# Comparison of hake species splits from observer data with the survey-generated splitting algorithm

**by**

**J.D. Gaylard and M.O. Bergh**
**OLRAC**
**Ocean and Land Resource Assessment Consultants**
**Suite 4, Silvermine House**
**Steenberg Office Park**
**Tokai 7945**

**September 2006**

## 1. Summary

Observer data collected by Capfish on commercial trawl voyages over the period November 2005 to May 2006 allow the split of hake catches into species to be calculated and compared to that predicted by the Gaylard and Bergh (2004) survey-based species splitting model. While some month to month deviation from the model predictions is evident, there appears to no significant bias on an annual basis.

## 2. Introduction

Gaylard and Bergh (2004) used survey data to develop a depth and size based algorithm for splitting hake catches into the component species *M. paradoxus and M. capensis.* In the absence of species information in the commercial catch and effort logbooks, this algorithm has been used since 2004 to produce species-specific catch and CPUE data for input into the stock assessments. Some concern over the application of the algorithm to commercial data is due to possible bias arising from the difference between commercial and survey fishing methods and the fact that surveys are conducted at specific times of year while commercial trawls occur throughout the year. With this in mind a study by Capfish was commissioned by the South African Deepsea Trawling Industry Association (SADSTIA) in which on-board observers record species composition and other relevant information on commercial voyages, with a view to comparing the observed composition with that predicted by the model. Capfish have provided us with data from 526 trawls, which we have used to make such a comparison. These data are regarded as being of a higher quality than the regular observer data, in which the quality control of the species identification information is of a lesser quality. OLRAC (2006) reports on the results of an analysis of the regular observer data and an assessment of the bias in the species splitting model in relation to these data.

## 3. Data

An initial observer dataset, covering 526 trawls from November 2005 to June 2006, has been supplied to the authors by Dave Japp of Capfish. The data include the start and end GPS location and depth of trawls, which we have averaged; date and time and sample length frequencies at 1cm intervals for each of *M. capensis* and *M. paradoxus.* We refer hereafter to this dataset as the “observer data”.

The spatial distribution of trawls are shown in Fig.1, showing a fairly even spread across the fishing grounds on both the West and South Coasts.

Fig.2 shows the distribution over depth at 50 meter intervals. Depths shallower than 200m are known to contain very few *M. paradoxus*, and depths greater than 400m are similarly known to contain few *M. capensis* so that the data of primary interest are in the intermediate “mixed zone” between 200m and 400m, which is occupied by both species. 256 (49%) of the observed trawls lie in this “mixed zone”.

Fig.3 shows the distribution of trawls by month, showing an even spread from November through May.

In addition we make use of the demersal trawl catch and effort dataset spanning years 1978 to 2003, supplied by MCM, which we refer to hereafter as the “commercial data”.

## 4. Methods

### 4. 1 Calculation of observed proportions in the observer data.

For each trawl length frequency sample in the observer data:

Each 1 cm length class is assigned a unit mass $w_L$ using the formula

$$w_L = 0.005L^{3.046} \qquad (1)$$

where $L$ is the length in cm.

The sample mass $M^{species}_{sizeclass}$ for each species and size class is summed over the corresponding length classes, i.e.

$$M^{s}_{small} = \sum_{L=21}^{42} w_L . N^s{}_L$$

$$M^{s}_{medium} = \sum_{L=43}^{57} w_L . N^s{}_L$$

$$M^{s}_{large} = \sum_{L=58}^{150} w_L . N^s{}_L \qquad (2)$$

where $N^s{}_L$ is the number of sampled fish of species *s* and length *L*

The observed proportion of *M. capensis* for each size class is calculated as

$$p^{obs}_{SC} = M^{cap}_{SC} \Big/ (M^{cap}_{SC} + M^{par}_{SC}) \qquad (3)$$

## 4.2 Calculation of predicted proportions in the observer data.

As per Gaylard and Bergh (2004) the predicted proportion of *M. capensis* in each trawl in the observer data is calculated by:

$$\bar{p}_s = \frac{1}{1+e^{B_s}} \qquad (4)$$

For trawls on the West Coast:

$$B_s = \kappa_s[d - (d^*_s + \alpha_y + \beta_L + 0.5\gamma_{summer})] \qquad (5)$$

where: $\kappa_s$ is the slope parameter for size category *s*,

$d$ is the trawl depth in metres,

$d^*_s$ is the shift parameter for size category *s*,

$\alpha_y$ is the year parameter for year *y*,

$\beta_L$ is the long-shore parameter for long-shore category *L*, and

$0.5\gamma_{summer}$ is the average of the summer and winter season factors.

The parameter values estimated are shown in Table 1.

The year parameter value used for all trawls is the "after 2003" value, which is an average of those for years 1999 to 2003.

For trawls on the South Coast:

$$B_s = \kappa_s[d - (d^*_s + \beta_L)] \qquad (6)$$

where: $\kappa_s$ is the slope parameter for size category *s*,

$d$ is the trawl depth in metres,

$d_s^*$ is the shift parameter for size category *s*, and

$\beta_L$ is the long-shore parameter for long-shore category *L*.

The parameter values estimated are shown in Table 2.

### 4.3 Implications of the observed monthly deviations for standardised CPUE trends.

We have made allowance for differences in proportions by month by including size-dependent month factors into the model.

On the West Coast, equation 5 is replaced by:

$$B_s = \kappa_s[d - (d_s^* + \alpha_y + \beta_L + 0.5\gamma_{summer} + \theta_{M,s})] \qquad (5^*)$$

and on the South Coast, equation 6 becomes

$$B_S = \kappa_S[d - (d_s^* + \beta_L + \theta_{M,s})] \qquad (6^*)$$

where: $\theta_{M,s}$ is the month parameter in month M for size category *s*, with no distinction made between coasts.

The month factors are estimated by fitting the model to the observer data, with the other parameters in the model being fixed at their original values.

The new model, including month factors, is then applied to the commercial data. Catch totals and GLM-standardised CPUE series for each species are generated using the new species splits and compared with those generated with the original splits. The GLM standardisation procedure used is the standard method used to produce CPUE inputs for the assessments as detailed in Glazer (2006).

## 5. Results and Discussion

Fig.4 presents, in each size class, a comparison of observed and predicted proportions of *M. capensis* for each of the 526 trawls plotted against depth. Note that the predicted proportions are not constrained to a single logistic curve because of differences in the model for different alongshore positions. These plots do not display any noticeable bias in the predicted proportions.

Fig.5 shows histograms of the residuals (observed minus model proportions) for each of the 526 trawls, and Fig.6 shows similar histograms but only for trawls in the 200m to 400m “mixed zone”.

Table 3 gives the total catch for all 526 trawls by size class and by species as (a) observed and (b) predicted using the model. The difference between observed and predicted species composition is illustrated in Figs 7 and 8. If the observed totals are taken as “true”, then the model can be said overall to have over-estimated the *M. capensis* component of the catch by 2.2%.

Table 4 compares the observed and predicted *M. capensis* component of the catch and the error by month. The table indicates that the model has overestimated the *M. capensis* catch from February through April and under-estimated it in November. This seems to suggest that the introduction of some dependence on month could add value to the model. It is particularly interesting then that the totals for the 7 months *in toto* are so accurately predicted, suggesting that by fortunate accident the survey months (January on the West Coast, April/September on the South Coast) represent “average” conditions with respect to the relative depth distributions of the 2 species. It should be noted too that the monthly sample sizes are small and some variance is to be expected. The seasonal patterns should be investigated in more detail once a full calendar year’s data have been gathered.

## 5.1 Implications of the observed monthly deviations

The month factors estimated by fits to the observer data are shown in Table 5. The observer data span months November through May so that the factors from months June through October are inestimable and assumed to be zero.

It should be noted that the number of trips each month is small and since trawls within each trip are generally in the same area, there is likely to be substantial within trip correlation of species split by trawl in accordance with the prevailing spatial distribution of the 2 species. Thus the between-month variance in species split from the observer data is likely to overestimate the corresponding variance in the fishery as a whole. The month factors emerging from the observer data should therefore be viewed as an extreme realization, and the "true" factors are likely to be somewhat smaller in magnitude than those estimated here.

The proportion of *M. capensis* in the annual catch totals from the commercial data, using the original and the modified splitting formulae (with month factors), are compared in Table 6 and Fig.9. The GLM-standardised CPUE series for each species, using the original and modified splitting formulae, are compared in Fig.10.

Neither the catch by species nor the CPUE trends by species appear to be substantially affected by the incorporation of month factors into the splitting formula.

# 6. Conclusions

While there is some evidence that the relationship between the hake species split and depth has some within–year pattern, it appears from the observer data received so far that the models as fitted to the survey data present an estimate of the species proportions that is reasonably accurate and without significant bias on an annual basis.

It should be noted however that the results are based on observer data spanning only 7 months. Further analysis should be conducted once a full year's observer data are available.

# 7. References

Gaylard, J.D. and M.O. Bergh. 2004. A species splitting mechanism for application to the commercial hake catch data 1978 to 2003. Unpublished MCM Demersal Working Group Document WG/09/04 D:H:21.

Glazer, J.P. 2006. Updated hake GLM-standardized CPUE series. Unpublished MCM Demersal Working Group Document WG/08/06/D:H:25

OLRAC. 2006. A comparison of the proportions of *Merluccius capensis* and *Merluccius paradoxus* in survey and observer data. Unpublished document.

**Table 1: Parameter values for substitution into equation (5) - the West Coast size-specific algorithm (Gaylard and Bergh, 2004).**

| | |
|---|---|
| **Size category values ($\kappa_s$)** | |
| $\kappa_{small}$ | 0.04722 |
| $\kappa_{medium}$ | 0.03325 |
| $\kappa_{large}$ | 0.02784 |
| **Depth parameter values ( $d^*_s$ )** | |
| $d^*_{small}$ | 177.46 |
| $d^*_{medium}$ | 282.76 |
| $d^*_{large}$ | 325.60 |
| **Year parameter $\alpha_y$** | |
| < 1985 | 14.04 |
| 1985 | 21.95 |
| 1986 | 13.52 |
| 1987 | 8.02 |
| 1988 | 0.50 |
| 1989 | 11.34 |
| 1990 | 32.73 |
| 1991 | 11.45 |
| 1992 | 21.14 |
| 1993 | 16.31 |
| 1994 | 4.84 |
| 1995 | 26.70 |
| 1996 | -6.6 |
| 1997 | 7.22 |
| 1998 | 5.25 |
| 1999 | 4.07 |
| 2000 | 5.25 |
| 2001 | 5.25 |
| 2002 | 21.51 |
| 2003 | 0.00 |
| After 2003 | 5.25 |
| **Longshore (latitude) factors ($\beta_L$)** | |
| North of 29°S | 0.00 |
| 29-30°S | -4.02 |
| 30-31°S | 4.81 |
| 31-32°S | 1.99 |
| 32-33°S | 5.75 |
| 33-34S | 14.93 |
| 34-35°S | 34.81 |
| South of 35°S | 36.27 |
| **Season factor** | |
| $\gamma_{summer}$ | -17.02 |

**Table 2: Parameter values for substitution into equation (6) - the South Coast size-specific algorithm (Gaylard and Bergh, 2004).**

| **Size category values ($\kappa_S$)** | |
|---|---|
| $\kappa_{small}$ | 0.09074 |
| $\kappa_{medium}$ | 0.03786 |
| $\kappa_{large}$ | 0.02085 |
| **Depth parameter values ( $d_S^*$ )** | |
| $d_{small}^*$ | 181.62 |
| $d_{medium}^*$ | 257.29 |
| $d_{large}^*$ | 386.85 |
| **Longshore (longitude) factors ($\beta_L$)** | |
| West of 21°E | 0.00 |
| 21-22°E | 18.92 |
| 22-23°E | -20.74 |
| 23-24°E | -33.63 |
| 24-25°E | -34.00 |
| 25-26°E | -11.64 |
| East of 26°E | 44.51 |

**Table 3. Observed and Predicted species components of the total catch for the 526 trawls in this study.**

| | **Large** | **Medium** | **Small** | **Total** |
|---|---|---|---|---|
| Catch (tons) | 13.92 | 10.98 | 6.95 | 31.85 |
| Observed *M. capensis* | 6.18 | 3.50 | 0.62 | 10.31 |
| Predicted *M. capensis* | 6.41 | 3.99 | 0.61 | 11.01 |
| Error | 0.23 | 0.48 | -0.01 | 0.70 |
| Error/Catch | 0.017 | 0.044 | -0.002 | 0.022 |
| Error/M. capensis | 0.037 | 0.138 | -0.019 | 0.068 |
| Error/M. paradoxus | 0.030 | 0.065 | -0.002 | 0.033 |

**Table 4. Observed and predicted *M. capensis* component of the catch and the percentage model error by size class and month.**

| | Number of trawls | observed *M. capensis* large | predicted *M. capensis* large | error large | | Observed *M. capensis* medium | Predicted *M. capensis* medium | error medium |
|---|---|---|---|---|---|---|---|---|
| **November** | 55 | 0.364550 | 0.228997 | -37% | | 0.011040 | 0.011776 | 7% |
| **December** | 55 | 0.715606 | 0.692144 | -3% | | 0.648423 | 0.645748 | 0% |
| **January** | 108 | 0.766527 | 0.765793 | 0% | | 0.296194 | 0.305850 | 3% |
| **February** | 79 | 0.739177 | 0.803018 | 9% | | 0.935124 | 1.090191 | 17% |
| **March** | 76 | 0.686917 | 1.030665 | 50% | | 0.555008 | 0.796047 | 43% |
| **April** | 76 | 0.244030 | 0.281205 | 15% | | 0.036412 | 0.093459 | 157% |
| **May** | 77 | 2.667461 | 2.612646 | -2% | | 1.020648 | 1.044074 | 2% |
| | | | | | | | | |
| | | **observed *M. capensis* small** | **predicted *M. capensis* small** | **error small** | | **observed *M. capensis* total** | **predicted *M. capensis* total** | **error total** |
| **November** | 55 | 0.001 | 0.002801 | 219% | | 0.376469 | 0.243574 | -35% |
| **December** | 55 | 0.097 | 0.097389 | 0% | | 1.461467 | 1.435282 | -2% |
| **January** | 108 | 0.010 | 0.008657 | -13% | | 1.072617 | 1.080300 | 1% |
| **February** | 79 | 0.302328 | 0.301000 | 0% | | 1.976629 | 2.194209 | 11% |
| **March** | 76 | 0.126844 | 0.115939 | -9% | | 1.368769 | 1.942651 | 42% |
| **April** | 76 | 0 | 0.000589 | | | 0.280442 | 0.375254 | 34% |
| **May** | 77 | 0.085221 | 0.084395 | -1% | | 3.773330 | 3.741115 | -1% |

**Table 5. Month factors estimated by fits to the observer data, by size class and month.**

| | Large | Medium | Small |
|---|---|---|---|
| **November** | 44.38 | -4.36 | -34.04 |
| **December** | 26.09 | 24.85 | 63.04 |
| **January** | -10.645 | -9.48 | 6.06 |
| **February** | -23.57 | -35.33 | -9.89 |
| **March** | -41.43 | -50.67 | 18.53 |
| **April** | -23.66 | -31.92 | -123.57 |
| **May** | -1.50 | -4.48 | -18.49 |
| **June to October** | 0 | 0 | 0 |

**Table 6. Proportion of *M. capensis* in the offshore trawl catch 1978 to 2003 using (a) the usual species split and (b) incorporating the month factors.**

| | **refcase** | **with additional month factors** | **difference** | | | **refcase** | **with additional month factors** | **difference** |
|---|---|---|---|---|---|---|---|---|
| 1978 | 0.198 | 0.197 | -0.001 | | 1991 | 0.187 | 0.193 | 0.006 |
| 1979 | 0.302 | 0.311 | 0.009 | | 1992 | 0.185 | 0.187 | 0.002 |
| 1980 | 0.235 | 0.249 | 0.013 | | 1993 | 0.123 | 0.132 | 0.009 |
| 1981 | 0.245 | 0.244 | -0.001 | | 1994 | 0.130 | 0.133 | 0.003 |
| 1982 | 0.268 | 0.274 | 0.006 | | 1995 | 0.139 | 0.141 | 0.002 |
| 1983 | 0.238 | 0.248 | 0.010 | | 1996 | 0.097 | 0.097 | 0.000 |
| 1984 | 0.269 | 0.261 | -0.007 | | 1997 | 0.107 | 0.110 | 0.003 |
| 1985 | 0.295 | 0.292 | -0.003 | | 1998 | 0.084 | 0.083 | -0.001 |
| 1986 | 0.217 | 0.220 | 0.003 | | 1999 | 0.112 | 0.112 | 0.000 |
| 1987 | 0.186 | 0.194 | 0.008 | | 2000 | 0.148 | 0.149 | 0.002 |
| 1988 | 0.208 | 0.212 | 0.004 | | 2001 | 0.105 | 0.104 | -0.001 |
| 1989 | 0.267 | 0.269 | 0.002 | | 2002 | 0.117 | 0.116 | -0.001 |
| 1990 | 0.306 | 0.299 | -0.008 | | 2003 | 0.089 | 0.087 | -0.002 |

**Figure 1.** **Spatial distribution of the 526 observed trawls in the study.**

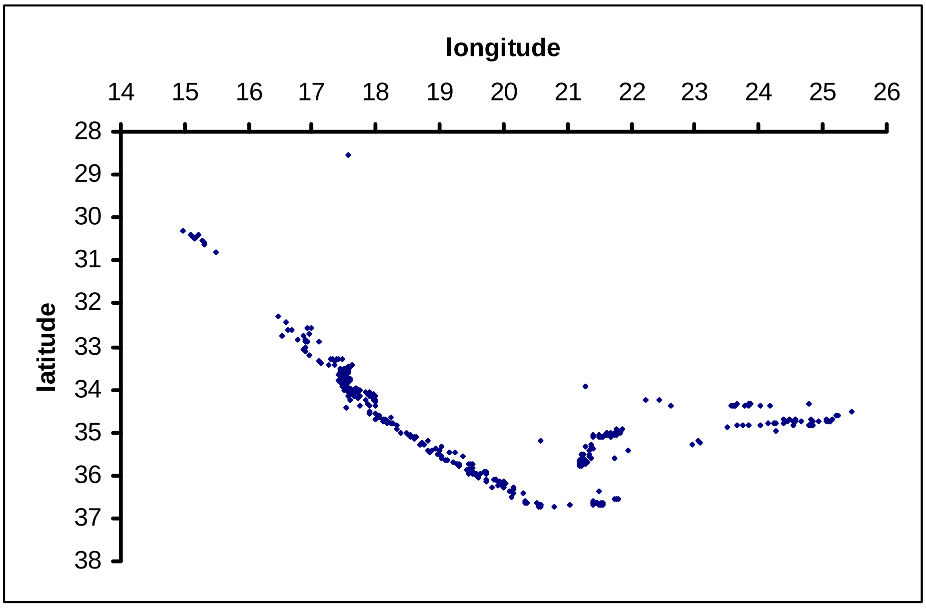



**Figure 2.** **Distribution over depth of the 526 observed trawls in the study.**

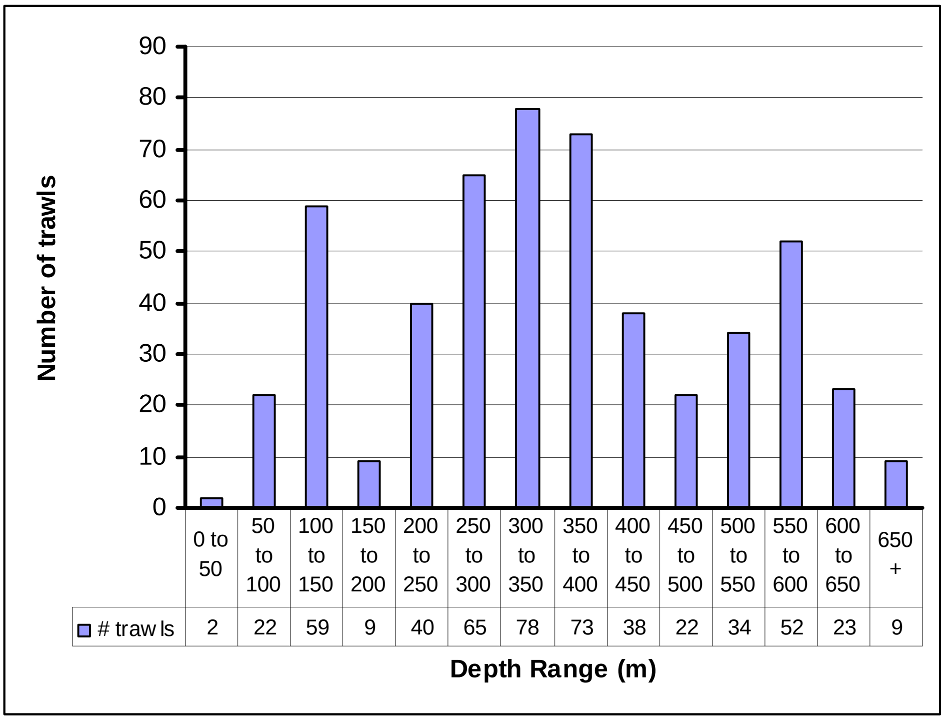


**Figure 3. Distribution over months of the 526 observed trawls in the study.**

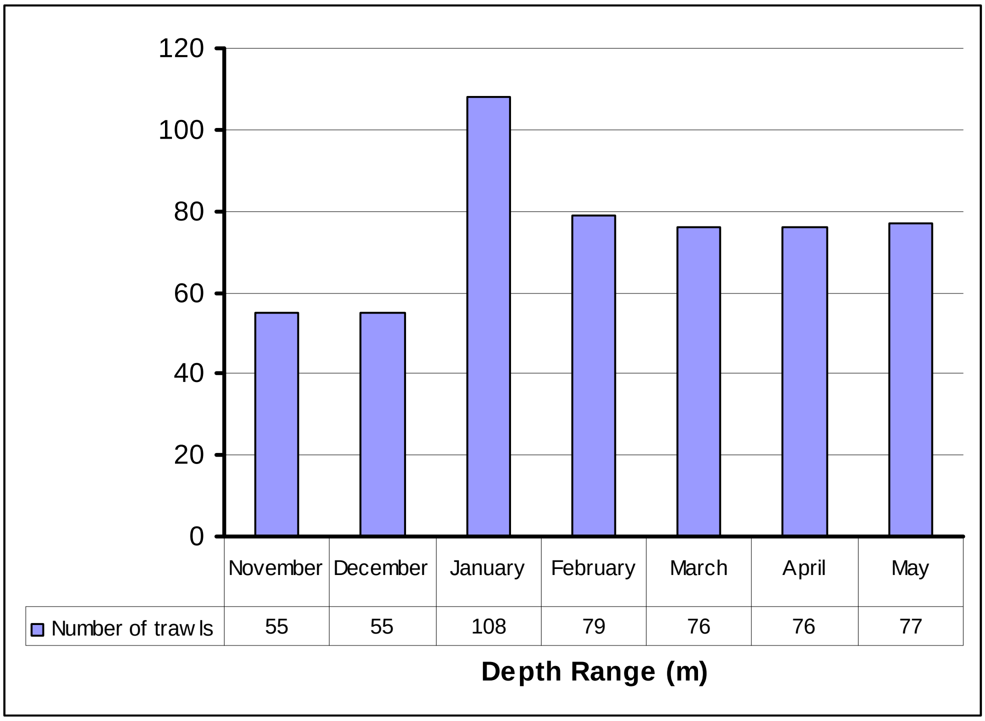


**Figure 4. Proportion of M. capensis in the hake catch as recorded by observers and as predicted by the species splitting model plotted against depth in three size classes.**

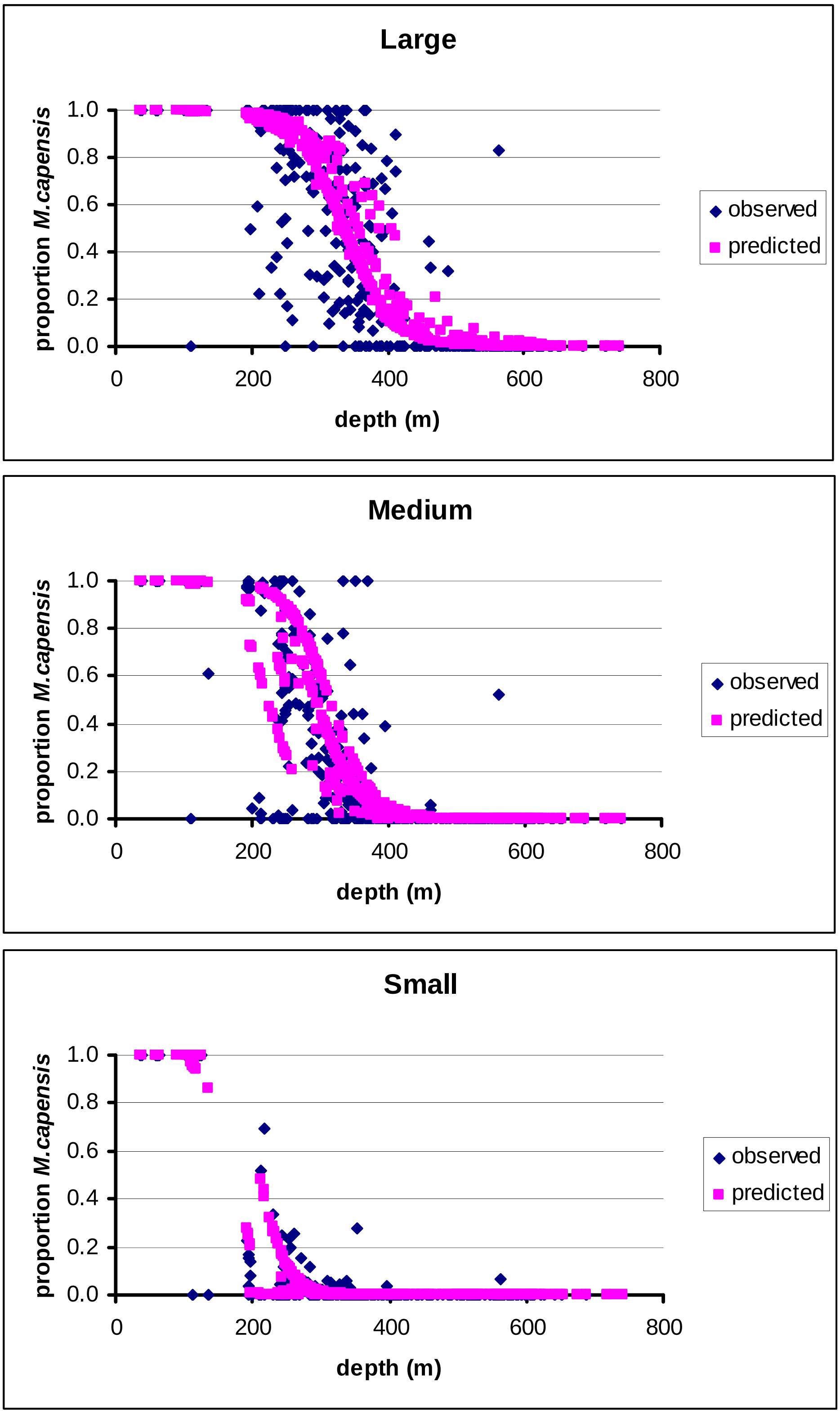


**Figure 5. Histograms of residuals (difference between observed an predicted proportions of *M. capensis*) by size class for all 526 trawls.**

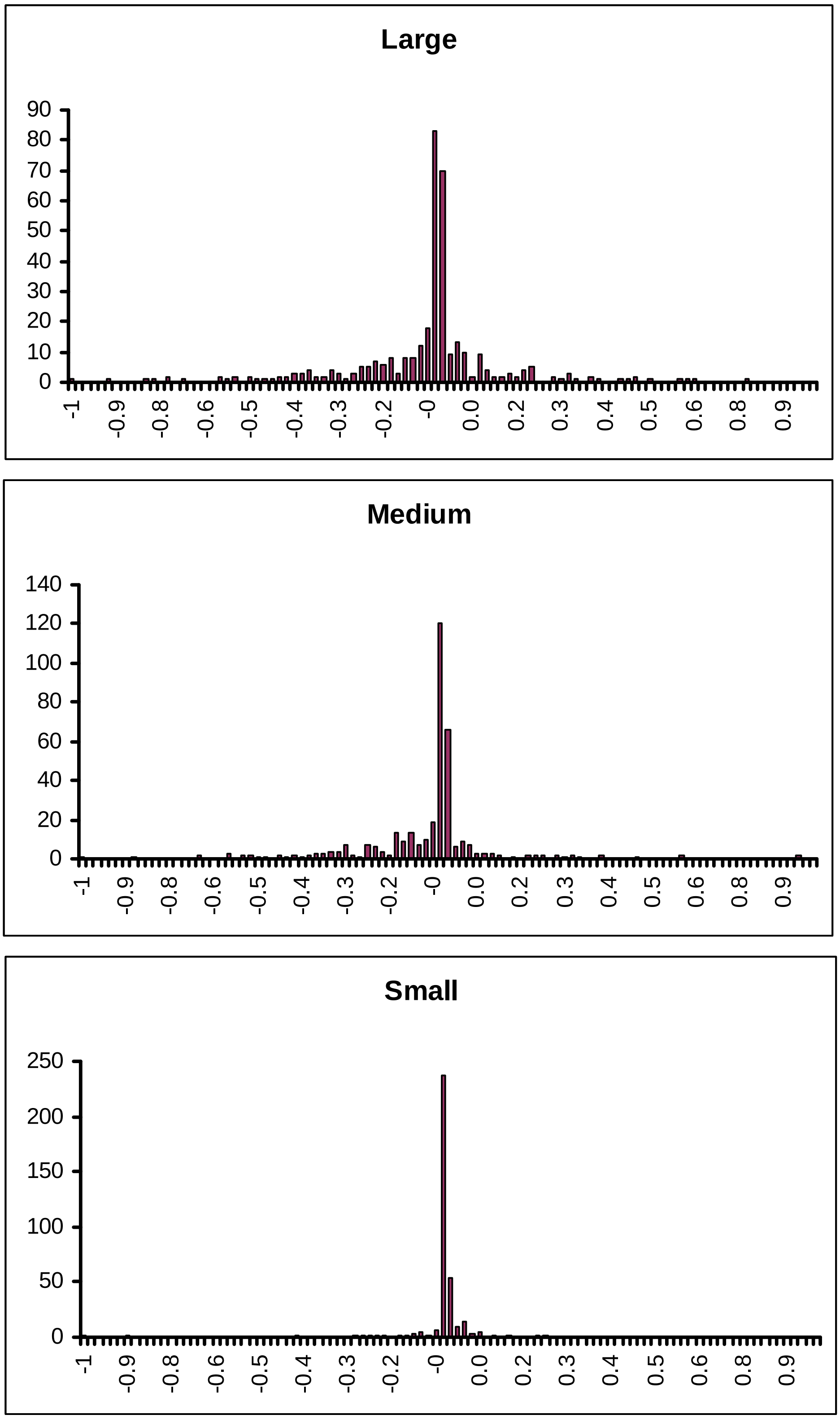


**Figure 6. Histograms of residuals (difference between observed an predicted proportions of *M. capensis*) by size class for the 256 trawls in the 200m to 400m depth range.**

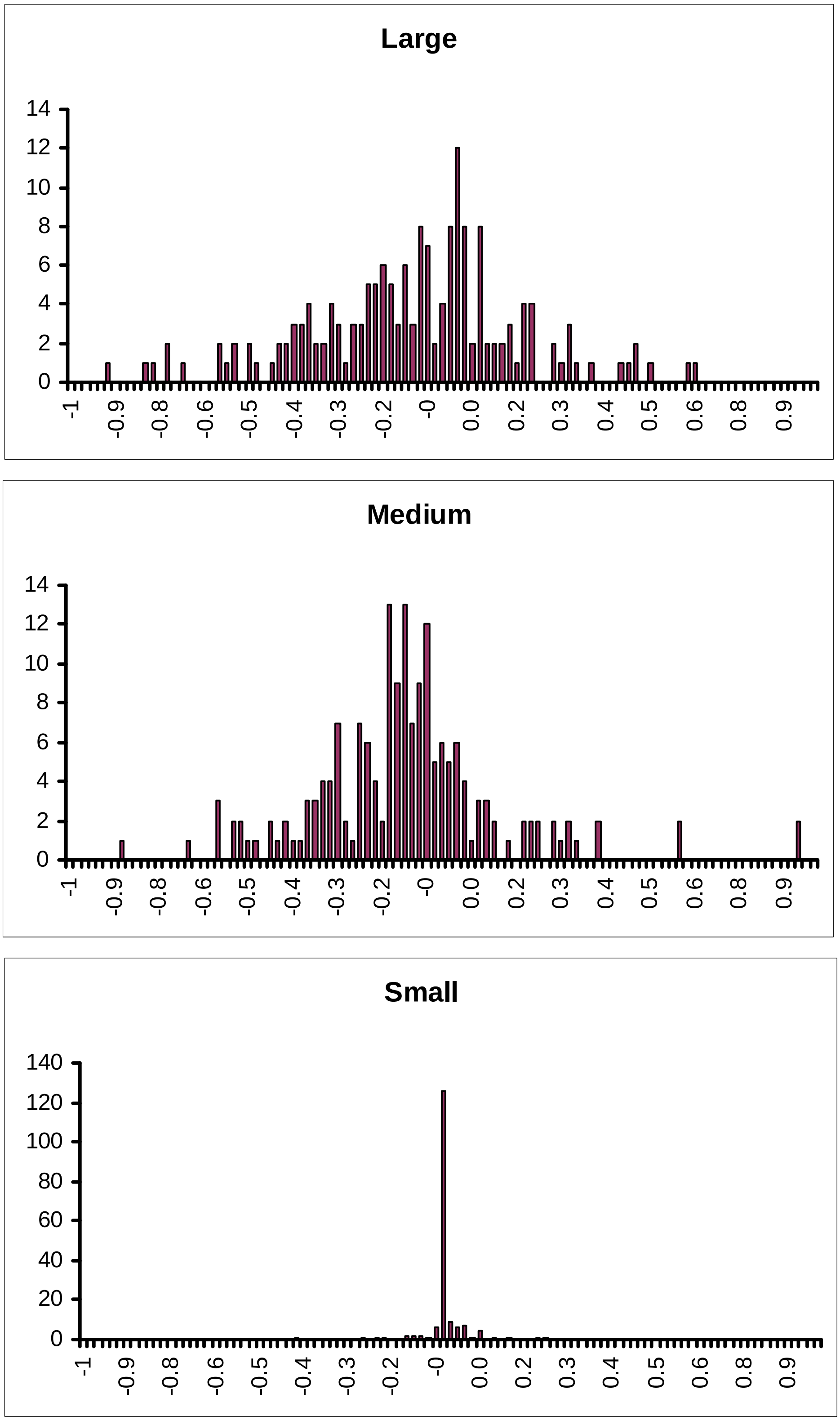
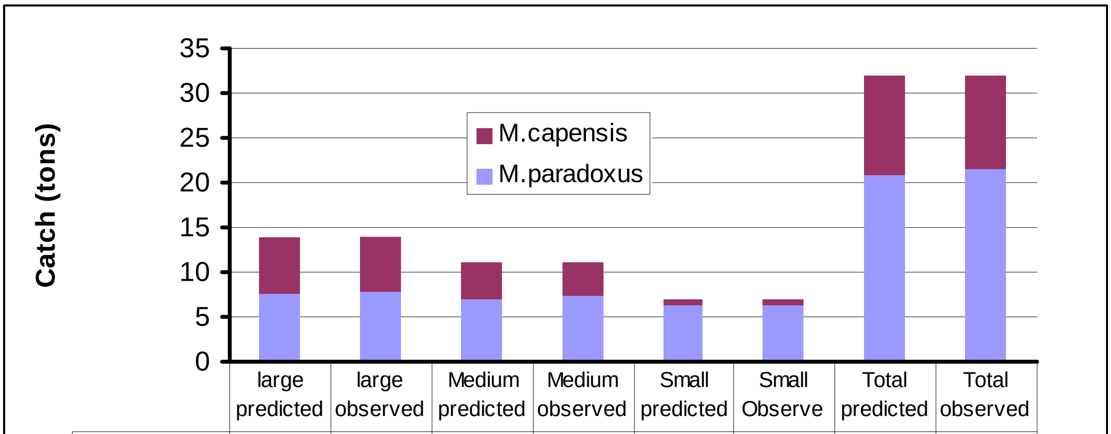


**Figure 7. The comparison of the predicted and observed species components of the *total catch* by size class from all 526 trawls in the study.**




| | large predicted | large observed | Medium predicted | Medium observed | Small predicted | Small Observe | Total predicted | Total observed |
|---|---|---|---|---|---|---|---|---|
| M.capensis | 6.41 | 6.18 | 3.99 | 3.50 | 0.61 | 0.62 | 11.01 | 10.31 |
| M.paradoxus | 7.51 | 7.74 | 6.99 | 7.48 | 6.34 | 6.33 | 20.84 | 21.55 |

**Figure 8. The error in species allocation of the total catch by comparison with the catch itself.**

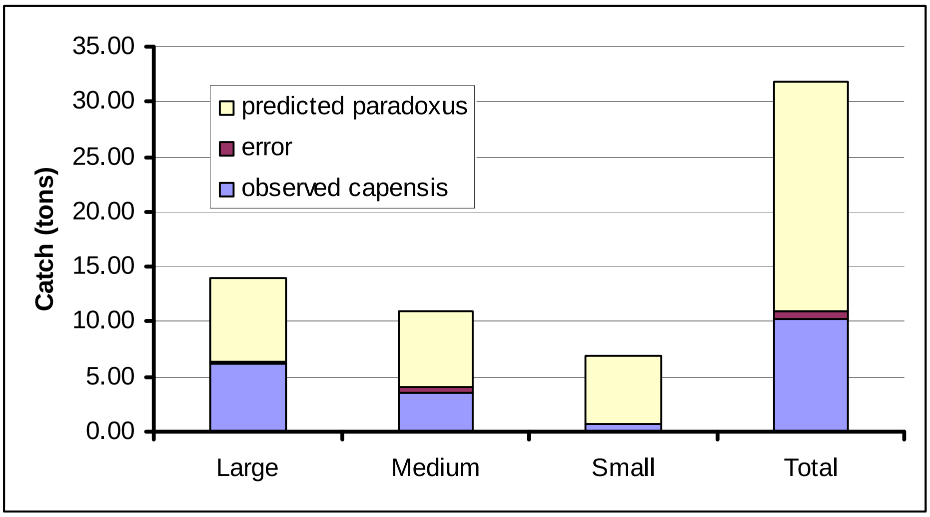


**Figure 9.** **Proportion of *M. capensis* in annual hake catches recorded in the commercial data (a) using the original species split formula and (b) using the modified formula including the month factors from Table 5.**

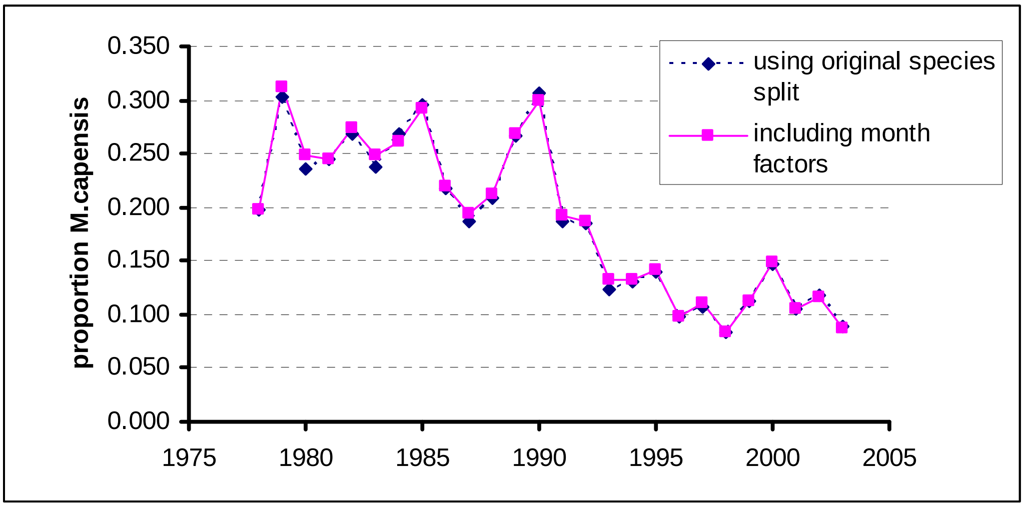



**Figure 10.** **GLM-standardised CPUE for *M. capensis and M. paradoxus* (a) using the original species split formula and (b) using the modified formula including the month factors in Table 5.**