CCSBT-ESC/0909/12

The aerial survey index of abundance: updated analysis methods and results

Paige Eveson
Jessica Farley
Mark Bravington

Prepared for the CCSBT 14th Meeting of the Extended Scientific Committee 5-11 September 2009, Busan, Korea

Table of Contents

Abstract 1
Introduction 1
Scientific aerial survey 2
  Field procedures 2
  Data preparation 3
  Search effort and SBT sightings 3
  Environmental variables 8
  Methods of analysis 10
  Results 11
Calibration experiment 13
  Field procedures 13
  Preliminary analysis and results 13
Summary 16
References 16
Acknowledgements 16
Appendix A – Methods of analysis 17
  Biomass per sighting (BpS) model 18
  Sightings per mile (SpM) model 19
  Combined analysis 20
Appendix B – CV calculations 21
Appendix C: Results and diagnostics 23
  Biomass per sighting (BpS) model 23
  Sightings per mile (SpM) model 27
  Evaluation of random effects 31

Abstract

The preliminary point estimate from the 2009 scientific aerial survey is lower than the 2008 estimate, and similar to the 2006 and 2007 estimates. Taking confidence intervals into account, the relative abundance estimates have remained similar since 2005, and are significantly lower than the average level in the mid-1990s (1993-1996).

The models used to analyse the aerial survey data in 2008 included a random effect for the 3-way interaction term between year, month and area. This year, the models were modified to include random effects for the 2-way interaction terms between year, month and area as well. This led to more stable model fits, as there are a number of 2-way strata for which little (or sometimes no) data exist, and the revised models accommodate these situations more effectively. The results obtained using the revised models are not significantly different to those obtained using the previous models.

In preparation for the contingency that future scientific aerial surveys may have only one observer in a plane, large-scale calibration experiments were undertaken between 2007 and 2009, with the primary purpose of comparing the number of sightings and estimated total biomass of SBT seen by the calibration plane (with one spotter) versus the survey plane (with two spotters). Analysis of the data show that, on average, the calibration plane made approximately half as many sightings as the survey plane. There was no significant difference in the results between years; however, there was a significant difference between spotter/spotter-pilot combinations (i.e., when the two dedicated spotters were swapped between the survey and calibration planes). Methods for using these results in future to "correct" the scientific survey analysis for having only one spotter are currently being explored.

Introduction

The index of juvenile Southern Bluefin Tuna (SBT) abundance based on a scientific aerial survey in the Great Australian Bight (GAB) is one of the few fishery-independent indices available for monitoring and assessment of the SBT stock. The aerial survey was conducted in the GAB between 1991 and 2000, but was suspended in 2001 due to logistic problems of finding trained, experienced spotters and spotter-pilots. (Note that the terms 'spotter' and 'observer' are used interchangeably). The suspension also allowed for further data analysis and an evaluation of the effectiveness of the survey. A decision to continue or end the scientific aerial survey could then be made on the merits of the data, in particular the ability to detect changes in abundance.

Analysis of the data was completed in 2003 and it showed that the scientific aerial survey does provide a suitable indicator of SBT abundance in the GAB (Bravington 2003). In the light of serious concerns about the reliability of historic and current catch and CPUE data and weak year classes in the late 1990s and early 2000s, this fishery-independent index is even more important (Anon 2008).

In 2005, the full scientific line-transect aerial survey was re-established in the GAB, and this survey has been conducted each year since. New analysis methods were developed, and all data were re-analysed. Based on these analyses, an index of abundance was constructed for the surveys in 1993-2000 and in 2005-2009.

In addition, in 2007 a large-scale calibration experiment was initiated with the primary purpose of comparing SBT sighting rates by one observer versus two observers in a plane (in light of the fact that future surveys may have only one observer in a plane). The data provided useful information about differences in sightings between observers (e.g., sightings made by one observer are often missed by another observer). However, it proved difficult to definitively estimate the effect of the number of observers on the index.

In 2008 and 2009 a new calibration experiment was designed and run in parallel with the full scientific aerial survey. This calibration experiment was designed to compare:

- the number of SBT sightings;
- and total estimated biomass of SBT observed;

by the single spotter plane versus the survey plane (with two spotters) over the same area and time strata.

This report summarises the field procedures and data collected during the 2009 season. The most recent methods for analysing the aerial survey data are described, and results are presented. Analysis of the calibration data is still underway, but preliminary results are presented.

Scientific aerial survey

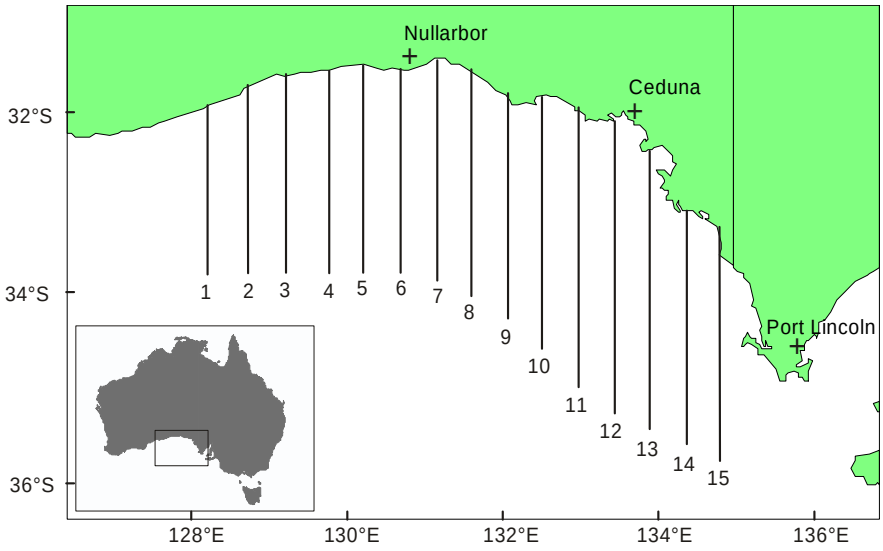
Field procedures

The 2009 scientific aerial survey was conducted in the GAB between 1 January and 31 March 2009. As for previous surveys (e.g. Eveson et al. 2008), the plane used was a Rockwell Aero Commander 500S. Three observers were employed for the survey, one spotter-pilot (who participated in each flight) and two spotters who alternated between the survey and calibration planes. Each spotter participated in approximately half of the line transect survey flights during the time that the calibration experiment was conducted (January and February – see Calibration Experiment section below) and only one spotter participated in the survey plane in March. The same observers employed for the 2007 and 2008 surveys were used throughout the 2009 survey.

The survey followed the protocols established for the 2000 survey (Cowling 2000) and used in all subsequent surveys, with respect to the area searched, plane height and speed, minimum environmental conditions, time of day the survey lines were flown, and data recording protocols. Fifteen north-south transect lines (Fig. 1) were surveyed. A complete replicate of the GAB consists of a subset of 12 (of the 15) lines divided into 4 blocks. The remaining 3 lines in a replicate (either: 1, 3 and 14, or 2, 13 and 15) were not searched, as SBT abundance is historically low in these areas and surveying a subset increases the number of complete replicate of the GAB in the survey. When flying along a line, each observer searched the sea surface from straight ahead through to 90° to their side of the plane (abeam of the plane) for surface patches (schools) of SBT. In the survey plane (i.e. a spotter pilot and spotter), a spotter would occasionally search both sides of the plane, if the other observer was temporarily unable to observe.

The 2009 field operation was very successful, largely due to the prevailing whether conditions in January and March and the availability of two planes. Five replicates of the GAB were completed, compared to 4.5 in 2008, and between 3.0 and 3.5 in the previous 3 years. The total flying time (transit and transect time) for the 2009 survey was 153.4 hours.

Figure 1. Location of the 15 north-south transect lines for the scientific aerial survey in the GAB.



Data preparation

The data collected from the 2009 survey were loaded into the aerial survey database and checked for any obvious errors or inconsistencies and corrections made as necessary.

In order for the analyses to be comparable between all survey years, only data collected in a similar manner from a common area were included in the data summaries and analyses presented in this report. In particular, only search effort and sightings made along north/south transect lines in the unextended (pre-1999) survey area and sightings made within 6 nm of a transect line were included (see Basson et al. 2005 for details). In cases where a sighting consisted of more than one school, then the sighting was included if at least one of the schools was within 6 nm of the line. We excluded secondary sightings and any search distance and sightings made during the aborted section of a transect line (see Eveson et al. 2006 for details).

Search effort and SBT sightings

A summary of the total search effort and SBT sightings made in each survey year is given in Table 1. This table and all summary information and results presented in this report, include only the data outlined in the previous section as being appropriate for analysis.

Table 1. Summary of aerial survey data by survey year. Only data considered suitable for analysis (as outlined in text) are included. All biomass statistics are in tonnes.[1]

Survey year	Total distance searched (nm)	Number SBT sightings	Total biomass	Average patches per sighting	Max patches per sighting	Average biomass per patch	Median biomass per patch	Max biomass per patch
1993	7603	130	12215	3.9	76	24.3	18.8	203
1994	15180	174	14967	3.3	23	26.3	21.5	245
1995	14573	179	21902	3.6	38	34.4	27.9	224
1996	12284	116	16499	4.1	46	34.6	27.4	147
1997	8813	117	9848	3.0	18	27.7	22.4	199
1998	8550	109	10270	2.3	21	40.4	20.3	948
1999	7555	56	3015	2.4	21	22.8	16.5	120
2000	6775	77	4797	2.6	17	23.9	19.9	100
2005	5968	80	5982	2.4	17	31.2	24.9	190
2006	5150	44	3954	2.0	8	46.0	30.9	263
2007	4872	42	3445	2.6	11	32.2	24.3	119
2008	7462	122	7952	3.5	24	18.8	13.1	308
2009	8101	154	8522	2.6	22	21.7	14.3	168

[1] The biomass statistics differ slightly from those reported in Table 3.1 of the 2008 Final Report to DAFF because the patch size estimates used in calculating these statistics have been corrected for differences between observers (see Appendix A, section 8.1). Observer differences are re-estimated each year using all available data and thus the corrected patch size estimates can change slightly.

The total distance searched increased slightly in 2009 compared to 2008, and was much greater than in 2005-2007. As noted above, this was primarily due to good flying conditions in January and March. Similar to 2008, the sightings rate (number of sightings per 100 nm) was higher than in previous survey years (Fig. 2), but the amount of biomass per nm was not as high as in early years (Fig. 2) because the median and mean biomass per patch was much smaller (Table 2; Fig. 3).

Also similar to 2008, almost all SBT sightings in 2009 were made in the eastern half of the survey area, with the distribution of sightings clustered particularly along transect lines 9-10 (Fig. 4). In addition, there were more sightings along the farthest east transect lines than in previous years.

Figure 2. Plots of a) total distance searched (i.e. effort) by year; b) biomass per mile by year; c) number of sightings per 100 miles by year. ***Note that these plots are based on raw data, which has not been corrected for environmental factors or observer effects.***

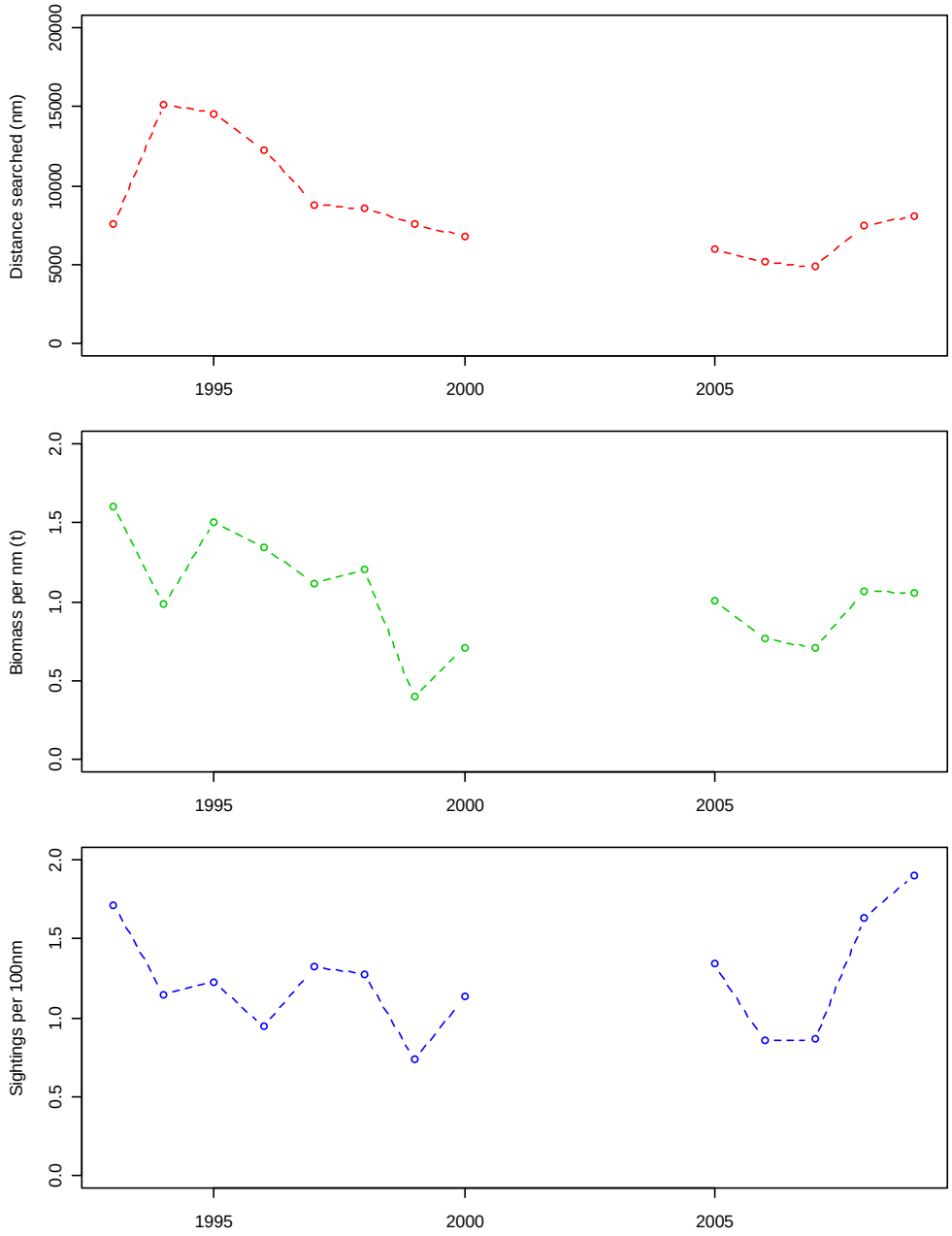


Figure 3. Frequency of SBT patch sizes (in tonnes) by survey year (excluding 1993 for display purposes).

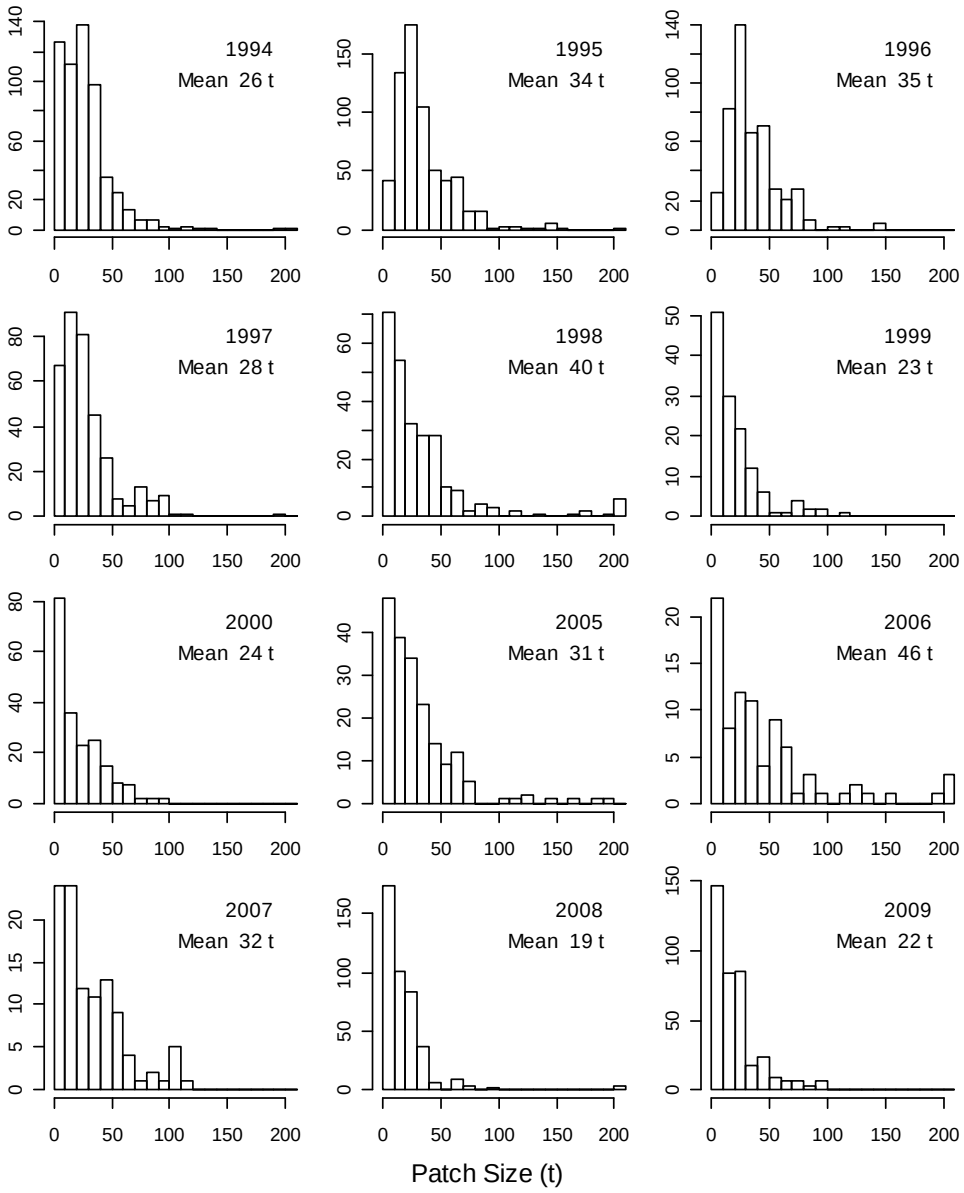
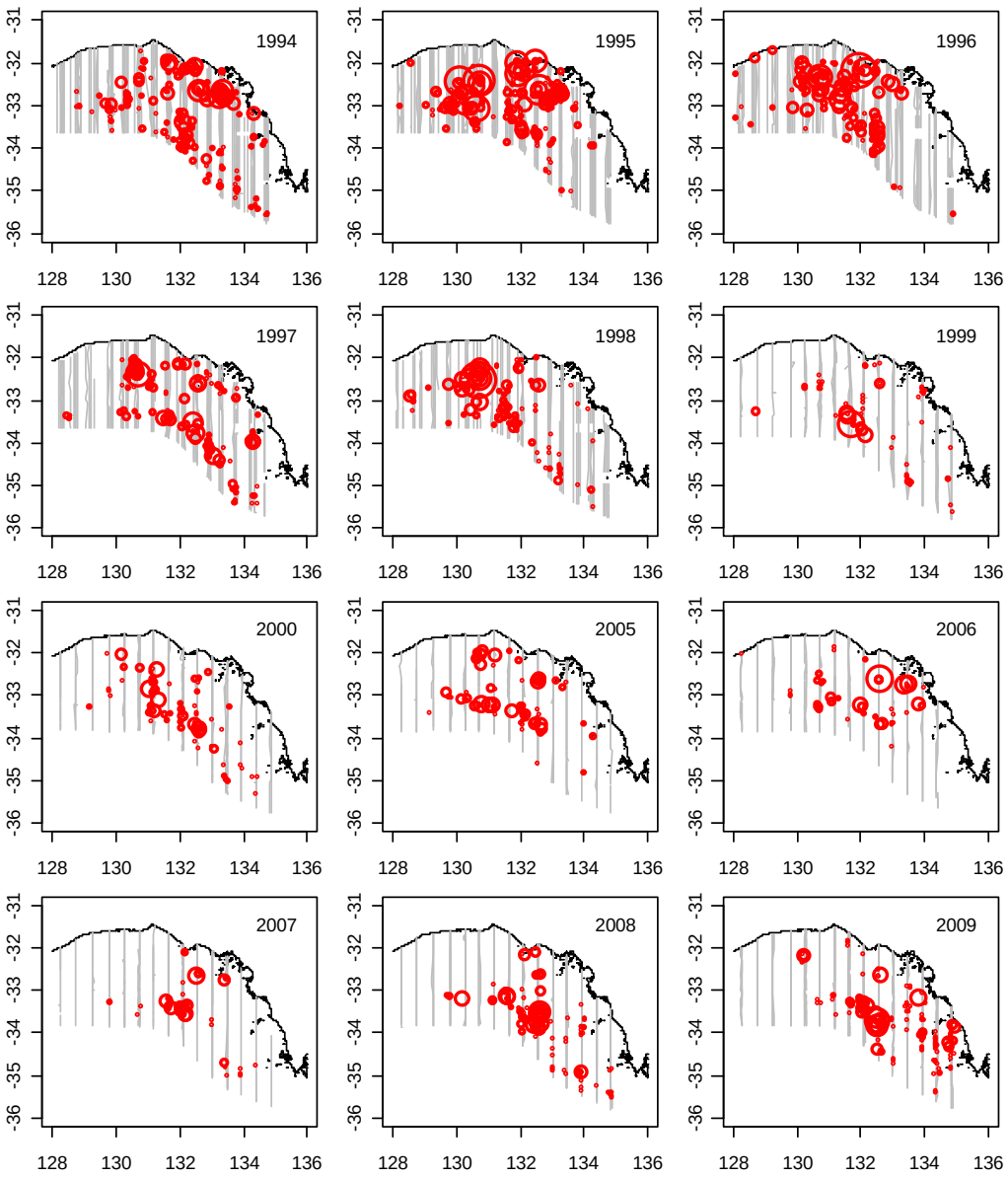


Figure 4. Distribution of SBT sightings made during each aerial survey year (excluding 1993 for display purposes). Red circles show the locations of SBT sightings, where the size of the circle is proportional to the size of the sighting, and grey lines show the north/south transect lines that were searched.



Environmental variables

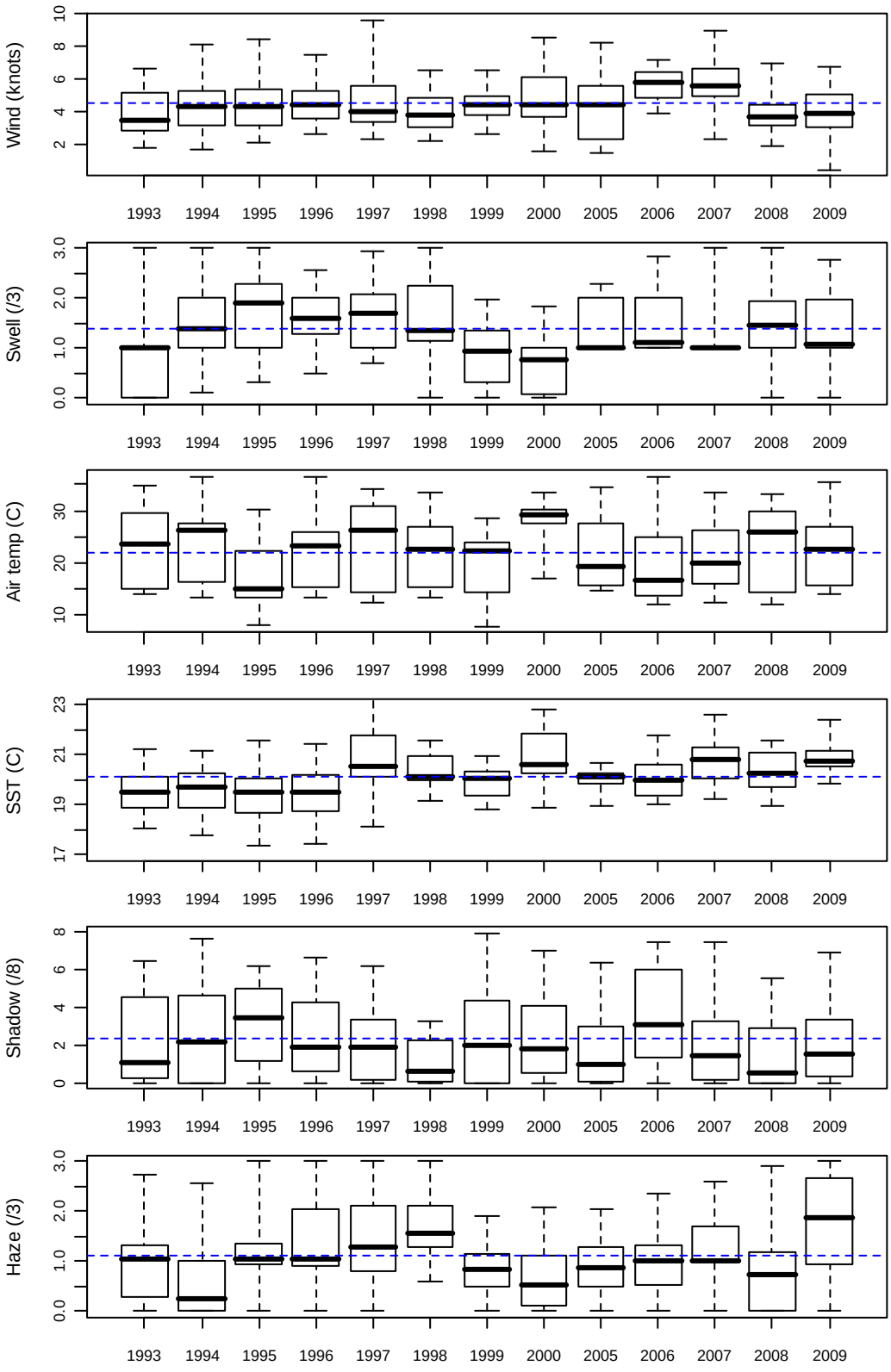
Table 2 and Figure 5 summarize the environmental conditions that were present during valid search effort in each survey year. All the environmental variables presented were recorded by the survey plane, with the exception of sea surface temperature (SST), which was extracted from the 3-day composite SST dataset produced by CSIRO Marine and Atmospheric Research's Remote Sensing Project (see Eveson et al. 2006 for more details).

Similar to 2008, the wind speed during the 2009 survey was lower on average compared to other survey years (Table 2; Fig. 5). SST was higher than average and less variable than in previous years, whereas the amount of haze tended to be higher (Table 2; Fig. 5). Sightings rates tend to be higher with low wind speed and high SST, so overall the conditions were above average in 2009 for making sightings.

Table 2. Average environmental conditions during search effort for each aerial survey year.

Survey year	Wind speed (knots)	Swell height (0-3)	Air temp (°C)	SST (°C)	Sea shadow (0-8)	Haze (0-3)
1993	3.9	0.8	24.4	19.6	1.8	0.9
1994	4.1	1.5	20.6	19.7	2.7	0.5
1995	4.4	1.7	18.7	19.6	2.7	1.1
1996	4.5	1.6	22.9	19.6	2.1	1.2
1997	4.1	1.7	25.3	21.1	1.6	1.3
1998	3.7	1.7	22.3	20.4	0.9	1.7
1999	4.1	0.9	22.0	19.9	2.9	0.7
2000	4.3	0.6	27.5	20.7	2.6	0.7
2005	4.7	1.5	21.7	19.8	1.6	0.8
2006	5.6	1.5	20.0	19.9	3.5	0.9
2007	5.8	1.3	21.6	20.8	2.0	1.3
2008	3.8	1.4	24.2	20.4	1.4	0.9
2009	3.8	1.4	22.2	21.0	2.1	1.7

Figure 5. Boxplots summarizing the environmental conditions present during valid search effort for each aerial survey year. The thick horizontal band through a box indicates the median, the length of a box represents the inter-quartile range, and the vertical lines extend to the minimum and maximum values. The dashed blue line running across each plot shows the overall average across all survey years.



Methods of analysis

In 2008, we were successful in implementing random effects versions of the models (Eveson et al. 2008). We continue to use random effects models this year, but with some amendments. Here, we give a brief description highlighting how the models have changed. Details of the analysis methods (which, for the most part, are the same as last year) can be found in Appendix A.

As in previous years, we fit generalized linear models to two different components of observed biomass—biomass per sighting (BpS) and sightings per nautical mile of transect line (SpM). We included the same environmental and observer variables in the models this year as in the past several years. Thus, the models can still be expressed as:

$$\log\mathbf{E}(\text{Biomass}) \sim \text{Year*Month*Area} + \text{SST} + \text{WindSpeed}$$

$$\log\mathbf{E}(\text{N\_sightings}) \sim \text{offset}(\log(\text{Distance})) + \text{Year*Month*Area} + \log(\text{ObsEffect}) + \text{SST} + \text{WindSpeed} + \text{Swell} + \text{Haze} + \text{MoonPhase}$$

Note that the term Year*Month*Area encompasses all 1-way, 2-way and 3-way interactions between Year, Month and Area (i.e., it is equivalent to writing Year + Month + Area + Year:Month + Year:Area + Month:Area + Year:Month:Area).

In 2008, we fit the 3-way Year:Month:Area interaction term as a random effect in both models, whereas the 1-way and 2-ways effects were fit as fixed effects. This year, we modified the models to treat the 2-way interaction terms as random effects as well. This decision was made because many of the 2-way strata have very few observations and were causing instabilities in the model fits when treated as fixed effects. Random effect models allow a coherent approach to inference even when some strata are data-poor. Estimates in data-rich strata are driven by the data themselves, but in strata where data are lacking – say a particular year and month – the estimate is adjusted towards the value that would be expected based on (i) other months in that year, and (ii) that month in other years. The extent of the adjustment depends on the amount of data available. A key advantage of using random effect models to make such adjustments (rather using an *ad hoc* approach) is that the uncertainty associated with the adjustments is correctly propagated through to the final abundance estimate.

Once the new models were fitted, the analysis proceeded as before. Specifically, the SpM and BpS model results were used to predict what the number of sightings per mile and the average biomass per sighting in each of the 45 area/month strata in each survey year would have been under standardized environmental/observer conditions. Using these predicted values, we calculated an abundance estimate for each stratum as 'standardized SpM' multiplied by 'standardized average BpS'. We then took the weighted sum of the stratum-specific abundance estimates over all area/month strata within a year, where each estimate was weighted by the geographical size of the stratum in nm^2, to get an overall abundance estimate for that year. Lastly, the annual estimates were divided by their mean to get a time series of relative abundance indices.

We emphasise that it is important to have not only an estimate of the relative abundance index in each year, but also of the uncertainty in the estimates. We used the same process as last year to calculate CVs for the indices, the details of which are repeated in Appendix B. Briefly, we first obtained standard errors (SEs) for the predicted values of 'standardized SpM' and 'standardized average BpS' in each year/area/month stratum. These were used to calculate SEs for the stratum-specific abundance estimates, which were in turn used to calculate SEs for the annual abundance estimates. Lastly, we applied the delta method to

determine SEs for the relative abundance indices. Note that CVs are given simply by dividing the SE of each index estimate by the estimate. We calculated confidence intervals for the indices based on the assumption that the logarithm of the indices follows a normal distribution, with standard errors approximated by the CVs of the untransformed indices

Results

Some results and diagnostics for the BpS and SpM model fits are provided in Appendix C. They suggest that the models fit reasonably well and that the random effects are behaving sensibly.

Figure 6 shows the estimated time series of relative abundance indices with 90% confidence intervals. The point estimates and CVs corresponding to Figure 6 are reproduced in Table 3. The point estimate for 2009 is lower than the 2008 estimate, and similar to the 2006 and 2007 estimates. Taking confidence intervals into account, the relative abundance estimates have remained similar since 2005, and are significantly lower than the average level in the period 1993-1996.

Figure 7 compares the results obtained this year (using models with 2-way and 3-way random effects) with those obtained last year (using models with only 3-way random effects). For most years the point estimates and confidence intervals are very similar, and the overall trends and conclusions remain the same.

Figure 6. Time series of relative abundance estimates with 90% confidence intervals.

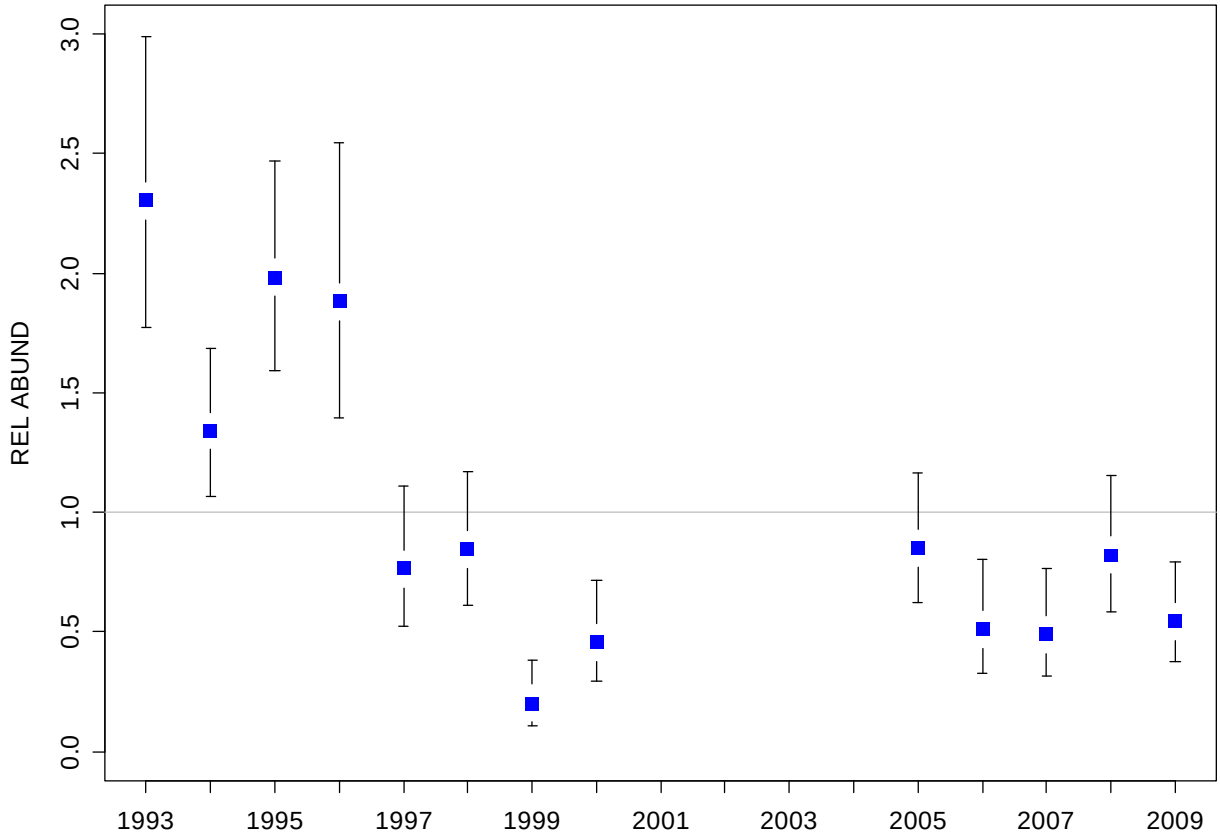
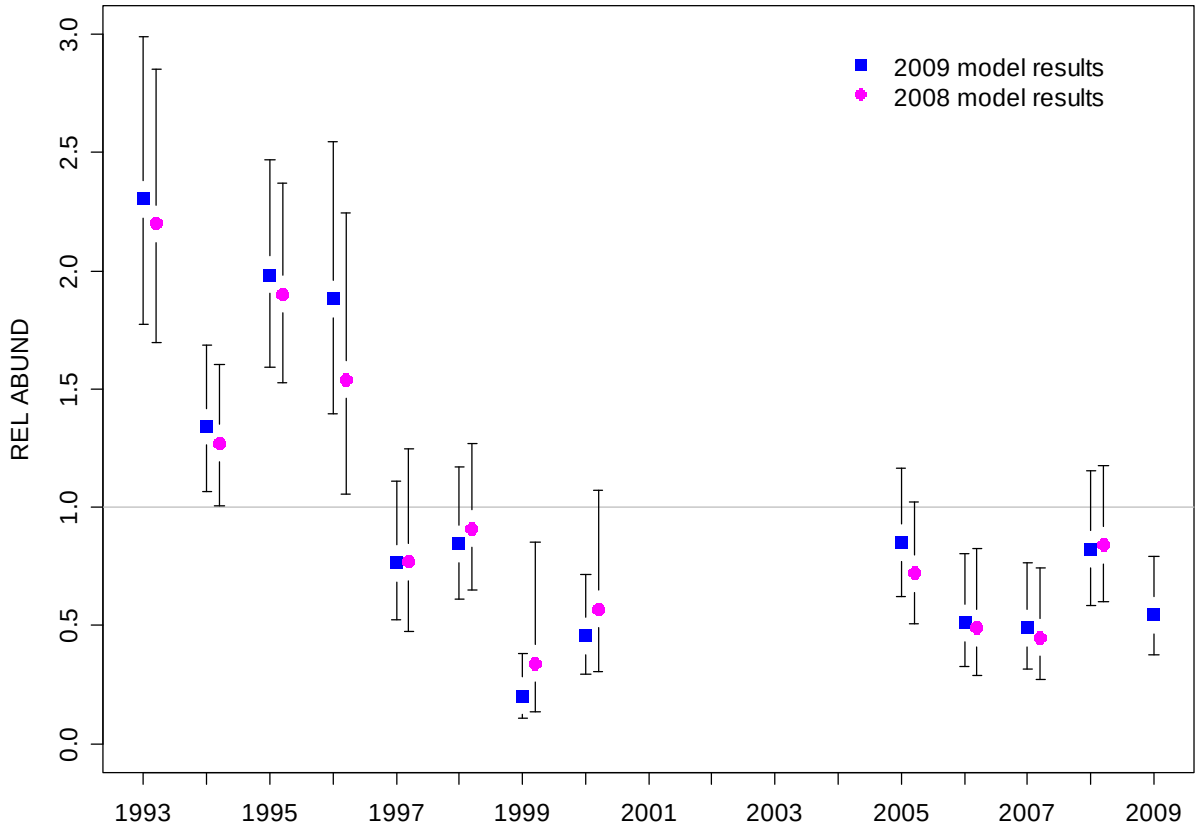


Table 3. Aerial survey relative abundance point estimates, standard errors, CVs, and lower and upper 90% confidence intervals (CI.05 and CI.95 respectively).

Year	Index	SE	CV	CI.05	CI.95
1993	2.30	0.37	15.9%	1.77	2.99
1994	1.34	0.19	14.0%	1.07	1.69
1995	1.98	0.26	13.3%	1.59	2.47
1996	1.88	0.34	18.3%	1.39	2.54
1997	0.76	0.17	22.8%	0.53	1.11
1998	0.85	0.17	19.6%	0.61	1.17
1999	0.20	0.08	38.9%	0.11	0.38
2000	0.46	0.12	27.3%	0.29	0.72
2005	0.85	0.16	19.0%	0.62	1.16
2006	0.51	0.14	27.5%	0.33	0.81
2007	0.49	0.13	26.8%	0.32	0.76
2008	0.82	0.17	20.7%	0.58	1.15
2009	0.54	0.12	22.7%	0.37	0.79

Figure 7. Comparison of relative abundance estimates and 90% confidence intervals obtained using the new (2009) random effects models versus the models used in last year's analysis.



Calibration experiment

Field procedures

The second component of the 2009 program was to undertake a calibration experiment in parallel with the scientific line-transect aerial survey. The design and primary goal of the 2009 survey were the same as for 2008 (Eveson et al. 2008). In brief, the studies were designed to compare the total number of SBT sightings and total estimated biomass of SBT seen by the calibration plane versus the survey plane over similar areas and time periods (as opposed to trying to "match up" individual sightings, as in the initial 2007 calibration experiment; Eveson et al. 2007). The experiments also provided additional data on observer variability, which may be able to help reduce this source of uncertainty in the aerial survey indices.

The calibration study was conducted in January and February 2009. A second plane (also a Rockwell Aero Commander) was chartered and a non-spotting pilot was employed. The experiment involved the second plane with one observer (and a non-spotting pilot) flying in tandem with the survey plane. Radio communication was maintained throughout the study to monitor the distance between planes and to communicate start/stop (see below). However, communication about SBT sightings did not occur at any time.

As mentioned in above, two dedicated spotters were swapped between the calibration and survey planes in January and February. The calibration plane followed the same protocols used by the survey plane, except the sole observer searched for patches of SBT from their side of the plane through 180° to the other side of the plane. This allowed for a direct comparison of the number of sightings and biomass recorded. The calibration plane recorded sightings as it flew along the transect lines either ahead or behind the dedicated survey plane. Both planes left the line as required to investigate each sighting and to estimate the number of patches and biomass in the sighting. The distance between the two planes varied depending on the situation. For example, if it was expected that the day was going to be 'long' (if the far western block was being surveyed, or it was expected that a larger number of SBT sightings could be encountered – given the weather conditions and location), then the two planes would take off at the same time but start on different lines. Once these lines were completed, the planes would swap lines. The lines were surveyed in the same direction approximately 1 hour apart. If a third line could be flown, it was flown in tandem with ~1 hr between planes. On 'short' days, the planes would take off approximately 30 minutes apart with the second plane following the first plane all day, remaining approximately the same time/distance apart. If the survey and calibration planes were too close to each other, the leading plane would instruct the other plane to stop searching and circle on the spot until given the signal to start searching again. In practice, this occurred very rarely.

Preliminary analysis and results

For each day of the calibration study in 2008 and 2009, we compared the number of SBT sightings and total biomass estimates made by the two planes (Table 4). In doing so, we only included data from sections of lines that were flown by both planes. The summary figures suggest that, with the exception of one day in 2009, the calibration plane consistently made fewer SBT sightings (and estimated less total biomass) than the survey plane. This was true regardless of which individual spotter was in the calibration plane (Table 4). Note that because the planes flew the same lines within a short time of one another, we do not need to

account for environmental conditions when comparing the number of sightings and biomass estimates between the two planes.

Table 4. Number of sightings and total biomass estimates (in tonnes) made by the survey plane versus the calibration plane during each day of the calibration experiment. Data are only included for line segments searched by both planes on the same day.

Year	Month	Day	Calib. spotter	Survey # sightings	Calib. # sightings	**Diff. in # sightings**	Survey biomass	Calib. biomass	**Diff. in biomass**
2008	1	10	50	27	17	**10**	3680	1879	**1801**
2008	1	12	3	0	0	**0**	0	0	**0**
2008	1	26	3	3	2	**1**	420	74	**346**
2008	1	27	3	0	0	**0**	0	0	**0**
2008	1	28	3	0	0	**0**	0	0	**0**
2008	1	4	50	0	0	**0**	0	0	**0**
2008	1	9	3	1	1	**0**	92	62	**30**
2008	2	16	3	1	0	**1**	0	0	**0**
2008	2	17	3	30	8	**22**	1276	846	**430**
2008	2	18	3	5	4	**1**	1	23	**-22**
2008	2	24	50	1	0	**1**	0	0	**0**
2008	2	26	50	0	0	**0**	0	0	**0**
2009	1	11	3	3	1	**2**	33	12	**21**
2009	1	18	3	0	0	**0**	0	0	**0**
2009	1	19	3	14	1	**13**	6	25	**-19**
2009	1	21	3	4	2	**2**	396	343	**53**
2009	1	24	50	4	2	**2**	270	36	**234**
2009	1	28	3	13	9	**4**	786	277	**509**
2009	1	6	3	0	0	**0**	0	0	**0**
2009	2	17	50	0	0	**0**	0	0	**0**
2009	2	18	50	14	7	**7**	141	53	**88**
2009	2	22	50	0	0	**0**	0	0	**0**
2009	2	25	50	4	6	**-2**	4	420	**-416**
2009	2	26	50	3	3	**0**	13	4	**9**
		Total:		127	63	**64**	7118	4054	**3064**

The number of patches per sighting and biomass per patch were, on average, estimated to be quite similar by the two planes in each of the years (Table 5). This fits our expectation because there is no reason for an observer's ability to estimate the number and size of patches to be affected by whether he is flying in the calibration plane or the survey plane.

Table 5. Comparison of summary statistics by year for calibration plane versus survey plane. Data are only included for line segments searched by both planes on the same day.

Year	Plane	Total distance searched (nm)	Number SBT sightings	Total biomass	Average patches per sighting	Average biomass per patch	Median biomass per patch
2008	Survey	3815	68	5428	3.5	23.3	21.8
	Calibrat.	3813	32	2777	3.3	27.2	18.4
2009	Survey	3524	59	1653	1.8	16.2	9.5
	Calibrat.	3514	31	1169	2.0	19.2	12.4

Although the summary statistics show that the calibration plane made fewer sightings than the survey plane, we require a statistical model to quantify this difference. Thus, we created a dataset with the number of sightings made by the calibration and survey planes in every year, month, day, area and transect line combination, which we refer to as a "sample unit". We then fit a generalized linear model to the number of sightings assuming an overdispersed Poisson error structure with a log link, and including "plane" (either survey or calibration) and "sample unit" as explanatory variables; i.e.,

$$\log\mathbf{E}(\text{N\_sightings}) \sim \text{Plane} + \text{SampleUnit}$$

Sample unit was fit as a random effect.

The factor "plane" came out as highly significant (p-value=0.000), confirming that the number of sightings differs significantly between the calibration plane and survey plane. An estimate of the average factor by which the number of sightings made by the calibration plane differs from the survey plane can be calculated using the estimated coefficients for "plane = calibration' and "plane = survey" as follows:

$$\exp(\text{coef\_calibration})/\exp(\text{coef\_survey}) = 0.496$$

Thus, on average, the calibration plane is expected to make approximately half as many sightings as the survey plane. We'll refer to this as the "calibration factor".

To see if the difference was consistent between years, we refit the model including year as a factor:

$$\log\mathbf{E}(\text{N\_sightings}) \sim \text{Plane}*\text{Year} + \text{SampleUnit}$$

The interaction between "plane" and "year" came out as insignificant (p-value=0.48). Furthermore, we calculated calibration factors for 2008 and 2009 and found them to be very similar (0.471 and 0.525, respectively).

We also wanted to see if the difference depended on which of the two dedicated spotters was in the calibration plane. Thus, we refit the model again including "spotter set" as a factor:

$$\log\mathbf{E}(\text{N\_sightings}) \sim \text{Plane}*\text{SpotterSet} + \text{SampleUnit}$$

where "spotter set" takes on two levels – either level 1 when dedicated spotter 1 is in the calibration plane and dedicated spotter 2 is in the survey plane, or level 2 when the dedicated spotters are reversed. In this case, the interaction between "plane" and "spotter set" came out as significant (p-value=0.0003), suggesting that the spotter does affect the results. We calculated the calibration factors for spotter set 1 and spotter set 2 to be 0.660 and 0.378, respectively.

If future surveys only have one spotter in the survey plane, then we need to be able to "correct" the SpM model for the fact that the number of sightings is expected to be fewer with only one spotter than with two spotters. Otherwise the results from future surveys will not be comparable to those from past surveys. We are currently investigating ways to use the above model results in this regard. Recall that for the SpM model, we first run a pair-wise observer analysis to estimate the relative sighting abilities of all observer pairs that have been involved in past and present surveys (see Appendix A). We may be able to estimate a relative sighting ability for a solo spotter based on the relative sighting ability estimates from when he flew as part of a pair (in past surveys), multiplied by his estimated calibration factor. For example, dedicated spotter 1 has flown as part of two different observer pairs in past and present surveys, with relative sighting ability estimates of 0.97 and 1.0. If we take the

average of these two relative sighting ability estimates and multiply it by the calibration factor of 0.660 for this spotter, this gives a relative sighting ability estimate for this spotter when flying solo of 0.651.

Summary

The preliminary point estimate from the 2009 scientific aerial survey is lower than the 2008 estimate, and similar to the 2006 and 2007 estimates. Taking confidence intervals into account, the relative abundance estimates have remained similar since 2005, and are significantly lower than the average level in the mid-1990s (1993-1996).

The models used to analyse the data were modified from last year to include random effects for the 2-way and 3-way interactions terms between Year, Month and Area (as opposed to just the 3-way interaction term). This led to more stable model fits since there are a number of 2-way strata for which little or no data exist, and the revised models can better handle these situations. The results obtained using the revised models did not differ significantly from those obtained using the previous models. It is recommended that the revised models be used as the standard method of analysis for estimating the aerial survey index in future.

References

Anonymous. 2008. Report of the Thirteenth Meeting of the Scientific Committee, Commission for the Conservation of Southern Bluefin Tuna, 5-12 September 2008, Rotorua, New Zealand.

Basson, M., Bravington, M., Eveson, P. and Farley, J. 2005. Southern bluefin tuna recruitment monitoring program 2004-05: Preliminary results of Aerial Survey and Commercial Spotting data. Final Report to DAFF, June 2005.

Bravington, M. 2003. Further considerations on the analysis and design of aerial surveys for juvenile SBT in the Great Australian Bight. RMWS/03/03.

Cowling, A. 2000. Data analysis of the aerial surveys (1993-2000) for juvenile southern bluefin tuna in the Great Australian Bight. RMWS/00/03.

Eveson, P., Bravington, M. and Farley, J. 2006. The aerial survey index of abundance: updated analysis methods and results. CCSBT-ESC/0609/16.

Eveson, P., Bravington, M. and Farley, J. 2007. Aerial survey: updated index of abundance and preliminary results from calibration experiment. CCSBT-ESC/0709/12.

Eveson, P., Bravington, M. and Farley, J. 2008. The aerial survey index of abundance: updated analysis methods and results. CCSBT-ESC/0809/24.

Acknowledgements

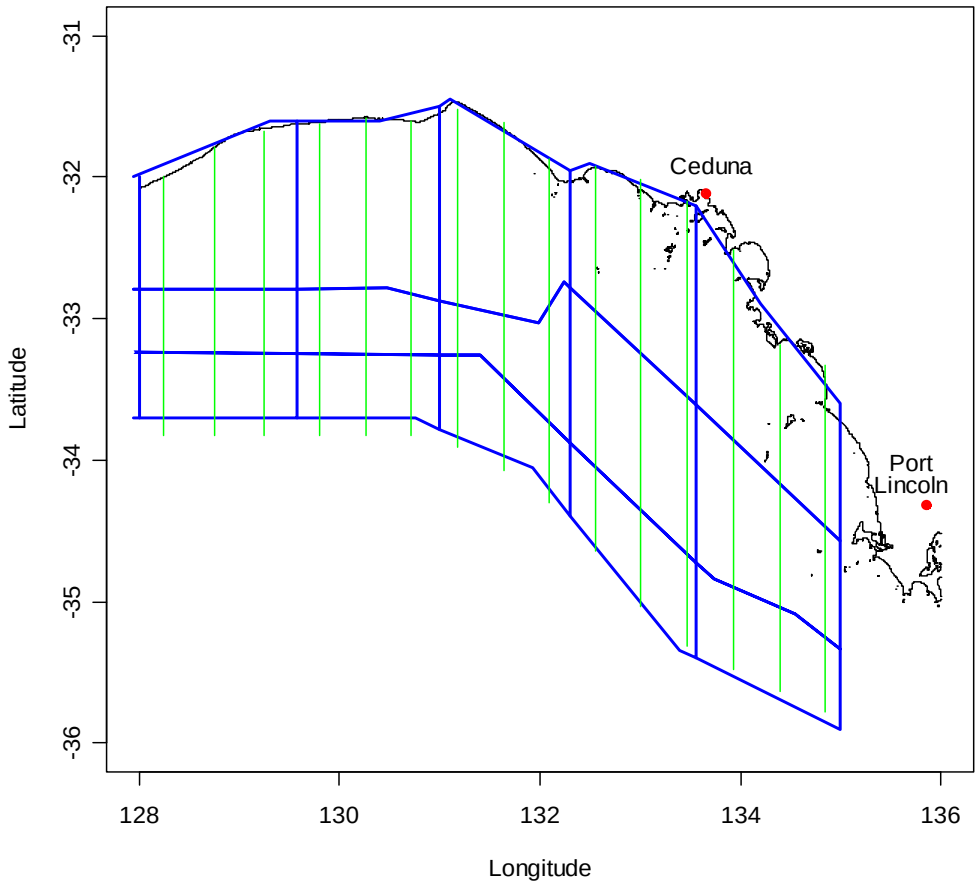
There are many people we would like to recognise for their help and support during this project. We would especially like to thank this year's spotters, pilots and data recorders; Andrew Jacob, Darren Tressider, Derek Hayman, John Veerhius, Thor Carter and Adam Vasey. We also thank the tuna fishing companies in Port Lincoln for their support of the project. This study was funded by AFMA, DAFF, Australian Industry, and CSIRO's Wealth from Oceans Flagship.

Appendix A – Methods of analysis

Separate models were constructed to describe two different components of observed biomass: i) biomass per patch sighting (BpS), and ii) sightings per nautical mile of transect line (SpM). Each component was fitted using a generalized linear model (GLM), as described below. Since environmental conditions affect what proportion of tuna are available at the surface to be seen, as well as how visible those tuna are, and since different observers can vary both in their estimation of school size and in their ability to see tuna patches, the models include 'corrections' for environmental and observer effects in order to produce standardized indices that can be meaningfully compared across years.

For the purposes of analysis, we defined 45 area/month strata: 15 areas (5 longitude blocks and 3 latitude blocks, as shown in Figure A1) and 3 months (Jan, Feb, Mar). The latitudinal divisions were chosen to correspond roughly to depth strata (inshore, mid-shore and shelf-break).

Figure A1. Plot showing the 15 areas (5 longitudinal bands and 3 latitudinal bands) into which the aerial survey is divided for analysis purposes. The green vertical lines show the official transect lines for the surveys conducted in 1999 and onwards; the lines for previous survey years are similar but are slightly more variable in their longitudinal positions and also do not extend quite as far south (which is why the areas defined for analysis, which are common to all survey years, do not extend further south).



Biomass per sighting (BpS) model

For the BpS model, we first estimated relative differences between observers in their estimates of patch size (using the same methods as described in Bravington 2003). As in Bravington (2003), we found good consistency between observers. In particular, patch size estimates made by different observers tended to be within about 5% of each other, except for one observer, say X, who tended to underestimate patch sizes relative to other observers by about 20%. The patch size estimates were corrected using the estimated observer effects (e.g. patch size estimates made by observer X were scaled up by 20%). Because the observer effects were estimated with high precision, we treated the corrected patch size estimates as exact in our subsequent analyses. The final biomass estimate for each patch was calculated as the average of the two corrected estimates (recall that the size of a patch is estimated by both observers in the plane). The final patch size estimates were then aggregated within sightings to give an estimate of the total biomass of each sighting. It is the total biomass per sighting data that are used in the BpS model.

The BpS model was fitted using a GLMM (generalized linear mixed model) with a log link and a Gamma error structure. We chose to fit a rather rich model with 3-way interaction terms between year, month and area. This is true not only for the BpS model but also for the SpM model described below. In essence, the 3-way interaction model simply corrects the observation (the total biomass of a sighting in the case of the BpS model; the number of sightings in the case of the SpM model) for environmental effects, which are estimated from within-stratum comparisons (i.e. within each combination of year, month and area).

In 2008, we fit the 3-way Year:Month:Area interaction term as a random effect in both models, whereas the 1-way and 2-ways effects were still fit as fixed effects. This year, we modified the models to include the 2-way interaction terms as random effects as well. This decision was made because many of the 2-way strata had very few observations and were causing instabilities in the model fits. Random effect models allow a coherent approach to inference even when some strata are data-poor. Estimates in data-rich strata are driven by the data themselves, but in strata where data are lacking – say a particular year and month – the estimate is adjusted towards the value that would be expected based on (i) other months in that year, and (ii) that month in other years. The extent of the adjustment depends on the amount of data available. A key advantage of using random effect models to make such adjustments (rather using an *ad hoc* approach) is that the uncertainty associated with the adjustments is correctly propagated through to the final abundance estimate.

Having decided on the overall structure, we then decided what environmental variables to include in the model. Based on exploratory plots and model fits, we determined the two environmental covariates that had a significant effect on the biomass per sighting were wind speed and, especially, SST (note that the selection of environmental covariates was undertaken as part of the 2006 analysis, and has not been repeated since then). Thus, the final model fitted was

$$\log \mathbf{E}(\text{Biomass}) \sim \text{Year}*\text{Month}*\text{Area} + \text{SST} + \text{WindSpeed}$$

where Year, Month and Area are factors, and SST and WindSpeed are linear covariate (note that E is standard statistical notation for expected value).

Sightings per mile (SpM) model

For the SpM model, we first updated the pairwise observer analysis described in Bravington (2003), based on within-flight comparisons of sighting rates between the various observers. This analysis gives estimates of the relative sighting abilities for the 18 different observer pairs that have flown at some point in the surveys. The observer pairs ranged in their estimated sighting rates from 72% to 97% compared to the pair with the best rate.

Although this analysis gives reasonable certainty about the relative ranking of different observer pairs, the data provide much less information about the relative efficiency; for example, even if it is clear from the data that A & B together would see more schools than C & D together under the same conditions, it is less clear whether A & B would see 100% more or only 10% more. If there was good certainty about the relative efficiencies, we could just include the estimates from the pairwise model as a known offset (i.e., as a predictor variable with known, rather than estimated, coefficients) when fitting the SpM model. However, because of the uncertainty in the relative efficiencies, we chose instead to include log-relative-efficiency as a covariate in the SpM model rather than as an offset, with the effect size (i.e., "slope") to be estimated. If the relative efficiencies from the pairwise analysis are correct, the slope estimate should be close to one. This approximation is not perfect, because there is still uncertainty about the relative rankings which we have ignored; in future, we hope to formally merge the pairwise model with the SpM model to correctly propagate all the uncertainty into the final CVs.

The data used for the SpM model were accumulated by flight and area, so that the data set used in the analysis contains a row for every flight/area combination in which search effort was made (even if no sightings were made). Within each flight/area combination, the number of sightings and the distance flown were summed, whereas the environmental conditions were averaged. The SpM model was fitted using a GLMM with the number of sightings as the response variable, as opposed to the sightings rate. The model could then be fitted assuming an overdispersed Poisson error structure[1] with a log link and including the distance flown as an offset term to the model (i.e. as a linear predictor with a known coefficient of one).

As we did for the BpS model, we included terms for year, month and area, as well as all possible interactions between them, in the SpM model, and we fitted the 2-way and 3-way interaction terms as random effects (see BpS model section). We determined what environmental variables to include in the model based on exploratory plots and model fits. A number of environmental covariates correlate highly with the number of sightings made (but not with each other) and came up as significant in the model fits. Again, SST was one of the most influential variables. The final model fitted was:

$$\log\mathbf{E}(N\_sightings) \sim offset(\log(Distance)) + Year*Month*Area + \log(ObsEffect) + SST + WindSpeed + Swell + Haze + MoonPhase$$

where Year, Month and Area are factors, MoonPhase is a factor (taking on one of four levels from new moon to full moon), and all other terms are linear covariates.

[1] Note that the standard Poisson distribution has a very strict variance structure in which the variance is equal to the mean, and it would almost certainly underestimate the amount of variance in the sightings data, hence the use of an overdispersed Poisson distribution to describe the error structure.

Combined analysis

The BpS and SpM model results were used to predict what the number of sightings per mile and the average biomass per sighting in each of the 45 area/month strata in each survey year would have been under standardized environmental/observer conditions[2]. Using these predicted values, we calculated an abundance estimate for each stratum as 'standardized SpM' multiplied by 'standardized average BpS'. We then took the weighted sum of the stratum-specific abundance estimates over all area/month strata within a year, where each estimate was weighted by the geographical size of the stratum in nm^2, to get an overall abundance estimate for that year. Lastly, the annual estimates were divided by their mean to get a time series of relative abundance indices.

[2] In our predictions, we used above average conditions, namely SST=21, wind speed =3, swell=1, haze=0, low cloud=0, moon phase=4 (full moon), and observer effect=1 (i.e. the 'best' observer pair).

Appendix B – CV calculations

This appendix provides details of how CVs for the aerial survey abundance indices were calculated.

Let $\hat{B}_{ijk}$ be the predicted value of BpS in year *i*, month *j* and area *k* under standardized environmental/observer conditions (see footnote 2 of main body), and $\hat{\sigma}\left(\hat{B}_{ijk}\right)$ be its estimated standard error. Similarly, let $\hat{S}_{ijk}$ be the predicted value of SpM in year *i*, month *j* and area *k* under the same environmental/observer conditions, and $\hat{\sigma}\left(\hat{S}_{ijk}\right)$ be its estimated standard error. Then,

$$\hat{A}_{ijk} = \hat{S}_{ijk}\hat{B}_{ijk}$$

is the stratum-specific abundance estimate for year *i*, month *j* and area *k*.

Since $\hat{B}_{ijk}$ and $\hat{S}_{ijk}$ are independent, the variance of $\hat{A}_{ijk}$ is given by

$$\begin{aligned} V\left(\hat{A}_{ijk}\right) &= V\left(\hat{S}_{ijk}\hat{B}_{ijk}\right) \\ &= V\left(\hat{S}_{ijk}\right)E\left(\hat{B}_{ijk}\right)^2 + V\left(\hat{B}_{ijk}\right)E\left(\hat{S}_{ijk}\right)^2 + V\left(\hat{S}_{ijk}\right)V\left(\hat{B}_{ijk}\right) \\ &\approx \hat{\sigma}^2\left(\hat{S}_{ijk}\right)\hat{B}_{ijk}{}^2 + \hat{\sigma}^2\left(\hat{B}_{ijk}\right)\hat{S}_{ijk}{}^2 + \hat{\sigma}^2\left(\hat{S}_{ijk}\right)\hat{\sigma}^2\left(\hat{B}_{ijk}\right) \end{aligned}$$

The annual abundance estimate for year *i* is given by the weighted sum of all stratum-specific abundance estimates within the year, namely

$$\hat{A}_i = \sum_j \sum_k w_k \hat{A}_{ijk}$$

where w_k is the proportional size of area *k* relative to the entire survey area ($\sum_k w_k = 1$).

If the $\hat{A}_{ijk}$'s are independent, then the variance of $\hat{A}_i$ is given by

$$V\left(\hat{A}_i\right) = \sum_j \sum_k w_k{}^2 V\left(\hat{A}_{ijk}\right)$$

Unfortunately, the $\hat{A}_{ijk}$'s are NOT independent because the estimates of BpS (and likewise, the estimates of SpM) are not independent between different strata. This is because all strata estimates depend on the estimated coefficients of the environmental/observer conditions, so any error in these estimated coefficients will affect all strata. Thus, we refit the BpS and SpM models with the coefficients of the environmental/observer covariates (denote the vector of coefficients by θ) fixed at their estimated values ($\hat{\theta}$). Note that θ contains the environmental/observer coefficients from both the BpS and SpM models; i.e. $\theta = (\theta_{\text{BpS}}, \theta_{\text{SpM}})$. The predictions of BpS and SpM made using the 'fixed environment' models should now be independent between strata, so the stratum-specific abundance estimates calculated using these predictions – which we will denote by $\hat{A}_{ijk}\left(\hat{\theta}\right)$ – should also

be independent between strata. Thus, we can calculate the variance of $\hat{A}_i$ conditional on the estimated values of the environmental/observer coefficients as

$$V\left(\hat{A}_i \mid \hat{\theta}\right) = \sum_j \sum_k w_k{}^2 V\left(\hat{A}_{ijk}\left(\hat{\theta}\right)\right)$$

where $V\left(\hat{A}_{ijk}\left(\hat{\theta}\right)\right)$ is calculated using the formula given above for $V\left(\hat{A}_{ijk}\right)$ but using the BpS and SpM predictions and standard errors obtained from the 'fixed environment' models.

To calculate the unconditional variance of $\hat{A}_i$, we make use of the following equation:

$$\begin{aligned} V\left(\hat{A}_i\right) &= E_\theta\left(V\left(\hat{A}_i \mid \theta\right)\right) + V_\theta\left(E\left(\hat{A}_i \mid \theta\right)\right) \\ &\approx V\left(\hat{A}_i \mid \hat{\theta}\right) + V_\theta\left(\hat{A}_i\right) \end{aligned}$$

where the first term is the conditional variance just discussed and the second term is the additional variance due to uncertainty in the environmental coefficients. The second term can be estimated as follows

$$V_\theta\left(\hat{A}_i\right) \approx \left(\frac{\partial \hat{A}_i}{\partial \theta}\right)' \mathbf{V}_\theta \left(\frac{\partial \hat{A}_i}{\partial \theta}\right)$$

where $\left(\frac{\partial \hat{A}_i}{\partial \theta}\right)$ is the vector of partial derivatives of $\hat{A}_i$ with respect to θ (which we calculated using numerical differentiation), and $\mathbf{V}_\theta$ is the variance-covariance matrix of the environmental coefficients[3].

Finally, the relative abundance index for year *i* is calculated as

$$\hat{I}_i = \frac{\hat{A}_i}{\sum_i \hat{A}_i}$$

Using the delta method, we can approximate the variance of $\hat{I}_i$ by

$$V\left(\hat{I}_i\right) \approx \left(\frac{\partial \hat{I}_i}{\partial \hat{A}_i}\right)^2 V\left(\hat{A}_i\right)$$

Then, the standard error of $\hat{I}_i$ is given by $\sigma\left(\hat{I}_i\right) = \sqrt{V\left(\hat{I}_i\right)}$

and the coefficient of variation (CV) of $\hat{I}_i$ is given by $\mathrm{CV}\left(\hat{I}_i\right) = \frac{\sigma\left(\hat{I}_i\right)}{\hat{I}_i}$.

[3] Recall that θ contains the environmental/observer coefficients from both the BpS and SpM models, so $\mathbf{V}_\theta = \begin{bmatrix} \mathbf{V}_{\theta_{\mathrm{BpS}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{\theta_{\mathrm{SpM}}} \end{bmatrix}$. The variance-covariance matrices for the individual models are returned from the model-fitting software.

Appendix C: Results and diagnostics

Biomass per sighting (BpS) model

Figure C1 shows plots of observed biomass per sighting versus the environmental covariates being included in the BpS model. From these plots, it appears that the size of a sighting tends to increase as SST increases, and decrease as wind speed increases (in a roughly linear fashion in both cases). The relationship with SST appears to be strongest, as supported by the model results (below).

Extract from the output produced by the software used to fit the model (the gam function in the R statistical package mgcv):

```
Family: Gamma
Link function: log

Formula:
Biomass ~ factor(Year) + factor(Month) + factor(Area) + SST + WindSpeed +
Y.M + Y.A + M.A + Y.M.A - 1
```

Parametric Terms:

Covariate	df	F	p-value
Year	13	2.02	0.016
Month	2	2.20	0.112
Area	14	2.71	0.001
SST	1	28.14	0.000
Wind Speed	1	0.78	0.378

```
R-sq.(adj) =  0.131   Deviance explained = 45.1%
GCV score = 1.7584  Scale est. = 1.5267    n = 1397
```

These results suggest that the covariate SST is highly statistically significant but that wind speed is not (recall that a comprehensive model selection process for determining which environmental covariates to include was last performed in 2006). Nevertheless, we prefer to include wind speed in the model because the plot in Fig C1 suggests a relationship does exist, and because, for the purposes of our analysis, including extra covariates should not have any negative effects.

Figure C2 shows some standard diagnostic plots for generalized linear models, and Figure C3 shows the residuals plotted against a number of factors. These plots do not suggest any major problems with the model fit. Ideally there should be no trend in the plots of the square root of the absolute residuals against the fitted values (i.e., lower half of Fig. C2, with left-hand side being on the link scale and the right-hand side being on the response scale); although there is a small kink revealed by a smooth through the data (red line), there is not a consistent increasing or decreasing trend.

Figure C1. Plots of observed biomass per sighting, on a log scale, versus the covariates included in the model; shown is the mean +/- 2 standard deviations.

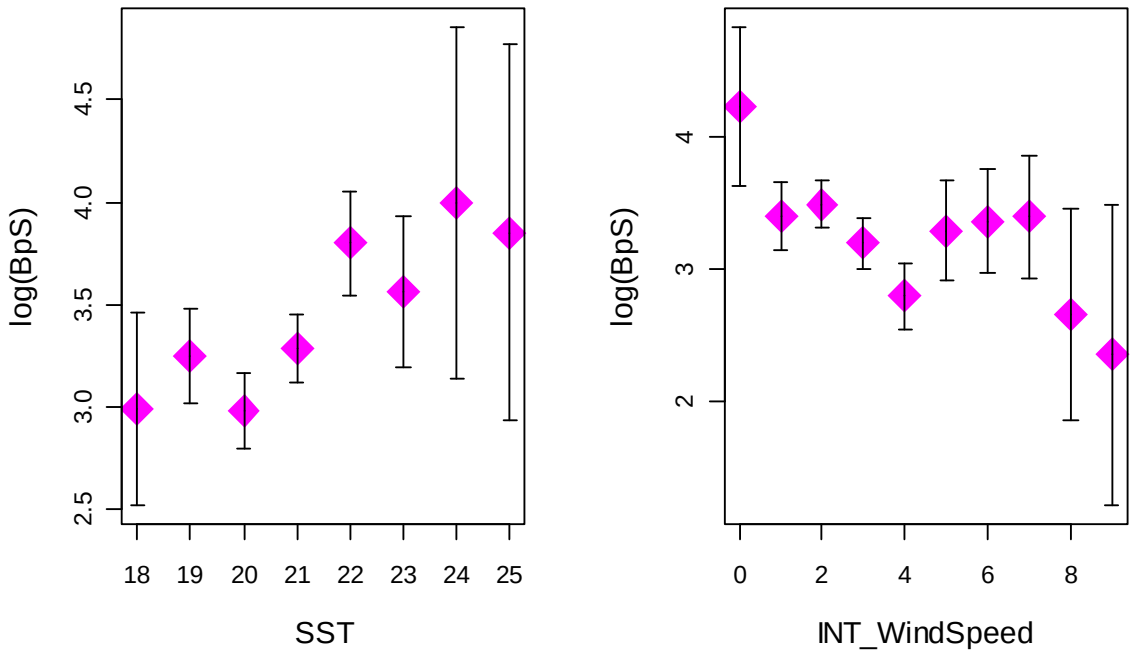


Figure C2. Standard diagnostic plots for biomass per sighting (BpS) model.

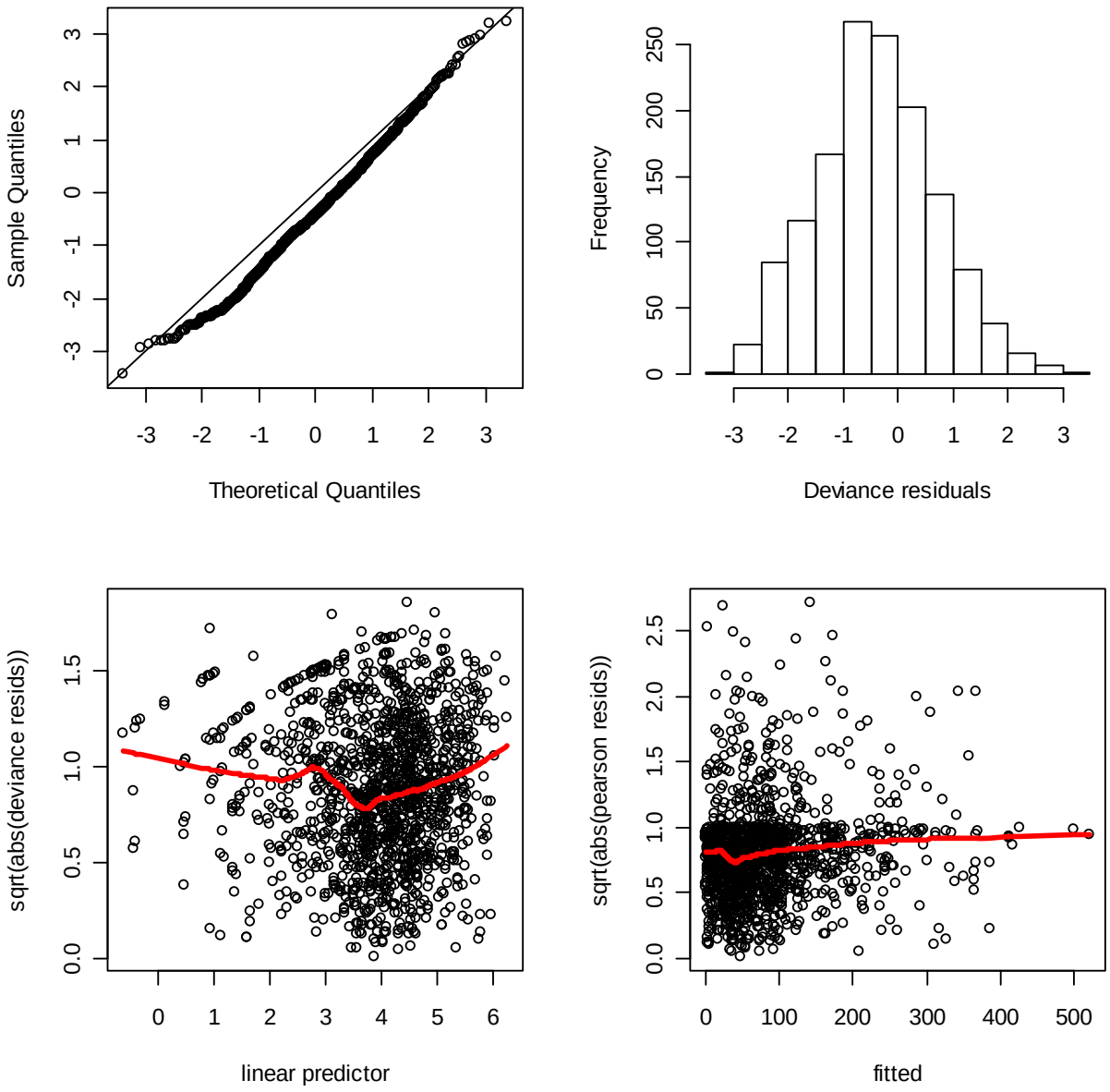
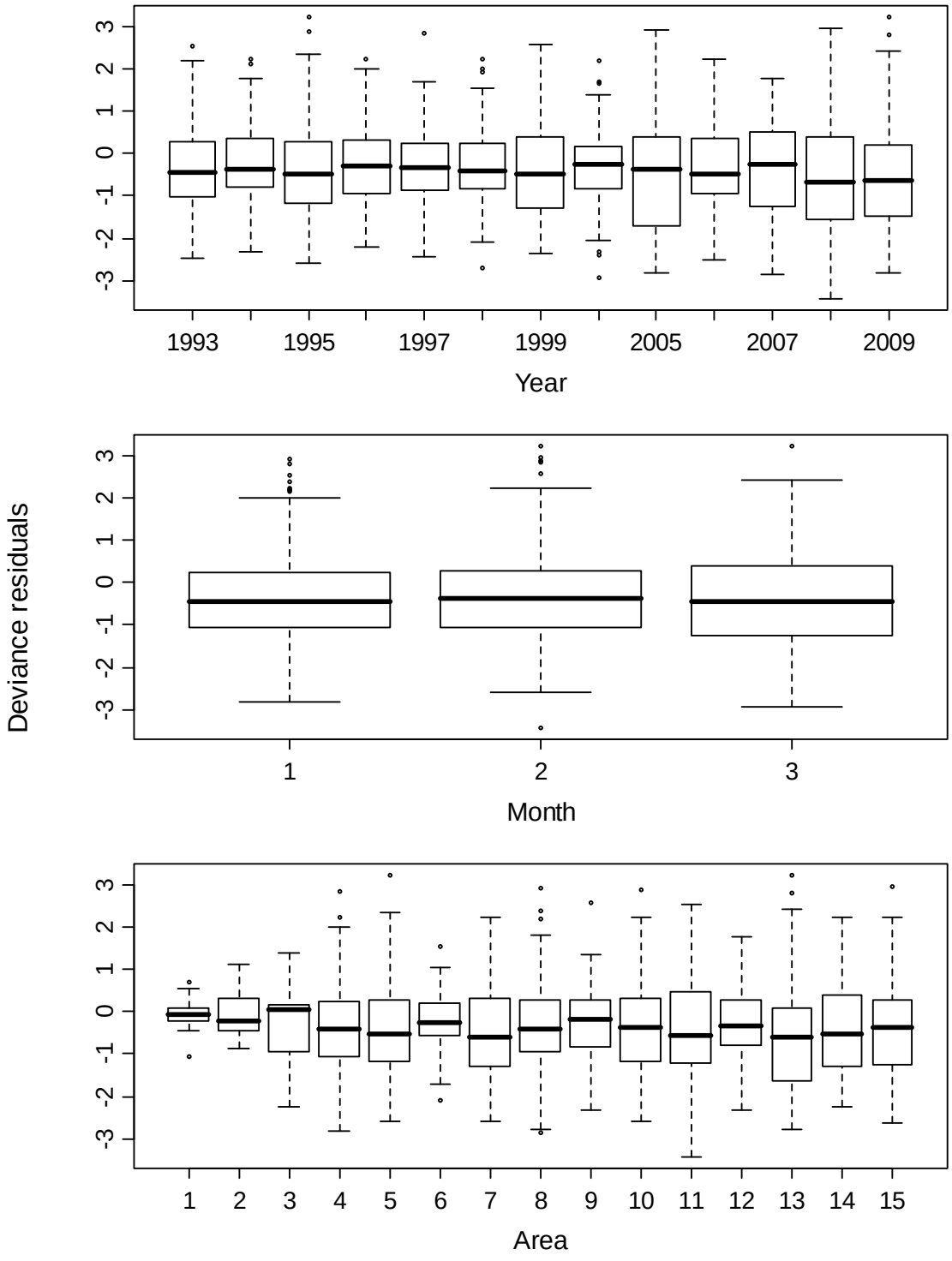


Figure C3. Boxplots of deviance residuals by year, month and area for biomass per sighting (BpS) model.



Sightings per mile (SpM) model

Figure C4 shows plots of observed number of sightings per mile versus the environmental covariates being included in the SpM model. These plots suggest that there is a strong tendency for the rate of sightings to increase as SST increases, and to decline as wind speed, haze and swell increase. With the exception of wind speed, the relationship appears to be linear, and this is even true for wind speed in the range of 1 to 7 knots (where most of the observations occur). Moon phase also appears to influence the sightings rate, with the rate being greatest when the moon phase is 1 (fraction of moon illuminated is 0-25%) or 4 (fraction of moon illumination is 75-100%). This suggests that the relationship between moon phase and tuna behaviour is complex.

Extract from the output produced by the software used to fit the model (the gam function in the R statistical package mgcv):

```
Family: quasipoisson
Link function: log

Formula:
N_sightings ~ offset(log(as.numeric(Distance))) + factor(Year) +
factor(Month) + factor(Area) + Y.M + Y.A + M.A + Y.M.A +
log(ObserverEffect) + AvgWindSpeed + AvgSST + AvgSwell + AvgHaze +
factor(MoonPhase) - 1

Parametric Terms:
```

Covariate	df	F	p-value
Year	13	7.96	0.000
Month	2	1.57	0.209
Area	14	2.12	0.009
log(ObserverEffect)	1	0.67	0.414
WindSpeed	1	142.45	0.000
SST	1	48.38	0.000
Swell	1	9.26	0.002
Haze	1	13.33	0.000
MoonPhase	3	1.60	0.187

```
R-sq.(adj) =  0.504   Deviance explained = 64.7%
GCV score =  1.574  Scale est. = 1.2776    n = 1301
```

The model results suggest that all of the environmental covariates included in the model are highly significant, with the exception of moon phase. These results are consistent with our expectations from Fig. C4. Although the observer effect term is not significant, we believe it is important to leave in based on principle because it may become significant as new observers enter into the survey.

Figure C5 shows some standard diagnostic plots for generalized linear models, and Figure C6 shows the residuals plotted against a number of factors. The Q-Q plot has a slight kink, but this does not appear too serious; otherwise there are no indications of problems with the model fit. The plots of the square root of the absolute residuals against the fitted values (i.e., lower half of Fig. C5, with left-hand side being on the link scale and the right-hand side being on the response scale) look a bit odd, but this is expected because we are modelling count data. A smooth line through these data is very flat, as desired, except for where it follows the residuals for the zero response values (i.e., where the observed number of sightings was zero).

Figure C4. Plots of observed sightings per mile, on a log scale, versus the covariates included in the model; shown is the mean +/- 2 standard deviations.

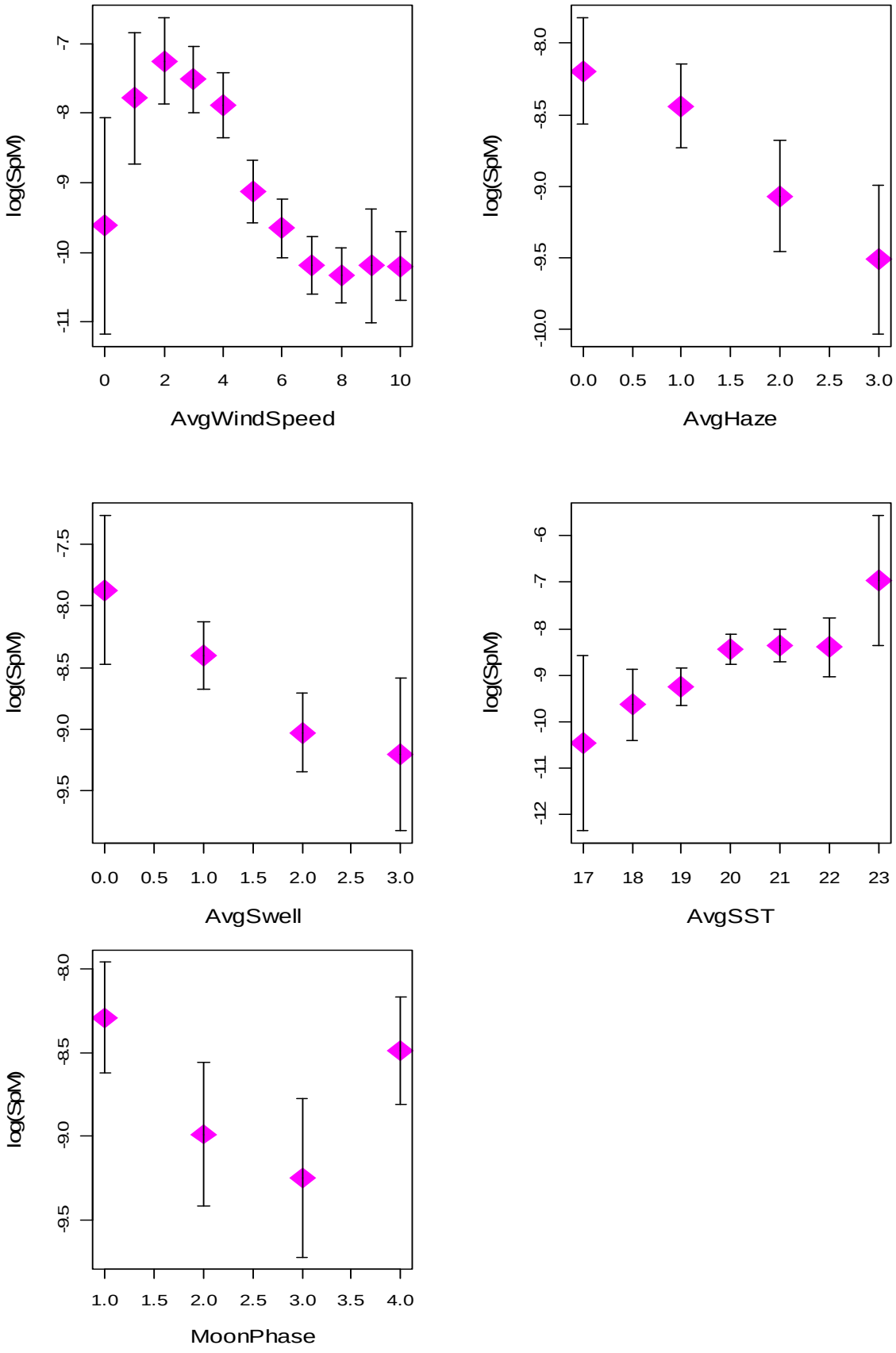


Figure C5. Standard diagnostics plots for sightings per mile (SpM) model.

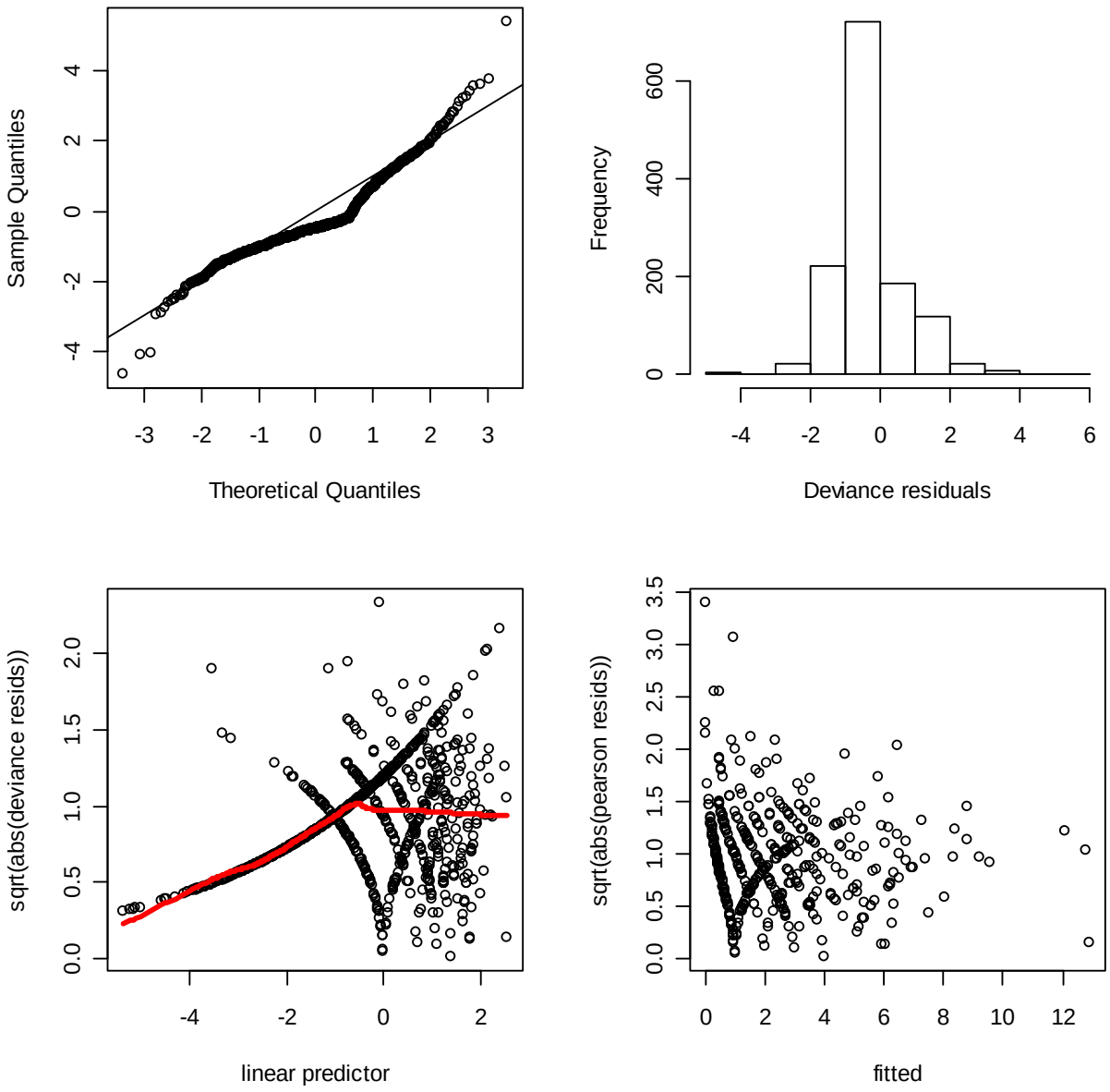
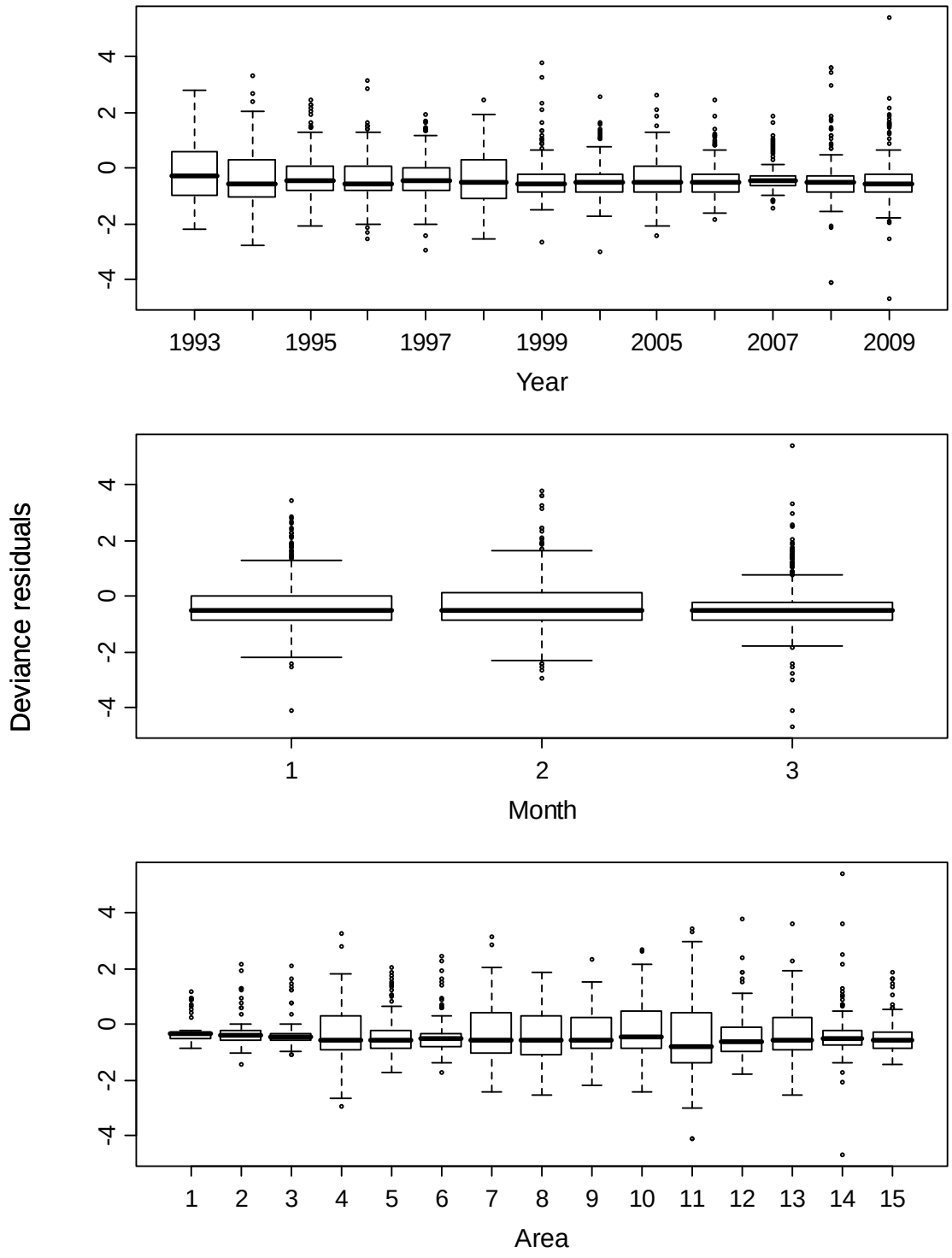


Figure C6. Boxplots of deviance residuals by year, month and area for sightings per mile (SpM) model.

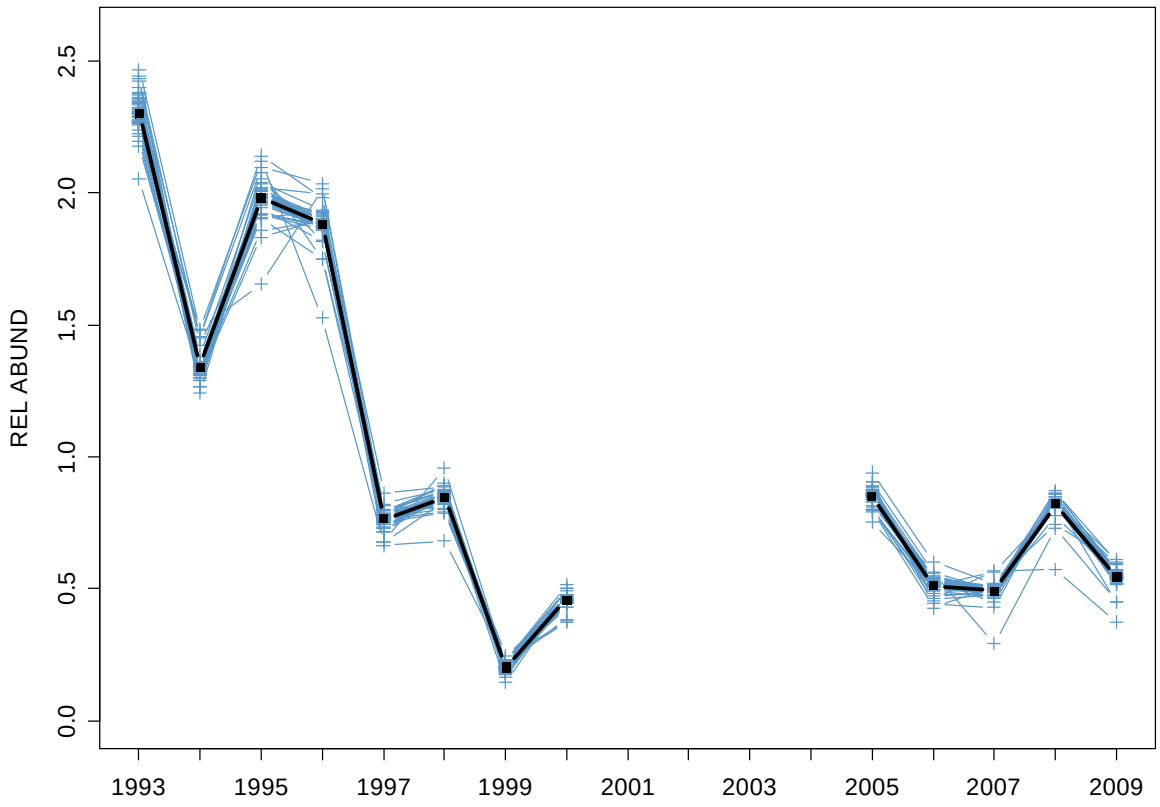


Evaluation of random effects

The diagnostics presented in the previous sections suggest that the BpS and SpM model fits are reasonably good. However, they are standard diagnostics for generalized linear models and do not directly assess the assumptions about the random effects. Methods for doing so are not well developed. Simple bootstrapping methods can fail with random effects models[4], and more sophisticated alternatives are obscure and difficult to implement.

Thus, we have attempted to evaluate the random effects by performing a type of jackknife. In particular, we deleted the observations from every month/area combination one at a time and refit the BpS and SpM models. A time series of relative abundance estimates was calculated from each set of refitted models, and these are plotted in Figure C7. The refitted estimates are all quite similar to the original estimates, with the mean and median of the refitted estimates within 1% of the original estimate in every year, suggesting that the random effects models are behaving sensibly.

Figure C7. Plot of the "jackknife" relative abundance estimates, calculated by leaving out the data from a particular month/area and redoing the analysis. Each blue line corresponds to a different month/area. The thick black line shows the original estimates (i.e., those in Figure 6 of the main report).



[4] Laird, N.M. and Louis, T.A. (1987) Empirical Bayes confidence intervals based on bootstrap samples. J. Amer. Stat. Assoc. 82: 739-750.

Also of interest is the impact that each of the random effect terms has on the model results. To evaluate, we set the coefficients for a particular random effect term equal to zero, then re-calculated the annual abundance indices, and repeated this for each random effects term. Table C1 shows the results, with the last row giving the mean percent change in the index over all years. It appears that the Year:Month term has the least impact overall.

Table C1. Estimates of relative abundance obtained by setting the coefficients of each of the random effect terms equal to zero. Y=Year, M=Month, A=Area

	Original	Random effect term set to zero			
Year	Index	Y:M	M:A	Y:A	Y:M:A
1993	2.30	2.19	2.41	2.81	2.42
1994	1.34	1.38	1.50	1.52	1.50
1995	1.98	2.05	1.94	1.81	1.96
1996	1.88	1.95	2.01	1.83	2.03
1997	0.76	0.77	0.77	0.74	0.78
1998	0.85	0.87	0.85	0.81	0.85
1999	0.20	0.21	0.22	0.23	0.22
2000	0.46	0.46	0.41	0.40	0.41
2005	0.85	0.90	0.95	0.89	0.93
2006	0.51	0.50	0.54	0.53	0.53
2007	0.49	0.51	0.43	0.41	0.43
2008	0.82	0.75	0.62	0.51	0.59
2009	0.54	0.47	0.35	0.52	0.34
Mean % change		4.3	10.3	11.3	10.7