AN ANALYSIS OF THE SMALL SCALE SURVEYS OF ANCHOVY ABUNDANCE AROUND ROBBEN AND DASSEN ISLANDS FROM 2009 TO 2013

William M.L. Robinson and Doug S. Butterworth
Marine Resource Assessment and Management Group (MARAM)
Department of Mathematics and Applied Mathematics
University of Cape Town, Rondebosch 7701, South Africa

Summary

The results from the small scale hydroacoustic surveys of the abundance of anchovy around Robben and Dassen islands over the 2009-2013 period are analysed under the assumption of a Gaussian form for the trends in density at each island over the course of the winter months. Based primarily on AIC_c, the model selected from amongst a number of variants has the same trend in abundance with year for the two islands, compatible with the assumption used by Robinson (2013) in his GLM analysis of the impact of closures to pelagic fishing around these islands on penguin recovery, though the data have limited power to distinguish possible deviations from that assumption. The abundance estimates from the island surveys, though compatible also with the May recruitment survey trends, show appreciably larger variance. This raises the question of whether these small scale surveys merit continuation, unless it would be possible to increase their frequency considerably during the winter months each year to improve the overall precision of the integrals over local abundance which they can provide – an option which seems logistically impractical given, for example, considerations of inclement weather over the period concerned.

Introduction

Small scale hydroacoustic surveys to determine anchovy around Robben and Dassen islands at various stages during (primarily) the winter months have been taking place since 2009. The purpose of these surveys has been to monitor anchovy abundance on a finer spatio-temporal scale than provided by the annual recruitment survey. This is to provide better insight into the levels of prey available annually to penguins during their breeding periods at Robben and Dassen islands. This has been hoped to assist in the interpretation of the results from the current feasibility study of alternating pelagic fishing closure around these islands, with a view towards determining whether such closures benefit penguin population recovery.

This paper provides an analysis of the results from these surveys in that context, through providing a time-integration of results for each year.

Data and Methods

Janet Coetzee kindly provided the anchovy abundance estimates and associated survey sampling standard errors for these surveys (see Table 1 and Figure 1), as well as this same information for the full area covered by the annual May recruit surveys (see Table 1 and Figure 3).

The basis of the method of analysis is to assume that the changes over time in the anchovy abundance in any year around an island follow a normal (Gaussian) curve. The details of the analysis are set out in the Appendix. For the most general form of the analysis attempted, the parameters of these Gaussian forms which relate to the magnitude of the peak abundance and the date on which this occurs vary freely amongst years and between the two islands. Only the spread parameter for these forms is the same from year to year, though potentially colony-dependent. The analysis considers a range of simplifications of this general (full) form of the model, with the best model being selected primarily on the basis of the AIC_c model selection criterion.

The main thrust of the analysis is to determine what inferences can be drawn about the annual abundances of anchovy available to penguins at these islands over the period of the surveys, and how these relate to the abundance estimates provided by the annual May recruit surveys. This has relevance, for example, to checking a key assumption made in the GLM analyses by Robinson (2013) of various statistics collected which relate penguin reproductive success to the anchovy abundances and catches in the near-vicinity of the two islands. This is that the abundances of anchovy available each year around each of the two colonies are (in expectation) in a proportion that stays constant over time (and is independent of the overall anchovy abundance).

Results and Discussion

Table 2 provides results for a large number of simpler forms of the full model mentioned above (Model 1), where these simplifications are described in detail in the Appendix. Figures 1a–1f provide fits to the small scale survey abundance indices for six of the twelve model variants. Importantly, in moving from Model 1 to Model 2, which assumes equal spread (σ_d^i) parameters for the distribution of abundance in time every year for the two islands and is AIC_c justified, the $\tilde{b}_y^i$parameters (see Appendix) become proportional to the time-integrals over the Gaussians forms for all the model variants following, and hence provide (on exponentiation) indices of integrated annual abundance in the near vicinity of island *i*.

The AIC_c values in Table 2 indicate that the data available do **not** support attempts to estimate a parameter (Δ) which separates the time when peak abundances occur at Dassen and Robben islands each year (Models 3, 5, 8 and 11, when compared to models which ignore this difference).

A minimum AIC_c is one of two criteria used to determine the selection of the best model. The other is the realism of the parameter estimates. The annual anchovy dynamics cycle is well known, with the bulk of the recruits passing southward along the South African west coast during the winter months. As such, estimates of parameters $\bar{d}_y^i$ (the day when the annual anchovy abundance peak occurs at colony *i*) which fall outside the April-to-August period, are regarded as unrealistic and grounds to reject the model. It so happens that this leads to the rejection of Model 10 (see Figure 1f) which has the lowest AIC_c value, but for which four of the $\bar{d}_y$ parameters fall outside this range.

The three best of the remaining models in terms of AIC_c are Models 6, 7 and 9, for which the results for the abundance related $\tilde{B} = \exp(\tilde{b})$ parameter are shown in Figure 2. Although Model 7 marginally shades Model 9 in AIC_c terms (their difference is that the former allows year-specific estimates of the date at which abundance peaks during the season), Model 7 is rejected for the same reason as Model 10 – estimates of these dates outside the realistic April-to-August range.

The final preferred model is thus Model 9, which sets annual abundances at Robben and Dassen to have the same ratio each year. This is preferred over both Model 6, which allows variation in this ratio, and Model 12 which treats the abundance of anchovy each year to be unchanged in expectation.

Figure 2a compares the estimates of the abundance-related $\tilde{B} = \exp(\tilde{b})$ parameter for Model 6, where these estimates are allowed to have different trends with year for the two islands. The point estimates for the three common years show broadly similar trends, but the associated variances are very high. Figure 2b repeats this for Models 7 and 9 for each of which Robben and Dassen are taken to have the same trend; these trends are to be compared with those from the May recruitment surveys shown in Figure 3. Estimates of the parameter values (other than those already listed in Table 2) for Models 7 and 9 are given in Table 3.

These comparisons are clearer in Figure 4, where these results are shown on the same plots with common normalisations to geometric means over 2010 to 2012. What is evident from these plots is compatibility with common trends from the small and larger scale surveys (though given the large CVs for the island results for Models 6 and 7, achieving compatibility in those cases is not a "strong" result), except for 2009 where the small scale survey around Robben gives low results compared to the May recruit survey. Recall that Model 9 is the preferred model, and there results do evidence compatibility more strongly.

Conclusions

The preferred model from the analysis is Model 9, for which the trends in abundance around Robben and Dassen are the same, as assumed for the GLM analyses of the impact of pelagic fishing around these islands on penguin recovery by Robinson (2013). Given the large CVs associated with the estimates from the small scale surveys, the small scale survey data clearly have limited power to distinguish deviations from this assumption, though nevertheless there is nothing in the results from the various models to suggest that this assumption could be appreciably incorrect.

The reason for these large CVs is related to the large estimates of additional variance (σ^{i}_{add}) forthcoming from these analyses (see Table 2). What this is indicating is that in addition to the largish sampling CVs for the small scale surveys, there is a larger still “process” error reflecting deviations from the normal distribution assumed to reflect the annual trend in abundance near an island over the penguin breeding season – presumably the result of the patchy nature of shoals of recruiting fish as they move down the west coast to the Agulhas bank, a feature which the larger-scale May recruit surveys are able to integrate over.

Unfortunately however, this indicates that there is little information content in these small scale surveys, which despite their greater frequency within a year, are indicated by these analyses to provide less precise estimates on an annual scale than the May recruitment surveys (see Figure 4b for Model 9 in particular). This must raise the question of whether these small scale surveys merit continuation, unless it would be possible to increase their frequency considerably during the winter months each year to improve the overall precision of the integrals over local abundance which they can provide. However Coetzee (pers. commn) advises that this seems logistically impractical given, for example, considerations of inclement weather over the period concerned

Acknowledgements

Janet Coetzee kindly provided the abundance estimates from the small scale survey data upon which these analyses are based. We thank Andre Punt for initial discussions on the form of analysis to apply.

Reference

Robinson WML. 2013. Modelling the impact of the South African small pelagic fishery on African penguin dynamics. PhD Thesis, University of Cape Town. 216 pp.

APPENDIX

Data

The following small-scale surveys have been conducted around Dassen and Robben islands:

Dassen Island: 2010 (3 surveys), 2011 (6), 2012 (4)

Robben Island: 2009 (6), 2010 (4), 2011 (6), 2012 (5), 2013 (1)

Basic model

The biomass of fish around Dassen and Robben islands is modelled as a Gaussian form each year which has three parameters: the maximum biomass each year, the day on which that maximum occurs, and the width of the Gaussian.

$$\ln B_{y,d}^{i,\text{obs}} = a^i + b_y^i - \frac{1}{2\left(\sigma_d^i\right)^2}\left(d - \bar{d}_y^i\right)^2 + \varepsilon_{y,d}^i \tag{1}$$

where:

i is the island around which the survey is conducted,

y is the year of the survey,

d is the number of days from 1 January to the survey date,

$B_{y,d}^{i,\text{obs}}$ is the pelagic fish biomass estimated from the survey,

$\bar{d}_y^i$ is the day which the abundance of fish around island i is at its maximum in year y,

σ_d^i characterizes the spread of the distribution of fish over time during the year, and

$\varepsilon_{y,d}^i$ is the error term, distributed as $N\left(0,\left(\sigma_{y,d}^i\right)^2\right)$, where $\left(\sigma_{y,d}^i\right)^2 = \left(\sigma_{y,d}^{i,\text{obs}}\right)^2 + \left(\sigma_{\text{add}}^i\right)^2$; the root mean square average over years of these values is denoted by $\bar{\sigma}^{\text{R}}$ or $\bar{\sigma}^{\text{D}}$ for Robben and Dassen islands respectively in Table 2.

The estimable parameters are a^i, b_y^i, σ_d^i, $\bar{d}_y^i$, and σ_{add}^i. The abundance parameters b_{2010}^{Dassen} and b_{2009}^{Robben} are set to zero (i.e. absorbed in the intercept terms a^i). The process error (or "additional variance") terms σ_{add}^i allow for the fact that the actual distribution each year is not exactly Gaussian.

The negative log-likelihood is:

$$-\ln L = \sum_{i,y,d}\left\{\ln \sigma_{y,d}^i + \frac{1}{2\left(\sigma_{y,d}^i\right)^2}\left[\ln B_{y,d}^{i,\text{obs}} - \left(a^i + b_y^i - \frac{1}{2\left(\sigma_d^i\right)^2}\left(d - \bar{d}_y^i\right)^2\right)\right]^2\right\} \tag{2}$$

The single 2013 survey is excluded from the analyses as it makes no meaningful contribution to the likelihood function. The full model has 18 parameters and 34 data points. The model can be simplified by reducing the number of parameters. The AIC_c scores are compared to judge which model is preferred. The following variations of the full model are considered:

$\sigma_d^i = \sigma_d$ Spread is the same for both islands.

$\bar{d}_y^{\text{R}} = \bar{d}_y^{\text{D}} + \Delta$ Robben maximum density occurs Δ days later than at Dassen.

$\bar{d}_y^{\mathrm{R}} = \bar{d}_y^{\mathrm{D}}$ Robben and Dassen maximum densities occur at the same time as each other each year.

$\bar{d}_y^{\mathrm{D}} = \bar{d}$, $\bar{d}_y^{\mathrm{R}} = \bar{d} + \Delta$ Maximum densities occur at the same time each year at a time Δ days apart.

$\bar{d}_y^i = \bar{d}$ The time at which Robben and Dassen densities are maximal occurs on the same day every year.

$b_y^i = b_y$ The same biomass difference occurs around the islands each year.

$b_y^i = b$ The biomass is the same each year at each island.

Models incorporating the following combinations of variations are considered:

1. Full model
2. $\sigma_d^i = \sigma_d$
3. $\sigma_d^i = \sigma_d$ $\bar{d}_y^{\mathrm{R}} = \bar{d}_y^{\mathrm{D}} + \Delta$
4. $\sigma_d^i = \sigma_d$ $\bar{d}_y^{\mathrm{R}} = \bar{d}_y^{\mathrm{D}}$
5. $\sigma_d^i = \sigma_d$ $\bar{d}_y^{\mathrm{D}} = \bar{d}$, $\bar{d}_y^{\mathrm{R}} = \bar{d} + \Delta$
6. $\sigma_d^i = \sigma_d$ $\bar{d}_y^i = \bar{d}$
7. $\sigma_d^i = \sigma_d$ $\bar{d}_y^{\mathrm{R}} = \bar{d}_y^{\mathrm{D}}$ $b_y^i = b_y$
8. $\sigma_d^i = \sigma_d$ $\bar{d}_y^{\mathrm{D}} = \bar{d}$, $\bar{d}_y^{\mathrm{R}} = \bar{d} + \Delta$ $b_y^i = b_y$
9. $\sigma_d^i = \sigma_d$ $\bar{d}_y^i = \bar{d}$ $b_y^i = b_y$
10. $\sigma_d^i = \sigma_d$ $\bar{d}_y^{\mathrm{R}} = \bar{d}_y^{\mathrm{D}}$ $b_y^i = b$
11. $\sigma_d^i = \sigma_d$ $\bar{d}_y^{\mathrm{D}} = \bar{d}$, $\bar{d}_y^{\mathrm{R}} = \bar{d} + \Delta$ $b_y^i = b$
12. $\sigma_d^i = \sigma_d$ $\bar{d}_y^i = \bar{d}$ $b_y^i = b$

To aid interpretation of the results, it is convenient to express abundances relative to the geometric mean around Robben Island over 2010 to 2012, i.e. re-parameterizing $b_y^i \to \tilde{b}_y^i$ as follows:

$$\tilde{b}_y^i = b_y^i - \frac{1}{3}\left(b_{2010}^{\mathrm{R}} + b_{2011}^{\mathrm{R}} + b_{2012}^{\mathrm{R}}\right)$$

For comparison with the May recruit survey estimates R, we take $r_y^{\mathrm{obs}} = \ln R_y^{\mathrm{obs}}$. The standard errors of r_y^{obs} are calculated as

$$\mathrm{SE}\left(r_y^{\mathrm{obs}}\right) = \sqrt{\ln\left\{1 + \left[\mathrm{CV}\left(R_y^{\mathrm{obs}}\right)\right]^2\right\}}$$

Re-parameterizing in terms of the 2010–2012 average,

$$\tilde{r}_y = r_y^{\mathrm{obs}} - \frac{1}{3}\left(r_{2010}^{\mathrm{obs}} + r_{2011}^{\mathrm{obs}} + r_{2012}^{\mathrm{obs}}\right)$$

The standard errors are then given by:

$$\left[\mathrm{SE}(\tilde{r}_y)\right]^2 = \begin{cases} \left[\mathrm{SE}\left(r_y^{\mathrm{obs}}\right)\right]^2 + \frac{1}{9}\left\{\left[\mathrm{SE}\left(r_{2010}^{\mathrm{obs}}\right)\right]^2 + \left[\mathrm{SE}\left(r_{2011}^{\mathrm{obs}}\right)\right]^2 + \left[\mathrm{SE}\left(r_{2012}^{\mathrm{obs}}\right)\right]^2\right\} & \text{for } y = 2009{,}2013 \\ \frac{4}{9}\left[\mathrm{SE}\left(r_y^{\mathrm{obs}}\right)\right]^2 + \frac{1}{9}\left\{\left[\mathrm{SE}\left(r_{2009}^{\mathrm{obs}}\right)\right]^2 + \left[\mathrm{SE}\left(r_{2013}^{\mathrm{obs}}\right)\right]^2\right\} & \text{for } y = 2010{,}2011{,}2012 \end{cases} \quad (3)$$

Table 1a: Small scale survey abundance estimates for areas around Robben and Dassen islands.

	Robben Island Day	Robben Island Biomass (MT)	Robben Island CV	Dassen Island Day	Dassen Island Biomass (MT)	Dassen Island CV
2009	95	1703	0.616			
	109	1004	0.650			
	124	4880	0.379			
	177	4456	0.270			
	208	16320	0.431			
	219	12996	0.417			
2010	157	64847	0.363	153	154182	0.478
	214	81621	0.322	201	146027	0.505
	245	41309	0.318	249	5179	0.459
	279	2585	0.540			
2011	67	49289	0.301	59	45381	0.235
	136	4406	0.203	130	28608	0.228
	178	48962	0.704	186	63749	0.562
	207	549	0.251	200	19172	0.500
	228	7556	0.838	222	9808	0.622
	269	657	0.350	262	3907	0.491
2012	86	41705	0.380	59	38496	0.330
	118	72711	0.350	116	20253	0.370
	199	159039	0.280	167	163258	0.340
	215	187249	0.320	206	42779	0.470
	249	31693	0.580			
2013	186	7159.16	0.322			

Table 1b: Recruit survey estimates up to Cape Infanta.

Year	Anchovy	CV
2008	1426705.18	0.202
2009	1306044.71	0.189
2010	1667994.16	0.267
2011	281260.18	0.283
2012	990378.35	0.138
2013	1164277.86	0.182

Table 2: Comparison of results from fitting different Models (see Appendix for Model descriptions).

Model	Parameters	data points	$-\ln L$	AICc	$\bar{\sigma}^{R}$	$\bar{\sigma}^{D}$	σ^{R}_{add}	σ^{D}_{add}	No. of $\bar{d}$'s	Not in Apr–Aug
Model 1	18	34	6.88	95.35	0.88	0.59	0.73	0.44	7	4
Model 2	17	34	6.90	86.04	0.87	0.60	0.74	0.44	7	3
Model 3	15	34	8.38	73.42	0.89	0.63	0.79	0.47	4	3
Model 4	14	34	8.41	66.93	0.89	0.62	0.80	0.47	4	3
Model 5	12	34	15.44	69.73	1.15	0.68	1.09	0.58	1	0
Model 6	11	34	16.28	66.57	1.19	0.70	1.12	0.60	1	0
Model 7	12	34	11.86	62.58	0.96	0.70	0.93	0.53	4	3
Model 8	10	34	18.66	66.88	1.28	0.73	1.25	0.64	1	0
Model 9	9	34	19.21	63.92	1.26	0.75	1.26	0.66	1	0
Model 10	10	35	14.37	57.90	0.92	0.90	0.85	0.77	5	4
Model 11	7	35	25.96	70.07	1.69	0.77	1.67	0.67	1	0
Model 12	6	35	27.27	69.54	1.77	0.80	1.73	0.69	1	0

Table 3: Further parameter estimates for Model 7 and Model 9.

	Model 7		Model 9	
	Estimate	Hessian s.d.	Estimate	Hessian s.d.
a^{R}	9.6	1.1	8.8	0.6
a^{D}	0.4	0.3	0.3	0.4
b_{2009}	-1.8	1.2	-2.1	0.7
b_{2010}	1.5	1.2	0.5	0.3
b_{2011}	-1.1	0.6	-0.7	0.3
b_{2012}	-0.3	0.7	0.3	0.3
d_{2009}	267.2	78.3		
d_{2010}	37.8	101.6		
d_{2011}	78.8	46.6		
d_{2012}	183.3	34.3		
d			135.1	13.2

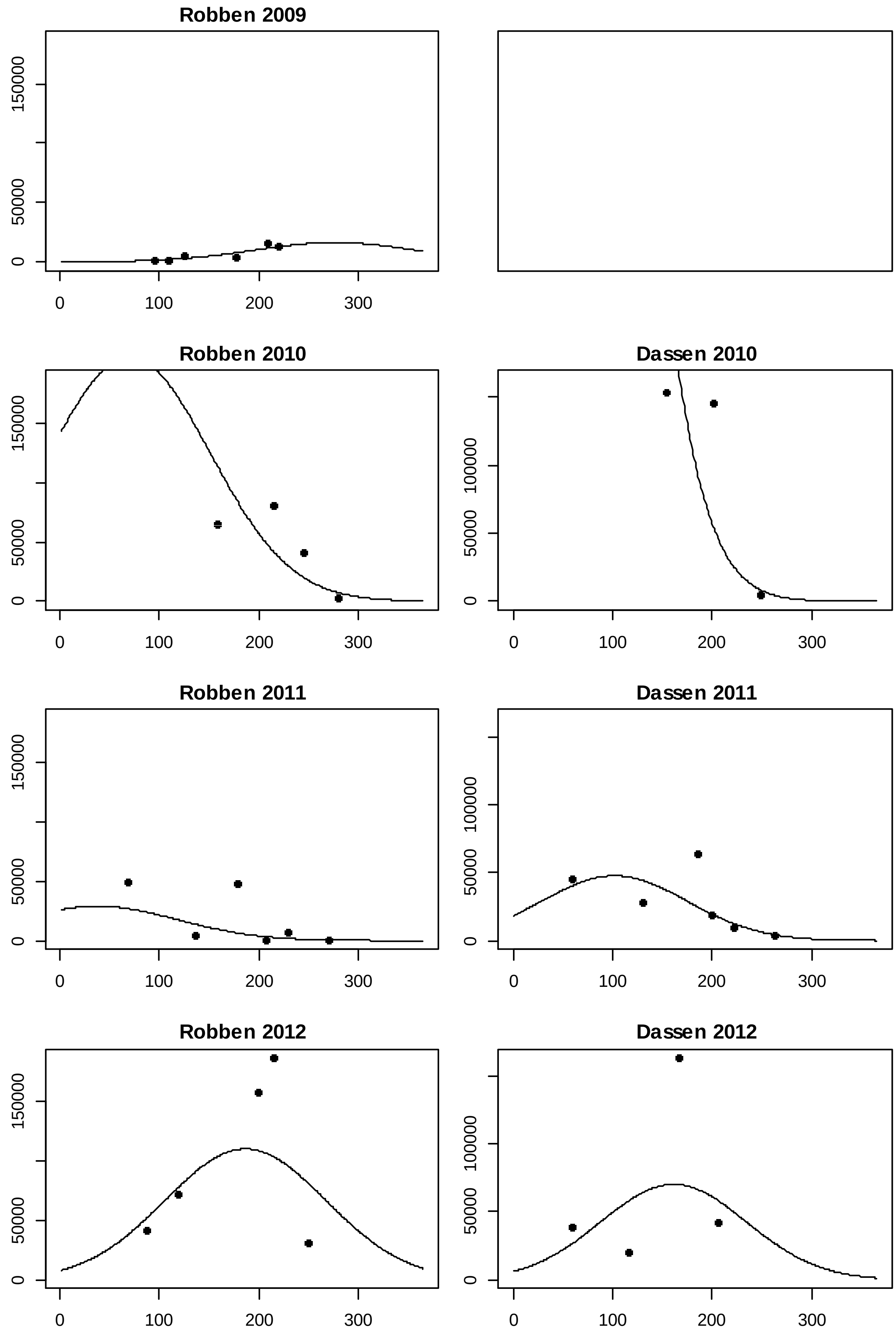

Figure 1a: Fits to the full model (all parameters estimated freely), i.e. Model 1.

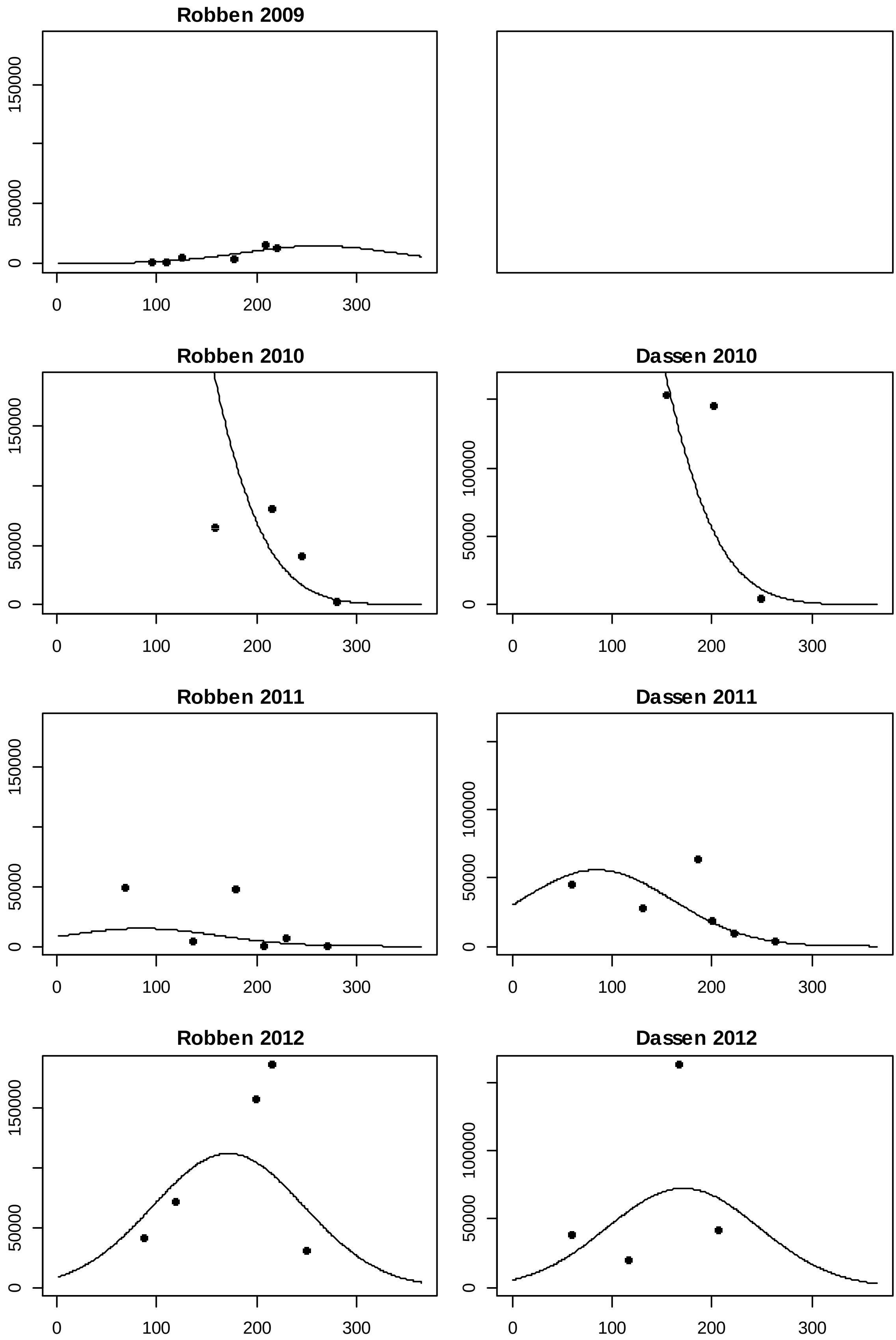

Figure 1b: Fits to Model 4 (the maximum densities at Robben and Dassen islands occur at the same time for each, though this time varies with year).

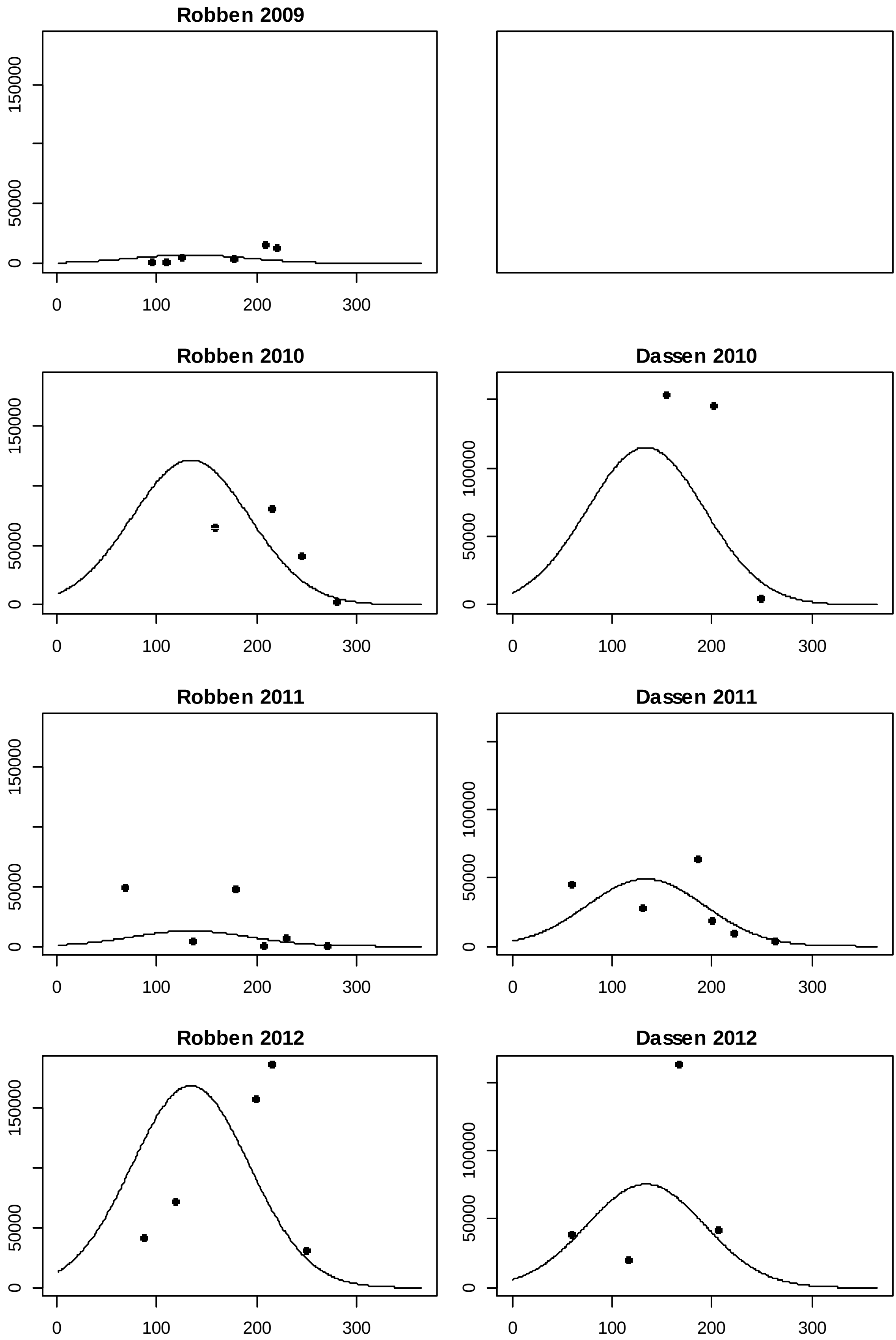

Figure 1c: Fit to Model 6 (the time at which densities at Robben and Dassen are maximal occurs on the same day every year).

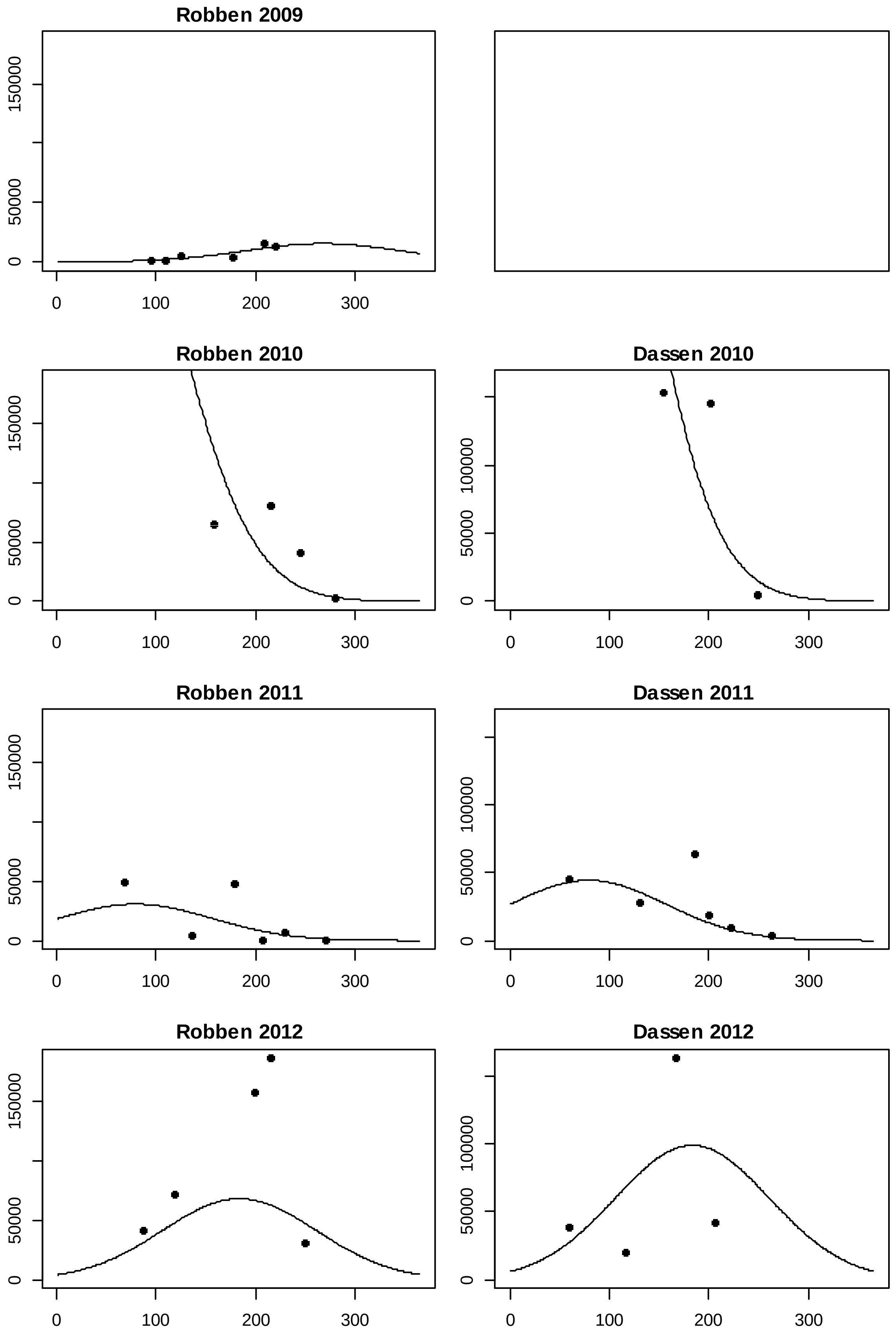

Figure 1d: Fits to Model 7 (as for Model 4, but with the same biomass difference between the models each year).

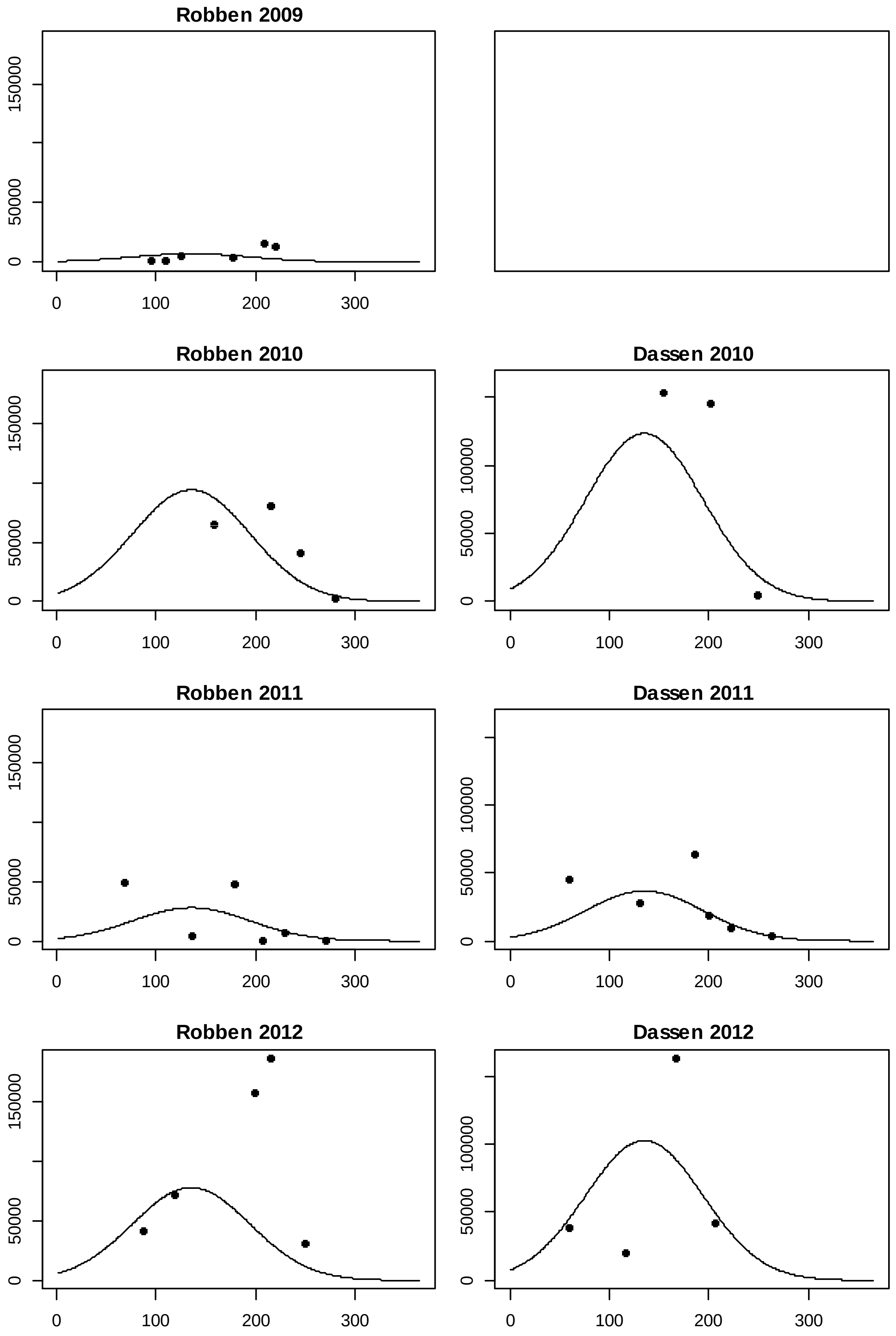

Figure 1e: Fits to Model 9 (as for Model 6, but with the same biomass difference between the models each year).

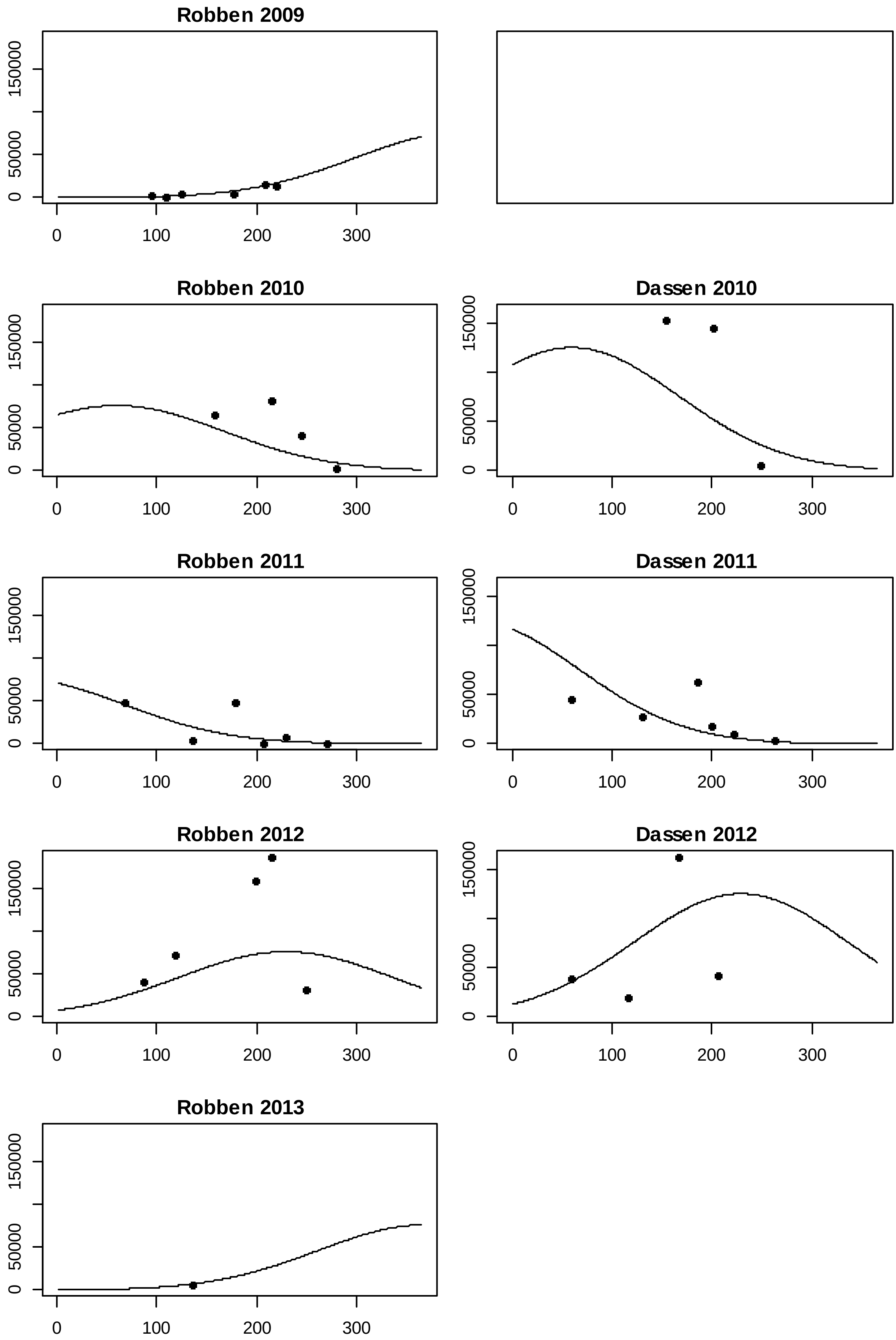

Figure 1f: Fits to Model 10 (as for Model 7, except that the biomass is the same each year at each island).

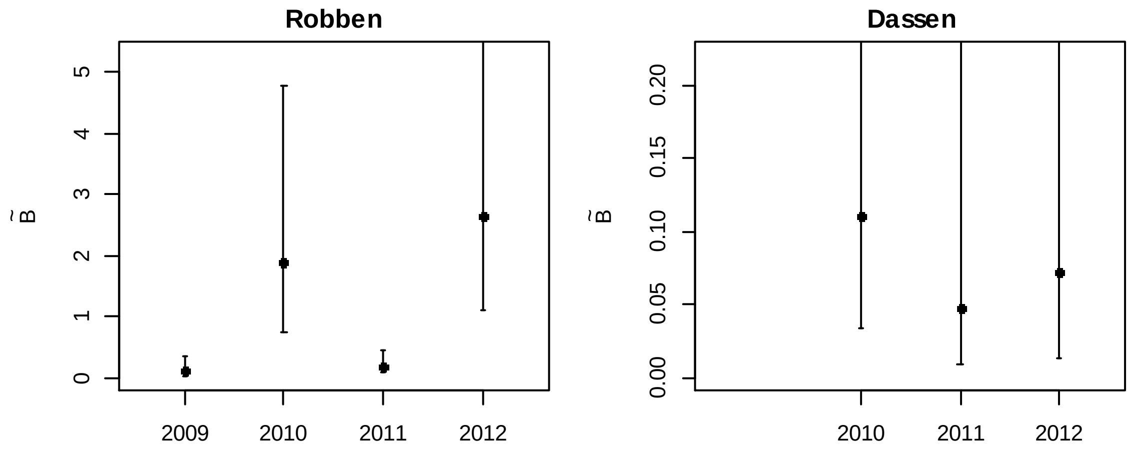

Figure 2a: Estimates of the annual abundance index $\tilde{B}$ for Robben and for Dassen islands for Model 6, renormalized to their geometric means over 2010 to 2012. The error bars here and below show 95% CIs based on the Hessian.

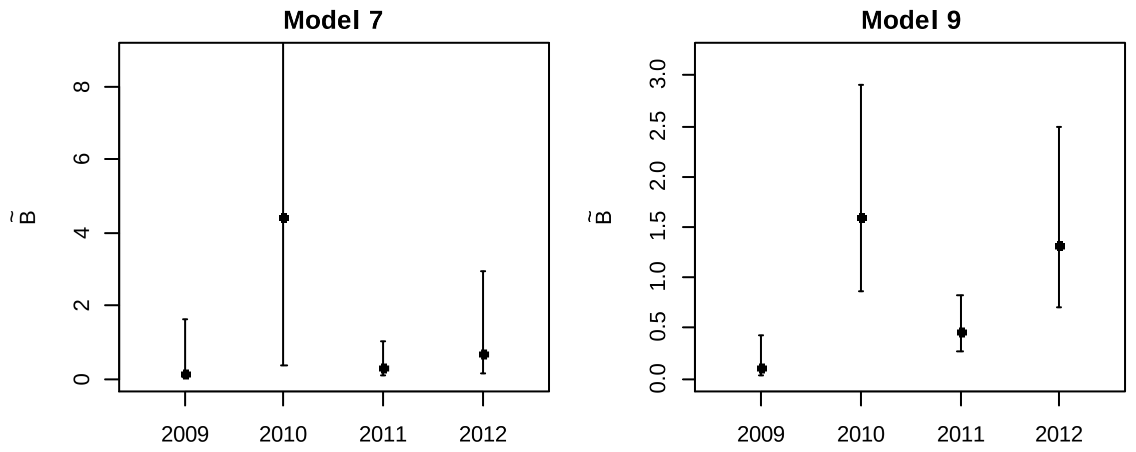

Figure 2b: Estimates of the annual abundance index $\tilde{B}$ renormalized to their geometric means over 2010 to 2012, for Model 7 and Model 9 for both of which Robben and Dassen islands share the same abundance trend.

Figure 3: May recruit survey estimates (up to Cape Infanta) $\tilde{R} = \exp(\tilde{r})$, renormalized to their geometric mean over 2010 to 2012 as detailed in the Appendix, together with 95% CIs.

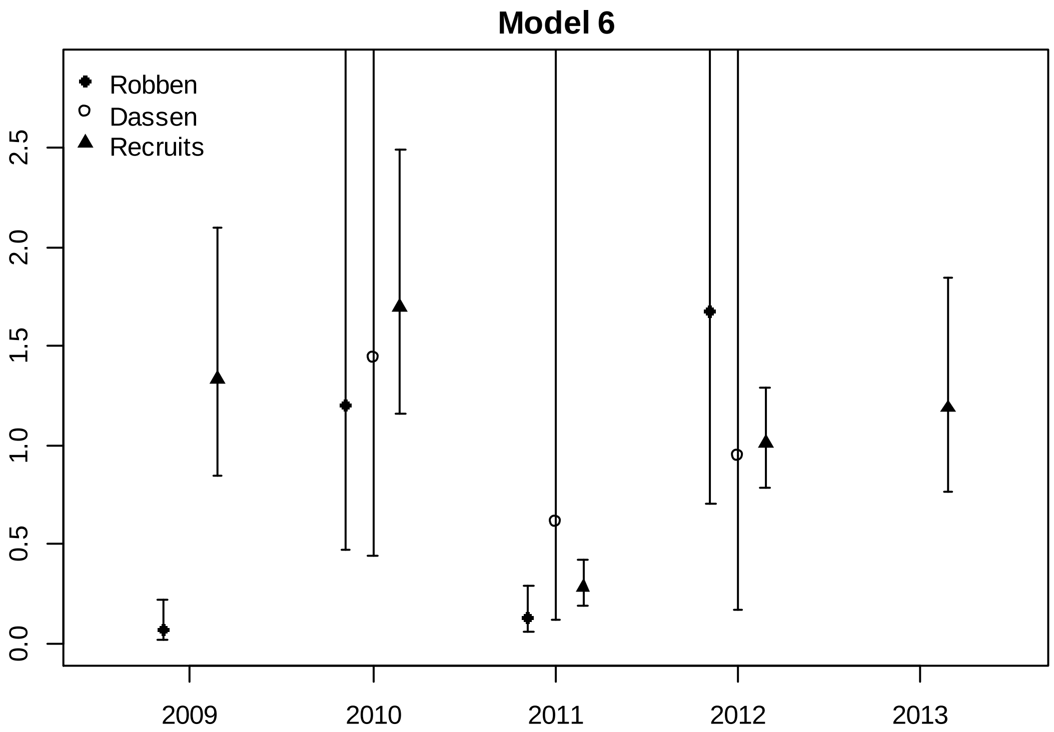

Figure 4a: Comparison of fish abundance indices at Robben and Dassen islands as estimated by the small scale surveys and the May recruit survey estimates, each normalised by their average abundance over 2010–2012.

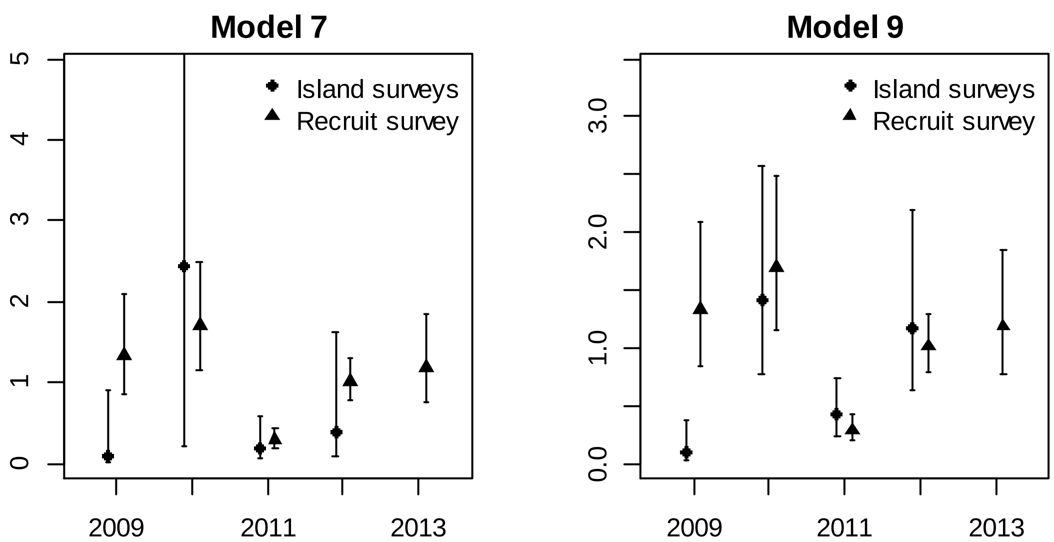

Figure 4b: Comparison of the fish abundance indices for Robben and Dassen islands (here assumed to have the same trend) as estimated by the small scale surveys $(\tilde{B})$ and the May recruit survey estimates $(\tilde{R})$, each normalised by their geometric means over 2010–2012, together with 95% CIs.