreases.

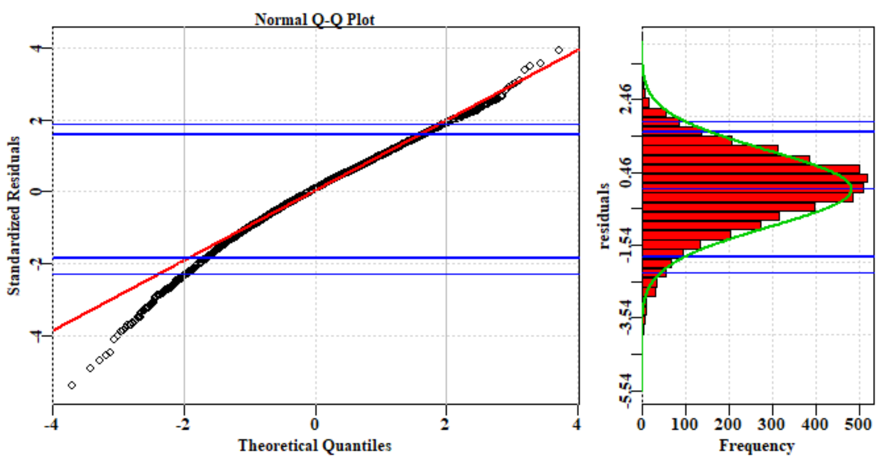

Figure 63. The distribution of the residuals from the optimum fit. The qqplot on the left-hand side indicates a reasonable fit to normality down to residuals of about -1.75. The histogram of residuals illustrates the 90% quantiles and so indicates that < 5% of records deviate from normality at the bottom end.

There is assumed to be a recognizable biomass in a given year which implies this is a discussion about a recognizable and repeatable stock structure. It also refers only to exploitable biomass as we are referring to CPUE, which, by definition applies only to the biomass available to exploitation rather than, for example, the mature or spawning biomass (which is more often associated with management reference points).

7.3.2 Increasing the Variability of Simulated Data

The reduction in variability of the residuals and of the predicted CPUE by record (**Figure 62**) as more factors are included reflects the fact that the standardization is attempting to track average effects of the various factors, especially the effects over years (over time). The more parameters fitted the more flexibility there will be for predicted values and hence residual values, which in turn should allow the variability of the residuals to decline as they should be able to match of the observed values more closely. As an aside, as more and more factors are added to the model then obviously the number of parameters fitted also increases. The extreme end-point would be to estimate a parameter for each observation, which would provide a perfect model fit but no summary concerning trends. One objective in such model fitting is to find the trade-off between the minimum number of parameters needed to be fitted that still provides a sufficiently good model fit to the data. This is where statistical criteria such as the Akaike's Information Criteria, or the adjusted-r^2 are used; each of these balances the number of parameters against a measure of the quality of model fit (Burnham and Anderson, 2002; Neter et al, 1996). This trade-off is especially important with such models when they are to be used for prediction but less so if they are merely to be used to describe the trends in CPUE. In the case of using the standardization to generate new simulated sets of CPUE records then factors should continue to be added until both the residuals and the predicted values exhibit relatively well-formed distributions (**Figure 62**).

For the approach of conditioning a simulation on data from a real fishery to provide realistic simulated CPUE data then a method of increasing the spread of the simulated data is needed following the initial standardization (compare the spread of observed log(CPUE) with the spread of the final fitted values in **Figure 59** and **Figure 62**).

The simplest approach to increasing the variation of the predicted log(CPUE) values would be to add normal random error to the predicted log(CPUE) values in each record from equation (15). The standardization outputs are already in log-transformed space so adding normal random error is equivalent to including log-normal errors on the back-transformed values. The mean of the added variation (noise) should be 0, and one needs to search for the standard deviation value used to produce the normal random values that generates a spread of predicted log(CPUE) that approximates the observed spread (**Figure 64**). If the CPUE values are required these can be obtained by back-transforming the log(CPUE) values using the exponential function.

For a simulation to be useful, after conditioning a statistical model of CPUE for an hypothetical fishery, there is also a need to know the biomass trajectory through time so that subsequent explorations of the simulated data can be related back to a known trajectory with known variation. In the examples used here the assumption is made that there is a linear relationship between CPUE and exploitable biomass. This assumption could be relaxed to examine the effect on analyses of hyper-stability (stable CPUE with declining biomass) or even hyper-sensitivity (CPUE changing more rapidly than biomass).

An iterative search can be used to find the optimum standard deviation value to give an average difference between the observed log(CE) and the simulated log(CE) that is as small as possible. This will involve numerous replicate trials with the final selection being dependent upon average behaviour.

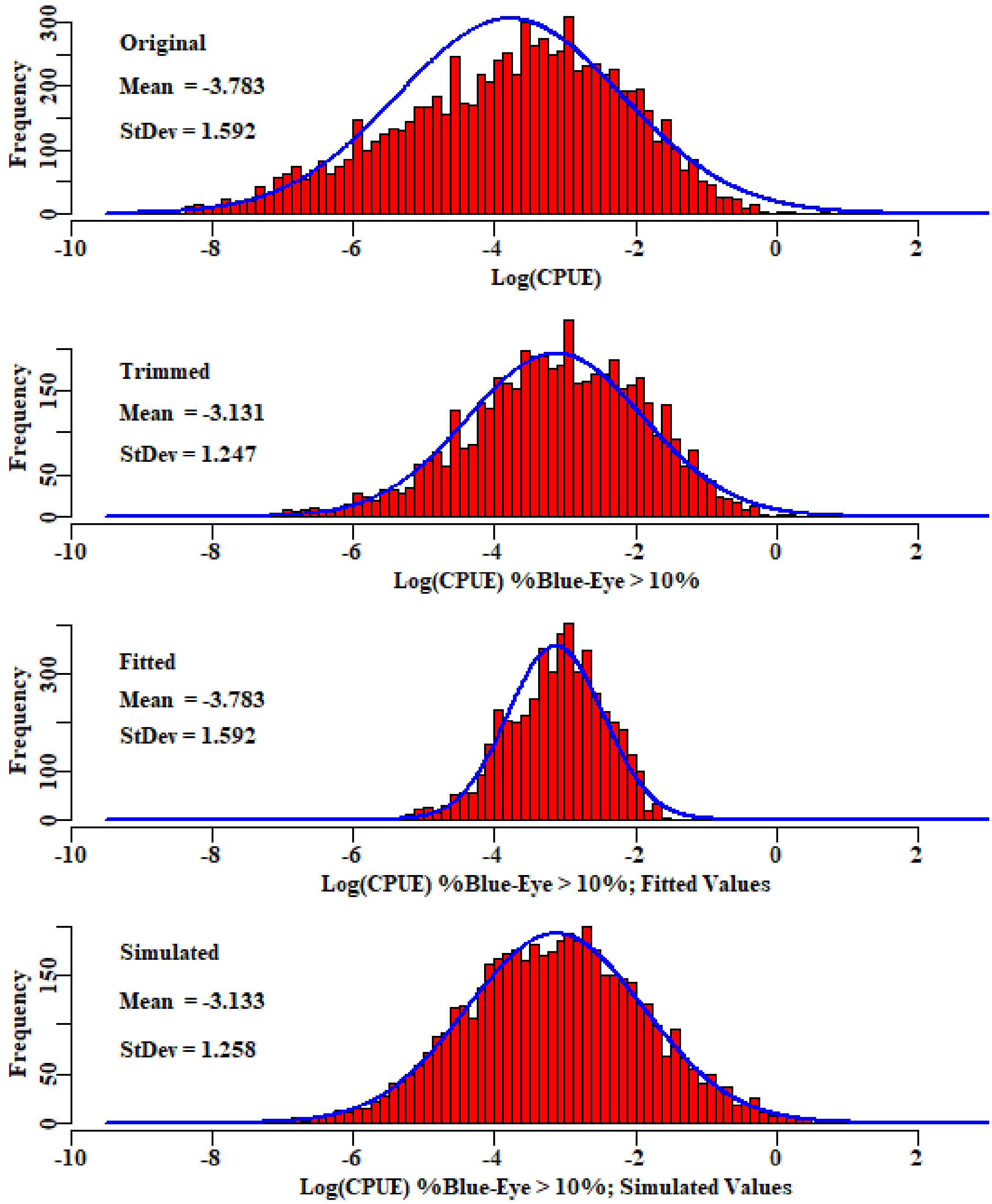

Figure 64. The top plot is of the log transformed CPUE for all Blue-Eye data from the auto-line fishery in the SESSF. The second plot is the same distribution but limited to records containing more than 10% Blue-Eye when considering both Blue-Eye and Pink Ling. This is the data used in the following analyses. The third plot is the fitted or predicted Log(CPUE) from the typical standardization, and the bottom plot is simulated by adding normal random variation (mean = 0, standard deviation = 0.984) to the fitted values.

7.4 Algorithm for Simulation of CPUE Data

7.4.1 Pseudocode

To condition a simulation on a real fishery (the specific example is given in *italics*):

1. load the data related to the specific fishery;
    a. *in the example take all Blue-Eye records*
2. select the particular data records to be used;
    a. *select for auto-line method, in SESSF zones 20, 30, 40, and 50, in depths >= 200m and <= 600m, between the years 2002 – 2016, with the proportion of the combined Blue-Eye and Pink Ling catch that is Blue-Eye being > 0.1, and finally, where each vessel reports from the fishery in > 1 year.*
3. conduct the full standardization using the optimum model: Model1;
    a. *use equation (16) LnCE ~ Year + Vessel + Month + Zone + DepCat + Month:Zone*
4. Extract the fitted values from the standardization;
5. Search for the standard deviation of the normal random errors to be added to the fitted values until it matches the observed standard deviation to at least 3 decimals places.
6. Alter the year parameters in the manner required to change the observed trend so as to be able to test for the sensitivity to such changes.
7. Before making any changes run the addition of random variation to the fitted values 1000 times and determine how often in each year predicted mean values exceed the 90^{th} percentile prediction intervals (in a one sided test no more than 50% of iterations should exceed these intervals.
8. Devise the scenarios to be tested and run each test at least 1000 times.

7.4.2 Scenarios Tested

In all cases the intention is to run each scenario 1000 times and determine whether the error rate (deviations above or below the expected prediction intervals) change relative to the natural error rate when no changes to the mean year parameters are made.

- Adding an upward trend to the last five years
- Adding a downward trend to the last five years
- Adding a year of exceptional lowered and increased CPUE in the middle of the time-series
- Generating log-normal fitted values but fitting standardization models using Gamma errors.

There does not appear to be an obvious means of simulating effort creep using an approach akin to that described here. What would be required is a way to increase the catchability rather than the parameter values for the year levels. In other words, use equation (14) or (15) to generate the same CPUE values even when the biomass values were reduced. Currently we have no algorithm for doing this.

7.5 Results

7.5.1 Increasing the Variance of Fitted Values

The optimum model fit to the available data involves using standard R functions as described in **6.5 Appendix: R code used to Conduct Standardizations**.

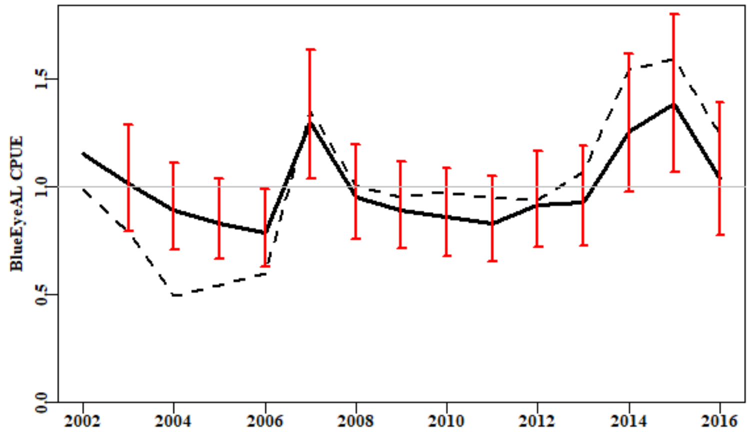

Figure 65. The optimum model fit using 4858 observations of Blue-Eye taken by auto-line in SESSF zones 20, 30, 40, and 50 in depths between 200 – 600 m from 2002 – 2016 (**Table 30**).

Table 30. The summary results for the sequence of models fitted to the Blue-Eye auto-line data. Even though Zone, DepCat, and the interaction term Month:Zone have very little influence on the final trend of yearly CPUE indices (**Figure 60**) they each improve the model fit markedly, especially the interaction term. The RSS is residual sum of squared, MSS is Model sum of squares, Nobs is number of observations, and adj_r2 is the adjusted r^2.

	Year	Vessel	Month	Zone	DepCat	Month:Zone
AIC	1734.392	1004.630	742.005	731.157	725.318	604.592
RSS	6899.640	5908.017	5571.818	5552.528	5527.622	5319.190
MSS	650.739	1642.362	1978.561	1997.851	2022.757	2231.189
Nobs	4858	4858	4858	4858	4858	4858
Npars	15	27	38	41	49	82
adj_r2	8.354	21.331	25.638	25.850	26.059	28.356
%Change	0	12.976	4.307	0.211	0.210	2.296

The fitted values of the log(CPUE) form a relatively clean normal distribution (**Figure 64**). The variation of these values can be increased appropriately, so as to condition the model onto the Blue-Eye auto-line fishery, by iteratively searching for a standard deviation that leads to the closest match between the standard deviation of the simulated log(CPUE) and that of the observed log(CPUE) values (**Figure 66**). The R-code used is provided in **7.7 Appendix: R-code used in Simulations**.

The fitted values are the basis of subsequent simulations. Each iteration of a simulation will entail adding normal random errors to the fitted values and then proceeding to any analysis.

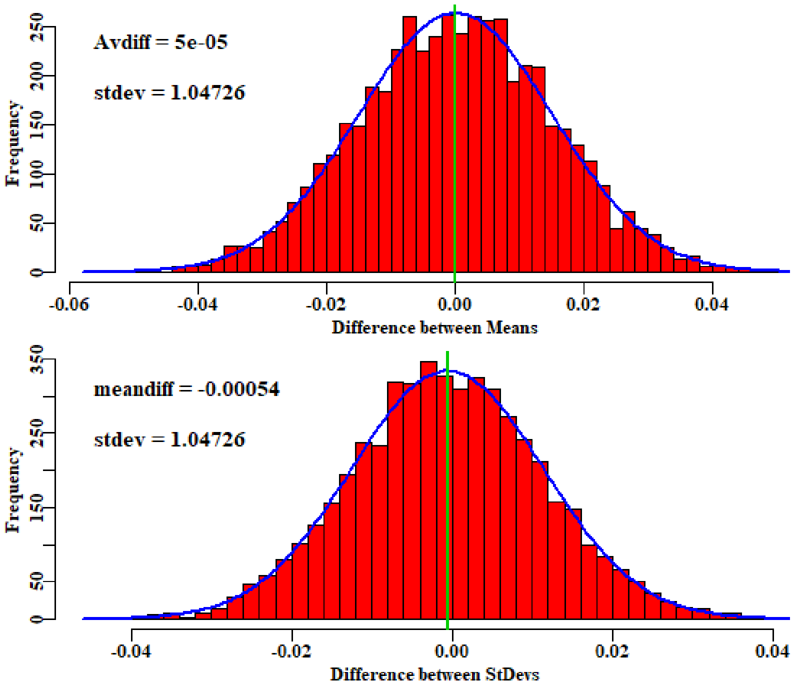

Figure 66. 5000 differences between simulated record by record log(CE) data and the original observed log(CE) data used to condition a statistical model. The mean difference of the overall means across 5000 replicates was approximately 0.0, and was very small between the spread of the differences, when the standard deviation of the normal random errors added to each record was set to 1.04726

7.5.2 Bias and Precision

Only including records where Blue-Eye constitute >10% of the combined Blue-Eye and Pink Ling catch is an approximate way of focussing the analysis on the targeted fishery for Blue-Eye. Thus, as an example, we will use the trimmed Blue-Eye data taken by auto-line described in **6.3.6 The Standardization of Blue-Eye** on page 96 and already used in the Introduction and Methods sections in this chapter (e.g. **Figure 43**, **Figure 59** and **Figure 62**).

Using the fitted values from the standardization of the trimmed data a comparison was made between the outcome when using the addition of random normal errors to the record-by-record fitted values in a Monte Carlo analysis and the outcome when using bootstrap samples from the residuals which were then added to the fitted values. In each case the same statistical model was then fitted, and the whole process repeated 1000 times. In all cases the bootstrap samples were the same size as the number of residuals (**Figure 67**). The outcomes in both cases were very similar and only slight differences were exhibited by just a few replicates.

When the quantiles from each set of 1000 replicates were compared only slight differences were found between them of generally less than 1% of the Monte Carlo value (**Table 31**). In subsequent analyses only the addition of random normal errors was used to generate simulated observed values.

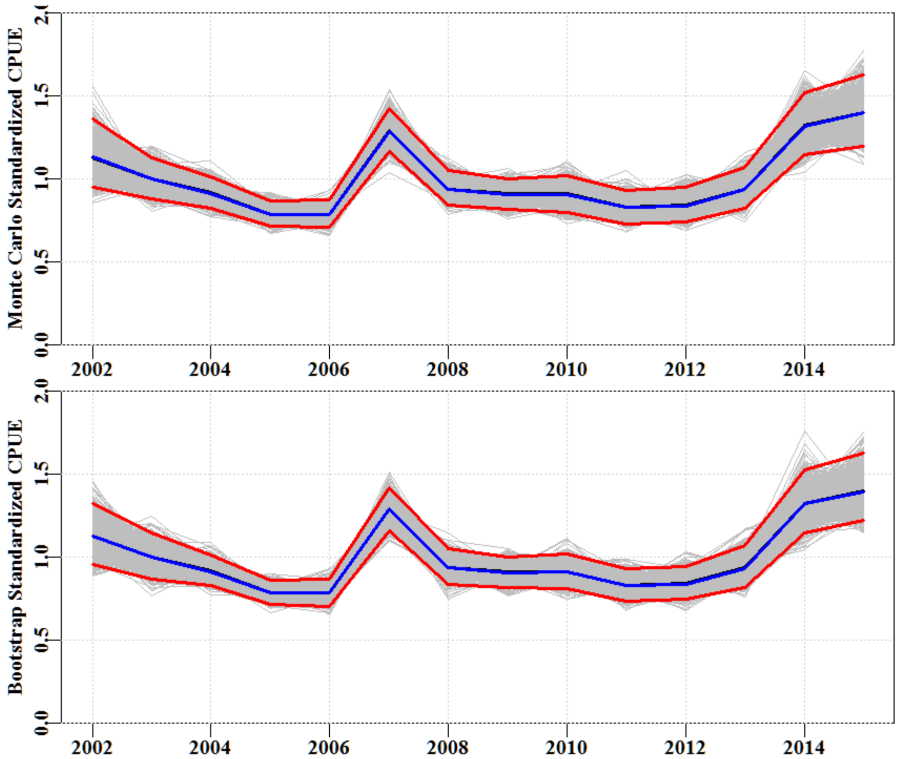

Figure 67. 1000 replicates of adding normal random error to the fitted record-by-record values from the standardization depicted in **Figure 65** to generate a simulated data set. Each of those data sets is then standardized and each replicate standardization is depicted by a grey line (top panel). The optimum model is a black line beneath the blue line, which is the 50% quantile of the 1000 standardizations. The red lines are the 95% quantiles. The lower panel is almost identical to the upper panel but the simulated data was obtained using a bootstrap process on the model residuals.

Table 31. The percentage difference obtained by subtracting the Monte Carlo quantiles from the bootstrap quantile and dividing by the bootstrap quantile for each of the quantiles estimated: 100 * (B – M)/B. The quantiles are listed in the top row at 2.5%, 5%, 50%, 95%, and 97.5%.

Year	0.025	0.05	0.5	0.95	0.975
2002	0.654	0.367	0.337	-0.727	0.796
2003	-0.223	0.189	0.412	0.318	-0.491
2004	0.256	0.023	0.094	0.134	0.020
2005	0.190	-0.113	0.332	0.028	-0.560
2006	-0.422	-0.827	-0.390	-0.840	-0.968
2007	-1.046	-1.534	-0.010	-0.612	-1.330
2008	-1.019	-0.719	-0.171	-0.345	-0.368
2009	-0.089	0.422	0.048	0.301	0.298
2010	-0.032	-0.049	-0.450	0.089	-0.178
2011	-1.115	-1.070	-0.235	0.358	1.707
2012	-0.010	-0.018	0.216	-0.631	-0.649
2013	0.111	0.295	0.375	-0.535	-1.024
2014	1.021	0.053	0.233	1.392	1.595
2015	-0.122	0.732	0.071	-1.281	0.078

7.5.3 Repeatability of Simulated CPUE Data

The original statistical model was copied and the variance of the fitted values was increased in the copy so as to maintain the original for on-going use (**Figure 66**). Then, to test for the natural error-rate the year parameters were multiplied by 1.0 (so as to make no change). The simulated data was then generated 1000 times and a new standardization fitted to that simulated-observed CPUE data. The proportion of replicates that exceeded the prediction interval was tabulated (**Table 32**) and the 90^{th} percentile bounds of the 1000 replicate analyses plotted to determine the overlap with the original standardization (**Figure 68**).

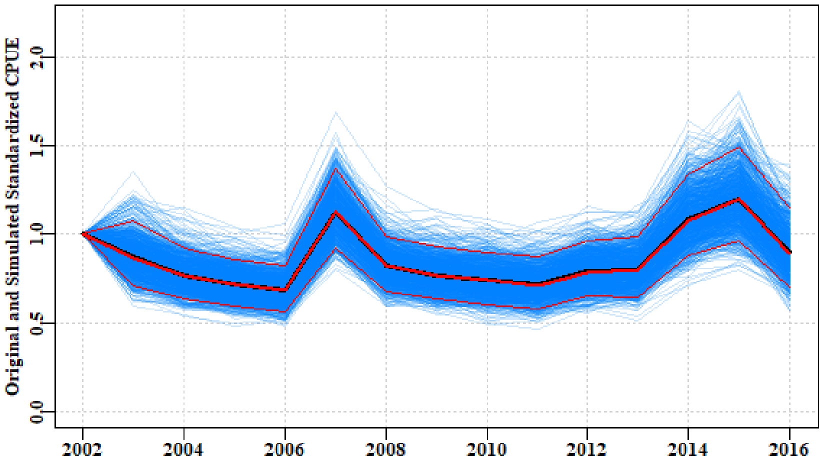

Figure 68. 1000 replicate standardization trajectories estimated from 1000 time-series of simulated-observed CPUE data which have had no changes made to the annual parameters. The density of blue reflects the density of lines. The black line is the original standardization. The red lines are the median of the 1000 replicates (thicker, on top of the black line) along with the 90^{th} prediction intervals (5% - 95% quantiles; expected range given the variation in the data).

Table 32. The proportion of 1000 replicate standardizations that lie above or below the 90^{th} percentile prediction intervals of the original standardization. The variance of the fitted values from which the simulated data was formed was randomly inflated by a normal random value with StDev 1.04726 (**Figure 66**). Mean of Low = 0.04657, mean of High = 0.44857

Year	Error Rate Low	Error Rate High	Multiplier	log(mult)
2002	0	0	1	0
2003	0.043	0.045	1	0
2004	0.048	0.039	1	0
2005	0.041	0.045	1	0
2006	0.044	0.044	1	0
2007	0.047	0.046	1	0
2008	0.045	0.055	1	0
2009	0.055	0.040	1	0
2010	0.046	0.043	1	0
2011	0.039	0.043	1	0
2012	0.045	0.038	1	0
2013	0.042	0.036	1	0
2014	0.042	0.049	1	0
2015	0.060	0.055	1	0
2016	0.055	0.050	1	0

The expected error rate when using the 90th prediction intervals is 5% above and 5% below (**Table 32**), so the statistical models are behaving as they should. The mean error rates vary depending on the pseudo random numbers used. But repeated trials vary with typical values between 0.044 – 0.053. As expected the median of the 1000 replicate standardizations essentially lays on top of the expected original mean trend (**Figure 68**).

7.5.4 Add a trend in Last Five Years

In the time-series from 2002 – 2016 an increasing trend was applied from 2012 – 2016 (**Table 33**).

Table 33. Empirical error rate from 1000 replicate samples where a model's fitted values have been randomly inflated by a normal random value with StDev = 1.04726. Mean values are no longer meaningful. The values refer to **Figure 69**.

Year	Lower Error Rate	Upper Error Rate	Multiplier	log(Multiplier)
2002	0	0	1	0
2003	0.057	0.057	1	0
2004	0.051	0.056	1	0
2005	0.039	0.045	1	0
2006	0.053	0.054	1	0
2007	0.052	0.058	1	0
2008	0.047	0.046	1	0
2009	0.046	0.046	1	0
2010	0.038	0.051	1	0
2011	0.043	0.042	1	0
2012	0.008	0.145	1.08	0.0770
2013	0.001	0.310	1.16	0.1484
2014	0.001	0.483	1.24	0.2151
2015	0	0.660	1.32	0.2776
2016	0	0.745	1.40	0.3365

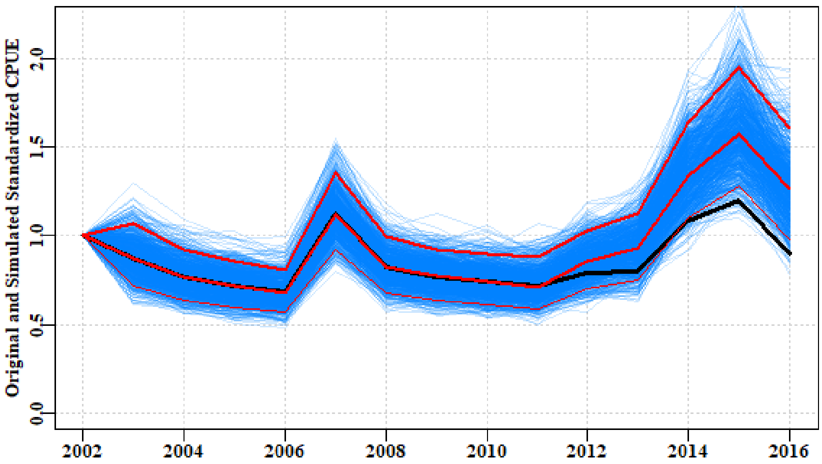

Figure 69. The effect on error rate of including an upward trend on the year parameters (see **Table 33**). The black line represents the original un-altered standardized means, which lie outside the 90th percentiles in 2015 and 2016..

The change in the general CPUE trend becomes apparent immediately in 2012 but only becomes significant (at $P = 0.9$) after 2014 (**Figure 69**) after which the 90th prediction interval lies above the original mean trajectory.

7.5.5 Add a Downward Trend to the Last Five Years.

In the time-series from 2002 – 2016 a decreasing trend of a smaller scale than the previous increasing trend was applied from 2012 – 2016 (**Table 34Table 33**).

Table 34. Empirical error rate from 1000 replicate samples where a model's fitted values have been randomly inflated by a normal random value with StDev = 1.04726. Mean values are no longer meaningful. The values refer to **Figure 70Figure 69**.

Year	Lower Error Rate	Upper Error Rate	Multiplier	log(Multiplier)
2002	0	0	1	0
2003	0.062	0.055	1	0
2004	0.056	0.054	1	0
2005	0.061	0.056	1	0
2006	0.062	0.055	1	0
2007	0.062	0.049	1	0
2008	0.052	0.056	1	0
2009	0.050	0.053	1	0
2010	0.054	0.054	1	0
2011	0.043	0.051	1	0
2012	0.124	0.013	0.94	-0.0619
2013	0.282	0.004	0.88	-0.1278
2014	0.473	0.003	0.82	-0.1985
2015	0.677	0	0.76	-0.2744
2016	0.776	0	0.7	-0.3567

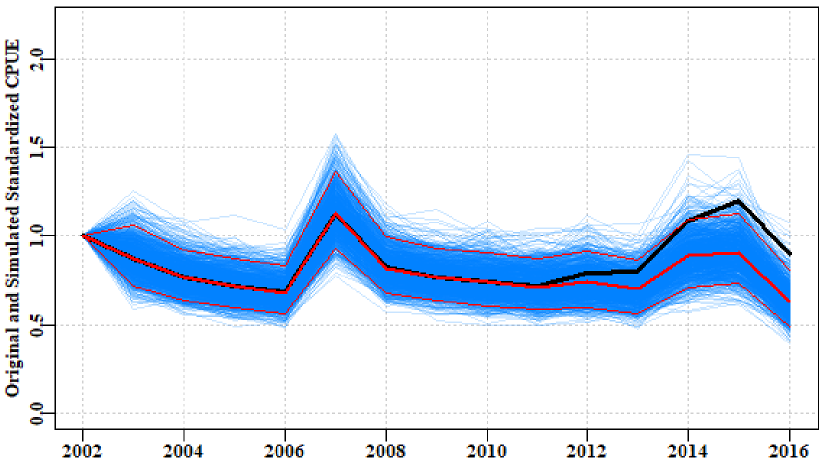

Figure 70. The effect on error rate of including a downward trend on the year parameters (see **Table 34**).

The effect of the downward trend on the year parameters is much more marked than the upward trend. To gain essentially the same impact the absolute change in the multiplier in the downward direction must be less. Note the maximum change of 1.4 upwards led to an upper error rate of about 0.745, whereas a change down to 0.7 led to an lower error rate of about 0.776 (compare **Table 33** with **Table 34**). This is simply a reflection of the log-transformation. The same pattern in the significant deviation from the original trend is also observed, only this time bellow the trend rather than above (**Figure 70**).

7.5.6 The Effect of Single Exceptional CPUEs

By placing exceptional annual CPUE values in the 2008 and 2012 positions within the time-series, with both reduced and increased values, the impact of single unusual events can be examined (**Table 35**; **Figure 71**).

Table 35. Empirical error rate from 1000 replicate samples where a model's fitted values have been randomly inflated by a normal random value with StDev = 1.04726. Mean values are no longer meaningful. The values refer to **Figure 71**.

Year	Lower Error Rate	Upper Error Rate	Multiplier	log(Multiplier)
2002	0	0	1	0
2003	0.047	0.045	1	0
2004	0.048	0.048	1	0
2005	0.048	0.044	1	0
2006	0.044	0.041	1	0
2007	0.049	0.046	1	0
2008	0.783	0.000	0.75	-0.2877
2009	0.054	0.044	1	0
2010	0.043	0.035	1	0
2011	0.047	0.045	1	0
2012	0.001	0.574	1.25	0.2231
2013	0.055	0.043	1	0
2014	0.051	0.051	1	0
2015	0.048	0.043	1	0
2016	0.037	0.052	1	0

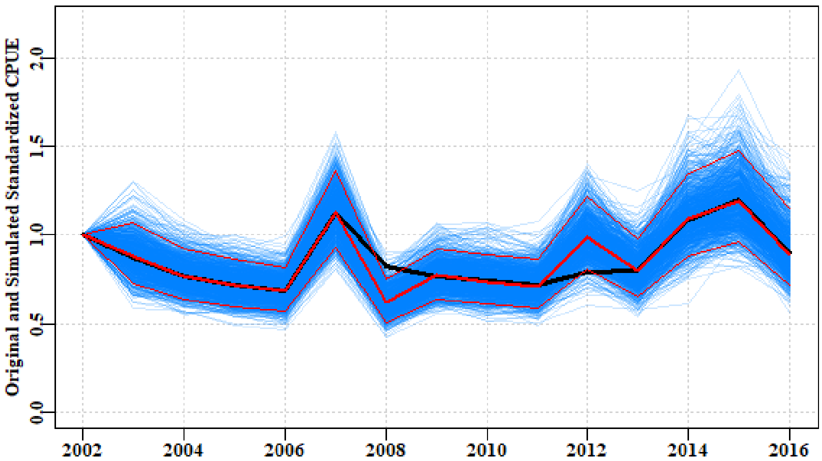

Figure 71. The effect on error rate of including a downward change in 2008 and an upward change in 2012 on the year parameters (see **Table 35**).

The effect of the singular events only influences the years with modified parameters, there is no lagged effects in subsequent years (**Table 35**). While it is possible to see the effect of the 25% deviation of the median line from the original standardization in the years altered, when considering 1000 replicates, the difference is large enough in both cases to push the 90% prediction interval of the simulated estimates beyond the original mean estimate (**Figure 71**).

7.5.7 Standardize using Gamma Errors

Randomized log-normal errors are included but with the multipliers all set to the one. The standardization, however, is conducted on the back transformed CPUE using Gamma errors in a GLM, using a log-link. 1000 replicates generated lower and upper error rates almost double that of the unadjusted log-normal standardizations (**Table 36**).

Table 36. Empirical error rate from 1000 replicate samples where the log-normal error structured data were standardized using Gamma errors on the CE data with a log-link. The mean estimates were 0.0982 and 0.09779 for the lower and upper error rates respectively. The values refer to **Figure 72**.

Year	Lower Error Rate	Upper Error Rate	Multiplier	log(Multiplier)
2002	0.000	0.000	1	0
2003	0.088	0.101	1	0
2004	0.092	0.099	1	0
2005	0.095	0.102	1	0
2006	0.084	0.106	1	0
2007	0.096	0.098	1	0
2008	0.104	0.111	1	0
2009	0.098	0.098	1	0
2010	0.099	0.106	1	0
2011	0.104	0.093	1	0
2012	0.106	0.093	1	0
2013	0.094	0.090	1	0
2014	0.103	0.094	1	0
2015	0.100	0.092	1	0
2016	0.113	0.085	1	0

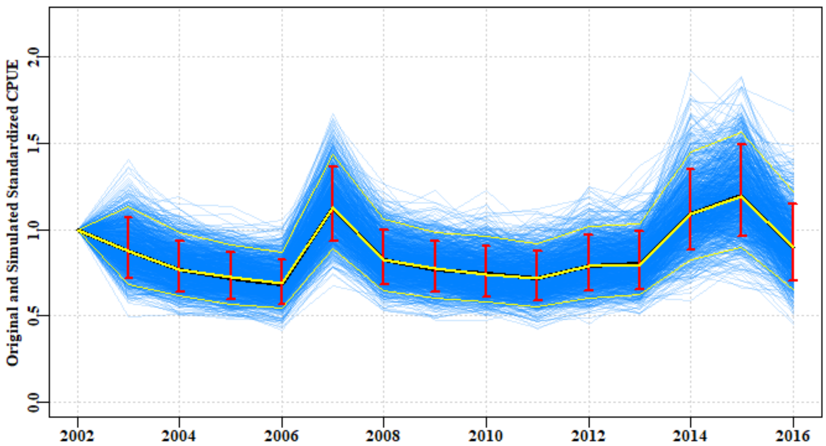

Figure 72. The effect on error rate of using Gamma errors on CE data with a log-link for the standardization (see **Table 36**). The Gamma fitted models are compared to the 90% prediction intervals for the original log-normal linear model (the red bars). The yellow lines are the 5%, 50% and 95% quantile of the simulated standardizations.

The general trend obtained from using the Gamma error model on the back-transformed CPUE data (using the log-link in the GLM) generates CPUE trajectories of the same overall general trend with trivial within year deviations. However, the Gamma error approach appears to be noisier when given log-normally distributed data with wider prediction intervals.

7.5.8 The Comparison of the Original with the Gamma Fitted Model

When the original log-normal standardization of Blue-Eye was compared with the Gamma error Generalized Linear Model on CPUE with a log-link there were large differences exhibited (**Figure 50**, on page 105). Once the data are trimmed of the lower proportion Blue-Eye shots (that mostly contain Pink Ling) the log-transformed CPUE data become rather more normally distributed and this has a marked effect on both standardizations.

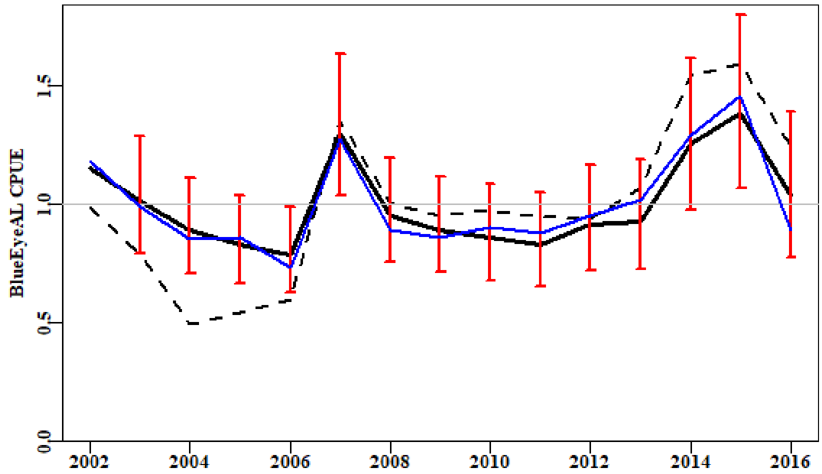

Figure 73. A comparison of the log-normal standard (black line with red confidence intervals) with the Gamma Error GLM fit on CPUE (blue line), illustrating the increased similarity between these two approaches when using more log-normally distributed log(CPUE).

There remain some differences in particular years, for example, 2013, 2015, and 2016. While the removal of records containing less than 10% of Blue-Eye relative to Pink Ling did have a normalizing effect on the combined data when the log(CPUE) is inspected by year (**Figure 74**) there are obvious deviations from normality especially in some years. Not all years with relatively low numbers of records and not particularly normal distributions (e.g. 2002) lead to discrepancies between the two trajectories. But those strongly affected by the rounding of catches or effort leading to strong spikes in the CPUE relatively often exhibit discrepancies between the Gamma Error model and the log-normal model.

Despite such deviations, the general trend remains the same. Where differences do arise is with the implied variability of the Gamma Errors simulation models, which is expressed as being more variable that the log-normally distributed model (which behaved appropriately under simulation (section 7.5.3 on page 132).

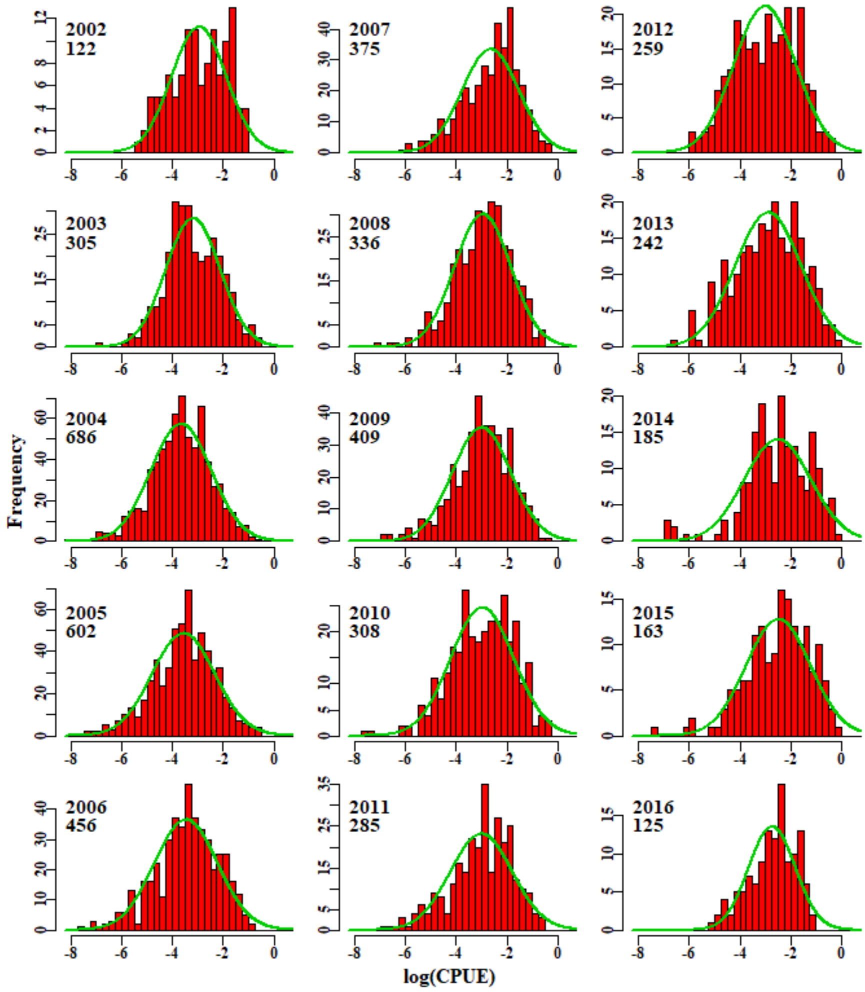

Figure 74. A histogram of each year's log(CPUE) distribution. Each plot is labelled with the year and the number of observations.

Table 37. Comparison of the original Gamma error and log-normal standardizations with the median of the simulated standardizations.

Year	OrigGamma	OrigLogN	medianSim	GammvsLog	LogvsSim
2002	1.2435	1.1522	1.1539	-7.93	-0.147
2003	1.0536	1.0086	1.0108	-4.46	-0.215
2004	0.9368	0.8875	0.8889	-5.55	-0.156
2005	0.8429	0.8296	0.8377	-1.61	-0.978
2006	0.8075	0.7881	0.7924	-2.46	-0.539
2007	1.3317	1.2993	1.3021	-2.49	-0.211
2008	0.9501	0.9493	0.9492	-0.08	0.009
2009	0.7330	0.8886	0.8922	17.51	-0.409
2010	0.7678	0.8561	0.8518	10.31	0.513
2011	0.8310	0.8271	0.8271	-0.47	0.004
2012	0.9397	0.9128	0.9122	-2.95	0.060
2013	0.9556	0.9266	0.9211	-3.13	0.597
2014	1.2946	1.2547	1.2548	-3.18	-0.007
2015	1.2469	1.3834	1.3742	9.87	0.667
2016	1.0653	1.0360	1.0317	-2.83	0.410

While there were relatively large deviations between the original log-normal and the Gamma error standardizations from a minimum of about -0.5% up to 17.5%, with the median simulated standardizations using Gamma errors the range was only from 0.004% up to -0.978%.

The simulated data from the log-normal standardization appears to be better behaved in terms of normality than the original (compare **Figure 75** with **Figure 74**). The inclusion of normal random errors onto the fitted values has retained some of the spikiness resulting from the rounding of the original data but has still smoothed the distributions in each year and improved their log-normality.

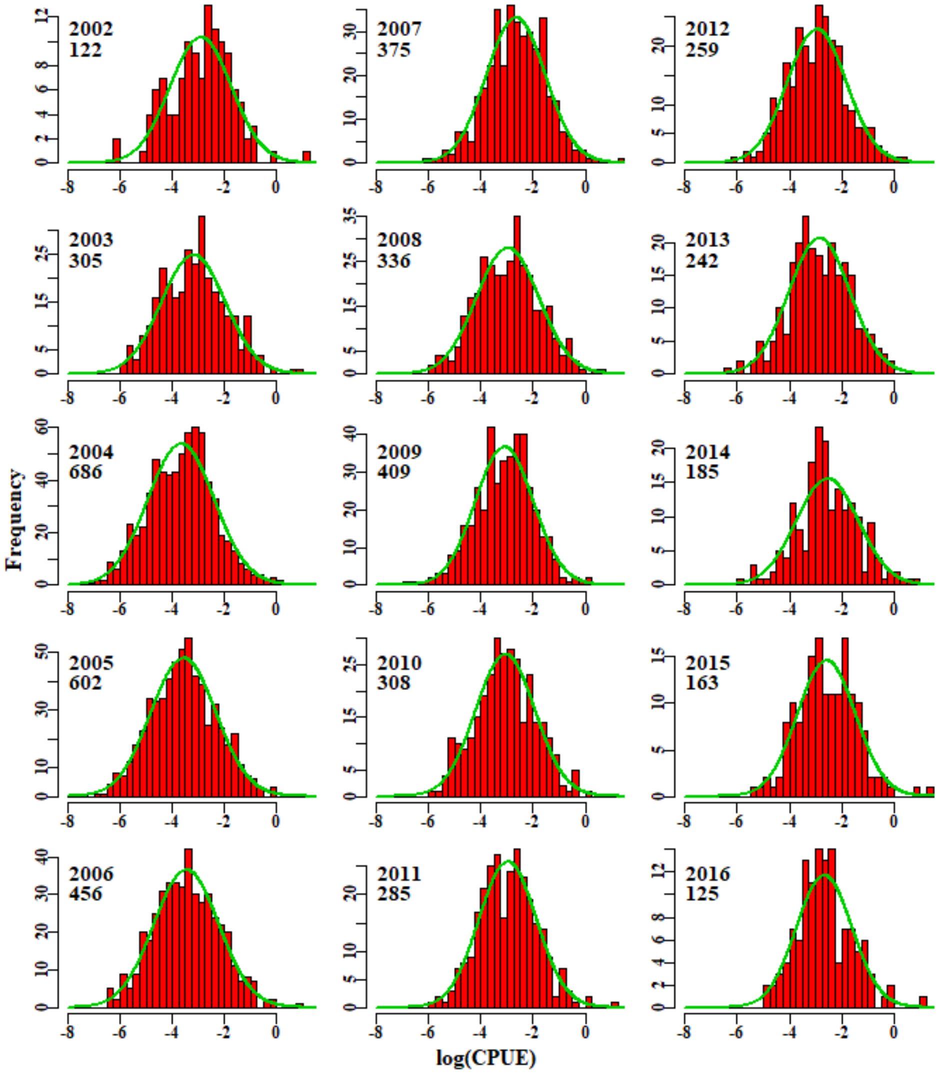

Figure 75. A histogram of a single simulation's log(CPUE) distribution for each year. Each plot is labelled with the year and the number of observations. The number of observations will always be the same as in the original data.

7.6 Discussion

The simulation of realistic CPUE data is more difficult that hoped but by modifying fitted yearly parameters and increasing the variation of the fitted values it is possible to generate plausible simulated CPUE data that follows a predetermined annual trajectory.

When the variability of the annual mean estimates is not shrunk by having 100,000s of records, as with Flathead, then relatively large deviations from the median trends are required to detect significant differences from the overall trend.

7.6.1 Repeatability of Simulated CPUE Data

When no changes are made to the data but the variation of the simulated (predicted) log-transformed CPUE values is expanded to match the variation in the original observations, when 1000 replicates are run the simulated data generated behaves as it is expected to with, on average, 90% of the replicates falling between the 90% prediction intervals. This implies that the simulation process does not introduce bias to the outcome and the precision of the annual mean estimates remains the approximately the same. This result is reflected in the comparison of the simulated data with bootstrap samples from the residuals of the original standardization.

This implies that when deviations from the original standardized trajectory, or the 90% prediction intervals, are observed following some treatment of the statistical model's annual parameters this is a reflection of that interference.

7.6.2 A Cumulative Change in Last Five Years

Both an increasing and a decreasing trend were applied to the estimated annual year parameters to determine the relative sensitivity to such trends. That the downward trend was more influential or sensitive than the upward trend is not surprising given that when dealing with log-transformed CPUE such changes are proportional (hence a 40% increase is smaller than a 40% decrease). This is why a 30% downward trend was able to generate about the same absolute change in error rate and lead to the median simulated mean exceeding the 90% prediction interval of the original CPUE trend as a 40% increasing trend. The change needed to be significantly different (in terms of breaching the 90% prediction intervals) was about an 18 - 20% change. But this would be idiosyncratic to the particular analysis being used in the illustration, being reflective of the original variation in the observed data.

This raises the issue of what it means to be significantly different. The classical notion of using the confidence with which the annual mean estimates are made when comparing mean trends becomes questionable when the precision of such estimates is a direct function of the number of observations and, as seen with flathead. When there are 270,000+ records across 30 years (an average of 9000 records a year) the classical confidence intervals are often so tight that they can be difficult to distinguish from the mean plotted point. This is why here we have been considering the spread of values observed and finding the 90th percentile interval (the prediction interval).

It means that what constitutes a significant difference between two CPUE trends is difficult to determine when an appropriate measure of variation around the mean estimates is unavailable. This issue stems in part from the serially correlated nature of CPUE data. The statistical analyses used assume each year's data is an independent sample and that the factors used in the standardization are independent. Both of these assumptions are incorrect. For CPUE to be useful as an index of relative abundance then it needs to reflect stock abundance in some way. As a stock's abundance is serially correlated through time this implies that any CPUE will also be serially correlated (in fact, this may not be the case with species having very short life-spans such as some squid and other small rapidly living species). It is also very common for correlations between factors in any standardization to sometimes be very high, which sometimes can be accounted for by interaction terms, but oft-times, when correlations are strongly non-linear, interaction terms are insufficient.

The independence of individual years is also illustrated by the lack of any lag effects following interventions on single year parameters. Unless such changes are due to external events that are not accounted for in the standardization, one would expect a surge in abundance to trail into subsequent years. On the other hand, it is possible to oceanographic events to influence the availability to some species and this can lead to a surge or a decline in CPUE in a particular year. In effect those simulations demonstrated that the years are indeed treated separately, and singular events, which affect availability rather than abundance, can be confusing when interpreting CPUE trends through time.

7.6.3 The Use of Gamma Errors on CPUE

When log-normally distributed simulated data was modelled using a Generalized Linear Model on CPUE with Gamma Errors and a log-link the median trend of the 1000 replicates was essentially the same as the original standardized trend when using the log-normal errors. This was despite the original Gamma Error model differing from the log-normal statistical model in minor ways (**Figure 73**). Importantly, when approximately log-normally distributed CPUE data is analysed using Gamma errors the overall variance exhibited by the outcome tends to be rather higher than if log-normal errors are used.

This increased variability may be a reflection of using the CPUE without log-transformation. While it is true that the Gamma Error statistical model uses a log-link to relate the mean or the predicted values with the mean of the observed values. The log-transformation acts directly to stabilize the variance of the CPUE as well as connecting the mean of the observations with the predicted values (Venables and Dichmont, 2004). Venables and Dichmont (2004) provide a detailed discussion of the differences, negatives, and positives of using either the “transformation approach” (log-normal errors) relative to using Gamma errors. As they conclude:

“In our experience, the transformation approach is often more realistic for catch rate data, particularly since the gamma distribution has a much thinner upper tail than the lognormal. Very fat upper tails are often a feature of catch rate distributions. Another way of looking at this is to note that the error term also acts multiplicatively on the response for the transformation model. For the gamma model, the fixed factors do so, but the error term, which is not simply added to the linear predictor in this case, does not. On the transformed scale, diagnostics are certainly simpler and easier to appreciate.” (Venables and Dichmont, 2004, p 324).

The Blue-Eye data, before the removal of the records containing less than 10% Blue-Eye relative to Pink Ling, illustrates well that as the observed data used strays further from the assumption of log-normality then the two approaches compared differ more and more. Even though differences between the two approaches remained when analysing the trimmed data the trends were closer to coincident than when using the untrimmed data. Close attention to the diagnostics for any analysis remains important.

With CPUE standardization, usual fisheries data has many features that compromise the assumptions behind any analysis. This is especially the case in mixed fisheries where the CPUE for a particular target species may be compromised by the fishers setting their gear in such a way as to capture a mixture of species rather than targeting a single species. When that factor is added to the many other sources of variation it becomes clear that how best to handle the CPUE standardization of mixed fisheries remains one of the major problems in CPUE standardization.

7.7 Appendix: R-code used in Simulations

See **6.5 Appendix: R code used to Conduct Standardizations** for the R-functions called in the following

```
wkdir <- getwd()
resdir <- paste0(wkdir,"/results/")
options("show.signif.stars"=FALSE,"stringsAsFactors"=FALSE,
        "max.print"=50000,"width"=240)
library(codeutils)
library(r4cpue)
library(r4maps)
source("C:/A_CSIRO/Rcode/CPUEExplore/utils/extra_utils.R")
source(paste0(wkdir,"/final_simulation_utils.R"))

# Use BluePink data -------------------------------------
spsB <- read.csv("bluepink.csv",header=TRUE)
dim(spsB)
head(spsB)
properties(spsB)
dim(spsB)
pickrec <- which((spsB$Year > 2001) & (spsB$Year < 2017) &
                    (spsB$Zone %in% c(20,30,40,50)) &
                    (spsB$propB > 0.1))
spsA <- droplevels(spsB[pickrec,])
dim(spsA)
# Count the number of years in the fishery
spsA <- addcount(spsA,"Vessel")
pickV <- which(spsA$count > 1)
sps1 <- droplevels(spsA[pickV,])
dim(sps1)
properties(sps1)
# Model1 Standardization ----------------------------------------------------
splabel <- "BlueEyeAL"
savefile <- FALSE
labelM <- c("Year","Vessel","Month","Zone","DepCat","Month:Zone")

years <- 2002:2016
table(sps1$Year)

sps2 <- makecategorical(labelM,sps1) # convert variables to factors
mods <- makemodels(labelM)
out1 <- standLM(mods,sps2,splabel)
out1$WhichM
model1 <- out1$optModel
lmgeo <- out1$Results[,1] # lm standardized geometric mean
answer1 <- getfact(out1,"Year")
rownames(answer1) <- years

plotprep()
plotstand(out1,bars=TRUE)

plotprep(height=7)
impactplot(out1,mult=4)


plotprep(width=7,newdev=FALSE)
qqdiag(model1)


# FIND BEST STDEV ---------------------------------------
# prepare simulation ------------------------------------------------
columns <- c(labelM[1:5],"Depth","Long","Lat","CE","LnCE")
sps0 <- sps1[,columns]
nsim <- length(sps0$LnCE)
sps0$sim <- NA
set.seed(12345)

# Find required sd ----------------------------------------------------
# Diff Obs vs Predicted
plotprep(height=4.5,newdev = FALSE)
stdev <- 1.047   # alter this by trial and error to find the closest matching sd
sps0$sim <- model1$fitted.values + rnorm(nsim,mean=0,sd=stdev)

par(mfrow=c(2,1),mai=c(0.25,0.45,0.05,0.05),oma=c(1.0,0,0.0,0.0))
par(cex=0.85, mgp=c(1.35,0.35,0), font.axis=7,font=7,font.lab=7)
```

```
bins <- seq(-9.0,2,0.125)
outh <- hist(sps0$LnCE,breaks=bins,col=2,border=1,main="")
ans1 <- addnorm(outh,sps0$LnCE,inc=0.1)
lines(ans1$x,ans1$y,lwd=2,col=5)
text(-9,150,"Observations",cex=1.1,font=7,pos=4)
text(-9,100,paste0("Mean  = ",round(ans1$stats[1],3)),cex=1.1,font=7,pos=4)
text(-9,50,paste0("StDev = ",round(ans1$stats[2],3)),cex=1.1,font=7,pos=4)
outf <- hist(sps0$sim,breaks=bins,col=2,border=1,main="")
ans2 <- addnorm(outf,sps0$sim,inc=0.1)
lines(ans2$x,ans2$y,lwd=2,col=4)
lines(ans1$x,ans1$y,lwd=2,col=5)
text(-9,150,"Fitted Values",cex=1.1,font=7,pos=4)
text(-9,100,paste0("Mean  = ",round(ans2$stats[1],3)),cex=1.1,font=7,pos=4)
text(-9,50,paste0("StDev = ",round(ans2$stats[2],3)),cex=1.1,font=7,pos=4)
mtext("Log(kg/hook)",side=1,outer=T,line=-0.1,font=7,cex=1.0)

cat(ans1$stats,range(sps0$LnCE),"\n")
cat(ans2$stats,range(sps0$sim),"\n")


# PLOT THE OPTIMUM STDEV ------
reps <- 5000
resultM <- numeric(reps)
resultSD <- numeric(reps)
origM <- ans2$stats[1]
origsd <- ans2$stats[2]
stdev <- 1.04726
for (i in 1:reps) {
   sim <- model1$fitted.values + rnorm(length(model1$fitted.values),mean=0,sd=stdev)
   outh4 <- hist(sim,breaks=seq(-12,4.0,0.1),plot=FALSE)
   ans4 <- addnorm(outh4,sim)
   resultM[i] <- origM - ans4$stats[1]
   resultSD[i] <- origsd - ans4$stats[2]
}

plotprep(height=5.0,newdev=FALSE)
par(mfrow=c(2,1),mai=c(0.45,0.45,0.05,0.05),oma=c(0.0,0,0.0,0.0))
par(cex=0.85, mgp=c(1.35,0.35,0), font.axis=7,font=7,font.lab=7)
outhM <- hist(resultM,main="",col=2,breaks=40, ylab="Frequency",
              xlab="Difference between Means")
ans <- addnorm(outhM,resultM,inc=0.0005)
lines(ans$x,ans$y,lwd=2,col=4)
out <- ans$stats
round(out,10)
abline(v=ans$stats[1],col=3,lwd=2)
ymax <- max(outhM$counts,na.rm=TRUE)
xmin <- min(outhM$breaks)
text(xmin,0.9*ymax,paste0("Avdiff = ",round(mean(resultM,na.rm=TRUE),5)),
     cex=1.1,font=7,pos=4)
text(xmin,0.7*ymax,paste0("stdev = ",stdev),cex=1.1,font=7,pos=4)

outhS <- hist(resultSD,main="",col=2,breaks=40, ylab="Frequency",
              xlab="Difference between StDevs")
ansS <- addnorm(outhS,resultSD,inc=0.0005)
lines(ansS$x,ansS$y,lwd=2,col=4)
outS <- ansS$stats
round(outS,10)
abline(v=ansS$stats[1],col=3,lwd=2)
ymax <- max(outhS$counts,na.rm=TRUE)
xmin <- min(outhS$breaks)
text(xmin,0.9*ymax,paste0("meandiff = ",round(mean(resultSD,na.rm=TRUE),5)),
     cex=1.1,font=7,pos=4)
text(xmin,0.7*ymax,paste0("stdev = ",stdev),cex=1.1,font=7,pos=4)


# DO SIMULATION --------------------------------------------------------
#modify model coeff -----
inmod <- model1

loccoeff <- grep("Year",names(inmod$coefficients))
ncoef <- length(loccoeff)
mult <- rep(1,14)
if (length(mult) != 14) stop(cat(length(mult)," mult is wrong length"))
inmod$coefficients[loccoeff] <- (inmod$coefficients[loccoeff] + log(mult))
mult <- c(1,mult)

stddev <- 1.04726
years <- 2002:2016
```

```
nyrs <- length(years)

# loop the simulation --------------
labelM <- c("Year","Vessel","Month","Zone","DepCat","Month:Zone")
splabel <- ("Blue_Eye_Sim")
yearfact <- getfact(model1,"Year")
original <- yearfact[,"Coeff"]
origse <- yearfact[,"SE"]
reps <- 1000

columns <- c("residSE","MultR2","AdjR2","Fval","df1","df2")
statout <- matrix(0,nrow=reps,ncol=length(columns),dimnames=list(1:reps,columns))
results1 <- matrix(0,nrow=reps,ncol=nyrs,dimnames=list(1:reps,years))
scaleres <- matrix(0,nrow=reps,ncol=nyrs,dimnames=list(1:reps,years))

for (i in 1:reps) {
   spsS <- sps1
   spsS$LnCE <- predict(inmod) + rnorm(length(spsS$LnCE),mean=0,sd=stddev)
   sps3 <- makecategorical(labelM,spsS) # convert variables to factors
   mods <- makemodels(labelM)
   outS <- standLM(mods,sps3,splabel,console=FALSE)
   modelS <- outS$optModel
   smod <- summary(modelS)
   statout[i,] <- c(smod$sigma,smod$r.squared,smod$adj.r.squared,
                    smod$fstatistic)
   outcoef <- getfact(modelS,"Year")
   results1[i,] <- outcoef[,"Coeff"]
   scaleres[i,] <- outcoef[,"Scaled"]
   if ((i %% 25) == 0) cat(i,"\n")
}

plotprep(width=7,height=5,newdev=FALSE)
plot(years,original,type="l",lwd=2,ylim=c(0,2.2),panel.first=grid(),
     xlab="",ylab="Original and Simulated Standardized CPUE")
Zmult <- -qnorm((1-(90/100))/2.0)
lower <- original * exp(-Zmult*origse)
upper <- original * exp(Zmult*origse)
for (i in 1:reps) lines(years,results1[i,],lwd=1,col=rgb(0,0.5,1,1/5))
lines(years,original,lwd=3,col=1)
qs <- apply(results1,2,quants)
lines(years,qs["50%",],lwd=2,col=2)
lines(years,qs["5%",],lwd=1,col=2)
lines(years,qs["95%",],lwd=1,col=2)
# arrows(x0=years[-1],y0=lower[-1],x1=years[-1],y1=upper[-1],
#        length=0.035,angle=90,col=2,lwd=2,code=3)
# abline(v=c(2007.8,2011.8),col=3)

pgtupper <- numeric(nyrs)
pltlower <- numeric(nyrs)
for (i in 2:nyrs) {
   pgtupper[i] <- length(which(results1[,i] > upper[i]))
   pltlower[i] <- length(which(results1[,i] < lower[i]))
}
proplower <- (pltlower/reps); mean(proplower[2:15])
prophigher <- (pgtupper/reps); mean(prophigher[2:15])
toXL(cbind(years,proplower,prophigher,mult,log(mult)))
```

8 Guidelines for using CPUE data in Assessments

8.1 Introduction

Much of the material in this concluding chapter should appear to be extremely obvious to any practitioner of statistical standardization. Nevertheless, it bears writing because very often in reports on CPUE standardization clear statements of tasks undertaken, data selection criteria used, assumptions made, diagnostics reviewed, and analyses conducted are not all explicitly included. It is possible that because these analyses eventually become treated as routine such requirements for details become relaxed. This is, of course, an error and not merely on the part of the analyst. It is a mistake because in all cases the analyses should always be defensible and the best way of defending an analysis is to allow it to be repeated. There is a growing move towards automated document generation (Xie *et al*., 2018) which can include the analytical code used to conduct analyses. For example, Haddon and Sporcic (2017) was generated using such an Rmarkdown file (sessf-cpue.Rmd). Anyone with access to the data files and that .Rmd file should be able to completely repeat the whole document. Despite this automation each species/fishery in Haddon and Sporcic (2017) is given individual treatment and customization as reflects the idiosyncrasies of each fishery. Such documents can be repeated and improved through constructive criticism each time the standardizations are needed to be repeated or updated with new data. The emphasis with such an approach is on the repeatability, it is important to understand that there remains a need to provide individual attention to each fishery to account for the very many differences and individualistic properties of each situation.

An important point that needs emphasis is that an appropriate analysis of CPUE data can take a great deal of time. The initial data exploration, the trialling of alternative statistical models, alternative error structures and the many different options available can involve many days of effort or longer. Often such time is not made available and this constitutes a threat to any such analysis. The trade-offs between the value of the fishery concerned and any risks an invalid analysis may lead to should be considered when conducting an analysis in a time-constrained situation.

8.1.1 First Step: Data Exploration

Attempting to write guidelines for using CPUE data in stock assessments is an ambitious and potentially condescending undertaking. Each fishery of substance invariably has idiosyncrasies that ideally require individual attention and special treatment. To generate guidelines to cover all eventualities may not even be possible. Nevertheless, an attempt will be made here which, it is hoped, could form the basis for others to improve upon, with the final aim of improving the standard of practice with regard to CPUE standardization. From what has been illustrated in previous chapters it seems clear that no single approach to conducting CPUE standardization is the “best” way, and perhaps the best guideline is to assert that any statement that claims that approach “XYZ” is always the best way is very likely wrong. A more congenial manner of stating that is to claim that generalities are difficult to make when discussing CPUE standardization.

Despite such difficulties if catch and effort data are available for a fishery then before their use for anything, the stock assessment scientist has a need to understand their strengths, weaknesses, and as much about their properties as time permits. Initially each data-set of fisheries catch and effort data needs to be treated in an exploratory and adaptive manner. Until its characteristics are better known decisions concerning about how best to use such data should not be made. It may be found that the data previously thought to form a coherent and consistent whole is, in fact, so poorly representative of the total fishery that its use needs to be restricted to only a sub-set of the fishery (which may not be useful for management). But, as always, it is better to know the limitations of one's data than to assume, perhaps rashly, that the available data will be informative about the fishery of interest.

It is already clear that very many factors other than relative abundance through time can influence observed CPUE therefore it is reasonable to state that before their use in a stock assessment a statistical standardization of the expected average catch rate per year is required. In many cases a standardization has little effect, but in others it has an enormous influence (**Figure 76**).

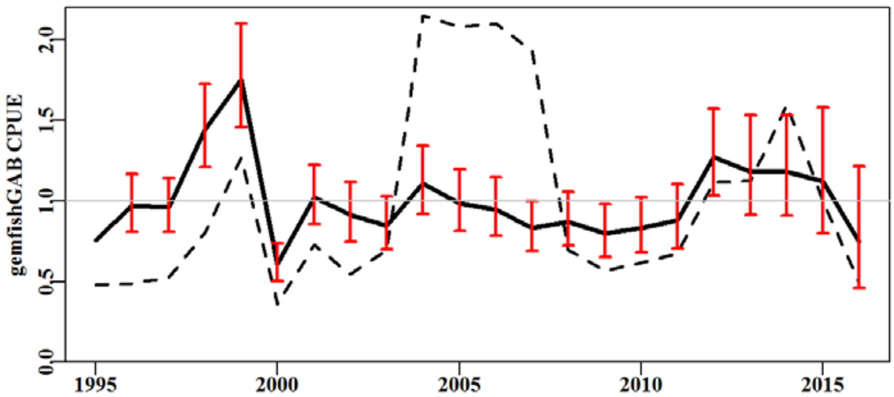

Figure 76. A plot comparing the unstandardized geometric mean CPUE with the standardized CPUE for western gemfish (*Rexea solandri*) in the Great Australian Bight (copied from Haddon and Sporcic, 2017). Note the massive increase in unstandardized CPUE between 2004 – 2007, the effect of which was removed by the standardization.

Prior to such an analysis however, it is also always best to know the assumptions inherent in the available data regarding coverage across the fishery, the representativeness of the reporting (is discarding an important factor). An important aspect of the process is simple data characterization.

8.1.2 Breaks in Time-Series

One advantage of using a repeatable document generating process is that it simplifies the inclusion of additional data analyses and characterizations. A strong assumption made in any time-series of data is that it is in fact a single time-series without any breaks. This can be examined by plotting different aspects of the data by year to identify any major changes in fisher behaviour. An example might be the distribution of the log(CPUE) in each year (see **Figure 74** on page 138). But this principle can be extended to every major factor in the fishery (see **Figure 77** for a similar plot of depth distribution by year). The catch by vessel by year can also be highly informative, although confidentiality constraints can often prevent such diagrams being made public they can still be generated for use by the analyst to aid in understanding the dynamics of a given fishery. The Blue-Eye auto-line fishery, for example, is dominated by just a few vessels and unless that is known some of the changes that occur through time cannot be understood.

CPUE is primarily a reflection of fishing behaviour so important properties to look for in a data exploration include:

- unexpected breaks in time-series, meaning large, biologically implausible or unexplained changes in CPUE,
- the influence of any large management changes (introduction of quotas, large closures, introduction of harvest strategies) on vessels or fishing behaviour,
- large changes in operating vessels or fishers
- consistent trends in depth, season, or location of fishing,
- large changes in associated species commonly taken with the primary species of interest,
- other factors of particular concern to particular fisheries.

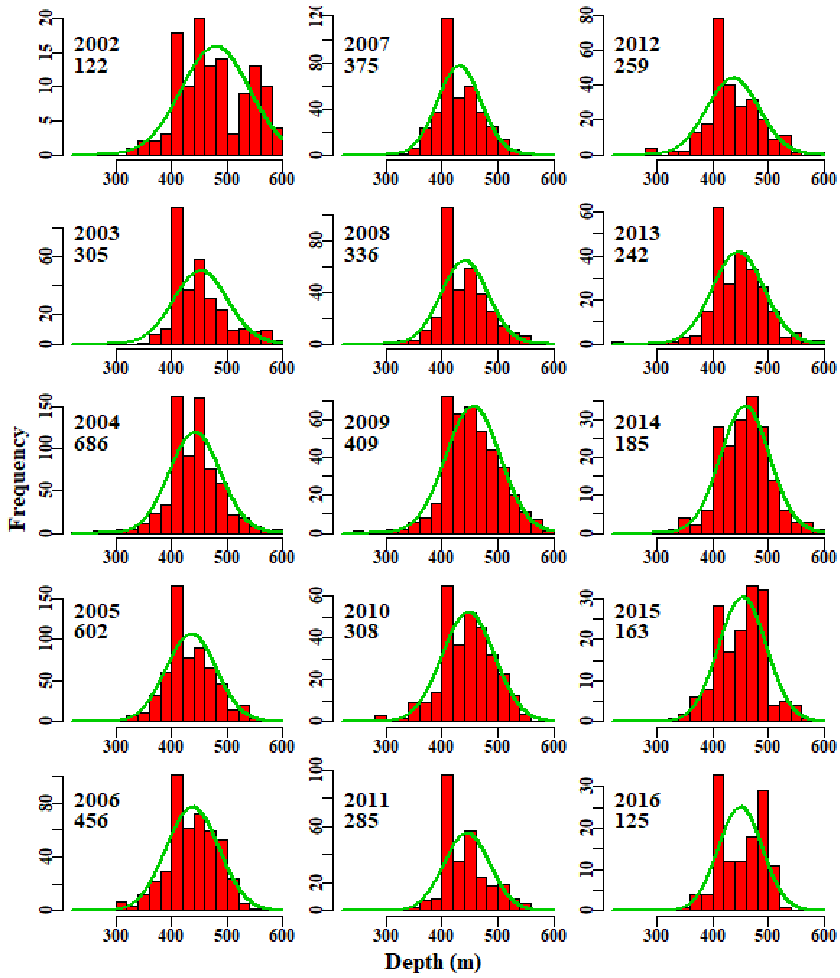

Figure 77. A histogram of the reported depth for each record in each year. Each plot is labelled with the year and the number of observations.

One idea behind such a data exploration is to understand the history of each fishery so as to identify and, ideally, understand, periods of change. It is only once the analyst has some understanding of a fishery that it becomes sensible to attempt to conduct a standardization that accounts for the more important factors affecting that fishery.

8.1.3 Try Alternative Approaches

As mentioned in Chapter 7, there are very many alternative approaches that can be used when standardizing CPUE data. When examining the literature (published and grey) it can appear that selecting which approach to use is at least partially dependent upon where the analysis is to take place (the approach taken appears to be influenced by whether there is a ‘tradition’ in the jurisdiction concerned), and who is doing the analysis (individuals tend to stick to methods they are familiar with). These are not necessarily bad reasons for picking an approach as the ability to communicate any results successfully can depend upon what the assessment groups have experienced in the past. However, before committing to any particular analytical direction it is best to attempt more than one approach so as to be able to examine the consequences of applying the different assumptions of each different approach. Ideally, if they all generate essentially the same result then the simplest to implement can be adopted. If they generate significantly different results then this can be considered an opportunity to learn more about how informative the data is about underlying changes in the stock. Alternatively, the different standardized trajectories can be used as alternative scenarios in the stock assessment of which they are a part. Fishery stock assessment invariably have many sources of uncertainty. The idea that there is a single standardized CPUE time-series that captures all of the uncertainty in the CPUE data is a simplification.

In practice, however, there is always limited time available for conducting standardizations and subsequent stock assessments. Fortunately, for most species that are sufficiently valuable to lead to a formal stock assessment, any standardization is likely to be repeated at regular intervals. This means that initially there may be sufficient data exploration to discover the standardization approach that leads to efficient, repeatable, and consistently interpretable results. Thus, in practice, with a fishery being assessed in detail for the first time, a full data exploration phase should occur and lead to the selection of a standardization strategy that can be applied consistently into the future (hence the value in a repeatable standardization and assessment document). Even though it is efficient to make such a selection, it remains a sensible strategy to review the selection of methods every few years (perhaps every five years) as the advent of new data may alter what constitutes the most effective method to use in standardizing the CPUE.

8.1.4 Standardization Diagnostics

Standardizations should include diagnostics and those should include but also go beyond the classical statistical diagnostics. The statistical diagnostics should at least be applied to the optimal statistical model selected to describe the available data. These include the classical plots (see **Figure 39** on page 97) and enhanced versions of these plots (see **Figure 63** on page 125). But, in addition, the plots of how the variation in the data is accounted for as factors are added (see **Figure 63** on page 125) and how any trend in the CPUE is altered by the addition of each new factor (see **Figure 60** on page 124; see also Bentley *et al*., 2012)

The factors available and those that are used are rarely independent of each other, which means that the order in which they are added to the statistical model can influence the proportion of the variance described by the total model that is accounted for by each factor. Automated step-wise algorithms that search for combinations of factors have some form of criterion associated with their use; often relating to variance accounted for. Unfortunately, the variance accounted for in the statistics does not always reflect the influence that a factor can have on the final trend in the standardized CPUE, though often they are correlated.

If the analyst is fortunate enough to have so many factors they do not know where to start, then some automated method of adding or subtracting factors to a model is a reasonable place to begin. However, in the final run the factors that have most influence on the trends, not necessarily the variance accounted for, should be entered into the model first followed by those with less influence. In many fisheries it turns out that most of the difference between the unstandardized and the standardized trends is accounted for by the first three or four factors after which little change tends to occur. Interaction terms can often account for a relatively high proportion of variation but their effect on the CPUE trend is what is important and often their influence is greatly lessened once the singular factors are included. Only be exploring these options can these insights become clear.

The optimal CPUE trend from the latest analysis should always be compared with the trend from at least the previous analysis (if there was one) to determine whether there have been any systematic changes occurring through time (a form of retrospective plot). In addition, this provides a double check that the database of CPUE data is giving the same observed data each year. Graphically, it is easier to make visual comparisons among trends by setting each time-series (of the same length) to a mean of 1.0 by dividing through the values in the time-series by the mean of the time-series:

$$CE = \frac{I_y}{\overline{I}}$$

where CE is the scaled time-series of CPUE, I_y is the standardized CPUE for each year y, and $\overline{I}$ is the mean of the time-series. This places the emphasis on any trend in the time-series rather than the absolute values. When comparing a new standardization with an older one then both need to be scaled to the mean across the same series of years.

A retrospective analysis on a given dataset will illustrate the effect of vessels coming and going from the fishery as well as other changes that might be more difficult to identify. This is especially informative when some of those vessels are major players in the fishery (**Figure 78**).

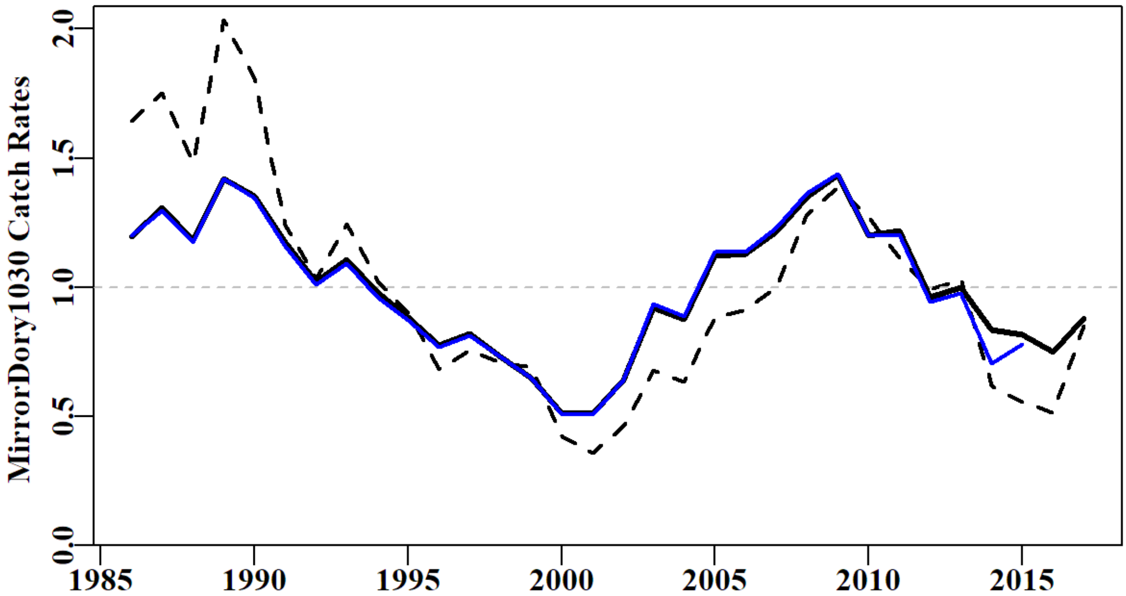

Figure 78. A comparison of the standardization for Mirror Dory in SESSF zones 10 – 30 for data to 2015 (blue line) and to 2017 (black line). The dashed line is the geometric mean or unstandardized trend.

In the SESSF the data generally becomes relatively stable after three or four years have passed. Differences are often due to extra data being entered into the catch effort data base.

8.2 Appendix: R-code used in Standardizations

```
#' @title histyear plots a histogram of a given variable for each year available
#'
#' @description histyear plots a histogram of a given variable for each year
#'    available
#'
#' @param x the data.frame of data with at least a 'Year' and pickvar present
#' @param Lbound leftbound on all histograms, defaults to -3.5
#' @param Rbound right bound on all histograms, defaults to 12.25
#' @param inc  the class width of the histogram, defaults to 0.25
#' @param pickvar which variable to plot each year default = 'LnCE'
#' @param years which variable name identifies the yaer column, default='Year'
#' @param varlabel what label to use on x-axis, default = 'log(CPUE)'
#' @param vline an optional vertical line to aid interpretation. If it is
#'    numeric it will be added to each plot
#' @param plots how many plots to generate, default = c(3,3)
#'
#' @return a matrix of the year, mean value, stdev, and N number of
#'    observations. It also plots a histogram for each year and fits a
#'    normal distribution to each one.
#' @export
#'
histyear <- function(x,Lbound=-3.5,Rbound=12.25,inc=0.25,
                     pickvar="LnCE",years="Year",varlabel="log(CPUE)",
                     vline=NA,plots=c(3,3)) {
   yrs <- sort(unique(x[,years]))
   nyr <- length(yrs)
   columns <- c("Year","maxcount","Mean","StDev","N","Min","Max")
   results <- matrix(0,nrow=nyr,ncol=length(columns),dimnames=list(yrs,columns))
   par(mfcol=plots,mai=c(0.25,0.25,0.05,0.05),oma=c(1.2,1.0,0.0,0.0))
   par(cex=0.75, mgp=c(1.35,0.35,0), font.axis=7,font=7,font.lab=7)
   for (yr in 1:nyr) {
      pick <- which(x[,years] == yrs[yr])
      outh <- hist(x[pick,pickvar],breaks=seq(Lbound,Rbound,inc),col=2,main="",xlab="",ylab="")
      mtext(paste0("  ",yrs[yr]),side=3,outer=F,line=-2,font=7,cex=0.9,adj=0)
      mtext(paste0("  ",length(pick)),side=3,outer=F,line=-3,font=7,cex=0.9,adj=0)
      if (is.numeric(vline)) abline(v=vline,col=4,lwd=2)
      if (pickvar != "catch_kg") {
         pickmax <- which.max(outh$counts)
         ans <- addnorm(outh,x[pick,pickvar])
         lines(ans$x,ans$y,col=3,lwd=2)
         results[yr,] <- c(yrs[yr],outh$mids[pickmax],ans$stats,
                           range(x[pick,pickvar],na.rm=TRUE))
      }
   }
   mtext("Frequency",side=2,outer=T,line=0.0,font=7,cex=1.0)
   mtext(varlabel,side=1,outer=T,line=0.0,font=7,cex=1.0)
   return(results)
} # end of histyear

#' @title qqplotout plots up a single qqplot for a lm model
#'
#' @description qqplotout generates a single qqplot in isolation from the
#'    plot of a model's diagnostics. It is used with lefthist to
#'    illustrate how well a model matches a normal distribution
#'
#' @param inmodel the optimum model from standLM or dosingle
#' @param title a title for the plot, defaults to 'Normal Q-Q Plot'
#' @param cex the size of the font used, defaults to 0.9
#' @param ylow the lower limit of the residuals
#' @param yhigh he upper limit of the residuals
#' @param plotrug a logical value determinning whether a rug is included
#'
#' @return currently nothing, but it does generate a qqplot to the current
#'    device
#' @export
#'
#' @examples
#' y <- rep(1:100,2)
#' x <- rnorm(200,mean=10,sd=1)
#' model <- lm(y ~ x)
#' dev.new(width=6,height=3.5,noRStudioGD = TRUE)
#' par(mai=c(0.45,0.45,0.15,0.05),font.axis=7)
#' qqplotout(model,ylow=-50,yhigh=50)
qqplotout <- function(inmodel, title="Normal Q-Q Plot", cex=0.9,
                      ylow=-5,yhigh=5,plotrug=FALSE)  {
   resids <- inmodel$residuals
   labs <- cex
   qqnorm(resids, ylab=list("Standardized Residuals", cex=labs, font=7),
        xlab=list("Theoretical Quantiles", cex=labs, font=7),
```

```
          main=list(title,cex=labs,font=7),ylim=c(ylow,yhigh))
   qqline(resids, col=2,lwd=2)
   grid()
   if (plotrug) rug(resids)
   abline(v=c(-2.0,2.0),col="grey")
}  # end of qqplotout

#' @title qqdiag generates a qqplot with a histogram of residuals
#'
#' @description qqdiag generates a qqplot with a complementary histogram of
#'     the residuals to illustrate the proportion of all residuals along the
#'     qqline. If the qqline deviates from the expected straigt line, which
#'     is red i colour to make for simpler comparisons, then the histogram
#'     enables one to estiamte what proportion of records deviate from
#'     normality. The zero point is identified with a line, as are the
#'     approximate 5% and 95% percentiles. In both cases > 5% is above or
#'     below the blue lines, with < 90% in between depending on the
#'     proportions in each class. To get a more precise estimate use the
#'     invisibly returned histogram values.
#'
#' @param inmodel the optimum model being considered
#' @param plotrug a logical term determining whether a rug is plotted on the
#'     qqplot.
#' @param bins defaults to NA, but can be set to a given series
#' @param hline Include some horizontal lines on the histogram. defaults to 0.
#' @param xinc the increment for tick marks on the xaxis of the histogram
#' @param yinc the increment for tick marks on the y-axis of the histogram
#' @param ylab the y-axis label for the histogram, defaults to 'residuals'
#'
#' @return plots a graph and invisibly returns the output from the histogram
#' @export
#'
#' @examples
#'
#' y <- rep(1:100,2)
#' x <- rnorm(200,mean=10,sd=1)
#' model <- lm(y ~ x)
#' dev.new(width=6,height=3.5,noRStudioGD = TRUE)
#' par(mai=c(0.45,0.45,0.15,0.05),font.axis=7)
#' qqdiag(model,xinc=1,yinc=10,bins=seq(-55,50,2.5))
qqdiag <- function(inmodel,plotrug=FALSE,bins=NA,hline=0.0,
                   xinc=100,yinc=1,ylab="residuals") {
   layout(matrix(c(1,2),ncol=2),widths=c(5,2.5))
   par(mai=c(0.45,0.45,0.15,0.05),oma=c(0.0,0,0.0,0.0))
   par(cex=0.85, mgp=c(1.35,0.35,0), font.axis=7,font=7,font.lab=7)
   resids <- inmodel$residuals
   qs <- quantile(resids,probs=c(0.025,0.05,0.95,0.975))
   if (!is.numeric(bins)) {
      loy <- min(resids); hiy <- max(resids)
      scale <- trunc(100*(hiy - loy)/35) / 100
      loy <- round(loy - (scale/2),2); hiy <- round(hiy + scale,2)
      bins <- seq(loy,hiy,scale)
   } else {
      loy <- min(bins); hiy <- max(bins)
   }
   qqplotout(inmodel,plotrug=plotrug,ylow=loy,yhigh=hiy)
   abline(h=qs,lwd=c(1,2,2,1),col=4)
   outL <- lefthist(resids,bins=bins,hline=0.0,yinc=yinc,xinc=xinc,
                    ylabel=ylab,width=0.9,border=1)
   abline(h=qs,lwd=c(1,2,2,1),col=4)
   ans <- addnorm(outL,resids)
   lines(ans$y,ans$x,lwd=2,col=3)
   return(invisible(outL))
}  # end of qqdiag

#' @title addnorm - adds a normal distribution to a histogram of a data set.
#'
#' @description  addnorm - adds a normal distribution to a histogram of a data
#'    set. This is generally to be used to illustrate whether log-transformation
#'    normalizes a set of catch or cpue data.
#' @param inhist - is the output from a call to 'hist' (see examples)
#' @param xdata -  is the data that is being plotted in the histogram.
#' @param inc - defaults to a value of 0.01; is the fine grain increment used to
#'    define the normal curve. The histogram will be coarse grained relative to
#'    this.
#' @return a list with a vector of 'x' values and a vector of 'y' values (to be
#'    used to plot the fitted normal probability density function), and a vector
#'    used two called 'stats' containing the mean and sandard deviation of the
#'    input data
#' @export addnorm
#' @examples
#' x <- rnorm(1000,mean=5,sd=1)
```

```r
#' dev.new(height=6,width=4,noRStudioGD = TRUE)
#' par(mfrow= c(1,1),mai=c(0.5,0.5,0.3,0.05))
#' par(cex=0.85, mgp=c(1.5,0.35,0), font.axis=7)
#' outH <- hist(x,breaks=25,col=3,main="")
#' nline <- addnorm(outH,x)
#' lines(nline$x,nline$y,lwd=3,col=2)
#' print(nline$stats)
addnorm <- function(inhist,xdata,inc=0.01) {
   lower <- inhist$breaks[1]
   upper <- tail(inhist$breaks,1)
   cw <- inhist$breaks[2]-inhist$breaks[1]
   x <- seq(lower,upper, inc) #+ (cw/2)
   avCE <- mean(xdata,na.rm=TRUE)
   sdCE <- sd(xdata,na.rm=TRUE)
   N <- length(xdata)
   ans <- list(x=x,y=(N*cw)*dnorm(x,avCE,sdCE),stats=c(avCE,sdCE,N))
   return(ans)
} # end of addnorm
```

9 Implications

The limitations of commercial CPUE data should become better appreciated as the findings in this report become better known. The remarkable fact remains that it is possible to obtain consistent and informative trends from such, at times, questionable data.

10 Recommendations

The final objective of this work was to write a reference manual on the application of the most robust CPUE standardization strategies for Australian fisheries. It should be clear that the range of fisheries in Australia (from benthic hand collected fisheries, to trawl fisheries, to pelagic purse-seine and lining fisheries) means that there is no single standard approach to CPUE standardization that will necessarily work well with every fishery. Nevertheless, it remains possible to write out a set of guidelines that will improve the defensibility of any conclusions drawn from CPUE standardizations as well as improve the presentation of results from such analyses to assessment groups and other interested parties. Many of the points to be made are included in Chapter 8 (starting on page 145).

10.1Documentation

One pillar of defensibility is complete and explicit documentation of all procedures used in any stock assessment or analysis so that the analysis can be repeated quickly and easily. However, most people interested in the results of an analysis focus primarily on the summary or abstract of results and only desire a brief document. Nevertheless, in the interests of openness and defensibility many of the recommendations below for more text, tables, and plots, should be included at least as supplementary materials in appendices. With the growth of electronic documents and reduction in the use of printed documents the size of the final document should not be an impediment to an improvement to how such analyses are presented. If a printed document is required, then the supplementary material need not be printed but should be referred to throughout the primary document.

In the case of CPUE standardization it is necessary:

1. Have an explicit section in any report on a standardization that focusses solely on the data selection and preparation processes and choices.
2. Describe and explain every choice in any data selection made.
3. Ideally tabulate and plot the distributions of catch, effort, CPUE, depth of fishing, month of fishing, and any other factors/variables included in the analysis to illustrate the quality of the data being used (helps identify whether there are outliers or there is rounding, or whether the data has unexpected properties, or just what those properties are).
4. Be explicit about the statistical models fitted, and how the model parameters (especially the year, or time-step, effects) are derived.
5. Be explicit about the assumptions behind the statistical distributions used in the statistical models.
6. Plot diagnostics relating to the statistical fit of the model to the data.
7. Identify and plot the relative influence of the different factors included in any analysis. Do not rely solely on the variance or deviance accounted for by each factor but also summarize the impact each factor has on the standardized CPUE trend.

Ideal Sensitivity Options
If enough time is available (albeit this is an unlikely scenario).

1. Apply the same statistical model structures but with different underlying statistical distributions to describe the residual structure (e.g. log-normal vs Gamma distributions). This tests for sensitivity to the basic assumptions used.
2. Apply different statistical model structures using the same statistical distribution for the residual errors structure to consider the sensitivity to model structure.
3. Conduct a retrospective analysis through at least the last half of the available years of data to search for consistency and/or for major changes in influences.

10.2 Further Development

How best to handle the analysis of CPUE from mixed fisheries is an issue that still requires further work. The possible solution explored when considering the auto-line fishery for Blue-Eye and Pink Ling appears to hold considerable promise. Such an initial exploration could form the basis of a fuller investigation involving different fisheries under different conditions. The analyses attempted appeared to solve some of the problems, but more attention is required to examine how such analyses might generate misleading results.

Further comparisons between GLMs and GAMs where more factors are treated as surfaces described by GAMs. Almost any analysis where there are multiple factors available, the number of alternative arrangements possible are great. The influence of fitting smooth surfaces rather than sub-dividing a continuous variable into different levels or categories should be explored further. Intuitively there would be a trade-off between the potential improvement possible and the proportion of erroneous or noisy data. How to make a decision as to when a data set is too noisy would require an empirical study of a number of available fisheries. The sensitivity of any analysis to mistakes in the data is also in need of further exploration. Nevertheless, the use of GAMs instead of categorization to facilitate the inclusion of non-linearities, especially in relatively data poor fisheries is a direction that may be beneficial.

It would be beneficial if Rmarkdown templates (or similar templates in other document generating systems) could be generated for general distribution so that individuals could adapt and modify them to suit the needs of each fishery/species, or jurisdiction. Even with modification this would nevertheless provide an opportunity to have a minimum specification for such analyses and simplify the application of many of the recommendations from this report.

11 Extension and Adoption

Many of the findings in this report have already been presented to the Resource Assessment Groups within the SESSF, and at other fishery meetings and various reviews around Australia. Some of the findings were built into an auxiliary R package used in a recent series of data-poor stock assessment workshops.

Other aspects of the material in this report has already been used to influence management and fishery monitoring in the SESSF. Chapter 5, for example, that makes a comparison between the fishery standardization results and the equivalent analyses from the biennial Fishery Independent Survey has been used in discussions regarding the utility of the SESSF FIS.

The SESSF standardization routines have also directly benefited from the work in this report. The analyses in the annual round of standardization analyses are now automated, even though each analysis is customized to suit the idiosyncrasies of each fishery. This facilitates and accelerated the production of the report (writing the interpretative text cannot be automated).

12 Project Material Developed

The primary output from this project is this current report.

13 Appendix 1: Staff

Rik Buckworth: CSIRO Oceans and Atmosphere, Brisbane
Natalie Dowling: CSIRO Oceans and Atmosphere, Hobart
George Leigh: Queensland Department of Primary Industries, Brisbane
Malcolm Haddon: CSIRO Oceans and Atmosphere, Hobart
David C. Smith: CSIRO Oceans and Atmosphere, Hobart

14 Appendix 2: Species Names

Name	CAAB	Code	Scientific
Bight Redfish	37258004	REB	*Centroberyx gerrardi*
Blue Grenadier	37227001	GRE	*Macruronus novaezelandiae*
Blue Warehou	37445005	TRT	*Seriolella brama*
Blue-Eye	37445001	TBE	*Hyperoglyphe antarctica*
Deepwater Flathead	37296002	FLD	*Platycephalus conatus*
Eastern Gemfish	37439002	GEM	*Rexea solandri*
Elephant Fish	37043000/1	ELE	*Callorhinchus milii*
Flathead	37296001	FLT	*Neoplatycephalus richardsoni*
Gummy Shark	37017001	SHG	*Mustelus antarcticus*
Jackass Morwong	37377003	MOR	*Nemadactylus macropterus*
John Dory	37264004	DOJ	*Zeus faber*
Mirror Dory	37264003	DOM	*Zenopsis nebulosus*
Ocean Jackets	37465006	LTC	*Nelusetta ayraudi*
Ocean Perch	37287001	REG	*Helicolenus percoides*
Orange Roughy	37255009	ORE	*Hoplostethus atlanticus*
Pink Ling	37228002	LIG	*Genypterus blacodes*
Redfish	37258003	RED	*Centroberyx affinis*
Ribaldo	37224002	RBD	*Mora moro*
Royal Red Prawn	28714005	PRR	*Haliporoides sibogae*
Saw Sharks	37023001/2	SAW	*Pristiophorus cirratus & nudipinnis*
School Shark	37017008	SHS	*Galeorhinus galeus*
School Whiting	37330014	WHS	*Sillago flindersi*
Silver Trevally	37337062	TRE	*Pseudocaranx dentex*
Silver Warehou	37445006	TRS	*Seriolella punctata*
Western Gemfish	37439002	GEM	*Rexea solandri*
Arrow Squid			*Nototodarus gouldi*
Commercial Scallops			*Pecten fumatus*

15 References

Bataille, B.C. and T.J. Quinn II (2004) Catch per unit effort standardization of the eastern Bering Sea walleye Pollock (*Theragra chalcogramma*) fleet. *Fisheries Research* **70:** 161 – 177

Begg, G.A. and J.R. Waldman (1999) An holistic approach to fish stock identification. *Fisheries Research* **43**: 35 – 44

Bentley, N., Kendrick, T.H., Starr, P.J., and P.A. Breen (2012) Influence plots and metrics: tools for better understanding fisheries catch-per-unit-effort standardizations. *ICES Journal of Marine Science* **69(1)**: 84-88

Bishop, J. (2006) Standardizing fishery-dependent catch and effort data in complex fisheries with technology change. *Reviews in Fisheries Biology and Fisheries* **16**: 21-38

Bishop J., and Die, D.J., and Wang Y-G. (2004) Analysing commercial catch and effort data from a Penaeid trawl fishery: A comparison of linear models, mixed models, and generalised estimating equations approaches. *Fisheries Research* **70**: 179–193

Burnham, K.P. and D.R. Anderson (2002) *Model Selection and Inference*. A Practical Information-Theoretic Approach. Second Edition Springer-Verlag, New York. 488 p.

Campbell, R.A (2015) Constructing stock abundance indices from catch and effort data: Some nuts and bolts *Fisheries Research* **161**: 109 – 130

Chen, C. (2000) Generalized additive mixed models. *Communications in Statistics - Theory and Methods* **29(5-6)**:1257-1271, DOI: 10.1080/03610920008832543

DAFF. (2007) *Commonwealth Fisheries Harvest Strategy Policy and Guidelines.* Australian Government Department of Agriculture, Fisheries and Forestry, Canberra, Australia. 55 pp.

Day, J. (2016) Tiger flathead (*Neoplatycephalus richardsoni*) stock assessment based on data up to 2015. CSIRO Oceans and Atmosphere, Presented at the SE RAG Nov 2016. 77p.

Dichmont, C. M., Deng, A. R., Punt, A.E., Venables, W. and M. Haddon. (2006) Management strategies for short-lived species: The case of Australia's Northern Prawn fishery 1. Accounting for multiple species, spatial structure and implementation uncertainty when evaluating risk. *Fisheries Research* **82**: 204-220.

Fay, G. (2007) Stock Assessment of Jackass Morwong (*Nemadactylus macropterus*) and RBC Calculations for 2007 Using Data up to 2005 pp 265 – 285 *in* Tuck, G.N. (ed.) 2007. *Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2006-2007. Volume 1: 2006.* Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 570 p.

Francis, R.I.C.C. (2011) Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences* **68**: 1124–1138.

Fulton, E.A., Smith, A.D.M., and A.E. Punt (2005) Which ecological indicators can robustly detect effects of fishing? *ICES Journal of Marine Science*, **62**: 540-551.

Fulton, E.A. (2010) Approaches to end-to-end ecosystem models *Journal of Marine Systems* **81**: 171-183

Garstang, W. (1900) The impoverishment of the sea - a critical summary of the experimental and statistical evidence bearing upon the alleged depletion of the trawling grounds. *Journal of Marine Biological Association of the United Kingdom*, **6**: 1-69.

Gavaris, S. (1980) Use of a multiplicative model to estimate catch rate and effort from commercial data. *Canadian Journal of Fisheries and Aquatic Sciences* **37**, 2272-2275.

Gorfine, H., Day, R., Bardos, D., Taylor, B., Prince, J., Sainsbury, K. and C. Dichmont (2008) *Rapid response to abalone virus depletion in western Victoria: information acquisition and reef code assessment*, Final report to Fisheries Research and Development Corporation Project No. 2007/066. The University of Melbourne.

Haddon, M. (1998) Standardization of the Catch/Effort data from the South-East Blue Grenadier fishery. *Blue Grenadier Fishery Assessment Group Paper.* Australian Maritime College, 9p

Haddon, M. (1998b) Standardization of the Catch/Effort data from the South-East Blue Warehou fishery. *Blue Warehou Fishery Assessment Group Paper.* Australian Maritime College, 10p

Haddon, M. (1999) Standardization of Catch/Effort data from the South-East Pink Ling fishery. *SEFAG Working Group Paper: June 1999*. Australian Maritime College. 16p.

Haddon, M. (2007) Standardized Commercial Catch-Effort Data for Selected Shelf and Slope Assessment Group Species for 1986 – 2005. Pp 31 – 84. *In* Tuck, G.N. (ed.) 2007. *Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2006 – 2007. Vol 1 2006.* Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 570 p.

Haddon, M. (2010) Catch rate standardization for data 1986 – 2008. Pp 21 – 191. *In* Tuck, G.N. (ed.) 2010. *Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2009. Part 2.* Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 428 p.

Haddon, M. (2014) Catch rate standardizations for selected species from the SESSF (data 1986 – 2012). Pp 57 – 275 *in* Tuck, G.N. (*ed*) (2014) *Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2013. Part 2.* Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 486p.

Haddon, M. (2015) Bight redfish (*Centroberyx gerrardi*) stock assessment using data to 2014/2015 CSIRO Oceans and Atmosphere, Presented at the GAB RAG Nov 2015. 40p.

Haddon, M. (2016) Deepwater Flathead (*Neoplatycephalus conatus*) stock assessment using data to 2015/2016. CSIRO Oceans and Atmosphere, Presented at the GAB RAG Dec 2016. 43p.

Haddon, M. (2016b) Blue-Eye Auto-Line and Drop-Line CPUE Characterization and Catch-per-Hook 1997 – 2015. CSIRO Oceans and Atmosphere, Presented at the SE RAG Nov 2016. 51p.

Haddon, M. (2016c) Blue-Eye (Hyperoglyphe antarctica) Tier 4 Analysis using Catch-per-Hook for Au-to-Line and Drop-Line from 1997 - 2015. CSIRO Oceans and Atmosphere, Presented at the SE RAG Nov 2016. 51p.

Haddon, M. (2017) DRAFT Orange Roughy East (*Hoplostethus atlanticus*) stock assessment using data to 2016. CSIRO Oceans and Atmosphere. 47 p.

Haddon, M. and K. Hodgson (2000) Standardization of Catch/Effort data from the South-East Spotted Warehou Trawl fishery – 1986 - 1999. *Spotted Warehou Assessment Group Paper* University of Tasmania. 14p.

Hardin, J.W. and J.M. Hilbe (2003) *Generalized Estimating Equations.* Chapman & Hall/CRC Press, Boca Raton, FL.

Hastie, T.J., and R.S. Tibshirani (1990) *Generalized Additive Models.* Chapman and Hall, London

Kimura, D.K. (1981) Standardized measures of relative abundance based on modelling log(c.p.u.e.), and their application to pacific ocean perch (*Sebastes alutus*). *Journal du Conseil International pour l'Exploration de la Mer*. **39**: 211-218.

Klaer, N.L. (1994) Methods for standardisation of catch/effort and data requirements. 86-90 *In*: Hancock, D.A (*ed*) *Population Dynamics for Fisheries Management*. Australian Society for Fish Biology Workshop Proceedings, Perth 24-25 August 1993. Australian Society for Fish Biology, Perth.

Klaer, N.L., and Tilzey, R.D.J. (1994) The multi-species structure of the fishery. In The South East Fishery. (ed R.D.J. Tilzey) pp. 72-94, Bureau of Resource Sciences, Commonwealth of Australia.

Knuckey, I., Koopman, M, and R. Hudson (2015) Resource Survey of the Great Australian Bight Trawl Sector 2015. Report to AFMA.

Knuckey, I., Koopman, M. and S. Boag (2017). Continuation of a fishery independent survey for the Southern and Eastern Scalefish and Shark Fishery — 2016. AFMA Project 2016/0802. Fishwell Consulting 54 pp.

Linnane, A., McGarvey, R., Feenstra, J. and P. Hawthorne (2015) Southern Zone Rock Lobster (*Jasus edwardsii*) Fishery 2013/14. SARDI Publication No. F2007/000276-9 Fishery Assessment Report to PIRSA Fisheries and Aquaculture. SARDI Aquatic Sciences. Adelaide, South Australia. 99p.

Little, L.R., Wayte, S.E., Tuck, G.N., Smith, A.D.M., Klaer, N., Haddon, M., Punt, A.E., Thomson, R., Day, J. and M. Fuller (2011) Development and evaluation of a cpue-based harvest control rule for the southern and eastern scalefish and shark fishery of Australia. ICES Journal of Marine Science. 68(8): 1699-1705.

MacCall, A. D. (2009) Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations. *ICES Journal of Marine Science* 66: 2267–2271.

Maunder, M.M. and A.E. Punt (2004) Standardizing catch and effort data: a review of recent approaches. *Fisheries Research* **70**: 141–149.

Maunder, M.N., and A.E. Punt (2013) A review of integrated analysis in fisheries stock assessment. *Fisheries Research* **142:** 61–74.

Mayfield, S., Ferguson, G., Carroll. J. and J. Dent (2014) Central Zone Abalone (*Haliotis laevigata* and *H. rubra*) Fishery. SARDI Publication No. F2007/000611-5 Fishery Assessment Report to PIRSA Fisheries and Aquaculture. SARDI Aquatic Sciences. Adelaide, South Australia. 77p.

McCullagh, P. and J.A. Nelder (1989) *Generalized Linear Models* 2nd Ed. Chapman & Hall/CRC, London 511 p.

McCulloch, C.E. and S.R. Searle (2001) *Generalized, Linear and Mixed Models.* Wiley, New York.

Methot and Wetzel (2013) Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management *Fisheries Research* **142**: 86-99

Murawski, S.A., Lange, A.M. and J.S. Idoine (1991) An analysis of technological interactions among Gulf or Maine mixed-species fisheries *ICES Marine Science Symposium* **193**: 237 – 252

Neter, J., Kutner, M.H., Nachtsheim, C.J, and W. Wasserman (1996) *Applied Linear Statistical Models*. Richard D. Irwin, Chicago. 1408 p.

O'Driscoll, R.; Doonan, I. (2015a) Review of the Southern and Eastern Scalefish and Shark Fishery (SESSF) Fishery Independent Surveys. Part 1: Commonwealth Trawl Sector. NIWA Client Report WLG2015-63.

O'Driscoll, R.; Doonan, I. (2015b) Review of the Southern and Eastern Scalefish and Shark Fishery (SESSF) Fishery Independent Surveys. Part 2: Great Australian Bight. NIWA Client Report WLG2015-69.

Pecl, G.T., Araújo, M.B., Bell, J.D., Blanchard, J., Bonebrake, T.C., Chen, I-C., Clark, T.D., Colwell, R.K., Danielsen, F., Evengård, B., Falconi, L., Ferrier, S., Frusher, S., Garcia, R.A., Griffis, R.B., Hobday, A.J., Janion-Scheepers, C., Jarzyna, M.A. Jennings, S., Lenoir, J., Linnetved, H.I., Martin, V.Y., McCormack, P.C., McDonald, J., Mitchell, N.J., Mustonen, T., Pandolfi, J.M., Pettorelli, N., Popova, E., Robinson, S.A., Scheffers, B.R., Shaw, J.D., Sorte, C.J.B., Strugnell, J.M., Sunday, J.M., Tuanmu, M-N., Vergés, A., Villanueva, C., Wernberg, T., Wapstra, E. and S.E. Williams (2017) Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* **355**: eaai9214

Peel, D., Bravington, M.V., Kelly, N., Wood, S.N. and I. Knuckey (2012) A Model-Based Approach to Designing a Fishery-Independent Survey. *Journal of Agricultural, Biological and Environmental Statistics* **18 (1)**: 1-21.

Punt, A.E. (1998) Preliminary VPA analysis for the Blue Warehou resource off Southern Australia *Blue Warehou Fishery Assessment Group Paper* 8p.

Punt, A.E. (2017) Some insights into data weighting in integrated stock assessments. *Fisheries Research* **192**: 52-65 http://dx.doi.org/10.1016/j.fishres.2015.12.006.

Punt, A.E., Day, J., Fay, G., Haddon, M., Klaer, N., Little, L.R., Privitera-Johnson, K., Smith, A.D.M., Smith, D.C., Sporcic, M., Thomson, R., Tuck, G.N., Upston, J., and S. Wayte (2018) Retrospective investigation of assessment uncertainty for fish stocks off southeast Australia. Fisheries Research 198: 117-128.

Punt, A.E., Haddon, M., Little, L.R. and G.N. Tuck (In Press) The effect of marine closures on a feedback control management 1 strategy used in a spatially-aggregated stock assessment: A case study based on pink ling in Australia. *Canadian Journal of Fisheries and Aquatic Sciences* **73**: dx.doi.org/10.1139/cjfas-2016-0017

Punt, A.E., Smith, A.D.M. and G. Cui (2002) Evaluation of management tools for Australia's South East Fishery 1. Modelling the South East Fishery taking account of technical interactions. *Marine and Freshwater Research* **53**: 615 – 629

Punt, A.E., Smith, D.C., Thomson, R.B., Haddon, M., He, X., and Lyle, J. (2001) Stock assessment of the blue grenadier *Macruronus novaezelandiae* resource off south-eastern Australia. *Marine and Freshwater Research* **51**, 701-717.

Quinn, T.J., and Deriso, R.B. (1999) *Quantitative Fish Dynamics*. Oxford, Oxford University Press

Rayns, N. (2007) The Australian government's harvest strategy policy. *ICES Journal of Marine Science* **64:** 596 – 598

Richards, L.J. (1991) Use of contradictory data sources in stock assessments. *Fisheries Research* **11:** 225–238.

Richards, L.J., and Schnute, J.T. (1992) Statistical models for estimating CPUE from catch and effort data. Canadian Journal of Fisheries and Aquatic Sciences 49, 1315-1327.

Russell, E.S. (1931) Some theoretical considerations on the "overfishing" problem. *Journal du Conseil International pour l'Exploration de la Mer*, **6**: 3-20.

Smith, A.D.M., Smith, D.C., Tuck, G.N., Klaer, N., Punt, A.E., Knuckey, I., Prince, J., Morison, A., Kloser, R., Haddon, M., Wayte, S., Day, J., Fay, G., Pribac, F., Fuller, M., Taylor, B., Little., R (2008) Experience in implementing harvest strategies in Australia's south-eastern fisheries. *Fisheries Research* 94: 373-379.

Smith, A.D.M., Smith, D.C., Haddon, M., Knuckey, I., Sainsbury, K.J., Sloan, S. (2014) Implementing harvest strategies in Australia: five years on. *ICES Journal of Marine Science* 71: 195-203.

Sporcic, M. and M. Haddon (2016) Statistical CPUE standardizations for selected SESSF species (data to 2015). Report prepared for the October 2016 SERAG meeting. CSIRO Oceans and Atmosphere, Hobart, 206 p.

Tilzey, R.D.J., and Klaer, N.L. (1994) Catch and Effort summary. In The South East Fishery. (ed R.D.J. Tilzey) pp. 46-71, Bureau of Resource Sciences, Commonwealth of Australia.

Tilzey, R.D.J. (*ed*) (1999) *The South East Fishery 1998*, Fishery Assessment Report compiled by the South East Fishery Assessment Group. Australian Fisheries Management Authority, Canberra. 199p.

Tuck, G.N. (*ed*) (2014) *Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2013. Part 1.* Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart. 313p.

Tuck, G.N. (*ed*) (2015) *Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2014. Part 1.* Australian Fisheries Management Authority and CSIRO Oceans and Atmosphere, Hobart. 170p.

Venables, W.N. and C.M. Dichmont (2004) GLMs, GAMs, GLMMs: an overview of theory for applications in fisheries research. *Fisheries Research* **70**: 315–333.

Venables, W.N. and B.D. Ripley (2002) *Modern Applied Statistics with S*. 4th Ed. Springer-Verlag, New York. 495p. ISBN 0-387-95457-0

Vieira, S., Perks, C., Mazur, K., Curtotti, R., and M. Li (2010) Impact of the structural adjustment package on the profitability of Commonwealth fisheries. *ABARE research report* 10.01**.** Canberra.

Vignaux, M. (1996) Analysis of spatial structure in fish distribution using commercial catch and effort data from the New Zealand hoki fishery. *Canadian Journal of Fisheries and Aquatic Sciences* **53**: 963-973.

Wayte, S.E. (ed.) 2009. Evaluation of new harvest strategies for SESSF species. CSIRO Marine and Atmospheric Research, Hobart and Australian Fisheries Management Authority, Canberra. 137 p.

Wayte, S.E. (2013) Management implications of including a climate-induced recruitment shift in the stock assessment for jackass morwong (*Nemadactylus macropterus*) in south-eastern Australia. *Fisheries Research* **142**: 47-55.

Xie, Y., Allaire, J.J., and G. Grolemund (2018) *R Markdown The Definitive Guide*. The R Series. CRC Press Chapman & Hall.

Ye, Y. and D. Dennis (2009) How reliable are the abundance indices derived from commercial catch–effort standardization? *Canadian Journal of Fisheries and Aquatic Sciences* **66**: 1169-1178

CONTACT US

t 1300 363 400
+61 3 9545 2176
e enquiries@csiro.au
w www.csiro.au

YOUR CSIRO

Australia is founding its future on science and innovation. Its national science agency, CSIRO, is a powerhouse of ideas, technologies and skills for building prosperity, growth, health and sustainability. It serves governments, industries, business and communities across the nation.

FOR FURTHER INFORMATION

Division/Unit Name
Jonathan Bates
t +61 3 9123 4567
e jonathan.bates@csiro.au
w www.csiro.au/businessunit-flagshipname