Standardised CPUE abundance indices for orange roughy off Namibia based on lognormal and delta-lognormal linear models

Anabela Brandão and Doug S. Butterworth

Marine Resource Assessment & Management Group (MARAM)
Department of Mathematics and Applied Mathematics
University of Cape Town
Rondebosch, 7701, Cape Town

February 2002

Abstract

GLM analyses are used to standardise the CPUE data for Namibian orange roughy. The possibility of there being a "learning" period of lower CPUE for a new vessel when it enters the fishery is taken into account. Alternative statistical approaches to deal with tows that record zero catch of orange roughy are considered. Further, to allow for areal expansion of the fishery at each aggregation, sub-aggregations are defined and CPUE trends estimated separately for each. Different methods for combining the results for the various sub-aggregations to provide a single index for an aggregation are considered. All aggregations show a downward trend in CPUE; however, the baseline CPUE standardisation method (comparable to that used in 2001) suggests an increase in CPUE in 2001 compared to 2000.

Introduction

In Brandão and Butterworth (2001) commercial CPUE data for orange roughy off Namibia were standardised by applying a Generalised Linear Model (GLM). These standardised CPUE indices of abundance are then used as an input to a population model to assess the state of the stock (Brandão and Butterworth 2001). In this paper the GLM of Brandão and Butterworth (2001) is updated using the re-entered data for commercial fishing with an extra year's data (for 2001). Two problems encountered in such analyses for this fishery are: i) a considerable number of tows with zero catches and ii) the areal distribution of effort shifting within and even beyond previously defined aggregations (especially notable for the *Johnies* aggregation). These problems are addressed in this paper as well.

The population model to assess the state of the orange roughy stock assumes that the fishery can be approximated by a single catch at the start of the year without compromising accuracy. However the peak time of the year when the orange roughy aggregate for spawning is in July, and therefore the time when most fishing effort is concentrated and is about the same as the time to which other abundance indices apply (acoustic survey and research swept area indices). Thus in this paper a "fishing year" is defined to be the period July to June when analysing the CPUE data.

The Models

Model to standardise the CPUE

Two models have been applied to the CPUE time series of data for Namibian orange roughy. The base case model is the GLM "full model" of Brandão and Butterworth (2001), which allows for the possibility that vessels might have different degrees of "effectiveness" in their first year in the orange roughy fishery compared to subsequent years. In this paper this model will be referred to as the "lognormal model". The second model, described by Lo *et al.* (1992) and Stone and Porter (1999), uses the delta-lognormal method to obtain standardised CPUE indices in the presence of tows with zero catch. This model will be referred to as the "delta-lognormal model".

The lognormal model

The lognormal model allows for possible differences in abundance trends in orange roughy in the various aggregations, and assumes the possibility that vessels might operate differently in their first year in the fishery, but have the same degree of “effectiveness” in all subsequent years. When this model was fitted to the re-entered data and with an extra year’s information, only the vessel *Whitby* showed a significant difference in its first two years of operation. Therefore only this vessel was differentiated from its first two years in the fishery and all subsequent years. This model is given by:

$$\ln(CPUE + \delta) = \mu + \alpha_{vessel} + \beta_{year} + \gamma_{month} + \lambda_{agg} + \eta_{year \times agg} + \varepsilon \qquad (1)$$

where:

μ is the intercept,

vessel is a factor with 9 levels associated with each of the vessels that have operated in the fishery:

Conbaroya Cuarto

Dantago

Emanguluko

Harvest Nicola

Hurinis

Southern Aquarius

Whitby (first year)

Whitby (second year)

Whitby (subsequent years),

year is a factor with 8 levels associated with the “fishing years” 1994–2001 (note: “1996”, for example, refers to the period July 1996 to June 1997),

month is a factor with 12 levels (January– December),

agg is a factor with 12 levels associated with the four aggregations and their sub-aggregations:

Johnies: *Johnies1*

Johnies2

Johnies3

Johnies4

Frankies: *21 Jump Street*

Frankies Flats

Frankies Outer

Three Sisters

Smifton

Rix: *Rix Inner*

Rix Outer

Hotspot,

$year \times agg$ is the interaction between year and aggregation (this allows for the possibility of different trends for the different sub-aggregations),

δ is a small constant added to the orange roughy CPUE to allow for the occurrence of zero CPUE values, and

ε is an error term assumed to be normally distributed.

Standardised CPUE time series for a given (sub)-aggregation are obtained by calculating:

$$CPUE_y = \exp\left[\mu + \beta_{year} + \lambda_{agg} + \eta_{year \times agg}\right] - \delta \qquad (2)$$

where in this application we are standardising on the vessel *Southern Aquarius* and on the month of *December*.

The delta-lognormal model

The delta distribution is often used in instances when there are a considerable number of zero observations, for which zero and non-zero data are consequently treated separately.

Final estimates of abundance are obtained from the product of the proportion and the mean of non-zero observations. For the delta-lognormal model, two lognormal linear models (GLMs) are fitted to the commercial CPUE data, one to estimate the proportion of tows for which there is a positive catch, and the other to estimate the standardised CPUE for orange roughy for tows that have a positive catch.

Relative abundance indices of orange roughy are then given by:

$$CPUE_y = CPUE_y^{+ve} CPUE_y^{zero} \tag{3}$$

where:

$CPUE_y^{+ve}$ is the standardised CPUE index for tows which have positive catches, and

$CPUE_y^{zero}$ is the standardised measure of the proportion of tows that have positive catches.

Standardised indices for the two components above were obtained by fitting lognormal linear models to each. The same lognormal model given in equation (1) was used for each individual component, i.e. the model to estimate the abundance of positive catches is given by:

$$\ln(CPUE^{+ve}) = \mu + \alpha_{vessel} + \beta_{year} + \gamma_{month} + \lambda_{agg} + \eta_{year \times agg} + \varepsilon \tag{4}$$

where the notation is as in equation (1), and the model to estimate the proportion of tows with a positive catch is given by:

$$\ln(CPUE^{zero} + 1) = \mu + \alpha_{vessel} + \beta_{year} + \gamma_{month} + \lambda_{agg} + \eta_{year \times agg} + \varepsilon \tag{5}$$

where a constant of 1 was added to allow for the logarithmic transformation when $CPUE^{zero}$ is zero. Investigation of residuals obtained from this model did not meet the assumption of normally distributed errors in equation (5). In the case of orange roughy tow data the proportion of tows with a positive catch is either “0” or “1” for an individual tow, and therefore an alternative model for the proportion positive assuming binomially distributed errors is considered as well, given by:

$$\ln\left(\frac{CPUE^{zero}}{1-CPUE^{zero}}\right) = \mu + \alpha_{vessel} + \beta_{year} + \gamma_{month} + \lambda_{agg} + \eta_{year \times agg} + \varsigma \qquad (6)$$

where

ζ is an error term assumed to be binomially distributed.

Standardised measures of the abundance of orange roughy in positive tows is estimated by calculating:

$$C\hat{P}UE_y^{+ve} = \exp\left[\hat{\mu} + \hat{\beta}_{year} + \hat{\lambda}_{agg} + \hat{\eta}_{year \times agg}\right]\psi_y^{+ve} \qquad (7)$$

where

ψ_v^{+ve} is a correction factor for bias (Lo *et al.* 1992), given by

$$\psi_y^{+ve} = g_m\left[\frac{m+1}{2m}\left(\hat{\xi}^2 - \hat{\xi}_{\hat{\theta}}^2\right)\right] \qquad (8)$$

where

$\hat{\xi}^2$ is the residual variance,

m is the degrees of freedom for the estimate of residual variance,

$\hat{\theta}$ is given by $\hat{\mu} + \hat{\beta}_{year} + \hat{\lambda}_{agg} + \hat{\eta}_{year \times agg}$,

$\hat{\xi}_{\hat{\theta}}^2$ is the variance of $\hat{\theta}$, and

$$g_m(t) = \sum_{p=0}^{\infty}\left[\frac{m^p(m+2p)}{m(m+2)\cdots(m+2p)}\left(\frac{m}{m+1}\right)^p \frac{t^p}{p!}\right]$$

where t is the argument of the function.

Similarly standardised measures of the proportion of positive catches of orange roughy is given by:

$$C\hat{P}UE_y^{zero} = \exp\left[\hat{\mu} + \hat{\beta}_{year} + \hat{\lambda}_{agg} + \hat{\eta}_{year \times agg}\right]\psi_y^{zero} - 1 \qquad (9)$$

where the correction for bias term ψ_y^{zero} is of the form of equation (8), except that the term θ is given by the estimates obtained from fitting either model (5) or model (6).

Model Implementation

In Brandão and Butterworth (2001) the definition of strata where those given in Brandão (1999), where the *Johnies* stratum was enlarged to accommodate the increased number of commercial tows that were deeper than 1000 m and further to the south-west. The most recent tow data again show this trend, with a considerable number of tows lying outside the *Johnies* stratum defined by Brandão (1999) and with relatively few tows within the original *Johnies* stratum defined by Branch and Roberts (1998). To take into account such "movement" of orange roughy within a known aggregation, the analyses in this paper take into consideration not only tows that lie within the inner strata of an aggregation, but also tows that take place in the outer strata of the aggregation. The levels of the factor for aggregations in the GLMs are then given as the various sub-aggregations. Figures 1 to 4 show the four aggregations *Johnies*, *Frankies*, *Rix*, and *Hotspot* and their sub-aggregations.

Commercial tow information inside the known aggregations of orange roughy in Namibia for the fishing years (July – June) 1994 to 2001, as provided by E. Johnsen has been used. The year 2001 is incomplete as this fishing year ends only in June 2002. Data until the end of October were available. A total of 14 766 data points was available for the analyses. Bottom distances were calculated from the GPS positions for each tow. For tows that did not have haul positions, but did have bottom time information, bottom distances were calculated by the following regression relationship:

Bottom distance [km] = *bottom time* [h] * 0.0883+0.4105.

GLM Results and Discussion

The lognormal linear model of equation (1) was fitted to the commercial CPUE data. In this instance, a value of δ taken to be 10% of the average of the orange roughy CPUE data (=81.19) was used. Examination of the results, especially the interaction terms between vessel and year, revealed that generally the only large effects observed occur in the first two years in which the vessel *Whitby* took part in the fishery. Given these results it was decided to include extra levels of the *Whitby* vessel factor to account for these first two years showing a different pattern from other years, and thus omitting a year-vessel interaction from the GLM.

When the CPUE series were standardised (equation (2)), the value of δ was originally taken to be 10% of the average orange roughy CPUE (δ = 81.19). This value was found to result in negative values of the standardised CPUE series in some instances. A value of δ = 0.1, which avoids this problem, was therefore adopted and all results shown in this paper are for this value.

The lognormal model (equation (1)) accounts for 32% of the total variation of orange roughy CPUE. Table 1 shows the parameter estimates obtained for the factor vessel. The lognormal model applied to tows with a positive catch (equation (4)) accounts for 41% of the total variation of orange roughy positive CPUE. A total of 12 201 tows have a positive catch. Tables 2 to 5 show the index of abundance provided by the lognormal model, the delta-lognormal model assuming binomial errors for the proportion positive, and the delta-lognormal model assuming lognormal errors for the proportion positive for each aggregation for each aggregation. Observations are not available for all years in all of the sub-aggregations. When the standardised CPUE indices from each individual sub-aggregation were combined to obtain a standardised CPUE index for each aggregation, three methods were used to deal with empty cells. The first method, referred to as the “zero” method, assumes that empty cells mean that there was no orange roughy in those areas for those years. The second method (“same”) assumes that although no observations were made, there was orange roughy. It is assumed that the same amount was present as at the first time an observation is made, or the same as last seen for subsequent years. The third method referred to as the “proportional” method makes the same assumption as the previous method except that now the amount is taken to be in the same proportion as that observed in another sub-aggregation for that year.

Figures 5 to 11 show the index of abundance provided by the lognormal model, the delta-lognormal model assuming binomial errors for the proportion positive, and the delta-lognormal model assuming lognormal errors for the proportion positive for each aggregation for each aggregation. For each aggregation a comparison is shown for the indices of abundance of orange roughy obtained by fitting the lognormal model the CPUE data for the three methods of combining the individual indices of the sub-aggregations. A comparison is also shown for the three models fitted to the CPUE data using the “zero” method of combining individual indices from sub-aggregations. For the aggregation *Johnies* there is not much difference between the three methods of combining individual indices (Table 2 and Fig. 5). In all aggregations, the indices obtained from fitting a lognormal model most resemble those obtained from fitting a delta-lognormal model assuming lognormal errors for the proportion positive. Differences between models and between methods of combining CPUE indices are most marked in the first few years of the series (mostly for pre 1997).

For the “zero” method with a lognormal model, all aggregations show an increase in CPUE in 2001 from that in 2000 (87% for *Johnies*, 98% for *Frankies*, 75% for *Rix*, and 132% for *Hotspot*).

Concluding Remarks

The results of the GLM provide a vessel factor that can be applied to adjust the research swept area estimate of 2000; in previous years the vessel *Southern Aquarius* was used to conduct the research swept area surveys, but in 2000 the new vessel *Emanguluko* was used.

Acknowledgements

A number of people have willingly provided data for this study. Assistance from Espen Johnsen, in particular, of NatMIRC is gratefully acknowledged, as is funding from the Namibian Deepwater Fishing Industry.

References

Branch, T.A. and Roberts, R.D. 1998. Swept-area estimates for Namibian orange roughy. Namibian Ministry of Fisheries and Marine Resources Document: WG/01/98/DWFWG/ORH:5.

Brandão, A. 1999. Annual swept-area estimates for Namibian orange roughy for the period 1995 to October 1998. Namibian Ministry of Fisheries and Marine Resources Document: DWFWG/WkShop/Jan99/doc19.

Brandão, A. and Butterworth, D.S. 2001. The application of GLM analyses to the CPUE data for Namibian orange roughy. Namibian Ministry of Fisheries and Marine Resources Document: DWFWG/WkShop/Mar01/doc1.

Lo, N.C., Jacobson, L.D., and Squire, J.L. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Can. J. Fish. Aqaut. Sci. 49: 2515–2526.

Stone, H.H. and Porter, J.M. (1999) Standardized CPUE indices for Canadian bluefin tuna fisheries based on commercial catch rates. International Commission for the Conservation of Atlantic Tunas, collective Volume of Scientific Papers: XLIX (2): 206–221.

Table 1. Parameter estimates for the vessel factor when the lognormal model (equation (1)) is fitted. The value of δ = 0.1 is chosen.

Vessel	**Vessel factor =** $e^{\alpha_{vessel}}$
Conbaroya Cuarto	0.590
Dantago	0.431
Emanguluko	0.763
Harvest Nicola	0.414
Hurinis	0.421
Southern Aquarius	**1.000**
Whitby (first year)	1.000
Whitby (second year)	1.000
Whitby (subsequent years)	1.698

Table 2. Standardised CPUE series (each normalised to their mean over the years considered) for the *Johnies* aggregation obtained by fitting the "lognormal model", the delta-lognormal model assuming binomial errors for the proportion positive, and the delta-lognormal model assuming lognormal errors for the proportion positive to the observed CPUE data for Namibian orange roughy. Three methods ("zero", "same" and "proportional" of dealing with years in which no observations were made in the sub-aggregations are considered.

Year	"Zero" method			"Same" method			"Proportional" method		
	Lognormal model	Delta-lognormal model (binomial errors)	Delta-lognormal model (lognormal errors)	Lognormal model	Delta-lognormal model (binomial errors)	Delta-lognormal model (lognormal errors)	Lognormal model	Delta-lognormal model (binomial errors)	Delta-lognormal model (lognormal errors)
1994	5.663	0.030	7.985	5.096	1.159	7.984	6.614	0.049	7.987
1995	0.662	3.528	0.010	1.041	3.159	0.011	0.773	5.792	0.0099
1996	0.343	0.941	0.0022	0.782	1.680	0.0031	0.400	1.545	0.0022
1997	0.720	2.237	0.0014	0.584	1.279	0.0003	0.115	0.392	0.0005
1998	0.141	0.376	0.003	0.114	0.215	0.0002	0.022	0.066	0.0001
1999	0.133	0.303	0.0002	0.108	0.173	0.0002	0.021	0.053	0.0001
2000	0.118	0.357	0.0002	0.095	0.204	0.0002	0.019	0.063	0.0001
2001	0.221	0.228	0.0002	0.179	0.131	0.0002	0.035	0.040	0.0001

Table 3. Standardised CPUE series (each normalised to their mean over the years considered) for the *Frankies* aggregation obtained by fitting the “lognormal model”, the delta-lognormal model assuming binomial errors for the proportion positive, and the delta-lognormal model assuming lognormal errors for the proportion positive to the observed CPUE data for Namibian orange roughy. Three methods (“zero”, “same” and “proportional” of dealing with years in which no observations were made in the sub-aggregations are considered.

Year	“Zero” method			“Same” method			“Proportional” method		
	Lognormal model	Delta-lognormal model (binomial errors)	Delta-lognormal model (lognormal errors)	Lognormal model	Delta-lognormal model (binomial errors)	Delta-lognormal model (lognormal errors)	Lognormal model	Delta-lognormal model (binomial errors)	Delta-lognormal model (lognormal errors)
1995	1.567	3.455	1.029	3.474	4.722	3.207	5.277	6.505	5.144
1996	3.353	2.153	3.642	2.556	1.887	2.650	1.224	0.417	1.324
1997	0.654	0.285	0.696	0.499	0.250	0.507	0.239	0.055	0.253
1998	0.300	0.085	0.406	0.229	0.074	0.295	0.109	0.016	0.147
1999	0.042	0.019	0.196	0.054	0.026	0.161	0.018	0.004	0.075
2000	—	—	—	0.081	0.022	0.114	0.001	0.001	0.041
2001	0.083	0.002	0.031	0.108	0.018	0.067	0.131	0.001	0.015

Table 4. Standardised CPUE series (each normalised to their mean over the years considered) for the *Rix* aggregation obtained by fitting the "lognormal model", the delta-lognormal model assuming binomial errors for the proportion positive, and the delta-lognormal model assuming lognormal errors for the proportion positive to the observed CPUE data for Namibian orange roughy. Three methods ("zero", "same" and "proportional" of dealing with years in which no observations were made in the sub-aggregations are considered.

Year	"Zero" method			"Same" method			"Proportional" method		
	Lognormal model	Delta-lognormal model (binomial errors)	Delta-lognormal model (lognormal errors)	Lognormal model	Delta-lognormal model (binomial errors)	Delta-lognormal model (lognormal errors)	Lognormal model	Delta-lognormal model (binomial errors)	Delta-lognormal model (lognormal errors)
1995	1.485	4.204	0.713	2.132	4.053	1.970	3.575	5.946	2.292
1996	0.329	0.227	0.612	1.347	0.931	1.914	0.791	0.322	1.965
1997	2.097	1.201	3.383	1.424	0.943	1.858	1.065	0.343	1.635
1998	1.293	0.780	1.267	0.878	0.612	0.696	0.656	0.222	0.612
1999	0.218	0.073	0.250	0.148	0.058	0.137	0.111	0.021	0.141
2000	0.575	0.375	0.560	0.390	0.295	0.307	0.292	0.107	0.271
2001	1.004	0.139	0.215	0.682	0.109	0.118	0.510	0.040	0.104

Table 5. Standardised CPUE series (each normalised to their mean over the years considered) for the *Hotspot* aggregation obtained by fitting the "lognormal model", the delta-lognormal model assuming binomial errors for the proportion positive, and the delta-lognormal model assuming lognormal errors for the proportion positive to the observed CPUE data for Namibian orange roughy. Three methods ("zero", "same" and "proportional" of dealing with years in which no observations were made in the sub-aggregations are considered.

Year	Lognormal model	Delta-lognormal model (binomial errors)	Delta-lognormal model (lognormal errors)
1994	6.017	7.1989	5.690
1995	1.483	0.7806	1.271
1996	0.228	0.0108	0.438
1997	0.057	0.0020	0.136
1998	0.050	0.0027	0.145
1999	0.073	0.0028	0.185
2000	0.028	0.0016	0.095
2001	0.065	0.0007	0.040

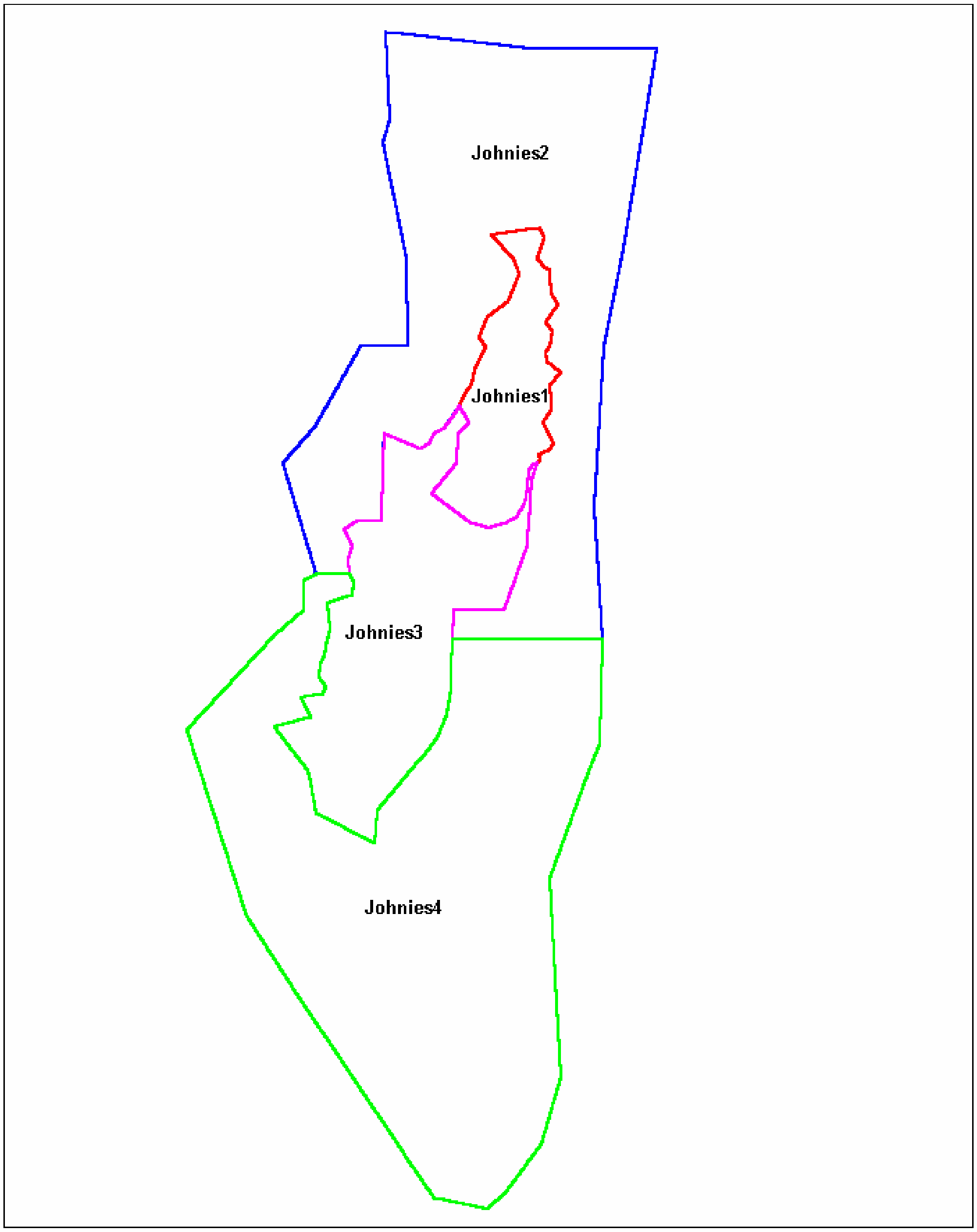

Figure 1. The *Johnies1*, *Johnies2*, *Johnies3*, and *Johnies4* strata that make up the *Johnies* aggregation.

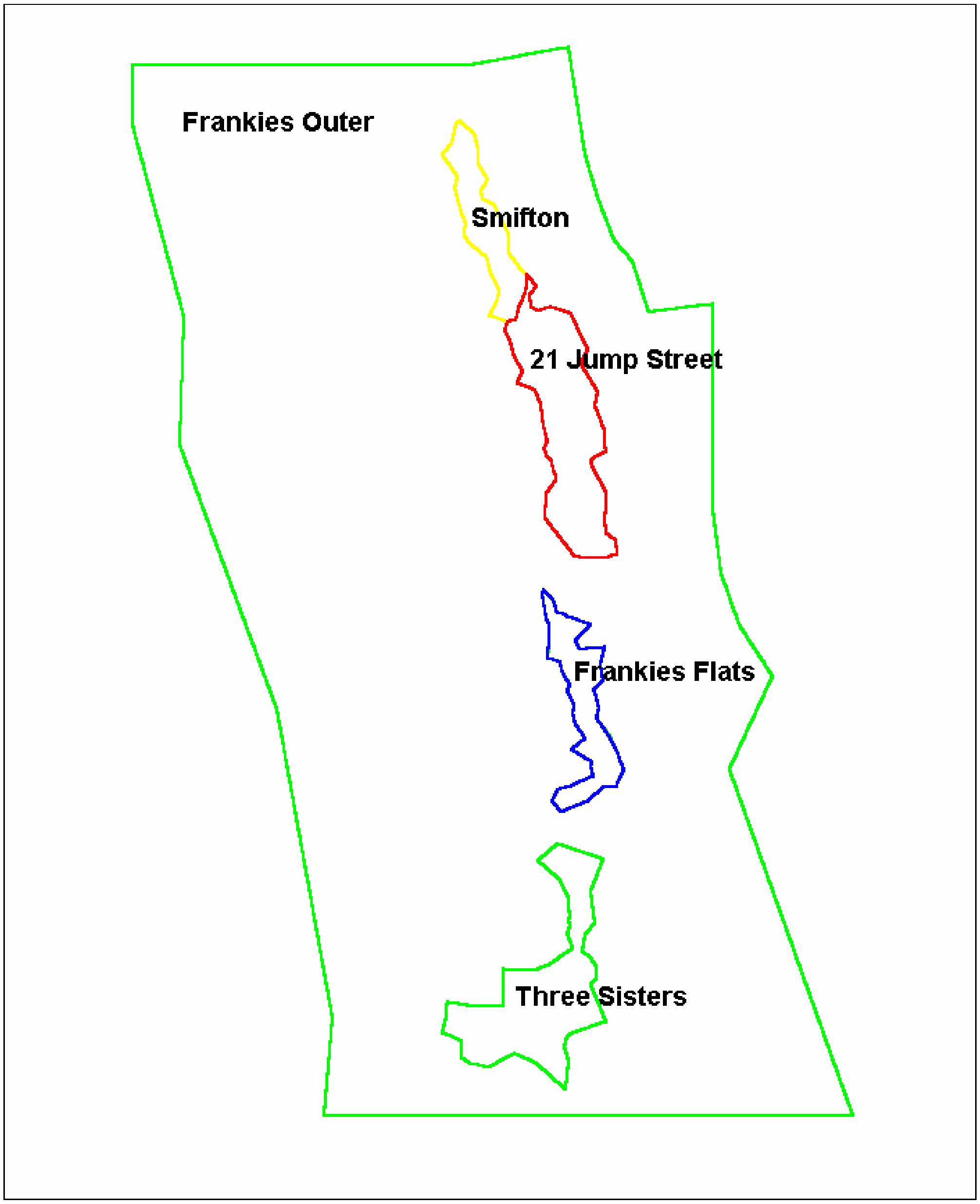

Figure 2. The *Frankies Outer*, *Smifton*, *21 Jump Street*, *Frankies Flats*, and *Three Sisters* strata that make up the *Frankies* aggregation.

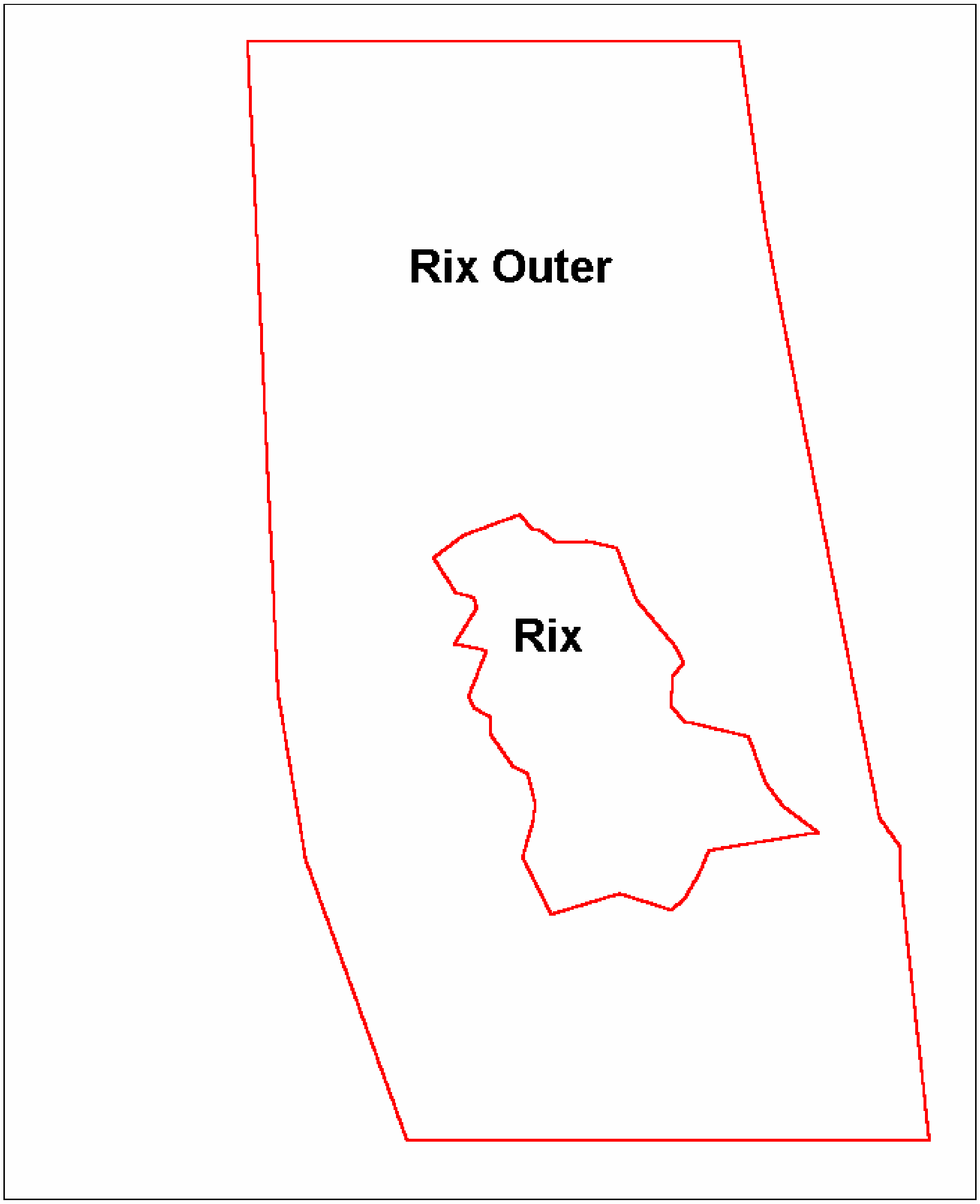

Figure 3. The *Rix* and *Rix* Outer strata that make up the *Rix* aggregation.

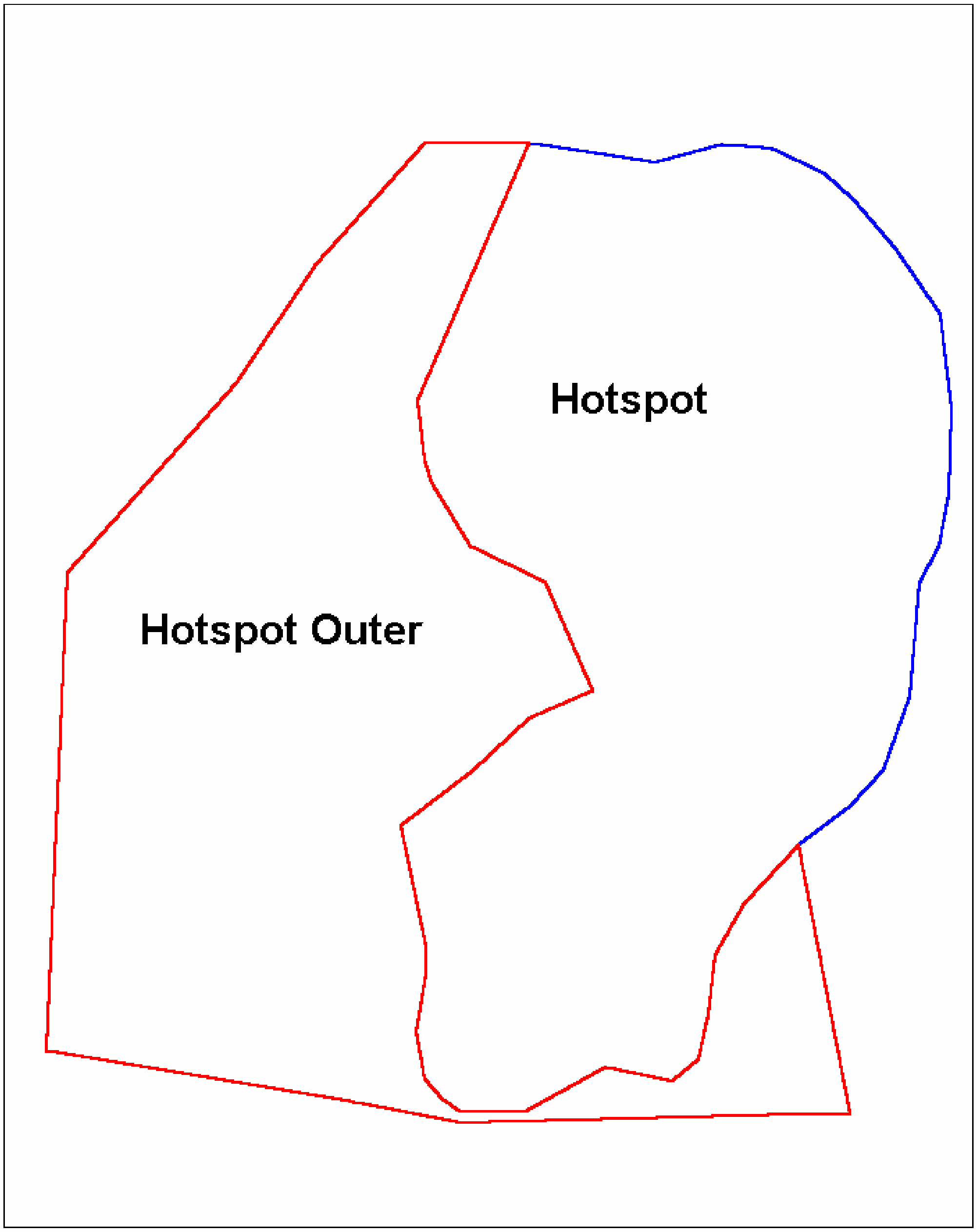

Figure 4. The *Hotspot* and *Hotspot Outer* strata that make up the *Hotspot* aggregation.

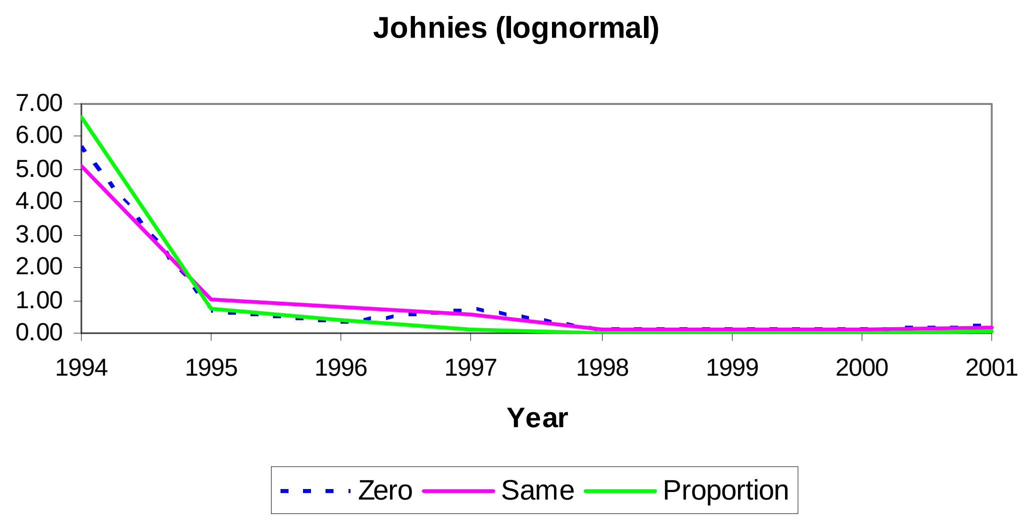

Figure 5. Index of abundance for the *Johnies* aggregation (normalised to its mean over the eight year period) for Namibian orange roughy obtained from fitting the lognormal model. Results are shown for the three methods of dealing with empty cells when combining the indices from sub-aggregations.

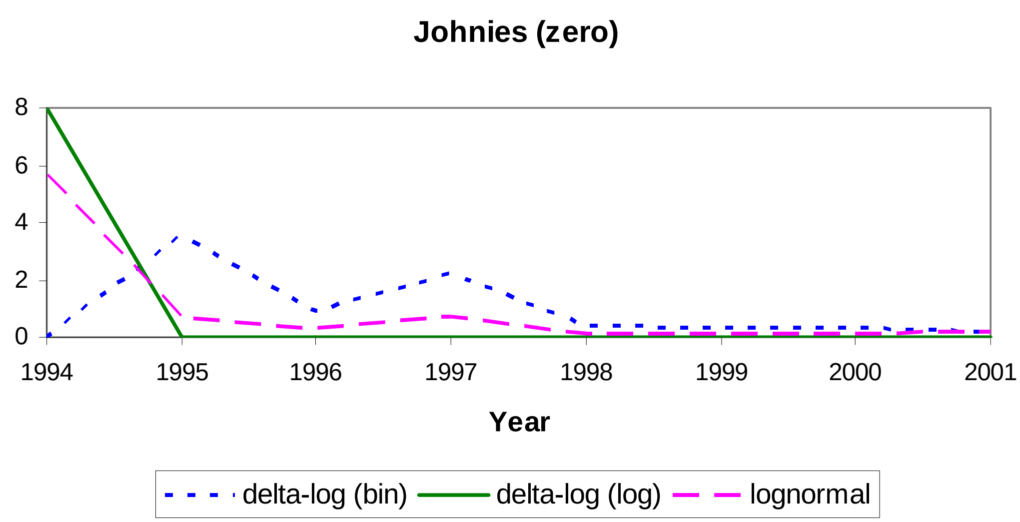

Figure 6. Index of abundance for the *Johnies* aggregation (normalised to its mean over the eight year period) for Namibian orange roughy obtained from fitting the lognormal model, the delta-lognormal model assuming binomial errors for the proportion positive, and the delta-lognormal model assuming lognormal errors for the proportion positive. Results are shown for the "zero" method of dealing with empty cells when combining the indices from sub-aggregations.

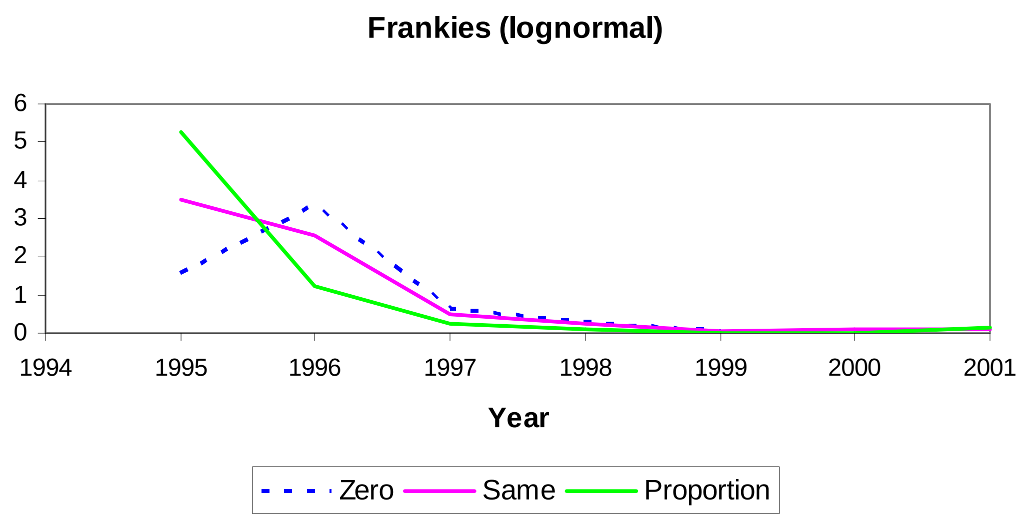

Figure 7. Index of abundance for the *Frankies* aggregation (normalised to its mean over the eight year period) for Namibian orange roughy obtained from fitting the lognormal model. Results are shown for the three methods of dealing with empty cells when combining the indices from sub-aggregations.

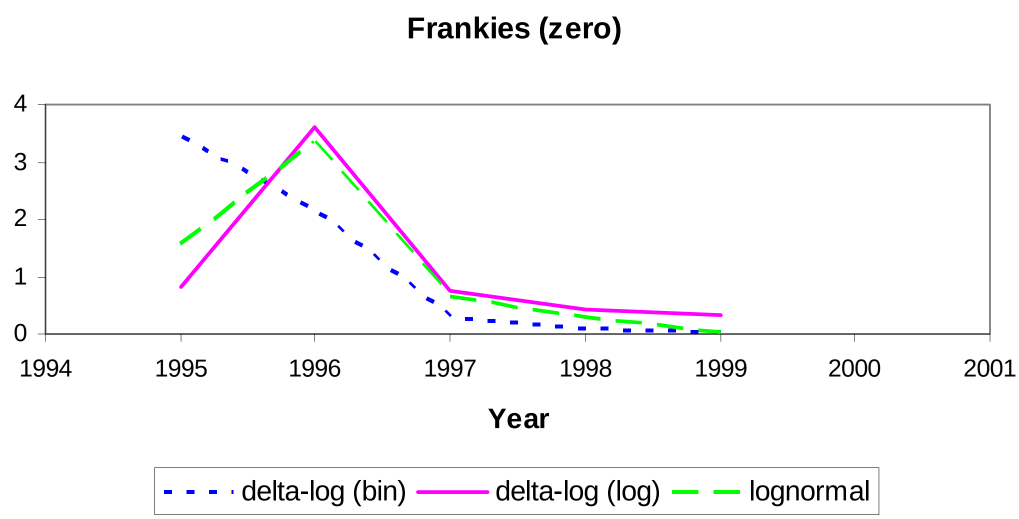

Figure 8. Index of abundance for the *Frankies* aggregation (normalised to its mean over the eight year period) for Namibian orange roughy obtained from fitting the lognormal model, the delta-lognormal model assuming binomial errors for the proportion positive, and the delta-lognormal model assuming lognormal errors for the proportion positive. Results are shown for the "zero" method of dealing with empty cells when combining the indices from sub-aggregations.

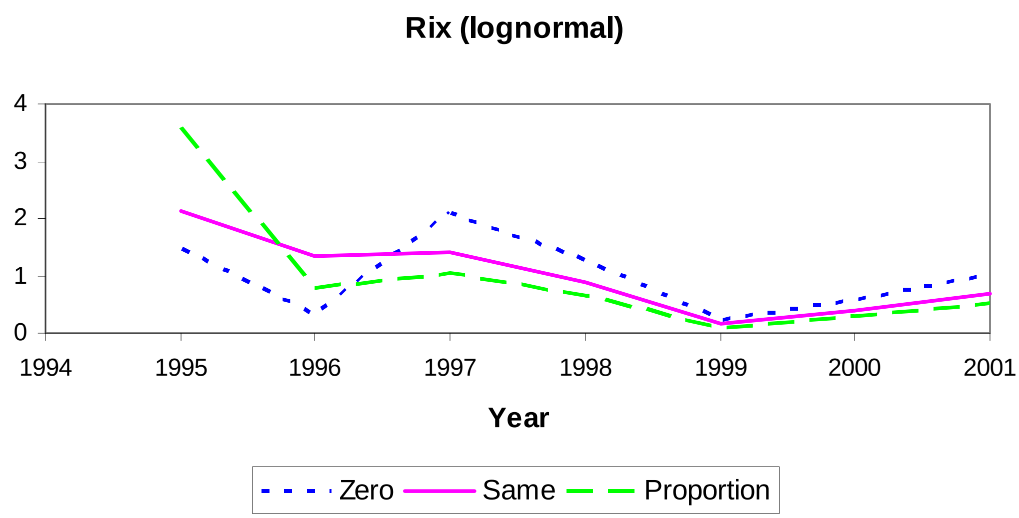

Figure 9. Index of abundance for the *Rix* aggregation (normalised to its mean over the eight year period) for Namibian orange roughy obtained from fitting the lognormal model. Results are shown for the three methods of dealing with empty cells when combining the indices from sub-aggregations.

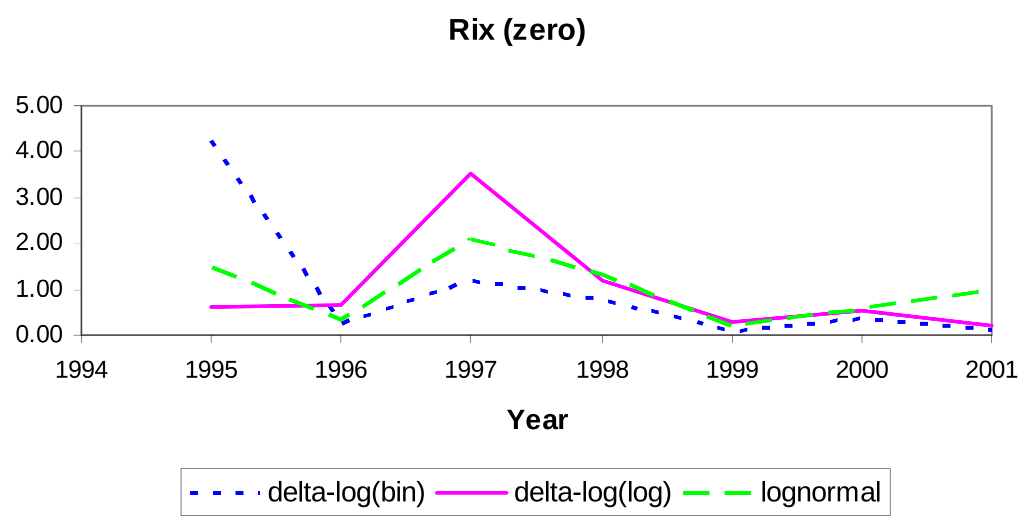

Figure 10. Index of abundance for the *Rix* aggregation (normalised to its mean over the eight year period) for Namibian orange roughy obtained from fitting the lognormal model, the delta-lognormal model assuming binomial errors for the proportion positive, and the delta-lognormal model assuming lognormal errors for the proportion positive. Results are shown for the "zero" method of dealing with empty cells when combining the indices from sub-aggregations.

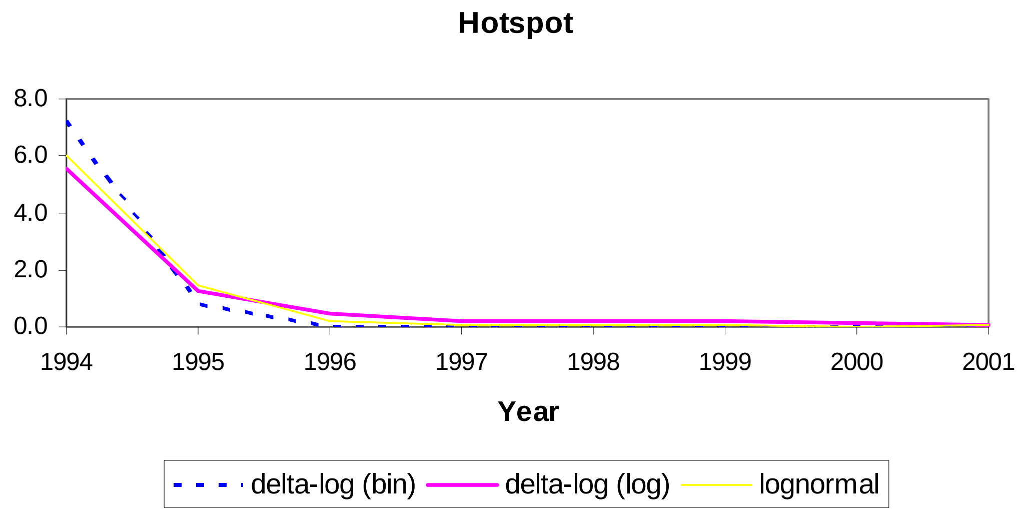

Figure 11. Index of abundance for the *Hotspot* aggregation (normalised to its mean over the eight year period) for Namibian orange roughy obtained from fitting the lognormal model. Results are shown for the three methods of dealing with empty cells when combining the indices from sub-aggregations.