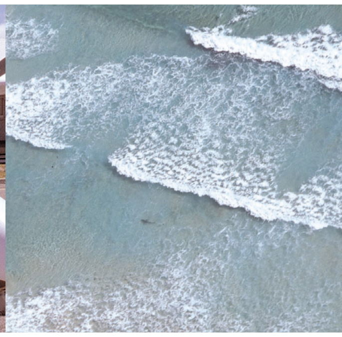
Development of a fishery independent index of abundance for juvenile southern bluefin tuna – **AND** – Improved fishery independent estimates of southern bluefin tuna recruitment through integration of environmental, archival tag and aerial survey data.






Development of a fishery independent index of abundance for juvenile southern bluefin tuna and improvement of the index through integration of environmental, archival tag and aerial survey data.

Ann Cowling • Alistair Hobday • John Gunn

CSIRO Marine Research, November 2002

Final report – FRDC Projects: 96/118 and 99/105

CSIRO
MARINE RESEARCH

FISHERIES RESEARCH & DEVELOPMENT CORPORATION

Cowling, Ann.
Development of a fishery independent index of abundance for juvenile southern bluefin tuna and improved fishery independent estimates of southern bluefin tuna recruitment through integration of environmental, archival tag and aerial survey data.

Bibliography.
Includes index.
ISBN 1 876996 41 2.

1. Southern bluefin tuna – Great Australian Bight (W.A. and S. Aust). 2. Fishery resources – Great Australian Bight (W. A. and S. Aust.). 3. Fish populations – Great Australian Bight (W.A. and S. Aust.). 4. Southern bluefin tuna – Research – Great Australian Bight (W.A. and S. Aust). I. Gunn, John, 1956- . II. Hobday, Alistair. J. III. CSIRO. Marine Research. IV. Title.

639.27783

Table of Contents

Contributing authors

Contributing authors (all employed by CSIRO Marine Research)

Ann Cowling	Project Results 6.1, 6.3, 6.4, 6.5
Alistair Hobday	Project Results 6.2, 6.3, 6.4
Jeremy O'Reilly	Project Results 6.3, 6.4
John Gunn	Project Results 6.3, 6.4
Mark Bravington	Project Results 6.6

1. Non-Technical Summary

1996/105	Development of a fishery independent index of abundance for juvenile southern bluefin tuna
and	
1999/118	Improved fishery independent estimates of southern bluefin tuna recruitment through integration of environmental, archival tag and aerial survey data

Principal Investigators:

Dr A. Cowling*, Dr A. Hobday, Mr J. Gunn
GPO Box 1538
CSIRO Marine Research
Hobart, Tasmania, 7001

*Current address:
Statistical Consulting Unit
Australian National University
ACT 0200

Objectives:

The two sets of objectives for the projects are presented as they appeared in original material proposing the research.

FRDC 1996/118:

1. To conduct an aerial survey for SBT over the Great Australian Bight each summer season from 1997 to 1999 and estimate various surface abundance indices.
2. To complete the statistical research required to:
   (a) incorporate environmental variables into the estimates;
   (b) incorporate estimates of the proportion of SBT at the surface under various environmental conditions;
   (c) reduce the sampling error in the estimates due to uncertainty in school size and fish size estimates.
3. To complete an evaluation of the usefulness of the indices of SBT abundance derived from the aerial survey.

FRDC 1999/105:

1. To conduct a range of statistical analyses of data from the archival tags and environmental and oceanographic archives to determine whether there are common responses in surfacing behaviour to environmental conditions through space and time
2. To conduct a range of statistical analyses of aerial survey data on surface distribution and surface abundance of juvenile SBT, and environmental and oceanographic archives to develop a spatial model of abundance which allows for environmental variation through space and time. This would include analyses of how environmental conditions affect the detect ability of surface schools from planes.
3. To develop an integrated analysis of abundance of SBT in the GAB incorporating the surfacing behaviour, surfacing abundance and spatial distribution models developed above.

Outcomes achieved:

This combined project has developed and improved a fishery-independent index of abundance for juvenile southern bluefin tuna (SBT) in the Great Australian Bight. It is hoped that this index will provide information on the status of the juvenile SBT population and in particular, the first early warning signs for recruitment failure for SBT globally. The domestic SBT industry has recognised the logistical problems associated with the survey, in particular the shortage of trained spotters. The industry has continued to support the development of an index, and recently supported the continuation of a spatially-reduced survey that solves many of the logistical problems. The importance of the juvenile abundance index is also acknowledged by the international management body for SBT, the Commission for the Conservation of Southern Bluefin Tuna (CCSBT). Recent scientific meetings of the CCSBT have continued to support the development and emphasised the importance of the index.

The parental stock of southern bluefin tuna (SBT) is at historically low levels, and there have been concerns in recent years about the risk of poor recruitment and the possibility of recruitment collapse. An index of juvenile SBT abundance is seen as a critical requirement for the effective management of SBT in both national and international contexts. To monitor the abundance of juvenile SBT and develop a fishery-independent index of juvenile abundance, Australia and Japan established a large-scale collaborative research program in 1993 (Recruitment Monitoring Program, RMP).

An aerial survey for SBT is one of the main projects in the RMP, and the data forms the basis of the juvenile SBT abundance index. These annual aerial surveys with comparable protocols have been conducted over the Great Australian Bight (GAB) for three months each summer since 1993. (Two earlier survey years, 1991 and 1992, differed in methodology and were unsuitable for some analyses). In this transect-based survey attempts are made to count the number and size of SBT schools, as well as estimate the size of fish within each school.

The FRDC supported the 1997 to 1999 surveys and development of various surface abundance indices using the full period of the consistent aerial survey (1993-2000). The abundance index has been refined through several stages, from spatial segregation of the school sightings in the GAB, to incorporation of behavioural effects and inclusion of environmental variables. In particular, research focused on identifying and incorporating environmental variables into the estimates, estimating the proportion of SBT at the surface under various environmental conditions and incorporating these estimates into the surface abundance estimates, and reducing uncertainty in the estimates arising from uncertainty in patch size and fish size estimates. An overall goal of the study was to evaluate the usefulness of the indices of SBT abundance, and in particular, the magnitude of changes in abundance that could be detected.

The early stages of the aerial survey analysis involved attempts to develop an index juvenile SBT in the GAB based on mean surface biomass density. It soon became apparent that many potential biases existed and as a result confidence in the estimates was low. In particular, efforts to validate the estimates of patch and fish size were undertaken using LIDAR experiment and spotter-validation experiments. Unfortunately, LIDAR was not able to provide an alternative technique for surveying SBT schools. A conclusion from the early work was that spotter estimates of patch and fish size varied considerably and that several age-classes of SBT occur within a single patch. As a result, robust estimates of abundance for each age-class were not considered possible. The initial indices had large CVs and as a result showed no evidence of a strong trend in juvenile biomass for the period 1993-2000.

Improvements to the indices and reduction of uncertainty were attempted by including environmental variables that may influence the abundance or detection of SBT in the aerial survey. Juvenile SBT are surface-orientated and non-randomly distributed in the GAB during the austral summer. In particular, SBT are clustered around the shelf break and inshore reefs, islands and rises, collectively known as lumps. Because of the greater detection frequency at these locations, the presence of SBT observed during the aerial survey was analysed with regard to the topographic characters of these features and local environmental variables. The focus on the regions of highest SBT abundance increased the “signal to noise ratio” and the goal was to identify environmental variables that could be included in future indices covering the whole survey region. Generalised linear models indicated non-linear relationships between the presence of SBT at topographic features and environmental and topographic variables, and models incorporating topography and the environment explained 40% and 28% of the deviance at the lumps and shelf break respectively. The significant environmental variables differed between locations; at the shelf were wind speed, swell, air temperature, and sea surface temperature (SST), while at the lumps they were wind speed and SST. Chlorophyll was important in some preliminary models, but insufficient temporal coverage for this variable prevented it being considered in the final models.

The ability to detect SBT in the aerial survey relies on schools being at the surface, a behaviour that may have temporal and spatial variation. Information on SBT surfacing behaviour obtained from archival tags in different regions of the GAB was expected to reduce the errors associated with the estimation of an abundance index derived from the aerial survey. Before deriving these surfacing behaviours, however, it was necessary to improve the estimates of position within the GAB. This would then allow different corrections for surfacing rates, and hence detection in the different areas covered by the aerial survey. In this process it was important to evaluate the performance of existing software with regard to position estimation. Improvements in the estimation of position were made and the improved position estimates used to attribute different SBT surfacing behaviours to different spatial regions.

The frequency of surfacing by juvenile SBT was determined using archival tag data. Depth and location information from archival tags deployed and recovered from juvenile SBT in the GAB during 1998 were used to investigate the relationship between surfacing and hence detectibility in an aerial survey, and environmental conditions. Several methods for classifying depth information into behavioural definitions related to surfacing were considered for four time-scales in different areas of the GAB. Generalised linear models were developed to explore the relationship between the surfacing measure (response) and environmental (covariates) variables in each of three areas covered by the aerial survey; insufficient archival tag data existed for a fourth area. A surface-oriented behavioural definition and the whole-day time-scale were chosen for final model selection based on preliminary analysis of the data. The important environmental covariates in the final models differed between areas, but the total set included the variables cloud cover, wind direction, barometric pressure, air temperature, moonphase, water temperature and chlorophyll. The final models for each area explained between 26.9% and 51.2% of the null deviance, although when compared with bootstrapped models using randomised data this was reduced to 19-25%. There was a significant difference in the SBT surfacing rates between the three areas examined; the range in the proportion of time at the surface was 35.9-55.3%. In future these surfacing behaviours should be examined in tags obtained from other years; until that time, the generality of the surfacing conclusions remains uncertain.

A number of features identified in the data since the first abundance index made the analyses and incorporation of environmental conditions more complex than originally thought. Examples of these complexities include identification of changes in detection rates due to changes in spotter personnel and experience, and variation in classifying school clusters. These features led to improved understanding of the accuracy of the analyses and the SBT surfacing processes. As a result, several different versions of indices of SBT abundance were compared, including biomass-based and presence-absence models. All the indices were robust to the period of time used to develop the underlying model. The indices developed in this section do show evidence of a decline in juvenile SBT abundance in the survey area over the period of the surveys (1993-2000). Comparison of all the abundance indices led to the conclusion that a presence-absence index was best as a long-term monitoring index due to its low CV's, relatively low cost and relative robustness to changes in aerial survey spotter personnel over time, however, this conclusion was tested a final time by a new member of the project team.

The potential utility of the different indices of relative abundance (biomass and presence-absence) was re-investigated by the new member of the project team. A different model underlying the index was developed and the resulting trends in the biomass index shown to be similar to previous models, although with slightly reduced CVs. Environmental and behavioral information were not included in this index, although future modifications would allow these to be incorporated. The most important limiting factor in this analysis was differences between the aerial survey spotters. However, results show that provided sufficient attention is given to spotter calibration, through protocols such as using pairs of spotters and through appropriate analyses, a useful and precise index of SBT abundance can be constructed. As time goes by and more data are collected, it is in principle possible to retrospectively improve the precision of past estimates by at least 5 percentage points (on a CV scale), reducing annual CVs below 30%. This precision is similar to the best fishery-independent surveys elsewhere in the world.

In overall conclusion, this project has identified the value of the aerial survey to monitor trends in the juvenile SBT population within the GAB. Significant progress has been made in the areas of developing the indices of surface abundance from the aerial surveys, understanding how environmental variation affects the estimates of surface abundance and SBT surfacing behaviour, and understanding SBT surfacing behaviour. The majority of the goals of the project were met, with some notable exceptions and modifications. While a combination of analytical, financial and logistical problems halted the survey in 2001, a spatially abbreviated survey was re-initiated in 2002 and 2003. Further development and continuation of the SBT index will require methods that allow these recent surveys to be included in the historical index. Inclusion of future and past tagging and environmental data as additional relationships are discovered should also improve the index. The influence of different survey effort and coverage is being explored in a new RMP project, based on an individual-based model, and will address in greater detail what change in SBT abundance can be detected by the survey under different levels of coverage. This is an important extension and builds on the results achieved in this project.

Keywords: southern bluefin tuna, ***Thunnus maccoyii***, fishery-independent survey, line transect methods, aerial survey, surfacing behaviour, environmental influences, Great Australian Bight

2. Acknowledgements

This study was assisted by the southern bluefin tuna industry, in particular the pilots, spotters, fishing companies, and management representatives. Additional financial support was also provided by AFMA, CSIRO Marine Research, Japan, and the Tuna Boat Owners Association. Specific acknowledgements for the particular project sections are included in each of the Project Results. Cover page produced by Lea Crosswell and report coordination by Toni Cracknell, CSIRO Marine Research.

3. Background

The background to both projects is presented as in the original proposals.

FRDC 1996/118

All recent assessments of southern bluefin tuna (SBT) indicate that the parental stock is at historically low levels. The current parental stock biomass has been judged to be below "commonly used scientific measures of biologically safe parental biomass" and there are concerns about the risk of poor recruitment and the possibility of recruitment collapse. There is also much uncertainty about whether the current catch level will allow for rebuilding of the SBT stock.

The current analytical assessment methods for SBT have a 4 to 5 year time lag in the estimates of the number of recruits, due to time lags in receiving catch data and the lack of a reliable index of juvenile abundance. In addition, there is much uncertainty about most recent estimates of recruitment as they are largely determined by the most recent juvenile catch rates. Therefore, current trends in recruitment remain one of the major unknowns in evaluating the status of this stock and its potential to rebuild under current catch rates. Moreover, there is no fishery independent information on stock or juvenile abundances. Lack of such information is a major limitation in evaluating the likelihood of stock rebuilding under current catch rates.

All recent scientific and management meetings, both under the previous informal trilateral arrangement and now under the Convention for the Conservation of Southern Bluefin Tuna, have considered the development of a fishery independent recruitment index of SBT to have a very high research priority. In response to this need, a developmental aerial survey program was started in 1990/91, and experimental surveys using line transect methods have been conducted annually during the fishing season in the Great Australian Bight since then.

In June 1993 a large scale five year collaborative program involving CSIRO and the Japanese National Research Institute of Far Seas Fisheries (NRIFSF) was established to monitor the abundance of juvenile SBT and develop a fishery independent index of juvenile abundance. The aerial survey project was established as one of the main projects in this program. Funding of the aerial survey part of this collaborative research program has come form a variety of Japanese and Australian sources; the Australian sources include CSIRO, SBTMAC, FRDC and FRRF. Each year a workshop is held to review and prioritise the collaborative research for the coming year. At the 1995 workshop, the aerial survey was reaffirmed as one of the highest priority projects.

During this project a great deal of data has been collected. The analysis of this data has increased our knowledge of SBT and their behaviour. We have revised some of our initial assumptions about the detection of SBT from planes and SBT distribution and behaviour. Our improved understanding of SBT distribution and surface abundance in the GAB will improve the final analysis of the data, but the greater complexity of the processes has slowed the development of the analysis. The development of this project has been delayed by a delay obtaining the archival tag data required by this project.

FRDC 1999/105

All recent assessments of Southern Bluefin Tuna (SBT) indicate that the parental stock is at historically low levels. The 5th CCSBT Scientific Committee meeting also considered continuing negative trends in the VPA estimates of recruitment, concerns that lead scientists to discuss the possibility of recruitment collapse in the SBT stock.

Although there is general agreement among scientist regarding parental biomass and recruitment trends indicated by VPAs, there is significant divergence in opinion as to whether current catch quotas will allow for rebuilding of SBT stock by 2020. Much of the uncertainty surrounds interpretation of the fishery data and targeting practices of the Japanese longline industry.

A significant source of uncertainty is the lack of reliable estimates of recruitment or the number of young fish that have entered the population in the past few years. It is these young fish that will form the parental stock in the future and provide the potential for rebuilding of the stock.

Current methods for assessing the status of SBT stock can only provide reasonable estimates of recruitment 6 years earlier than the most recent catch data. Thus, in 1997 we can reasonably estimate the levels of recruitment in 1990 but not after this. The uncertainty about what has happened to recruitment levels since 1991 severely limits our ability to predict future stock trends.

Recognising the limitations of available methods, and the necessity for accurate and timely estimates of recruitment, SBT scientists and managers have consistently emphasised the extremely high priority of recruitment monitoring research since 1988. The rationale behind this has been that a fishery independent index of abundance of new recruits would allow scientists and managers to gauge the success of management strategies designed to rebuild the stock in almost “real time”. Equally, if recruitment collapse occurred it would show first in the abundance of juvenile SBT in the Australian fishery.

In July 1993, CSIRO and Japan’s National Research Institute for Far Seas Fisheries commenced a five year collaborative research program (RMWS) aimed at developing methods for estimating and routinely monitoring SBT recruitment in the Great Australian Bight (GAB).

The primary methods developed by the collaborative research program is an aerial survey. The annual surveys are statistically designed and based on line transect methodology. The most recent analyses of data from the last 6 years of aerial surveys have shown a significant decline in the surface abundance of juvenile SBT in the GAB.

However, the aerial survey analyses do not yet adequately account for the possible effects of intra- and inter annual variation in environmental conditions and the surfacing behaviour of SBT. Thus it is possible that the apparent variation in surface abundance(i.e. the significant decline over the last 6 years) is due to environmental and behavioural variation rather than indicating a true decline in abundance of juvenile SBT in the GAB.

This proposal seeks to use existing data from archival tags, oceanographic and meteorological archives and the aerial survey to better define the links between biological factors (such as feeding, migration, schooling), environmental factors (such as wind speed and direction, sea and air temperature barometric pressure, current speed and direction) and the surfacing behaviour of SBT.

This proposal builds on initial analyses which are partially funded by FRDC 96/118. Investigations to date have shown that there are definite associations between environmental conditions and surface abundance which can be statistically modeled, and also that statistical models linking surfacing behaviour and environmental conditions can be developed. However, the initial analyses also suggest that the problem is more complex than originally envisaged and that to understand and model the mechanisms underlying the appearance of juvenile SBT in the GAB, input from a wider range of scientists will be necessary.

4. Need

The need for the two projects is reported as in the original proposals, with minor clarification.

FRDC 1996/118

Analyses of the SBT aerial survey data collected to date indicate that the data can be used to provide estimates of the number of schools, total biomass and biomass by cohort for fish at the surface with reasonable coefficients of variation. The estimates have started to provide an initial useful comparison with VPA results, as estimates of some cohorts from the two methods began to overlap in 1995. However, there are still a number of research problems that need to be addressed in order to evaluate whether these estimates can provide a reliable index of juvenile abundance. The problems are associated with the unknown variability in the proportion of schools at the surface, the proportion of juveniles within the Bight, environmental effects on detectability of surface schools and tuna surfacing behaviour, and the reliability of estimates of fish and school sizes.

The biggest source of uncertainty and perhaps the biggest source of variation in the analyses of aerial surveys to date, is that no account is taken of the variability in the proportion of schools at the surface. If the proportion of schools at the surface varied little from year to year, this would not be a problem. However, surfacing behaviour of SBT appears to be strongly influenced by environmental conditions. Although the aerial survey is only conducted under weather conditions favourable to detection of tuna at the surface, the aerial surveys to date have encountered substantial inter-annual differences, with sea-surface temperatures being perhaps the most important and variable. The variation in the proportion of surface schools must be accounted for to improve the interpretation of the aerial survey results. This issue will be a major focus of the research over the next four years.

This research will develop an integrated statistical model based on the recent and growing body of data on surfacing behaviour of SBT in the Great Australian Bight acquired from archival tags together with detailed environmental data collected in the aerial survey as well as from other sources. In addition, research using recently available laser technology (airborne LIDAR systems) that detects schools below the surface of the water will be conducted to try to estimate the proportion of surface schools.

Research is also needed to improve the reliability of the results including improvements in the estimates of school size, fish size, the effects and interactions of environmental factors on the detection and size of surface schools, and statistical methods for obtaining the variances of the estimates.

Finally, the current developmental time series of aerial survey indices must be extended and improved. Without such an extension, it would not be possible to evaluate whether the aerial survey can provide a useful index of abundance. With the results from 1995/96 and the three additional years covered by this proposal, the aerial survey index will overlap the VPA estimates of recruitment for seven cohorts. This overlap will provide the basis for a statistical analysis of the aerial survey results as an index of recruitment, which is to be conducted as part of the current research proposal.

The results of this project have led to some reassessment of the needs.

A) The estimates of biomass by cohort are now assessed to be insufficiently reliable to compare with VPA results, and are no longer reported in the results of the analyses.

B) Incorporation of surfacing behaviour into the surface abundance estimates is not necessary under one of the models being investigated at present. Under this model, the results and interpretation of the surface abundance analysis is greatly strengthened by the model of surfacing behaviour as it provides independent verification of the surface abundance model. Under other methods of estimation of surface abundance, it is necessary to incorporate surfacing rates.

FRDC 1999/105

The parental biomass of SBT remains at historically low levels, there is evidence from CPUE and VPA analyses that recruitment has continued to fall throughout the 1990's, and there is significant disagreement within the CCSBT Scientific Committee on the prediction of future population levels. It is therefore essential to know more about the recruitment dynamics of SBT, and in particular to reduce the uncertainty in the aerial survey estimates of surface abundance of juveniles in the GAB. These remain the only fishery-independent source of abundance data on SBT, and as such their importance cannot be overstated.

To reduce the uncertainty in the current aerial survey estimates we need to investigate how environmental factors and surfacing behaviour influences what is seen during aerial surveys. If current levels of uncertainty in the indices are reduced by incorporating these sources of variation, the value of the indices to the CCSBT will be substantially increased.

There is large variation in estimated surface abundance between survey replicates within a season and between seasons but there are also large differences between years in environmental conditions in the GAB (eg sea surface temperature), which confound the interpretation of changes in the index between years. It is possible that the apparent decline in surface abundance over the last 6 years is due to environmental and behavioural variation rather then indicating a true decline in the abundance of juvenile SBT.

To adequately understand how environmental variation and the resulting behavioural responses of SBT affect the recruitment indices we require thorough analyses of:

1. the surfacing behaviour in SBT and its relationship with environmental variables, migrations patterns and possibly also feeding behaviour.
2. the relationship between surface abundance (i.e what the aerial survey detects) and environmental variables,
3. the spatial variation in abundance of SBT in the GAB (incorporating both data on surface abundance from the aerial survey and data from archival and conventional tagging experiments).

The proposal project will use all existing data, collected over almost a decade, with funding from industry, FRDC, CSIRO and Japan. These data are an invaluable resource. The integration of behavioural, environmental and abundance data into an improved estimate of the surface abundance of SBT is listed under Priority 2 and 3 of SBTMAC Research Priorities.

5. Objectives

FRDC 1996/118

1. To conduct an aerial survey for SBT over the Great Australian Bight each summer season from 1997 to 1999 and estimate various surface abundance indices.
2. To complete the statistical research required to:
   - incorporate environmental variables into the estimates;
   - incorporate estimates of the proportion of SBT at the surface under various environmental conditions;
   - reduce the sampling error in the estimates due to uncertainty in school size and fish size estimates.
3. To complete an evaluation of the usefulness of the indices of SBT abundance derived from the aerial survey.

FRDC 1999/105

1. To conduct a range of statistical analyses of data from the archival tags and environmental and oceanographic archives to determine whether there are common responses in surfacing behaviour to environmental conditions through space and time
2. To conduct a range of statistical analyses of aerial survey data on surface distribution and surface abundance of juvenile SBT, and environmental and oceanographic archives to develop a spatial model of abundance which allows for environmental variation through space and time. This would include analyses of how environmental conditions affect the detect ability of surface schools from planes.
3. To develop an integrated analysis of abundance of SBT in the GAB incorporating the surfacing behaviour, surfacing abundance and spatial distribution models developed above.

6. Project Results

Project Overview

This final report covers two projects, FRDC 1996/118 and 1999/105. The goal of both projects was to develop a fishery-independent index of abundance for juvenile southern bluefin tuna (SBT). Improvements in the development of the index, through incorporation of additional information and a more rigorous analytical approach, led to a decision to combine both projects, as the results would be best integrated in a complete package.

These project results are presented as a series of sections that describe the methodological approach, findings, and subsequent implications for the development of an abundance index. The first results section reports on preliminary analyses of the aerial survey data without inclusion of modifying behavioural or environmental variables. The second results section describes identification of the important environmental variables, through consideration of SBT abundance patterns at the locations where SBT are detected most frequently in the aerial survey. Behavioural information from archival tags deployed on SBT within the GAB can also be used to improve the abundance index. The third results section describes how estimates of tag position are improved, such that data on SBT behaviour can be used to determine the influence of the environment on surfacing rates, which are crucial to describe detection of the fish from the air. The fourth results section describes development of algorithm-based descriptions of surfacing behaviour derived from the archival tag depth data and the resulting surfacing rates in the GAB. The penultimate results section describes the initial attempts to synthesize the behavioural and environmental influences on SBT abundance with regard to the abundance index. The final results section makes best use of all the data and approaches and presents the overall analytical conclusion with regard to the index. This sixth results section should be read in preference to all the others, as it summarizes and presents the "last word" (for now) on the utility of the fishery-independent juvenile SBT abundance index. In the overall conclusion, these results are summarised with regard to the utility of the fishery-independent index for SBT management. In particular, this project has identified the value of the aerial survey to monitor trends in the juvenile SBT population within the GAB.

6.1 Development of a fishery independent index of abundance for juvenile SBT

Ann Cowling

Abstract

This section discusses some of the early stages of the aerial survey analysis and the attempts to estimate the abundance index of juvenile SBT in the GAB. A brief description of the aerial survey method and some preliminary analyses are provided. The survey is limited by the introduction of trainee spotters in the latter two years of the survey, as they have an unknown lower detection rate. In particular efforts to validate the estimates of patch size and school size were undertaken using LIDAR experiment and spotter-validation experiments. LIDAR was not able to provide an alternative technique for surveying SBT schools. Biases between spotters exist, but can be corrected. A conclusion of this work is that several age-classes of SBT occur within a single patch and so estimates of abundance for each age-class are not possible. There appears to be no strong trend in juvenile biomass for the period 1993-2000.

Introduction

All recent assessments of southern bluefin tuna (SBT) indicate that the parental stock is at historically low levels. The current parental stock biomass has been judged to be below “commonly used scientific measures of biologically safe parental biomass” and there are concerns about the risk of poor recruitment and the possibility of recruitment collapse. There is also much uncertainty about whether the current catch level will allow for rebuilding of the SBT stock.

The current analytical assessment methods for SBT have a 4 to 5 year time lag in the estimates of the number of recruits, due to time lags in receiving catch data and the lack of a reliable index of juvenile abundance. In addition, there is much uncertainty about most recent estimates of recruitment as they are largely determined by the most recent juvenile catch rates. Therefore, current trends in recruitment remain one of the major unknowns in evaluating the status of this stock and its potential to rebuild under current catch rates. Moreover, there is no fishery independent information on stock or juvenile abundances. Lack of such information is a major limitation in evaluating the likelihood of stock rebuilding under current catch rates.

All recent scientific and management meetings, both under the previous informal trilateral arrangement and now under the Convention for the Conservation of Southern Bluefin Tuna, have considered the development of a fishery independent recruitment index of SBT to have a very high research priority. In response to this need, a developmental aerial survey program was started in 1990/91, and experimental surveys using line transect methods have been conducted annually during the fishing season in the Great Australian Bight (GAB) since then.

In June 1993 a large scale five-year collaborative program involving CSIRO and the Japanese National Research Institute of Far Seas Fisheries (NRIFSF) was established to monitor the abundance of juvenile SBT and develop a fishery-independent index of juvenile abundance. The aerial survey project was established as one of the main projects in this program. Funding of the aerial survey part of this collaborative research program has come form a variety of Japanese and Australian sources; the Australian sources include CSIRO, SBTMAC, FRDC and FRRF. Each year a workshop is held to review and prioritise the collaborative research for the coming year.

The aerial survey data have been analysed in several different ways since the survey began, and different analyses lead to different conclusions about possible trends in surface abundance of SBT. A major goal of this project is to investigate explanations for the different trends given by different approaches, with the aim of deciding the most appropriate method of analysis. The main index of abundance estimated from the surveys is mean surface biomass density during the survey period in the survey area. The statistical methodology used in the analysis each year is described in [R4], [R6], [R8], [R10] and [R11].

Several reviews of the project have been held to discuss the survey data, its analysis and interpretation. In 1997/1998 an internal review of the aerial survey was conducted. The project collaborators, CSIRO and NRIFSF, held a full review of the project in September 1999; [R1] and [R9]. A further workshop involving international experts was held in February 2000 in Port Lincoln, [R2]. This workshop recommended a method for incorporating environmental conditions into the survey [R4]. This method allows annual surface abundance estimates to be estimated for standardised environmental conditions, allowing direct year-to-year comparison of the estimates. However, further work is needed to understand certain environmental associations before we can be confident of the results of these analyses. These associations are developed in subsequent sections of the report.

Incorporation of estimates of the proportion of SBT at the surface under various environmental conditions

Changes in the proportion of SBT at the surface and hence detectable during an aerial survey will potentially bias estimates of abundance. To allow correction of surfacing rates in abundance models requires information on the proportion of time that fish spend at the surface. Analysis of data from archival-tagged SBT in the GAB will allow estimation of surfacing rates in different spatial, temporal and environmental conditions. Archival tag development and deployment was not funded from this project, although the development of the tag technology has been proceeding during this project component. There were delays in receiving data from archival-tagged SBT in the GAB (as fish with functioning tags were rarely recaptured), which has delayed the development of this stage of the project. The manufacturer of the 1993 to 1995 archival tags had major production problems in 1996 and 1997 and no new tags were made. As a consequence, a new source of archival tags was located and these tags were first placed in the field in 1998. Data from these tags show that the location of SBT can be determined much more accurately compared with earlier model archival tags. A total of 325 archival tags were released between June 1993 and March 1995. To date, 61 of these tags have been returned. However, only seven tags contained reasonable amounts of data for the location and periods covered by the aerial survey. Five of these seven tags were analysed in [R14] using weather observations from the Ceduna weather station as explanatory variables, together with SST from the tags, and moon phase.

In the analysis of these five archival tags [R14], the most effective explanatory variable was found to be time-of-day allowing for four different 24-hourly patterns during a

lunar cycle, each pattern lasting for a week. During the week of the full moon, SBT tended to spend little time on the surface, compared to the other weeks. They also appear to spend more time on the surface when the SST is high. Preliminary fitted models had little explanatory power. This may be because the weather at sea may have little correlation with that at Ceduna weather station.
Another 198 archival tags were released between January 1998 and February 2000. A number of the tags returned to date contain useable data during the aerial survey period. The classification of surfacing behaviour was begun using a more detailed classification of surfacing behaviours than the previous analyses. These classification schemes are updated in a subsequent section of the report, and not in this section.

Incorporation of environmental variables into the estimates

Larger quantities of SBT are detected during the surveys during warm calm conditions. Therefore, there is a need to adjust the survey estimates for the between year and within year differences in environmental conditions. A first method of estimating surface abundance, which incorporates various environmental conditions in a statistical modeling approach, was introduced in the 1998 aerial survey report [R6].

The 2000 Port Lincoln aerial survey workshop ([R2]) agreed that statistical modeling using individual transect lines as the unit of analysis was likely to be an effective analytical approach. A complicating factor is that many transect lines are searched during the surveys without detecting any SBT. In a model-based approach, the large number of zero observations must be incorporated appropriately. This is done using a two-stage model with environmental, spatial and temporal covariates. The first step is to model the probability of presence or absence of SBT and the second is to model the biomass given that SBT were detected. This is a commonly used statistical approach and is more fully described and developed in [R4], using the half line as the unit.

This paper discusses the development of an integrated statistical model based on the recent and growing body of data on surfacing behaviour of SBT in the GAB acquired from archival tags together with detailed environmental data collected in the aerial survey and from other sources. In addition, research using recently available laser technology (airborne LIDAR systems) that detect schools below the surface of the water was conducted to estimate the proportion of surface schools. Detailed description of the statistical methods and results are not provided, as they have been superseded by results presented in the latter sections of this report. Instead, an overview is presented, and references are provided to other documents which completely describe these early analysis stages.

Methods

Development of the aerial survey technique for SBT over the Great Australian Bight each summer season from 1997 to 1999 and estimation of various surface abundance indices

The area of the GAB searched during the surveys lies between 128^{o}E and 135^{o}E, running from the coast to about the 700-800 m depth contour just off the continental shelf. Fifteen equally spaced north-south transect lines are searched during the surveys (**Figure 1)**. Two planes fly in the surveys with two spotters in each plane. During each flight, information is collected about detected schools of SBT. Environmental data (wind speed and direction, air temperature, swell, haze, glare) are also collected at specified intervals during the flights.

The survey takes place over the three months from January to March each year on days when the weather conditions are suitable for survey operations (wind speed less than 10 knots). Each plane is able to search two or three lines per day. Thus, one survey replicate takes between three days and one month to complete, depending on the weather conditions. The survey is replicated four to eight times per season. Further details of the survey design, implementation and methodology are given in [R4], [R6], [R8], [R10] and [R11].

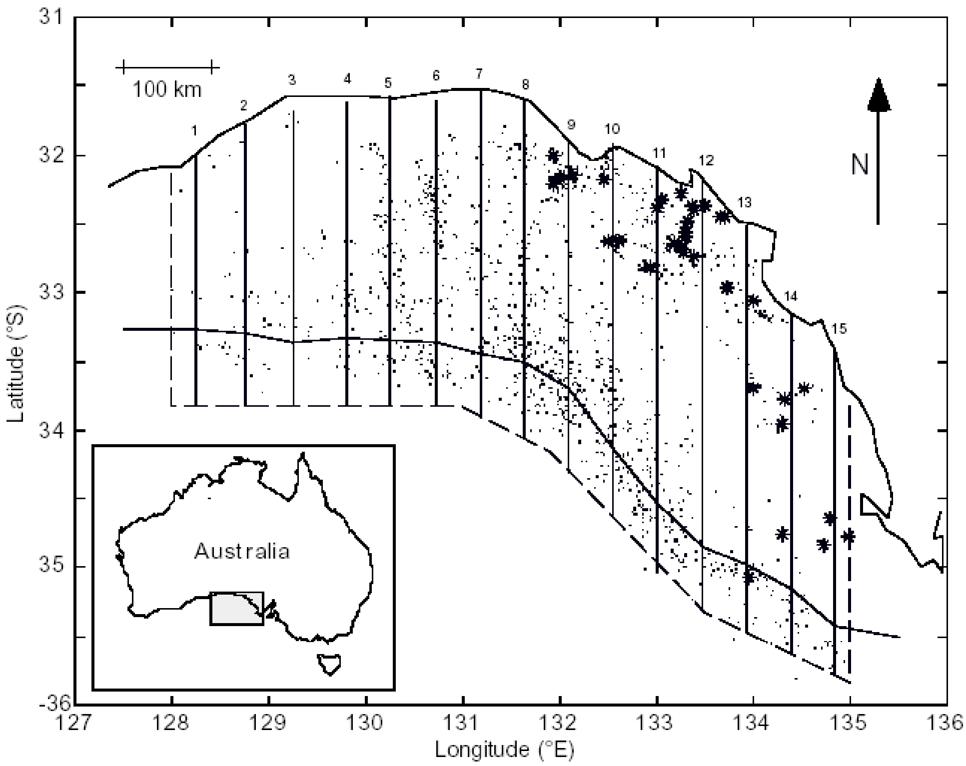


Figure 1: Transect lines for 1999 and 2000 aerial surveys. The shelf break and lumps (*) within the survey area are indicated.

Reduction of the sampling error in the estimates due to uncertainty in school size and fish size estimates

A source of error in the aerial survey data is the estimation of school size and fish size. These errors are important for the development of a biomass-based index of abundance. In this project these errors were investigated with a LIDAR experiment, and with two-plane spotter validation experiments.

LIDAR Experiment (1997)

LIght Detection And Ranging uses the same principle as RADAR, and is a remote sensing technology developed by the US military. The LIDAR instrument transmits light out to a target. The transmitted light interacts with and is changed by the target. Some of this light is reflected / scattered back to the instrument where it is analysed. The change in the properties of the light enables some property of the target to be determined. The returned light can also be analysed to provide an image of the reflecting object. LIDAR has been successfully used in fish survey elsewhere in the world (e.g. [R15], http://swfsc.nmfs.noaa.gov/frd/FY99%20Program%20Review/CRA33.htm).

In 1997 experiments using a LIDAR carried in the survey plane were conducted with the aim of determining whether LIDAR technology can be used to

A) Estimate the size of schools of SBT
B) Estimate the size of fish within schools
C) Detect sub-surface schools of SBT that are not detectable to the spotters.

Two experts from Arete Associates (Tucson, AZ, USA) brought LIDAR equipment to Port Lincoln as part of this project. The LIDAR experiments are described in detail in [R3] and [R12] and at http://swfsc.nmfs.noaa.gov/prd/dsweb/PDFs/swr-99-02.pdf

Spotter-validation experiments

In 1998, 1999 and 2000, validation experiments were designed and conducted to collect independent patch size and fish size estimates to attempt to calibrate the spotters' patch size and fish size estimates. In these experiments, the spotters in several planes simultaneously estimated the patch size and fish size of the same patch. This was repeated for a number of different patches. One plane led and identified isolated patches for the study. When a suitable patch was identified, the lead plane called the other planes to that patch on the radio. The lead plane maintains the lowest altitude and is easily followed by the higher planes. The following planes keep a safe distance from the lead plane but close enough to quickly get to the same patch. The plane at the highest altitude confirms that all planes are looking at the same patch. Further details are given in [R4], [R6] and [R8].

Results

Development of the aerial survey technique for SBT over the Great Australian Bight each summer season from 1997 to 1999 and estimation of surface abundance indices

Annual reports summarising the results of each year's fieldwork, description of the development in the analytical methods and updated indices of abundance have been produced every year since this project commenced; [R4], [R6], [R8], [R10], [R11]. The search effort and sighting rates are summarised in **Table 1**.

Table 1: Search effort and sighting rates of SBT in the aerial survey for the entire GAB; 1993-2000.

	1993	1994	1995	1996	1997	1998	1999	2000
Number of survey replicates completed	4	8	8	7	5	5	4	4
Total flying time (hrs)	213	405	438	332	287	297	238	177
Total time searching (hrs)	112	215	206	173	129	124	77	51
Total distance searched (effort) (nm)	10174	20261	20793	18243	12799	11937	7499	5960
# SBT sightings	267	289	295	186	189	146	56	82
# SBT sightings/100 nm	2.62	1.43	1.42	1.02	1.48	1.22	0.75	1.38

Alternative methods to estimate the abundance of SBT based on the number of observed patches during the survey lead to differing results at this time. The strip transect method of analysis is relatively simple, but does not adjust for differences in weather conditions between years. The second method of analysis, using a statistical modeling approach, allows an understanding of the relative importance of the different environmental variables which effect presence/absence and biomass of SBT. In models used to date it has been assumed that the apparent increase in surface abundance of SBT in higher air temperatures (AT) is related to the association between AT and sea surface temperature (SST). Therefore only AT or SST has been included in previous models. In such analyses including AT, there is a significant decline in presence/absence of SBT between 1993 and 2000. However, when SST is substituted for AT in these models, there is no significant decrease. It is necessary to study the SST/AT relationship further to determine whether AT, SST or both should be included in models of estimated SBT abundance (see **Section 6.5**).

In 1999 and 2000, changes within the South Australian SBT industry meant that only one trained spotter was available to spot in each survey plane. As a result, the project team began to train young spotters to work in future aerial surveys. The effect of using trainee spotters is not clear –they do detect fewer schools than a trained spotter, but it is not clear exactly how much less. Consequently there is greater uncertainty in the survey results for these last two years. The results of the strip transect analysis are shown in **Figure 2**, with upper and lower limits for the estimates for the years 1999 and 2000. The method of calculating the limits is explained in more detail in [R4]. The results of the strip transect analysis indicate that there has been no major increase or decrease in the surface biomass density of SBT between 1993 and 2000.

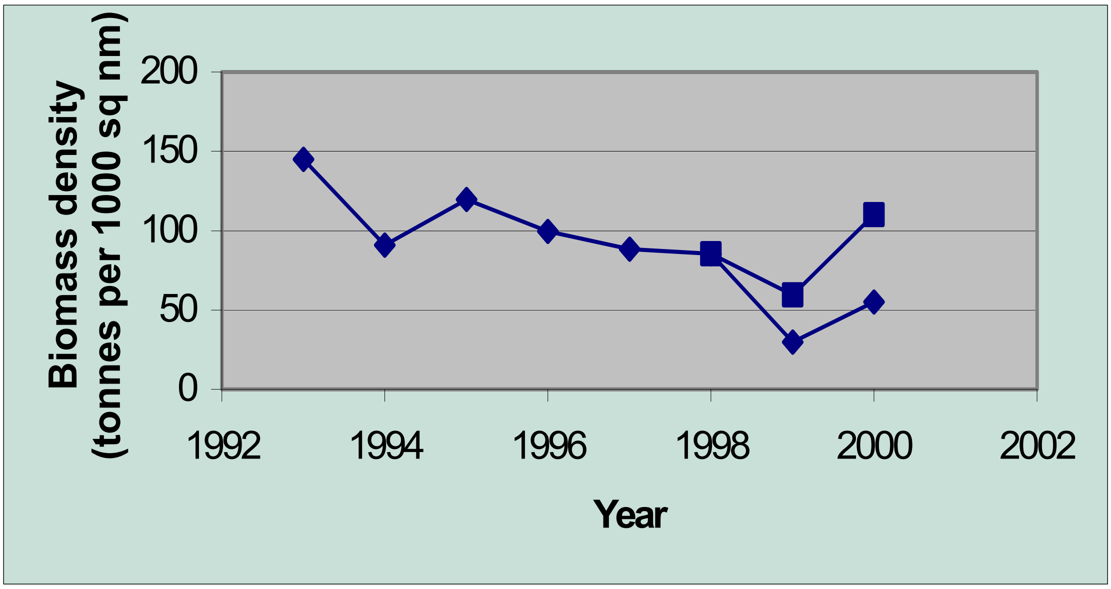


Figure 2. Annual estimates of surface biomass density of juvenile SBT in the GAB, 1993-2000, using the strip transect method of analysis.

Reduction of the sampling error in the estimates due to uncertainty in school size and fish size estimates

LIDAR experiments

The results of the LIDAR trial are given in [R3] and [R12]. They show that while LIDAR has the potential to measure fish size and patch size, the technology has not yet advanced sufficiently to allow school size and fish size to be routinely measured under field conditions. Further work is required on increasing the resolution and developing real time processing of the data. SBT are not highly reflective fish, especially when seen from above, which makes them difficult to see with a laser unless the resolution is improved sufficiently to considerably increase the contrast. Further technological development of the LIDAR instrument was not in the scope of this project.

Spotter-validation experiments

Fish size

A comparison of the fish size estimates given by the spotters in 2 planes in 1998 and 3 planes in 1999 and 2000 are shown in **Figures 3-6**. In each of these years there is little consistency in fish size estimates between spotters. This may be because of the large range of fish sizes within a patch and the short glimpse of fish obtained while circling the patches. This is further evidence that an alternative method of estimating abundance by age-class should be developed if abundance by age-class is to be used in an index.

Further complication regarding estimating the size of fish in a patch is the recent finding that patches are mixed with regard to age and size. For example, analysis of the age composition of SBT caught within individual patches in conventional tagging experiments carried out in the GAB between 1991 and 1997 [R9, p38-39] showed that patches of fish do not comprise a single age of fish as had previously been assumed. Therefore the method used in the aerial survey analyses to estimate abundance by age class in [R10] and [R11] will contain substantial error, as these methods involve attributing a single age to each sighting of SBT based on the estimated dominant age class. Thus, in the future estimates by age class will not be given.

Patch size

A comparison of the patch size estimates given by the spotters in 2 planes in 1998 and 3 planes in 1999 and 2000 are shown in **Figures 7-10**. There is remarkable consistency in patch size estimates between spotters. The correlation between the 1998 estimates is 0.78, and the correlation between the 1999 estimates is between 0.83 and 0.93 for the different pairs of spotters. Although one spotter’s estimates are generally higher than those of the other spotters, because they are so highly correlated, they can be adjusted to a common level each year.

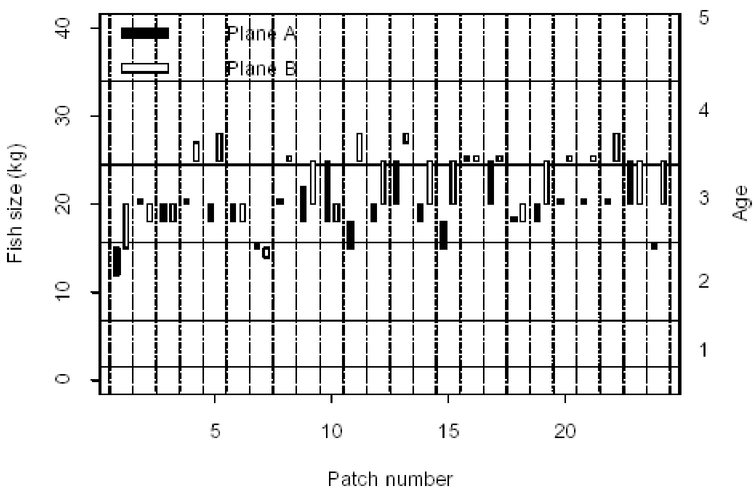


Figure 3: Spotters' fish size estimates, 2 plane experiment, 1998

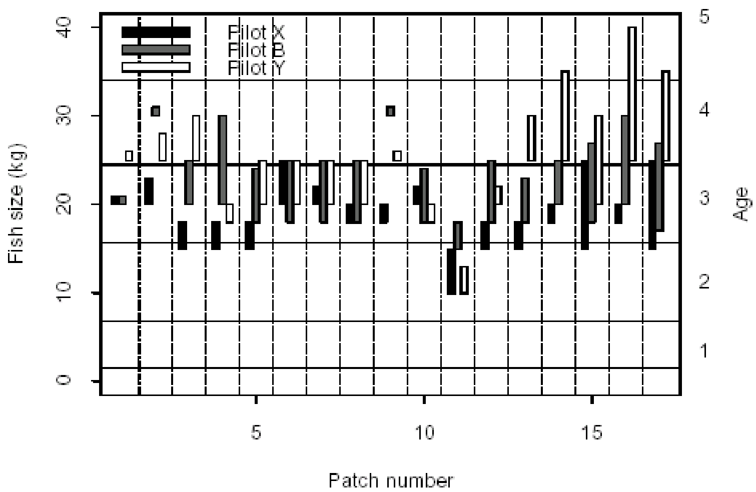


Figure 4: Spotters' fish size estimates, 3 plane experiment, 1999

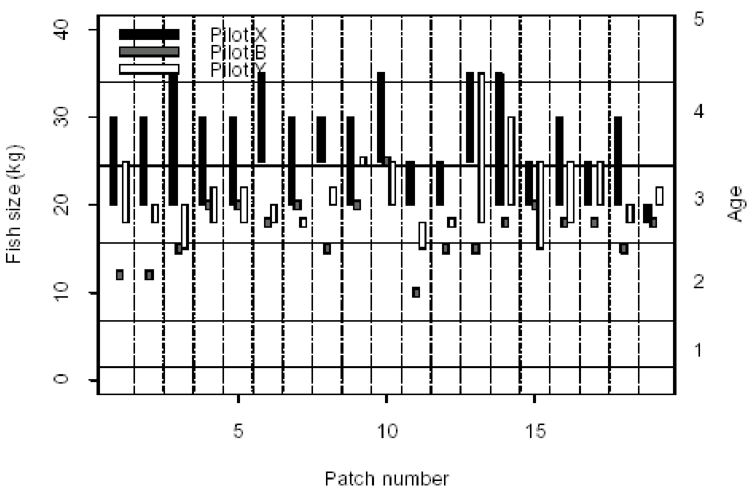


Figure 5: Spotters' fish size estimates, 3 plane experiment, 9 March 2000

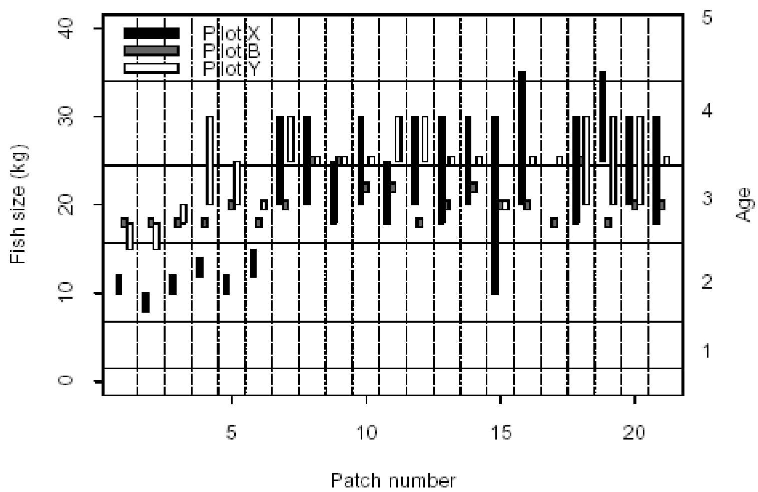


Figure 6: Spotters' fish size estimates, 3 plane experiment, 10 March 2000

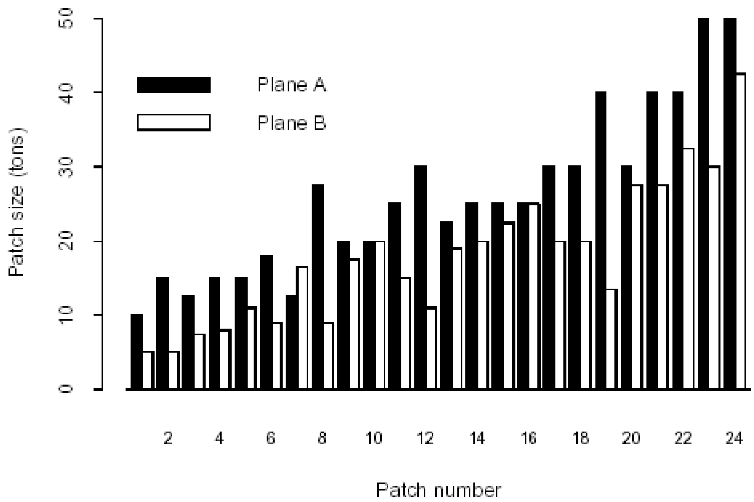


Figure 7: Spotters' patch size estimates, 2 plane experiment, 1998

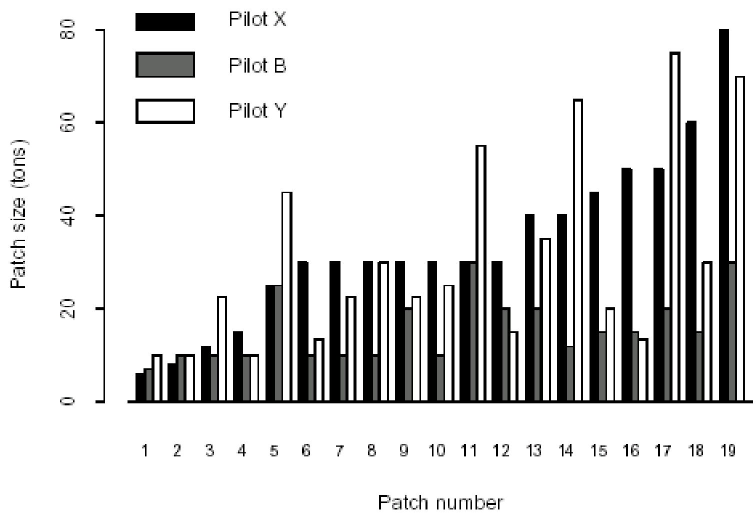


Figure 8: Spotters' patch size estimates, 3 plane experiment, 1999

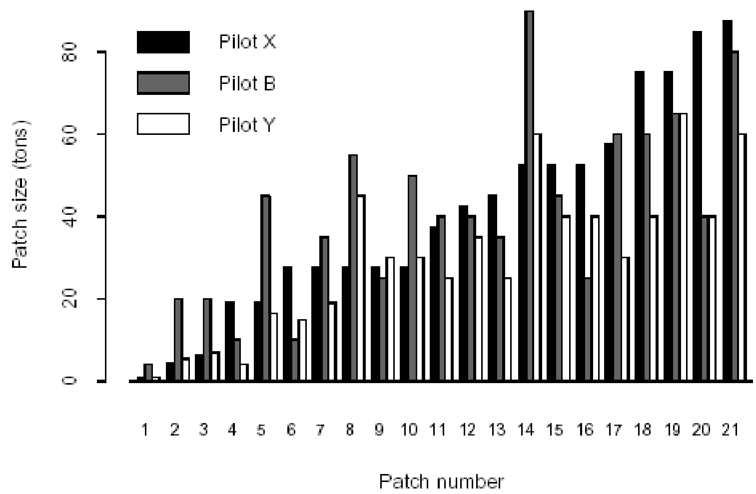


Figure 9: Spotters' patch size estimates, 3 plane experiment, 9 March 2000

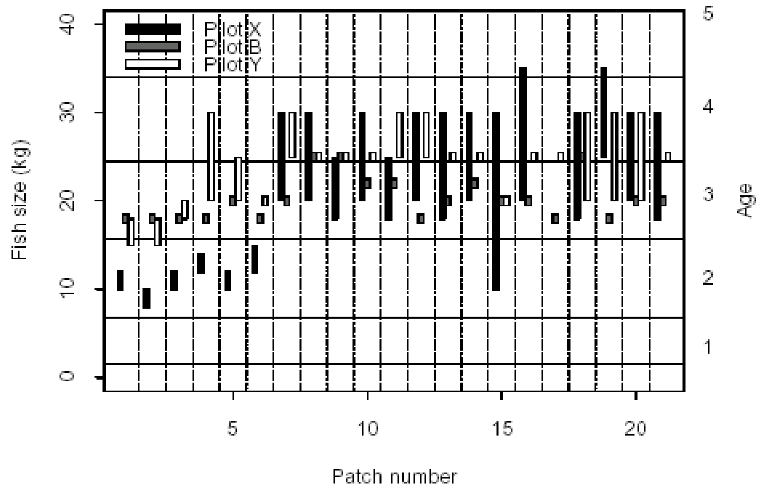


Figure 10: Spotters' patch size estimates, 3 plane experiment, 10 March 2000

Discussion

During this stage of the project a great deal of data has been collected. The analysis of this data has increased knowledge of SBT and their behaviour and led to a revision of assumptions about the detection of SBT from planes and about SBT distribution and behaviour. This improved understanding of SBT distribution and surface abundance in the GAB will improve the final analysis of the data, but the greater complexity of the processes has limited analysis at this stage.

Evaluation of the usefulness of the indices of SBT abundance derived from the aerial survey to date

The strip transect method of analysis is relatively simple, but does not adjust for differences in weather conditions between years. The second method of analysis, using a statistical modeling approach, allows an understanding of the relative importance of the different environmental variables which effect presence/absence and biomass of SBT. Analyses of the survey data have resulted in estimates of the number of schools, total biomass and biomass by cohort for fish at the surface with reasonable coefficients of variation. The estimates have started to provide an initial useful comparison with VPA results, as estimates of some cohorts from the two methods began to overlap in

1995. However, there are still a number of research problems that need to be addressed in order to evaluate whether these estimates can provide a reliable index of juvenile abundance. The problems are associated with the unknown variability in the proportion of schools at the surface, the proportion of juveniles within the GAB, environmental effects on detectability of surface schools and tuna surfacing behaviour, and the reliability of estimates of fish and school sizes.

Further research is needed to improve the reliability of the results including improvements in the estimates of school size, fish size, the effects and interactions of environmental factors on the detection and size of surface schools, and statistical methods for obtaining the variances of the estimates. These efforts are discussed in subsequent sections of this final report.

The biggest source of uncertainty and perhaps the biggest source of variation in the analyses of aerial surveys to date is that no account is taken of the variability in the proportion of schools at the surface. If the proportion of schools at the surface varied little from year to year, this would not be a problem. However, surfacing behaviour of SBT appears to be strongly influenced by environmental conditions. Although the aerial survey is only conducted under weather conditions favourable to tuna surfacing, the aerial surveys to date have encountered substantial inter-annual differences, with sea-surface temperatures being perhaps the most important and variable. The variation in the proportion of surface schools must be accounted for to improve the interpretation of the aerial survey results.

Finally, the current time series of aerial survey indices must be extended and improved. Without temporal extension, it will not be possible to evaluate whether the aerial survey can provide a useful index of abundance. Although results from the full survey period (1993-2000) will overlap the VPA estimates of recruitment for seven cohorts, and this overlap could provide the basis for a statistical analysis of the aerial survey results as an index of recruitment, the inability to reliably estimate the size of fish in a patch will prevent this original project objective from being achieved.
The results from this project to date have shown that there are problems in the estimation of biomass by age class. The low reliability of these estimates means that only the estimates for pooled age classes are considered realistic, however, such pooled estimates will have more limited use. The indices derived from the surveys to date provide a quantitative measure of surface abundance in the GAB. The results suggest that there may have been some increase or decrease in abundance since 1993, but the results since 1993 do not show any major change in abundance. A recruitment collapse has not occurred.

Conclusions

The results of this project to date have shown

- the estimates of biomass by cohort are now considered to be insufficiently reliable for comparison with VPA results, and are no longer reported in the results of the analyses.
- incorporation of surfacing behaviour into the surface abundance estimates is not necessary under one of the models being investigated at present. Under this model, the results and interpretation of the surface abundance analysis is greatly strengthened by the model of surfacing behaviour as it provides independent

verification of the surface abundance model. Under other methods of estimation of surface abundance, it is necessary to incorporate surfacing rates.

The results achieved to date in this project show that the majority of the goals of the project are likely to be achievable, but that to attain them at a level in which most of the information is extracted from the data will require further work after the completion of this project. This project has therefore been extended for a further two years. Subsequent sections of the final report will cover these developments.

Acknowledgments

This research is part of a collaborative research project to monitor the abundance of Southern Bluefin Tuna that is being conducted between Australia (through the CSIRO) and Japan (through the National Research Institute of Far Seas Fisheries of the Japan Fisheries Agency). The research was funded by the Australian government and tuna industry through AFMA and FRDC, and the Japanese government through JAMARC. We thank Song Chen, the scientist in charge of the aerial surveys in 1993-94; Barry Brady, Derek Hayman, Neville Harding, Lyall Jaensch, Andrew Jakobs, Milo Mislov, Kevin Warren, Tony Warren, Simon Bey, Chris McGown, Lana Schultz, Bob Stanley, Diane Stanton, Paul Waide and Mardi van der Weilen for their dedicated survey spotting and recording work in the 1993-2000 surveys; James Forbes, Helen Galea, Colin Millar and Jeremy O'Reilly of CSIRO Marine Laboratories for providing programming support to the project from 1993-2000; Jeff Cordell, Erik Madsen, Alex Papij and Matt Sherlock of CSIRO Marine Laboratories for providing electronic assistance; and Kevin Williams and Deidre Waddington for their support to the survey.

References

[R1] Anon (1999). Aerial Survey Design Workshop. Caloundra, 1999.

[R2] Anon (2000). Report of the Aerial Survey Workshop. Port Lincoln, 2000.

[R3] Cowling, A (1997) Report on the LIDAR trials for 1997. RMWS/97/17. 1997 Recruitment Monitoring Workshop. Hobart.

[R4] Cowling, A (2000) Data analysis of the aerial surveys (1993-2000) for juvenile southern bluefin tuna in the Great Australian Bight. RMWS/00/3. 2000 Recruitment Monitoring Workshop. Hobart.

[R5] Cowling, A, Gunn, J and O'Reilly, J (1999). Integrated Analysis Project: a basis for improved annual estimates of SBT recruitment to the GAB. RMWS/99/10. 1999 Recruitment Monitoring Workshop. Caloundra.

[R6] Cowling, A and Millar, C (1998). Data analysis of the aerial surveys (1991-1998) for juvenile southern bluefin tuna in the Great Australian Bight. RMWS/98/3. 1998 Recruitment Monitoring Workshop. Hobart. Report for the CCSBT Scientific Meeting. CCSBT/SC/9807/8.

[R7] Cowling, A, Hobday, A, O'Reilly, J and Gunn, J (2000). 1999-2000 Integrated Analysis Project Report. RMWS/00/6. 2000 Recruitment Monitoring Workshop. Hobart.

[R8] Cowling, A and O'Reilly, J (1999). Data analysis of the aerial surveys (1993-1999) for juvenile southern bluefin tuna in the Great Australian Bight. RMWS/99/3. 1999 Recruitment Monitoring Workshop. Caloundra.

[R9] Cowling, A and O'Reilly, C (1999). Background to review of aerial survey. Caloundra, 1999.

[R10] Cowling, A, Polacheck, T and Millar, C (1996). Data analysis of the aerial surveys (1991-1996) for juvenile southern bluefin tuna in the Great Australian Bight. RMWS/96/4. 1996 Recruitment Monitoring Workshop. Hobart. Report for the CCSBT Scientific Meeting. CCSBT/SC/96/21.

[R11] Cowling, A, Polacheck, T and Millar, C (1997). Data analysis of the aerial surveys (1991-1997) for juvenile southern bluefin tuna in the Great Australian Bight. RMWS/97/4. 1997 Recruitment Monitoring Workshop. Hobart. Report for the CCSBT Scientific Meeting. CCSBT/SC/9707/21.

[R12] Griffis, A (1997) Report on the 1997 LIDAR trial. RMWS/97/7. 1997 Recruitment Monitoring Workshop. Hobart.

[R13] Kloser, R (1998). Assessment of acoustics as a research tool to determine the distribution, biomass and behaviour of southern bluefin tuna schools and their prey in the Great Australian Bight. RMWS/98/6. 1998 Recruitment Monitoring Workshop. Hobart.

[R14] Millar, C and Cowling, A. (1997) SBT Surfacing Behaviour. RMWS/97/26. 1997 Recruitment Monitoring Workshop. Hobart.

[R15] Schoen C. and J. Sibert. 1996. Feasibility of Dual Mode Lidar for Pelagic Fish Surveys, SOEST Pub. 96-02, JIMAR 96-301

6.2 The influence of topography and environment on presence of juvenile SBT in the Great Australian Bight

Alistair J. Hobday

Abstract

Juvenile southern bluefin tuna (*Thunnus maccoyii*, SBT) are surface-orientated and non-randomly distributed in the Great Australian Bight (GAB) during the austral summer. In particular, SBT are clustered around the shelf break and inshore reefs, islands and rises, collectively known as lumps. The presence of SBT observed during an aerial survey in the GAB (1991-2000) was analyzed with regard to the topographic characters of these features and local environmental variables. A SBT attraction region around these features was derived and related to feature-specific topographic characters. The attraction region at the shelf break spanned an average distance of 23 km to the north and south of the break, and in total covered 11% of the survey area and contained 20% of all SBT sightings. Shelf break topographic characters were not significantly related to the attraction distances. Eighteen of 36 lumps examined were attractive to SBT, with an average attraction radius of 5.2 km. Fifteen percent of all sightings were inside the lump attraction areas which covered just 1.2% of the survey area. There were no significant differences between the topographic characters of attractive and unattractive lumps; however, depth and isolation were significant terms in a multiple linear regression explaining 55% of the variation in the size of the attraction region for attractive lumps. Generalised linear models indicated non-linear relationships between the presence of SBT at topographic features and environmental and topographic variables, and models incorporating topography and the environment explained 40% and 28% of the deviance at the lumps and shelf break respectively.

Introduction

Tunas are wide-ranging pelagic species found in most temperate and tropical oceans (e.g. Sund et al. 1981), however; their distribution is not uniform. The distribution of pelagic species, like tuna, can be influenced by spatially fixed topographic features and temporally variable oceanographic conditions either alone or in combination (Boehlert 1987; Borcard et al. 1992; Hooker et al. 1999). In particular, associations between seafloor topography and biology have been found in many regions and for many taxa (e.g. Boehlert 1987; Dower et al. 1992). These relationships may exist because topographic features such as seamounts (Boehlert 1987), shelf breaks (Young et al. 1996; Gremillet et al. 2000), reefs (Genin et al. 1994), and islands (Hunt et al. 1996; Kleiber and Hampton 1994) can aggregate marine species from a variety of trophic levels. Physical processes at topographic features may lead to concentration and retention of nutrients and increased primary production (Roden 1987; Wolanski and Hamner 1998; Chapman and Haidvogel 1992). This may in turn concentrate relatively passive lower trophic level consumers (Freeland 1994; Genin and Lonsdale 1989), and so on until the apex predators are also aggregated in these regions (Croxall and Prince 1996; Gremillet et al. 2000). Aggregation of individuals from higher trophic levels, such as tuna, may be related to concentrations of lower trophic levels (food), rather than the feature itself or its retentive circulation (Holland et al. 1999). Despite this general understanding, the underlying cause behind aggregation of tuna at topographic features remains elusive.

Fishers and scientists have also recognised relationships between the local environment and the abundance of fish, independent of topography. A problem in understanding environment-fish relationships is that although fish live in a three dimensional environment, the human interaction is at the two-dimensional interface between the ocean and the atmosphere. This has generally restricted the environmental information used to explore biological relationships to surface variables, such as sea surface temperature (SST). Rigorous investigation of environmental relationships has also been limited by the spatial and temporal coverage of the physical data. Although temporally matched in-situ measurements may be the preferred data for exploring environmental influences, remote sensing offers the potential for identifying environmental relationships on a wider temporal and spatial scale, and when in situ data is unavailable (Lasker 1978). Since the late 1960's planes have provided remotely-sensed surface ocean information used in fishery studies (e.g. Hynd 1968, 1969; Laurs et al. 1977). When satellites began to carry sensors, SST and ocean color information became available on even larger scales (Fiedler et al. 1984; Laurs et al. 1984; Myers 1984). High explanatory power from environmental and topographic relationships alone is not assured, however, as fish also respond to prey, predators and conspecifics, all of which can influence spatial and temporal distribution.

Southern bluefin tuna *Thunnus maccoyii* (SBT) is a long-lived, highly mobile fish found throughout most of the southern hemispheres temperate oceans (Caton 1991). It is believed to comprise a single stock with one known breeding area in tropical waters south of Indonesia (Grewe et. al. 1997; Farley and Davis 1998; Polacheck et al. 2000). Young-of-the-year juveniles migrate southward along the west coast of Australia, and then east to the Great Australian Bight (GAB). Juveniles between ages 1-5 years are found in the GAB during the austral summer (December – April) (Polacheck et al. 2000) where they form large schools visible at the ocean surface and are targeted by a surface fishery. The surface-orientated behaviour of juvenile SBT allows detection of schools by commercial spotting pilots who direct fishing activity, and aerial surveys designed to assess stock status (Chen et al. 1995, Cowling and Millar, 1998).

The GAB is the only known summer location for juvenile SBT (Caton 1991), indicating the general environment is attractive, however, there is little understanding of how the local environment influences the distribution within the GAB. Within the GAB concentrations of SBT are observed at topographic features such as the shelf break and inshore reefs. The time-varying environment might also influence patterns of abundance at these topographic features. Information about the response of SBT to topographic and environmental factors is important in interpreting changes in local abundance such that better predictions for more efficient catch or better stock assessment for this commercially important species are possible. The goal of this study was to identify the important topographic characters and local environmental variables and explore their relationship with SBT aerial survey sightings in the GAB. This understanding is important in the development of a fisheries-independent juvenile SBT abundance index for the GAB, as it may lead to reduction in the uncertainty of abundance estimates in the index (Chen et al. 1995).

Methods

Biological data

Sightings of SBT were recorded during dedicated aerial surveys in the GAB coordinated by CSIRO in the months January to March between 1991 and 2000 (Chen et a. 1995; Cowling and Millar, 1998). These surveys covered the shelf break, shelf area and inshore topographic features, an area of approximately 158,628 km^2 (**Figure 1**). A sighting was a single detection of SBT at a location, and may have included a number of schools or patches of fish at the surface. Analysis of the sighting patterns was undertaken at the shelf and inshore features because of the higher concentration of sightings at these locations.

GAB bathymetry

Bathymetric data for the GAB shelf break and inshore features were taken from the Australian Geological Survey Organization (AGSO) Bathymetric 30-second Grid for the Australian Region.

Shelf break

The shelf break is often defined as the 200 m depth contour (e.g. Nybakken, 1997), however, a more rigorous definition is the depth at which the slope changes most rapidly at the junction of the continental shelf and the continental slope. Bathymetric data were extracted along the 15 transect lines (shelf segments) of the aerial survey to determine the location of the shelf break. The shelf break was identified along each line as the point where depth changed most rapidly (i.e. where the minimum second derivative of depth occurred). The position (longitude and latitude), depth, gradient, rate of change of gradient, and the overall slope between 100 and 300 meters depth at each location were used as predictor variables in subsequent analyses.

Lumps

"Lumps" is the generic term used by fishers in the GAB to describe a variety of seafloor topographic features, which were the location for historical SBT pole and line fisheries. In this study lumps were all the non-shelf topographic features, including bottom rises that do not break the surface, islands, and reefs. Lumps are concentrated in the eastern half of the survey area (**Figure 1**). SBT fishers provided the locations of 21 lumps to aid the initial design of the aerial survey (pers. comm. Ann Cowling, CSIRO). From inspection of marine charts an additional 15 lump locations similar to the nominated features were included in the analysis (**Table 1).** Each location was used as the center of a 20' x 20' (37 x 37 km) bathymetric data box extracted from the AGSO dataset. Processing and visualization in Matlab identified the highest point of each feature and refined the location according to the bathymetric data. The location (latitude and longitude), distance from the transect line, distance from the mainland, average distance from all other lumps (isolation), depth of the highest point of the lump, height of the lump, and maximum, average, and variation in depth in the box, were used as predictor variables in subsequent analyses.

Attraction regions

Because different sized features might influence tuna over different spatial scales, regions of influence for tuna sightings were calculated for each topographic feature. This attraction region is defined here as the distance within which more SBT sightings were recorded than expected if sightings were randomly distributed with distance from the feature. The attraction distances were found using the sightings for all years, as described below.

Attraction distance: Shelf break

The distance of each sighting from the closest point in the shelf break, linearly interpolated between the closest two shelf break points on the 15 transect lines, was calculated. Sightings could be to the north or to the south of the shelf break, and the attraction distance was calculated in each direction. Sightings in each north-south strip (parallel and centered on the transect lines) were ranked by distance from the shelf break. The attraction distance at each transect point of the shelf break was the distance at which the distance between ranked sightings increased to more than three times the mean distance between ranked sightings. The mean distance between ranked sightings was calculated using all sightings within 100 km to the north and south of the shelf break within that transect strip, respectively.

Attraction distance: Lumps

The distance of each sighting from each lump was calculated and ranked by distance for each lump. The attraction distance was defined as the distance at which the distance between ranked sightings increased to more than twice the mean difference for all sightings within 50 km. If the initial distance between sightings exceeded twice this mean difference, or did not exceed twice the mean within the distance considered, the attraction distance was defined as zero.

The frequency of sightings each year within individual and combined lump and shelf attraction areas was also calculated. In addition to attraction distance, a second measure of attractiveness, the percentage of all flights past a feature that detected SBT, was calculated and is hereafter referred to as "success". Attractive lumps or shelf segments by this measure were defined as those with success greater than the mean success for all lumps or shelf break segments respectively.

Environmental data

The aerial survey lines were divided into segments during the survey, and environmental variables recorded for each segment regardless of whether SBT were detected. These variables included swell, wind speed, and air temperature. Data from all flight segments that came within 10 km of the lumps or the shelf break were extracted from the aerial survey database maintained at CSIRO. A distance of 10 km was chosen based on maximum attraction distances at the features. Additional data used to characterize the local environment for these close flights included remotely-sensed SST, surface colour (a proxy for chlorophyll a, CHL), and sea surface height (SSH), and moon illumination.

The SST and CHL data were one kilometer spatial and daily temporal resolution. The SST, available between 1993 and 2000, may be biased slightly low (0.3-0.4°C) compared to in situ measurements, whereas the accuracy of the CHL estimates, available between 1997 and 2000, is within 50% of in situ measurements (Polovina et al. 2000; pers. comm. Chris Rathbone, CSIRO). The mean value and gradient (max-min) of SST and CHL in the 10 km region around the feature were matched to each close flight. A front-finding algorithm adapted from Cayula and Cornillon (1992) was used to locate fronts from SST and CHL data. The distance from the feature to the nearest SST and CHL front, and its length, gradient, and value was found for each close flight.

Sea surface height was obtained from satellite altimetry data from TOPEX/Poseidon, with 10 day temporal and 0.25° (~27 km) spatial resolution. Data from 1993-1998 were processed by CSIRO to correct for tides and the SSH from the closest date for each close flight were obtained from these composite images.

Daily moon illumination data were obtained from the US Navy website (http://aa.usno.navy.mil/AA/). The ratio of the moon's area illuminated by direct sunlight to its total area, multiplied by 100, is the percent of the moon's surface illuminated.

Analyses

The relationships between sightings, topographic and environmental variables were explored in a variety of Matlab routines. T-tests were used to compare the topographic properties of attractive and unattractive lumps and shelf segments using both measures of feature attractiveness. Linear regression was used to explore individual relationships between predictor and response variables at the shelf break and lumps. Multiple linear regression (SYSTAT) was also used to determine the suite of significant topographic variables related to attractiveness for lumps and the shelf break. An alpha of 0.05 was used as the significance level in all tests.

Generalised additive models (GAMs) and generalised linear models (GLMs) were used (S-plus software package) to analyze relationships between environmental and topographic variables and the presence or absence of SBT at the lumps and shelf break for all flights. The GAM technique is a non-parametric generalization of GLM, which is in turn a parametric generalization of multiple linear regression methods (Hastie and Tibshirani 1990; Bigelow et al. 1999). In cases where data was missing from a flight record, the record was excluded. Two models were developed, as different topographic characters existed at the lumps and shelf break and different environmental variables may be important. Because of different periods of satellite data availability, and hence exclusion of records, three periods were considered,

1. 1993-2000, SST available (no CHL or SSH variables).
2. 1993-1998, SST and SSH available (no CHL variables).
3. 1998-2000, SST and CHL available (no SSH variables)

In the development of each model, a GAM was first constructed using the continuous and discrete predictor variables available for each period. The presence or absence of a sighting close to a lump or the shelf break segment formed the Bernoulli response variable included in the GAM. In order to make explanatory terms in the model additive, the logit (log odds) link function (log(P/(1-P)) was included in all models. Consequently it was the log odds that were predicted by the fitted model. The range of each predictor variable was restricted if outliers influenced the fitted GAM regression. The shape of the fitted regression for each continuous predictor variable in the resulting GAM was used to select the order of a polynomial describing that variable in a GLM. The entry order of the terms was varied in the GLM process, which did not affect the overall variance explained, but did alter the significance of terms in the model. The minimum number of significant terms was retained in the final GLM. A psuedo-R^2 coefficient, the fraction of the total deviance explained by the model, was used as a measure of the explanatory importance of the final model (Maury et al. 2001). This value was compared to the distribution of 1000 psuedo-R^2 values obtained if the presence/absence values were randomly assigned to the set of predictor variables for each flight included in the model. This re-sampling preserved the correlation structure of the predictor variables, and tested the hypothesis that the observed model was an improvement over a random association of sightings with the specific environments for flights past the topographic features.

Finally, the environmental conditions at the attractive and unattractive topographic features were compared to determine if differences in the environment might explain differences in attractiveness of the features for SBT.

Results

A total of 1731 SBT sightings were made during the aerial survey months between 1991 and 2000 in the GAB. Sightings were most common in water depths between 50 and 250 meters, and were concentrated at the shelf break and nearshore lumps (**Figure 1**). The observed depth distribution of SBT did not differ from a depth distribution obtained for the same number of random locations through the survey area **(Figure 2**), suggesting SBT were not selecting waters of a particular depth within the GAB.

Shelf and lump topographic characters

The shelf break on each of the transect lines ranged between 140 and 190 meters in depth (**Figure 3 B**). The slope between 100 and 300 meters was steepest in the west (**Figure 3 A)**, however, the gradient at the shelf break did not have any trend with longitude (**Figure 3 C**). The rate of change in gradient increased to the east **(Figure 3 D).** The longitudinal contrast in the patterns between the slope (100-300 m) measure and the change in gradient measure exists because in the west bottom depth changes more consistently between 100 and 300 meters over a shorter horizontal distance, whereas in the east there is a wide shelf area between 100 and 300 m depth, with a more sudden increase in depth at the shelf break. The lump characters had little systematic pattern with regard to geographic location, and are described simultaneously with the attraction distances in the following section.

Attraction regions and topographic relationships

Examples of the calculation of the SBT attraction region are shown in **Figure 4**. Calculation of the attraction region was somewhat sensitive to the exact form of the definition. In general, exceeding the mean difference between ranked sightings by two to three times resulted in similar numbers of lumps (**Figure 5 A**) and shelf segments (**Figure 5 D)** with attraction distances. The mean attraction distance was similar for the lumps when defined as the distance where the difference between ranked sightings exceeded the mean difference by 2-4 times (**Figure 5 B),** whereas the mean shelf attraction distance increased with an increase in the threshold (**Figure 5 E).** The total number of sightings within the attraction distances at the lumps (**Figure 5 C**) and the shelf break (**Figure 5 F**) reflected this pattern. Increasing the distance over which the mean difference between ranked sightings was calculated also affected the results, and exploration with a variety of values led to a distance of 100 km for the shelf break and 50 km for the lumps being chosen. Larger distances are unrealistic, as the scale of influence over SBT is unlikely to be of such a magnitude, whereas smaller distances at the lump and the shelf break resulted in attraction distances similar to the total distance considered.

There was no significant correlation between the magnitude of the two measures of attractive/unattractive features for the lumps ($F_{1,16} = 0.95$, $p < 0.35$, $R^2 = 0.05$) or for the shelf break ($F_{1,13} = 0.45$, $p < 0.52$, $R^2 = 0.034$). Twelve of the same lumps were identified as attractive using the attraction distance definition (n_{total}=17) and the success>mean(success) definition (n_{total}=15) (**Table 1, Figure 6).** All but one of the shelf-break segments were attractive by the first definition (**Figure 6 A)**, whereas using success>mean(success) divided the shelf into five attractive and 15 unattractive segments (**Figure 6 B**).

Shelf Break

The SBT attraction distances were not significantly different north and south of the shelf break ($t_{24} = 0.64$, $p < 0.53$) (**Table 2, Figure 6 A)**. The total attraction distance at the shelf break averaged 23.2 km, and a total of 345 sightings were observed within the combined shelf attraction regions (**Table 2).** This number represents 20% of all the sightings, in an area 10.7% (17,004/158,628 km^2) of the total survey area. Overall, there was little interannual variation in which shelf segments contained most of the sightings made within the attraction regions (**Table 2)**. Shelf segment 9 had an average of 24% of all sightings in the attraction regions of the shelf over all years, followed by segment 12 with 14%, and segments 10 and 11 each with 11%.

The topographic characters of the shelf break where the aerial survey transect lines crossed were all non-significantly related to the size of the attraction distance (**Figure 7 A-F).** The attraction distance at the shelf break increased non-significantly with shallower depth (**Figure 7 C**, R^2=0.20) and lower gradients (**Figure 7 D** R^2=0.13). A more rapid change in gradient at the shelf break was associated with an increased attraction region (**Figure 7 E)**. This apparent mismatch between the pattern with respect to the gradient and the rate of change of the gradient indicates an increased attraction region at the shelf break was associated with a sharp and sudden drop-off from a relatively flat shelf. This is confirmed by the shape of the relationship between the attraction distance and the total slope between the depths 100 and 300 m (**Figure 7 F,** R^2=0.11). The success of flights close to the shelf break was highest for the middle portion of the survey region (**Figure 7 G)**, with about 16% of flights recording a tuna sighting. The mean success for all segments was 7.59%; five middle segments had a success greater than the mean, and could also be considered attractive to SBT (**Table 2)**. The only significant linear relationship with shelf topographic measures using success as the measure of attractiveness was the slope between 100 and 300 m ($F_{1,13}$=6.692, $p<0.05$, R^2=0.34) (**Figure 7 H)**.

There were no significant predictors of the attraction distance around the shelf break in a multiple linear regression analysis, reflecting the lack of significant single linear predictors. When the success of flights across the shelf break was used as the response variable, the only significant predictor was the slope between 100 and 300 m, which produced an identical result to the linear regression using just this variable.

Lumps

Of the 36 lumps examined, 18 had an attraction region, although Lump 7, Hamburger Hill, was not included in subsequent analyses as it was within the shelf attraction distance (**Figure 6 A)**. The mean attraction distance was a radius of 5.18 km from a lump (range: 1.83 - 11.68 km) **(Table 3)**. A total of 259 sightings (15% of all sightings) were observed within the attraction region of these lumps, in an area 1.21% (1,919/158,628 km^2) of the total survey area. As for the shelf break segments, there was little change between years in which lump attraction region contained most of the sightings within the attraction regions (**Table 3)**. Lump 21, Yatala, averaged 26% of all sightings in the attraction regions, followed by lump 13, Nuyts Reef, (19%) and lump 18, West of Ward, (12%).

There were no significant differences in the topographic characters between lumps that were and were not attractive to SBT using the attraction distance measure of attractiveness (t-tests, all p>0.16) (**Figure 8 A-J**). In general, attractive lumps were slightly taller (**Figure 8 D,** mean feature height 48 m vs. 44 m), located in areas of less variable topography (**Figure 8 G**, mean standard deviation in depth, 2.27 vs. 2.88), and further offshore (**Figure 8 H,** mean distance to mainland 42.3 km vs. 32.8 km) than unattractive lumps. The relationship between topographic characters and lumps was explored in more detail for the attractive lumps using linear regression (**Figure 9 A-J)**. There was no significant relationship between the attraction distance and geographic location, distance to the mainland, or distance to the transect lines ($p>0.10$, $R^2<0.15$) (**Figure 9 A B H I)**. Lumps close to the surface had significantly larger attraction distances than more submerged lumps (Regression $F_{1,15}=9.03$, $p<0.009$, $R^2=0.38$) (**Figure 9 C)**. Larger attraction distances were also significantly positively related to lump size (Regression $F_{1,15}=10.47$, $p<0.006$, $R^2=0.41$) (**Figure 9 F)** and isolation (Regression $F_{1,15}=7.56$, $p<0.015$, $R^2=0.33$) (**Figure 9 J)**. The attraction distance declined non-significantly with shallower average and maximum depth around the lump ($R^2=0.22$ and $R^2=0.16$) (**Figure 9 D E)**, and with increasing heterogeneity in bottom depth ($R^2=0.20$) (**Figure 9 G)**.

The mean success of flights past lumps was 21.85% (**Table 1)**, and 15 lumps had success greater than the mean, and could be considered attractive using this definition. Comparing the topographic characters of attractive and unattractive lumps using this measure (**Figure 8 A-J)**, showed attractive lumps were significantly west of unattractive lumps (**Figure 8 A**, $t_{15,20} = 3.46$, $p < 0.002$), and further offshore (**Figure 8 H**, $t_{15,20} = 2.44$, $p < 0.02$). No other characters were significantly different. Linear regression showed the success of close flights was significantly for lumps in the west ($F_{1,33} = 10.08$, $p < 0.004$, $R^2=0.23$ **Figure 9 K)** with a maximum success of about 60% at Yatala. Success was also significantly higher for flights past the more offshore lumps ($F_{1,33} = 8.82$, $p < 0.006$, $R^2=0.21$) (**Figure 9 M, Figure 6 B)**.

Multiple regression analysis showed only isolation and depth were significant in a backward stepwise model describing the size of the attraction region around the lumps ($F_{2,14}= 8.70$, $p<0.005$, $R^2=0.55$). When success of flights close to lumps was used as the response variable, the same variables identified individually, longitude and distance to the mainland were significant predictors in a model explaining 50% of the variance ($F_{2,32}= 15.8$, $p<0.0001$, $R^2=0.50$).

Environment at the shelf and lumps for close flights

Data from 1991-1992 were excluded altogether from this portion of the analysis because little in situ environmental data was collected during these survey years. Between 1993-2000 a total of 693 flights crossed the shelf break, and SBT were detected within 10 km of the shelf break on 197 (28%) of these flights, whereas SBT were detected on 299 of the 918 (33%) flights which passed within 10 km of a lump. These subsets were analyzed separately with regard to the presence of SBT and the environment and topographic characters at each feature type. Cloud cover in the GAB led to spatial and temporal gaps in the satellite coverage, affecting CHL more than SST, and SSH least of all, and differences in the deployment of the satellites affected the period of data availability.

Environment and topography at the shelf and lumps

Preliminary multivariate analyses, using GAMs and GLMs, of relationships between environmental and topographic variables and the presence/absence of SBT at the lumps and shelf for the three time periods showed some variables were rarely or never significant predictors. In particular, CHL and SSH variables were not significant predictors. In the case of CHL predictors, this may be due to the limited period of data availability, and this variable should be included in future studies with greater temporal coverage. Thus, only a single model based on data available from 1993-2000, but not including CHL or SSH predictor variables, is reported here for both the shelf break and the lumps.

Shelf Break

The final GLM for the shelf break, based on a trimmed subset of 604 observations, included seven significant variables and had a psuedo-R^2 of 0.28. This compared with the mean and maximum psuedo-R^2 of the re-sampled observations of 0.033 and 0.06 respectively (**Figure 10 A**). The continuous variables were all fitted with 3rd order polynomials, except longitude was approximated with a 2nd order polynomial (**Figure 11**). The significant topographic predictors for sightings at the shelf break were the longitude and the slope between 100 and 300 m. There was a higher probability of seeing SBT in the central portion of the shelf break, and in regions where the slope was moderate. The probability of sightings decreased with an increase in both wind speed and swell. Air temperature was significant, but most of the effect was due to an increase in probability at low temperature; above 15°C there was little signal. A slight increase in sighting probability occurred with an increase in SST, which was the weakest variable in the model. The probability of sightings was high in 1993, 1997 and 1998, and declined in the most recent years.

Lumps

The final GLM for the lumps, based on a trimmed subset of 793 observations, included eight significant variables, and had a psuedo-R^2 of 0.40. This compared with the mean and maximum psuedo-R^2 of the re-sampled observations of 0.031 and 0.06 respectively (**Figure 10 B**). The continuous variables were all fitted with 3rd order polynomials (**Figure 12**). There were five significant topographic variables included in the final model. There was a higher probability of sightings in the west (longitude), a lower sighting probability at intermediate distances to the mainland and higher probability both near and far from the mainland. Bigger lumps (lump height) were associated with higher sighting probabilities, isolation from other lumps reduced the sighting probability, and an intermediate depth below the surface was associated with the highest sighting probability (lump depth). The two environmental variables were SST and wind speed, which were positively and negatively associated with increased sighting probability respectively. The sighting probability at the lumps decreased from January to March, whereas at the annual scale the probability of sighting SBT has decreased since a high in 1993-1994.

Environment at attractive features

The environmental variables were compared between the attractive and unattractive features defined using the success measure. (Results were similar using the attraction distance measures that are not presented here). This provides further insight into why some features were more attractive. For example, mean SST was higher at the attractive lumps, whereas at the shelf break, the mean SST did not differ (**Figure 13 G)**. In both locations there was less variability in SST around the attractive features (**Figure 13 H).** Interestingly, mean wind speed was lower at attractive lumps (**Figure 13 A**), suggesting SBT may have been more easily detected at "attractive features". Air temperature was higher at the attractive lumps, but not at the shelf break (**Figure 13 B**). Although CHL was not examined in the multivariate analyses due to lack of temporal coverage, it was lower at attractive lumps and shelf break segments (**Figure 13 E).** As noted previously, this variable deserves close attention when additional data becomes available.

Discussion

Juvenile southern bluefin tuna did not appear to favour waters of a particular depth in the GAB aerial survey area, as observed depths were consistent with the depth distribution obtained using random locations. There were, however, areas with a greater proportion of tuna sightings than expected if the distribution was random and these areas were the focus of this study.

Two measures of SBT attractiveness were considered for two classes of features, lumps and the shelf break. Attraction regions, estimated using the first method, contained a higher portion of sightings then expected on the basis of the total area surveyed for both feature classes. The attraction region was smaller around lumps, perhaps because of the smaller size of these features compared to the shelf break. It was not possible to calculate the feature attraction distance for individual years due to lack of data, but interannual variation in the size of the attraction regions is possible. Using the second measure of attractiveness, sighting success of close flights, showed the most attractive lumps were more attractive than the most attractive part of the shelf break. This alternative to the attraction distances measure showed similar results, although some different topographic variables were identified as significant by each method.

Although this analysis concentrated on the shelf break and lumps, it is also possible to calculate attraction distances for SBT sightings independent of bathymetric features, using each sighting as the potential centre of attraction. Many of the same attraction areas identified in the lump and shelf break analyses were found (Hobday unpublished data), independently supporting an association between topography and SBT. There were also areas not associated with lumps or the shelf break which were attractive to SBT, indicating attraction to an area can exist in the absence of topographic features; a pattern which should be considered further.

No shelf break topographic characters were significantly related to the attraction distance at the break. In fact the pattern of SBT sightings and attraction distances appeared to be more related to the geographic location of the shelf break. The influence of topographic characters in explaining patterns of SBT sightings at the lumps was demonstrated when only the attractive lumps were considered, with height and isolation of the lump explaining most variation in the magnitude of the attraction distance. The success measure identified longitude and distance offshore as important for success at the lumps. Multiple linear regressions also identified these variables and explained about 50% of the variance for both measures. These results indicate the temporally-constant spatially-fixed characters of the lumps and shelf break alone cannot explain patterns of tuna sightings in the GAB.

Inclusion of the temporally-variable environmental variables for flights past the bathymetric features was expected to increase the explanatory power of models describing the sighting patterns. For example, the shelf break may be a common location for thermal fronts and the cross-frontal transfer of nutrient rich offshore slope waters with the thermally stratified, nutrient poor shelf waters may stimulate higher biological production (e.g. Podesto et al. 1993) and attract SBT. Associations with surface features such as SST fronts are known for a variety of pelagic species (e.g. Podesto et al. 1993; Polovina et al., 2000). Seasonal movements of young northern bluefin tuna have been linked to the seasonal movement of SST fronts (Humston et al. 2000), and the large-scale migration of SBT outside the GAB may occur along frontal regions (Gunn unpublished data). Other SST features, such as spatial SST gradient, distance to the nearest front, and persistence of fronts are reported to influence swordfish (Podesto et al. 1993) and yellowfin tuna (Maury et al. 2001) CPUE in the Atlantic. The first two measures of frontal activity were used in this study, but were not significant predictors of SBT presence within the GAB.

The GLM approach identified topographic and environmental variables related to the probability of detecting SBT at the lumps and the shelf break during the aerial survey. Recall for SBT presence to be detected, fish had to be sighted at the surface. Fish not sighted can be present, or not sighted and not present; if sighted they were obviously present. Thus, an understanding of the SBT relationship with topography and the environment is impeded by the possible failure to detect fish when they actually were present, and the environment and topography were "suitable". In light of this, the significant variables from the GLM analysis can be considered in four categories. The first set includes variables which likely influence the detection of SBT at the surface, particularly swell and wind speed at the shelf break and wind speed at the lumps. As wind speed and swell increased, the probability of sighting SBT declined, perhaps because the fish were deeper or could not be seen from the plane due to surface disturbance. The spatial homogeneity of variables like wind and swell across the GAB make it unlikely SBT change geographic location in response to patterns in these variables.

The second set of variables identified includes the fixed topographic characters; a number of these were important at both locations. Longitude was the common variable at both feature types, this is likely to be a proxy for some other unknown variable, unless SBT actually "prefer" or "home" to a specific longitude. The remaining topographic characters indicate how bathymetry in the GAB influences the SBT distribution. At lumps for example, SBT appear to favor clusters (less isolated) of tall (height) offshore (distance to mainland) lumps.

The third category of variables within the GLMs comprises the environmental measures. SST was the most important significant variable at the lumps (partial variance explained), but the least important significant variable at the shelf break. At both feature types, increased SST was associated with an increased sighting probability, and was higher at the attractive features, suggesting SBT prefer warmer water. They do not occur in the warmest part of the GAB (north-west), however, indicating other variables are also important. The only other significant environmental variable, out of more than 20 considered, was air temperature at the shelf break. An increase in sighting probability with air temperature only occurred at low temperatures, and likely reflects the cooler seasonal conditions before SBT have arrived in the GAB. Chlorophyll and associated measures were not significant in the preliminary models, perhaps due to the short (3 year) period of data availability. Chlorophyll may be a proxy for productivity and tuna prey availability (e.g. Lehodey et al. 1998; Young et al. 2001) and should be considered in future work. Sea surface height variables were also non-significant. These were included in preliminary models as they identify patterns of water circulation, such as upwelling zones and cold and warm core rings. At larger spatial scales, SSH variables may be more important predictors of tuna and other pelagic species (e.g. Polovina et al. 2000).

The final category of variables in the two GLMs included the temporal variables, year and month. At both feature types, there has been a decline in SBT sighting probability over the years considered, with 2000 a year of low probability at both features. The trends are not the same, however, as lumps in the years 1993 and 1994 had the highest sighting probabilities, whereas at the shelf break, the years 1993 and 1998 had the highest probabilities. This pattern may reflect different abundance in the two locations. At the lumps, sighting probability was lower in March than in the two previous months, which is consistent with SBT leaving the inshore features in preparation for exiting the GAB. Similar movement patterns are detected in archival tag data (J. Gunn, pers. comm.).

Although the sighting probability may be interpreted as an index of abundance, a change in sighting probability may be due to a change in environmental conditions. Recall again, that only fish at the surface can be detected from the air. If SBT have different behaviours under different environmental conditions, then apparent changes in abundance might reflect different surfacing behaviours and detection probabilities in a survey. Analysis of archival tag data is currently addressing this question. Finally, SBT of different ages may have different habitat preferences and surfacing behaviours. Changes in the age structure of SBT spending the summer in the GAB might result in different sighting probabilities. Analyses of SBT size and age within the GAB have not indicated any trends which might lead to the pattern observed here (Stanley or Polacheck pers comm), however, otoliths have yet to be completely analyzed for the period considered in this study.

The overall explanatory power of the GLMs was similar to those developed for other species and systems (e.g. Bigelow et al. 1999; Maury et al. 2001), and considering the other influences on SBT distribution, reasonable. The distribution of SBT sightings with regard to the environment and topography was non-random, as demonstrated by the re-sampling procedure. Overall, the limited number of environmental variables that were significant predictors indicates the environment has a minor influence on the distribution of juvenile SBT within the GAB. In fact, the whole GAB environment may be suitable, and other factors, such as prey availability may be more important once SBT arrive. Attractive lumps and shelf break segments differed from unattractive features with regard to both environmental and topographic characters, but the reason why these variables were "preferred" remains conjectural. The explanatory power of the final models was higher for the lumps than for the shelf break, suggesting uncertainty in an abundance index could be reduced more using sightings at the attractive lumps. Future aerial surveys may not need to cover such a large area to provide an adequate index of abundance, and when coupled with studies using tagging technologies, uncertainty may be further reduced, allowing trends in juvenile SBT abundance to be detected.

Most of the environmental variables considered were surface-associated, and although likely correlated with subsurface conditions, might be poor measures of what really influences SBT. Inclusion of subsurface measures, such as thermocline temperature and mixed layer depth, might have explained more variance in presence. Simply including different surface-related variables is unlikely to lead to a dramatic increase in the explanatory power of models, as most environmental variables are correlated with each other. Alternative technologies such as archival and acoustic tags can also deliver information about sub-surface environments and associated SBT behaviour, and will be important in future research. The influence of prey on distribution and abundance should also be a target for future efforts to reduce the uncertainty in a fishery-independent abundance index for juvenile SBT in the GAB.

Acknowledgements

Jeremy O'Reilly extracted the data on sightings, and was invaluable with regard to data analysis and processing. Jason Hartog extracted bathymetry data from the AGSO dataset. Madeline Cahill and David Griffin (CSIRO, Marine Research) provided altimetry data. Satellite CHL and SST data were provided by Chris Rathbone and Paul Tildesley (CSIRO, Marine Research). Ann Cowling, Jeremy O'Reilly and John Gunn provided discussion and analysis suggestions. This project was supported by FRDC grant (1999/105) to Ann Cowling and John Gunn, and by JAMARC, CSIRO Marine Research, and the Australian Fisheries Management Agency as part of the Japan/Australia SBT Recruitment Monitoring Program.

References

Bigelow, K. A., Boggs, C. H., He, X., 1999. Environmental effects on swordfish and blue shark catch rates in the US North Pacific longline fishery. Fisheries Oceanography 8 (3), 178-198.

Boehlert, G. W. (1987). A review of the effects of seamounts on biological processes. In B. H. Keating, P. Fryer, R. Batiza, & G. W. Boehlert (Eds.), Seamounts, Islands and Atolls (pp. 319-334). Washington DC: American Geophysical Union.

Borcard, D, Legendre, P., Drapeau, P. 1992. Partitioning out the spatial component of ecological variation. Ecology 73:1045:1055.

Caton, A. E. (1991). Review of aspects of southern bluefin tuna biology, population and fisheries. In R. S. Shomura, J. Majkowski, & S. Langi (Ed.), Proceedings of the first FAO Expert Consultation on interactions of Pacific tuna fisheries, (pp. 296-343). Noumea, New Caledonia:

Cayula, J.-F., Cornillon, P. 1992. Edge Detection Algorithm for SST Images. Journal of Atmospheric and Oceanic Technology 9 (1), 67-80.

Chapman, D. C., Haidvogel, D. B. 1992. Formation of Taylor caps over a tall isolated seamount in a stratified ocean. Geophysical and Astronomical Fluid Dynamics 64 31-65.

Chen, S., Cowling, A., Polacheck, T. 1995. Data analysis of the aerial surveys (1991-1995) for juvenile southern bluefin tuna in the Great Australian Bight. 1995 Southern Bluefin Tuna Recruitment Monitoring Workshop report. RMWS/95/6, 57 pp. CSIRO Marine Laboratories Hobart

Cowling, A., Millar, C. (1998). Data analysis of the aerial surveys (1991-1998) for juvenile southern bluefin tuna in the Great Australian Bight. Southern bluefin tuna recruitment monitoring and tagging program. Report of the tenth workshop. 1998; RMWS/98/3.

Croxall, J. P., Prince, P. A. 1996. Cephalopods as prey. I. Seabirds. Philosophical Transactions of the Royal Society of London B Biological Sciences 351 (1343), 1023-1043.

Dower, J., Freeland, H., Juniper, K. 1992. A strong biological response to oceanic flow past Cobb Seamount. Deep-Sea Research, Part A 39 (7-8), 1139-1145.

Farley, J. H., Davis, T. L. O. 1998. Reproductive dynamics of southern bluefin tuna, *Thunnus maccoyii*. Fishery Bulletin 96 (2), 223-236.

Fiedler, P. C., Smith, G. D., Laurs, R. M. 1984. Fisheries applications of satellite data in the eastern North Pacific. Marine Fisheries Review 46 (3), 1-13.

Freeland, H. 1994. Ocean circulation at and near Cobb Seamount. Deep-Sea Research, Part A 41 1715-1732.

Genin, A., Gal, G., Haury, L. 1994. Copepod carcasses in the ocean. II. Near coral reefs. Marine Ecology Progress Series 123 65-71.

Genin, A., Noble, M., Lonsdale, P. F. 1989. Tidal currents and anticyclonic motions on two North Pacific seamounts. Deep-Sea Research, Part A 36 (12), 1803-1815.

Gremillet, D., Wilson, R. P., Wanless, S., Chater, T., 2000. Black-browed albatrosses, international fisheries and the Patagonian Shelf. Marine Ecology Progress Series 195 269-28.

Grewe, P. M., Elliott, N. G., Innes, B. H., Ward, R. D. 1997. Genetic population structure of southern bluefin tuna (*Thunnus maccoyii*). Marine Biology. 127: 555-561.

Hastie, T. J., Tibshirani, R. J. 1990. Generalised Additive Models. London: Chapman and Hall.

Holland, K. N., Kleiber, P., Kajiura, S. M. 1999. Different residence times of yellowfin tuna, *Thunnus albacares*, and bigeye tuna, *T. obesus*, found in mixed aggregations over a seamount. Fishery Bulletin 97 (2), 392-395.

Hooker, S. K., Whitehead, H., Gowans, S. 1999. Marine protected area design and the spatial and temporal distribution of cetaceans in a submarine canyon. Conservation Biology 13 (3), 592-602.

Humston, R., Ault, J. S., Lutcavage, M., Olson, D. B., 2000. Schooling and migration of large pelagic fishes relative to environmental cues. Fisheries Oceanography 9 (2), 136-146.

Hunt, G. L., Coyle, K. O., Hoffman, S., Decker, M. B., Flint, E. N. 1996. Foraging ecology of short-tailed shearwaters near the Pribilof Islands, Bering Sea. Marine Ecology Progress Series 141 1-11.

Hynd, J. S. 1968. How sea surface temperature maps aid tuna fishermen. Australian Fisheries Newsletter May 23-29.

Hynd, J. S. 1969. Isotherm maps for tuna fishermen. Australian Fisheries July 13-22.

Kleiber, P., Hampton, J. 1994. Modeling Effects of FADs and Islands on Movement of Skipjack Tuna (*Katsuwonus pelamis*): Estimating Parameters from Tagging Data. Canadian Journal of Fisheries and Aquatic Sciences 51 (12), 2642-2653.

Lasker, R., 1978. The relation between oceanographic conditions and larval anchovy food in the California current; identification of factors contributing to recruitment failure. Rapp. P.-v. Reun. Cons. int. Explor. Mer. 173 212-230.

Laurs, R. M., Fiedler, P. C., Montgomery, D. R. 1984. Albacore tuna catch distributions relative to environmental features observed from satellites. Deep-Sea Research 31 (9), 1085-1099.

Laurs, R. M., Yuen, H. S. H., Johnson, J. H. 1977. Small-scale movements of albacore, *Thunnus alalunga*, in relation to ocean features as indicated by ultrasonic tracking and oceanographic sampling. Fishery Bulletin 75 (2), 347-355.

Lehodey, P., Andre, J.-M., Bertignac, M., Hampton, J., Stoens, A., Menkes, C., Memery, L., Grima, N., 1998. Predicting skipjack tuna forage distributions in the equatorial Pacific using a coupled dynamical bio-geochemical model. Fisheries Oceanography 7 (3-4), 317-325.

Maury, O., Gascuel, D., Marsac, F., Fonteneau, A., De Rosa, A.-L., 2001. Hierarchical interpretation of nonlinear relationships linking yellowfin tuna (*Thunnus albacares*) distribution to the environment in the Atlantic Ocean. Canadian Journal of Fisheries and Aquatic Sciences 58 458-469.

Myers, D. G. 1984. Satellites aid fish-finding in WA, possibly beyond. Australian Fisheries February 2-4.

Nybakken, J. W. 1997. Marine biology: an ecological approach. 4th ed. Addison Wesley Longman, Inc., Menlo Park, Calif. 481 pp.

Podesta, G. P., Browder, J. A., Hoey, J. J. 1993. Exploring the association between swordfish catch rates and thermal fronts on U. S. longline grounds in the western North Atlantic. Continental Shelf Research 13 (2/3), 253-277.

Polacheck, T., Gunn, J., Davis, T., Kalish, J., Hausknecht, C. 2000. An Overview of the Fishery for Southern Bluefin Tuna. Background Document for the Indonesia Australia Workshop on Shark and Tuna: future collaboration including research and fisheries management cooperation Bali, Indonesia Workshop, 2-3 March 2000. In CSIRO.

Polovina, J. J., Kobayashi, D. R., Parker, D. M., Seki, M. P., Balzas, G. H., 2000. Turtles on the edge: movement of loggerhead turtles (*Caretta caretta*) along oceanic fronts, spanning longline fishing grounds in the central North Pacific 1997-1998. Fisheries Oceanography 9 (1), 71-82.

Roden, G. I. 1987. Effect of seamounts and seamount chains on ocean circulation and thermohaline structure. In B. H. Keating, P. Fryer, R. Batiza, & G. W. Boehlert (Eds.), Seamounts, Islands and Atolls (pp. 335-354). Washington DC: American Geophysical Union.

Sund, P. N., Blackburn, M., Williams, F., 1981. Tunas and their environment in the Pacific Ocean: a review. Oceanography and Marine Biology Annual Review 19 443-512.

Wolanski, E., Hamner, W. M., 1988. Topographically controlled fronts in the ocean and their biological influence. Science 241 (4862), 177-181.

Young, J. W., Bradford, R. W., Lamb, T. D., Lyne, V. D. 1996. Biomass of zooplankton and micronekton in the southern bluefin tuna fishing grounds off eastern Tasmania, Australia. Marine Ecology Progress Series 138 1-14.

Young, J. W., Bradford, R. W., Lamb, T. D., Clementson, L. A., Kloser, R., Galea, H., 2001. Yellowfin tuna (*Thunnus albacares*) aggregations along the shelf break off south-eastern Australia: links between inshore and offshore processes. Marine and Freshwater Research 52 463-474.

Table 1. Names and locations of the GAB lumps. Success is the percentage of flights within 10 km of the lump which detected SBT. Attractive lumps by definition 1 were those with an attraction distance, whereas definition 2 identified lumps with success greater than mean success. * Lump 7 was defined as attractive using definition 1, but was within the attraction region of the shelf break, and so was not included in analyses.

					Attractive Lumps	
Lump	Name	Longitude (°E)	Latitude (°S)	Success (%)	Def. 1	Def. 2
1	Bell Point	133.05	32.33	3.45		
2	Cannon Reef	133.24	32.68	56.76	√	√
3	Inside Cannon	133.40	32.75	34.62	√	√
4	Outside Cannon	133.27	32.71	56.76	√	√
5	Daros	133.73	32.96	15.15		
6	Fowlers Point lump	132.46	32.18	28.26	√	√
7	Hamburger Hill	133.96	35.08	20.00	*	
8	210 from Hart	132.93	32.83	27.45		√
9	Hart	133.17	32.65	48.28		√
10	North Sceale Bay	133.99	33.06	17.65	√	
11	Nuyts 9m lump	132.00	32.16	42.86		√
12	Nuyts 12m lump	131.93	32.21	48.39	√	√
13	Nuyts Reef	132.13	32.15	51.22	√	√
14	Rocky Island	134.73	34.83	9.76		
15	West of Rocky lump	134.30	34.75	0		
16	S E St Francis	133.39	32.75	34.62	√	√
17	South Sceale Bay	134.00	33.06	17.65		
18	West of Ward	134.00	33.70	51.11	√	√
19	West St Francis lump	133.00	32.39	4.76		
20	West Yatala	132.50	32.63	21.43		
21	Yatala	132.61	32.63	60.61	√	√
22	D'Entrecasteaux Reef	131.93	32.00	23.08	√	√
23	St Francis Is (SFI)	133.33	32.49	2.86		
24	Masillion Is (SFI)	133.30	32.57	16.67	√	
25	Fenelon Is (SFI)	133.29	32.59	44.44	√	√
26	Lacy Is	133.38	32.40	0		
27	Lacy Reefs	133.36	32.37	0		
28	Evans Is	133.50	32.38	0		
29	Franklin Is 1	133.65	32.46	0		
30	Franklin Is 2	133.70	32.45	0		
31	Purdie Island Group	133.25	32.28	2.7		
32	Greenly Is	134.80	34.65	0		
33	Whidbey Is, 4 Hummock	135.05	34.77	0		
34	Pearson Is	134.29	33.96	27.27	√	√
35	Flinders Is	134.51	33.71	3.13		
36	Ward Is	134.32	33.78	15.79	√	
Mean				21.85	n=17	n=15

Table 2. Attraction distance and number of sightings within the attraction region at the shelf break. The numbers of sightings by year within the total attraction distance of each shelf point are also shown. NaN: no attraction distance found for these shelf points. Success is the percent of all close flights which detected SBT. Attractive segments using definition 1 were all those segments with an attraction distance, whereas definition 2 identified the segments with success>mean(success).

	North of shelf break		South of shelf break		North and south of shelf break		Sightings inside all the shelf attraction regions by year (%)											Success (%)	Attractive segment
Shelf Point	Attraction Distance (km)	n	Attraction Distance (km)	n	Attraction Distance (km)	n	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	Mean		Definition 2
1	NaN	0	NaN	0	NaN	0	0	0	0	0	0	0	0	0	0	0	0.00	2.67	
2	16.3	5	NaN	0	16.3	5	0	0	4.55	0	0	0	0	2.22	7.14	0	1.39	3.33	
3	38.5	12	23.16	2	61.65	14	6.25	0	7.58	7.02	6.90	4.17	0	0	0	5.00	3.69	3.91	
4	7.77	1	13.66	6	21.43	7	6.25	0	3.03	3.51	0	0	5.56	0	0	0	1.84	5.44	
5	9.21	6	5.34	3	14.55	9	0	0	1.52	0	0	0	11.11	8.89	0	0	2.15	6.84	
6	0.99	2	3.71	4	4.7	6	0	0	0	0	13.79	0	2.78	2.22	0	0	1.88	6.56	
7	22.15	18	2.64	2	24.8	20	0	0	1.52	0	6.90	4.17	8.33	11.11	7.14	35.00	7.42	5.83	
8	6.29	10	2.55	5	8.84	15	0	5.26	3.03	0	0	12.50	11.11	4.44	7.14	5.00	4.85	8.46	√
9	13.07	38	19.58	50	32.64	88	12.5	42.1	21.21	35.09	17.24	29.17	8.33	31.11	28.57	15.00	24.03	15.29	√
10	3.76	8	16.79	31	20.55	39	0	34.21	13.64	0	10.34	37.50	0	6.67	0	10.00	11.24	16.10	√
11	18.92	21	14.5	32	33.42	53	6	10.53	21.21	14.04	27.59	0	19.44	13.33	0	0	11.21	12.31	√
12	10.94	5	29.79	36	40.73	41	0	0	13.64	7.02	17.24	12.50	11.11	11.11	42.86	25.00	14.05	9.20	√
13	6.56	3	29.98	23	36.54	26	6.25	7.89	3.03	22.81	0	0	11.11	6.67	0	0	5.78	6.88	
14	2.99	2	25.22	19	28.21	21	31.25	0	6.06	10.53	0	0	11.11	2.22	0	5.00	6.62	5.86	
15	NaN	0	4.29	1	4.29	1	0	0	0	0	0	0	0	0	7.14	0	0.71	5.10	
Mean	12.11	10.15	14.71	16.46	23.24	23.07												7.59	n=5
Total		131		214		345	16	38	66	57	29	24	36	45	14	20			

Table 3. Attraction distance and number of sightings within the attraction distance for each lump which was attractive to SBT, using attraction distance as a measure of attractiveness. Lumps which were not attractive by this definition are not shown in the table.

			Sightings inside attraction region of the lumps by year (%)										
Lump	Attraction Distance (km)	n sightings	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	Mean
2	3.8	21	0.00	6.67	3.57	3.77	15.69	18.18	12.90	0.00	0.00	0.00	6.08
3	1.99	3	0.00	0.00	0.00	3.77	1.96	0.00	0.00	0.00	0.00	0.00	0.57
4	3.22	13	0.00	0.00	0.00	9.43	7.84	4.55	9.68	0.00	0.00	0.00	3.15
6	6.28	16	4.35	10.00	10.71	5.66	1.96	9.09	6.45	0.00	16.67	0.00	6.49
7	11.68	11	4.35	10.00	3.57	9.43	0.00	0.00	0.00	7.69	0.00	0.00	3.50
10	3.95	4	0.00	0.00	3.57	3.77	1.96	0.00	0.00	0.00	0.00	0.00	0.93
12	1.93	5	0.00	3.33	0.00	3.77	1.96	4.55	0.00	0.00	0.00	0.00	1.36
13	7.59	48	30.43	36.67	17.86	13.21	9.80	27.27	9.68	15.38	33.33	0.00	19.36
14	6.36	8	4.35	0.00	7.14	1.89	0.00	0.00	9.68	0.00	16.67	0.00	3.97
16	2.37	3	0.00	0.00	0.00	3.77	1.96	0.00	0.00	0.00	0.00	0.00	0.57
17	1.83	3	0.00	0.00	3.57	1.89	1.96	0.00	0.00	0.00	0.00	0.00	0.74
18	6.31	31	8.70	16.67	10.71	7.55	15.69	0.00	9.68	46.15	0.00	0.00	11.51
21	6.53	45	17.39	13.33	21.43	20.75	11.76	22.73	12.90	7.69	33.33	100.00	26.13
22	6.91	11	8.70	3.33	0.00	0.00	7.84	13.64	0.00	7.69	0.00	0.00	4.12
24	4.44	5	8.70	0.00	0.00	0.00	5.88	0.00	0.00	0.00	0.00	0.00	1.46
25	2.31	5	8.70	0.00	0.00	0.00	5.88	0.00	0.00	0.00	0.00	0.00	1.46
34	7.79	17	4.35	0.00	7.14	5.66	7.84	0.00	19.35	7.69	0.00	0.00	5.20
36	7.99	10	0.00	0.00	10.71	5.66	0.00	0.00	9.68	7.69	0.00	0.00	3.37
Mean	5.18	14.39											
Total		259	23	30	28	53	51	22	31	13	6	2	

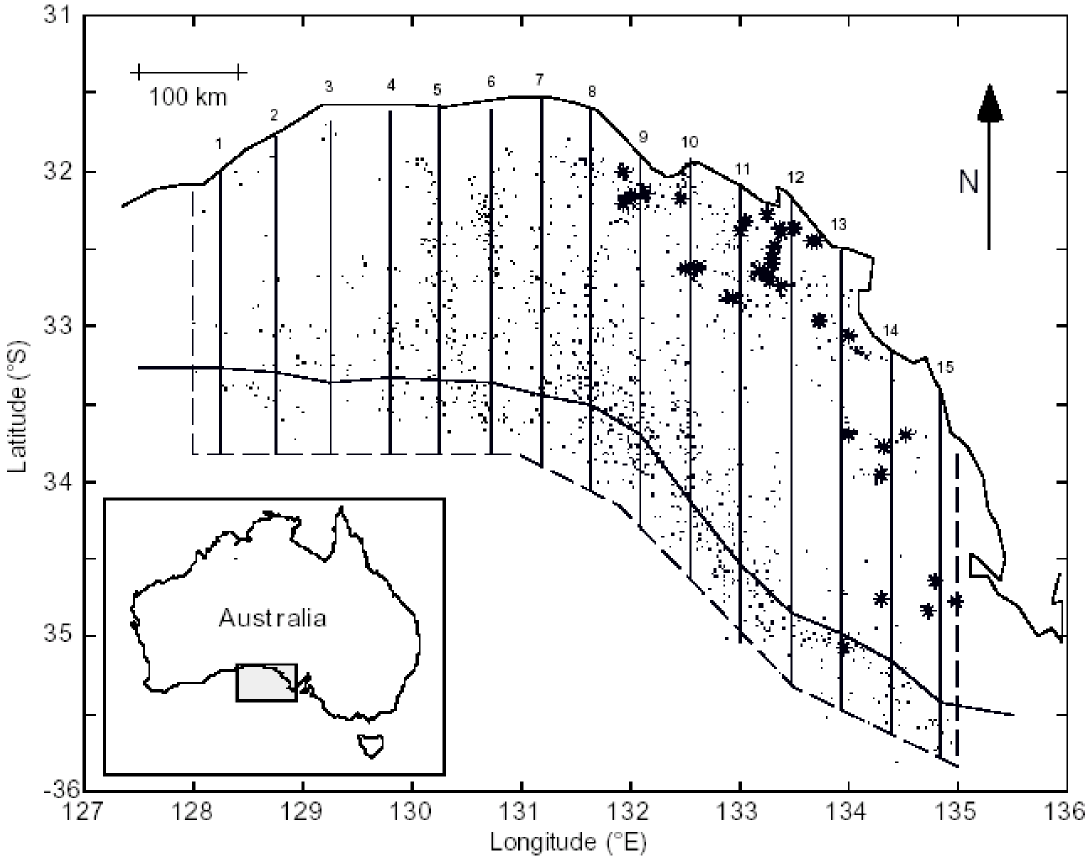


Figure 1. Location of the 15 aerial survey lines (1999-2000) in the GAB. In earlier years lines were located around these north-south lines, but varied from flight to flight. The locations of the lumps (*) and the shelf break (solid line) are shown, together with the location of all the SBT sightings (dots) within the aerial survey area (dashed lines).

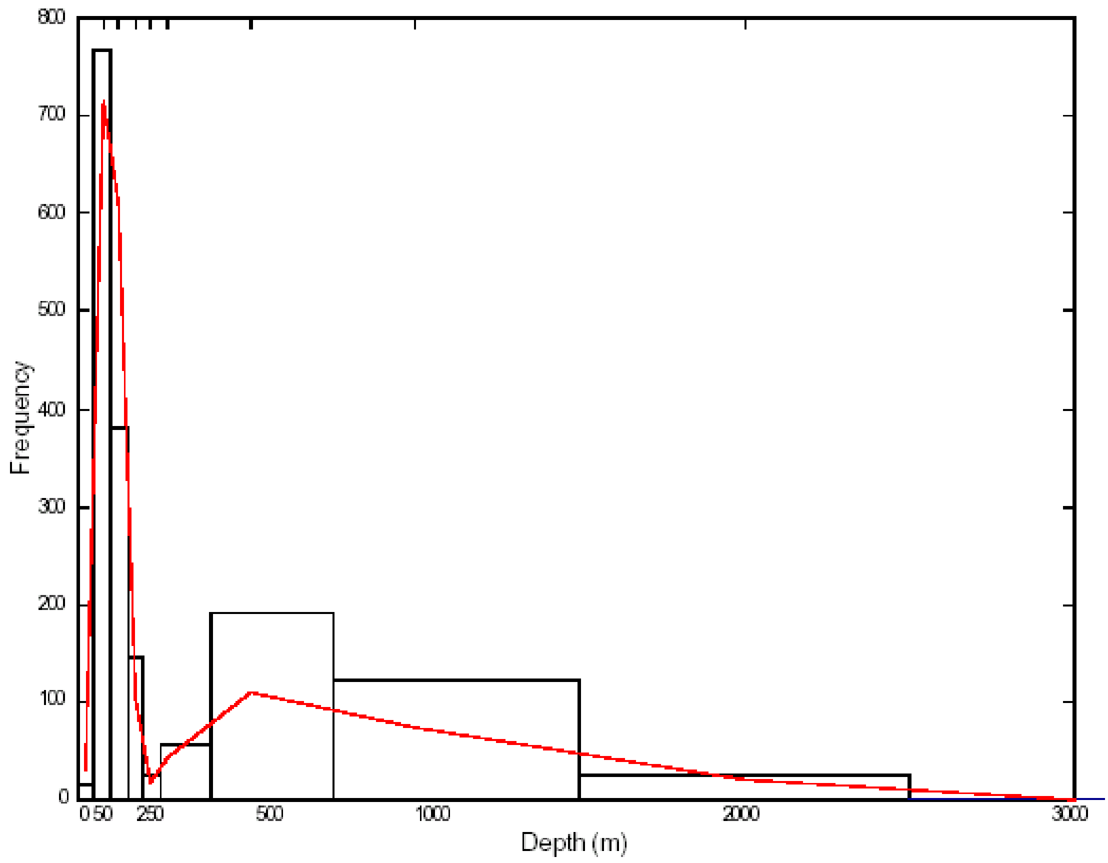


Figure 2. Observed depth distribution of SBT sightings (bars) compared to depths at random positions through the survey area (line).

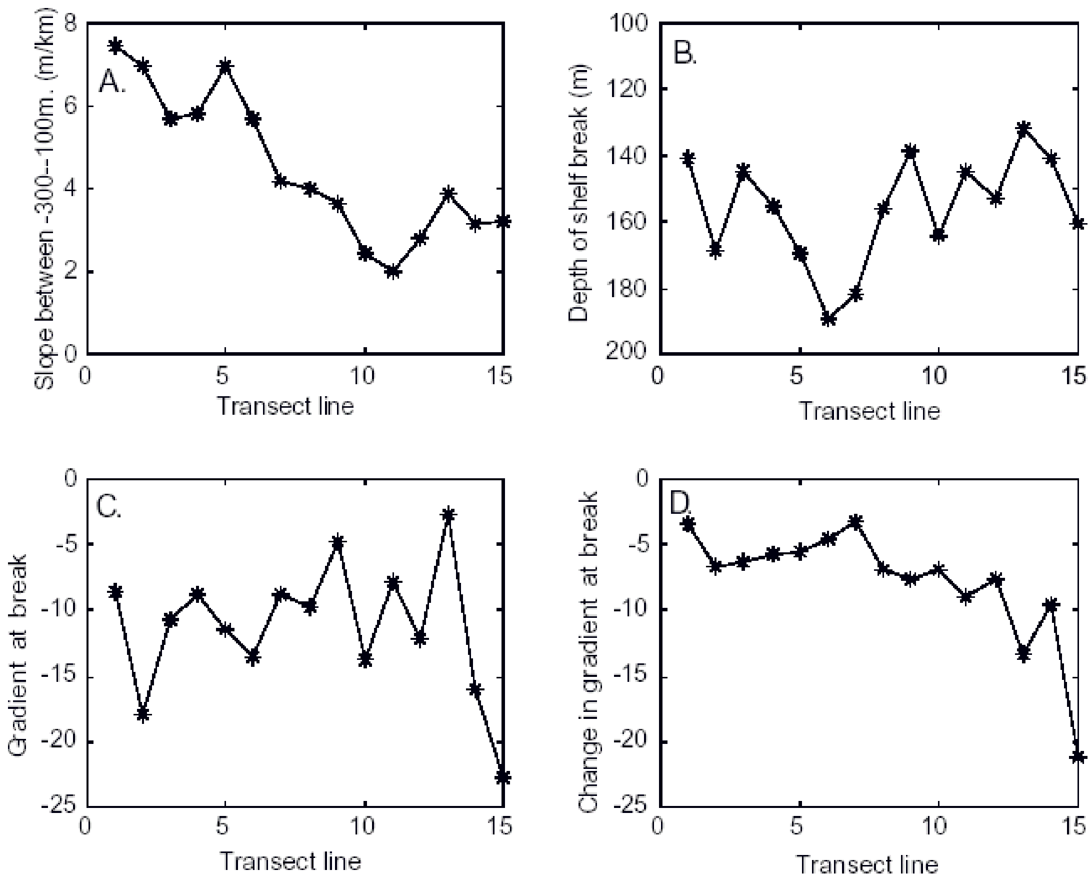


Figure 3. Shelf break characters on each aerial survey transect line.

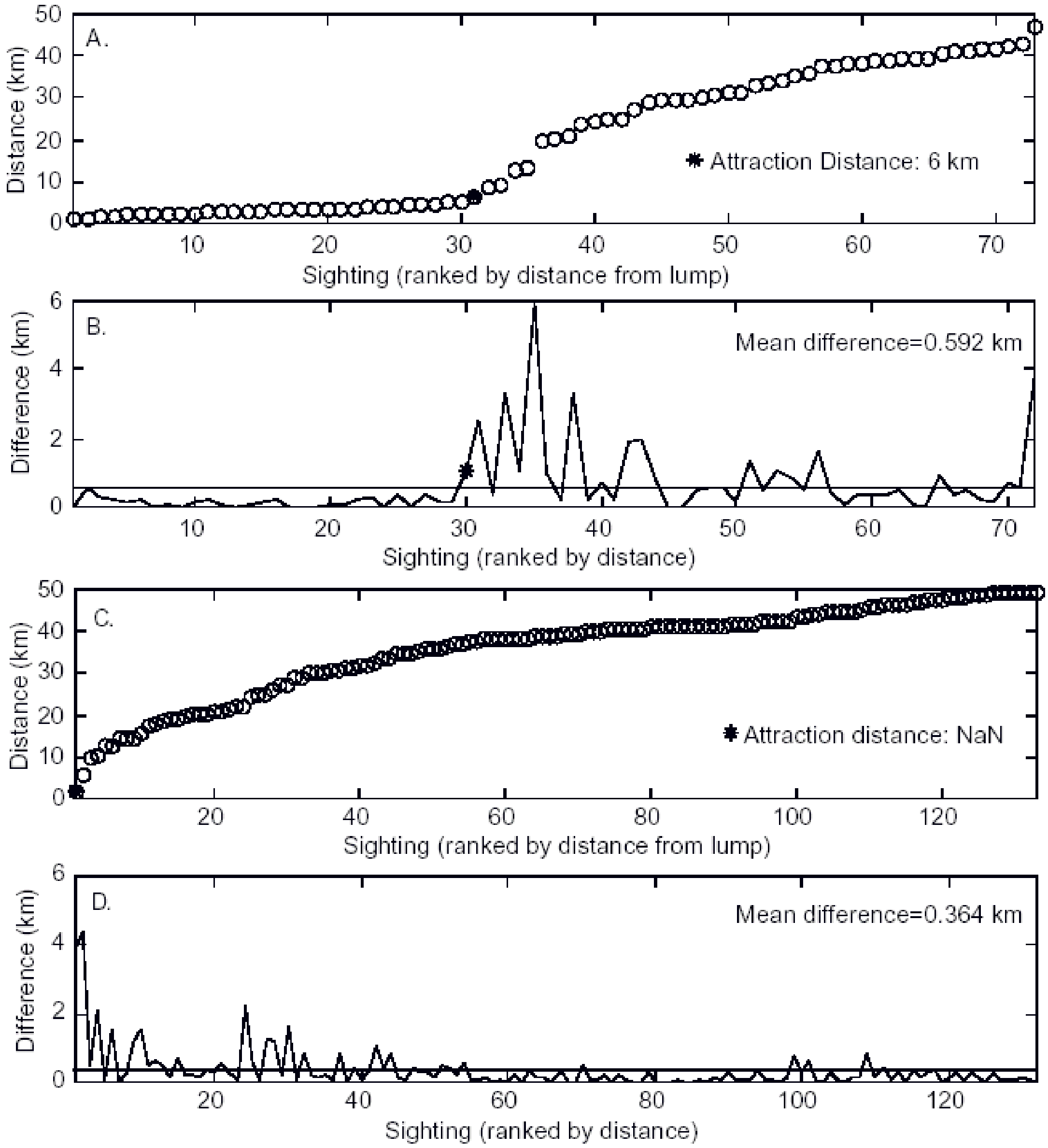


Figure 4. Examples of the attraction distance estimation method. A and B. Lump where there was an attraction distance found (Lump 18, West of Ward) C and D. Lump where an attraction distance of zero was found (Lump 1, Bell Point). A and C. Sightings ranked by distance from the lump. B and D. Difference between ranked sightings as a function of ranked observation number. The mean difference is shown with the horizontal line. The attraction distance is defined as the distance at which the mean difference between sightings is exceeded by a factor of two. The star indicates the attraction distance.

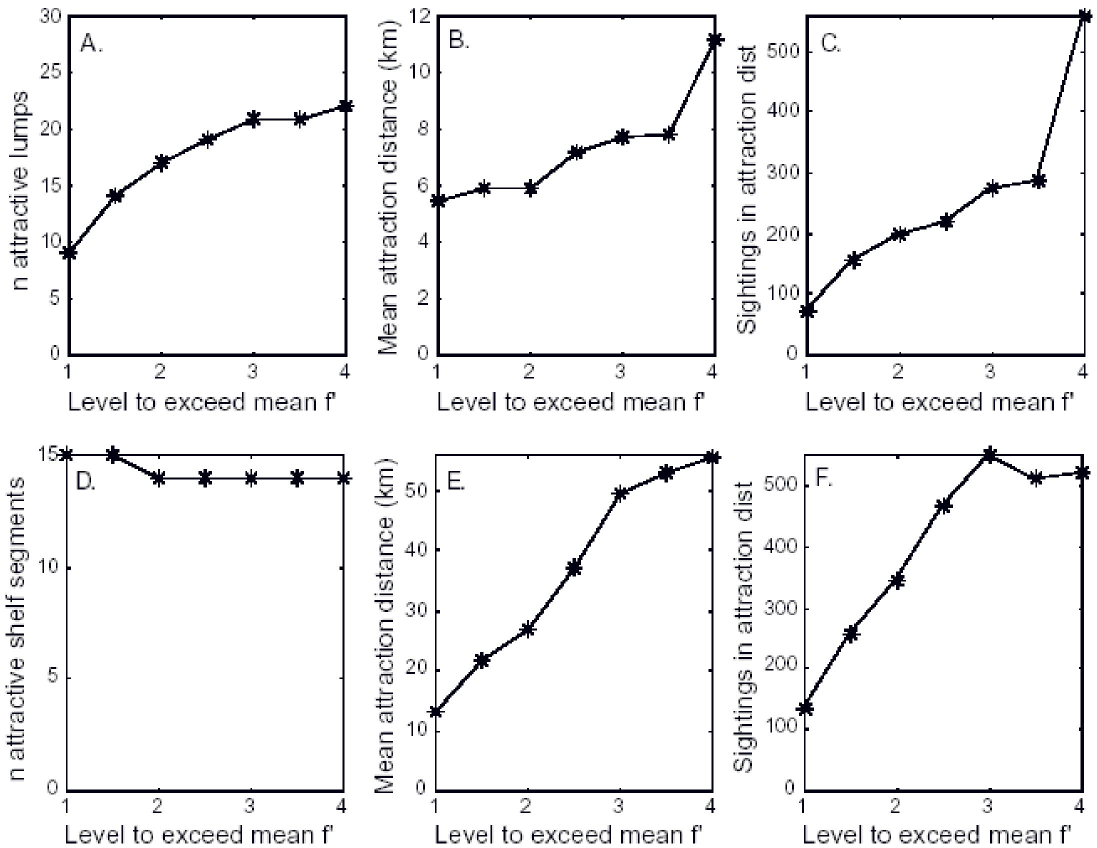


Figure 5. Attraction distance. Influence of the form of the definition used to determine the attraction distance. The level to exceed the mean difference between ranked sightings ranges from 1x the mean difference to 4x the mean difference A. Lumps. Number of lumps with an attraction distance, as a function of the level above the mean difference to define the end of the attraction region. B. Lumps. Mean attraction distance for the lumps, as a function of the level exceeded. C. Lumps. Number of sightings within the total attraction distance for the lumps, as a function of the level exceeded. D. Shelf break. Number of shelf segments with an attraction distance, as a function of the level above the mean difference used to define the end of the attraction region. E. Shelf break. Mean attraction distance for the shelf segments, as a function of the level exceeded. F. Shelf break. Number of sightings within the total attraction distance for the shelf segments, as a function of the level exceeded.

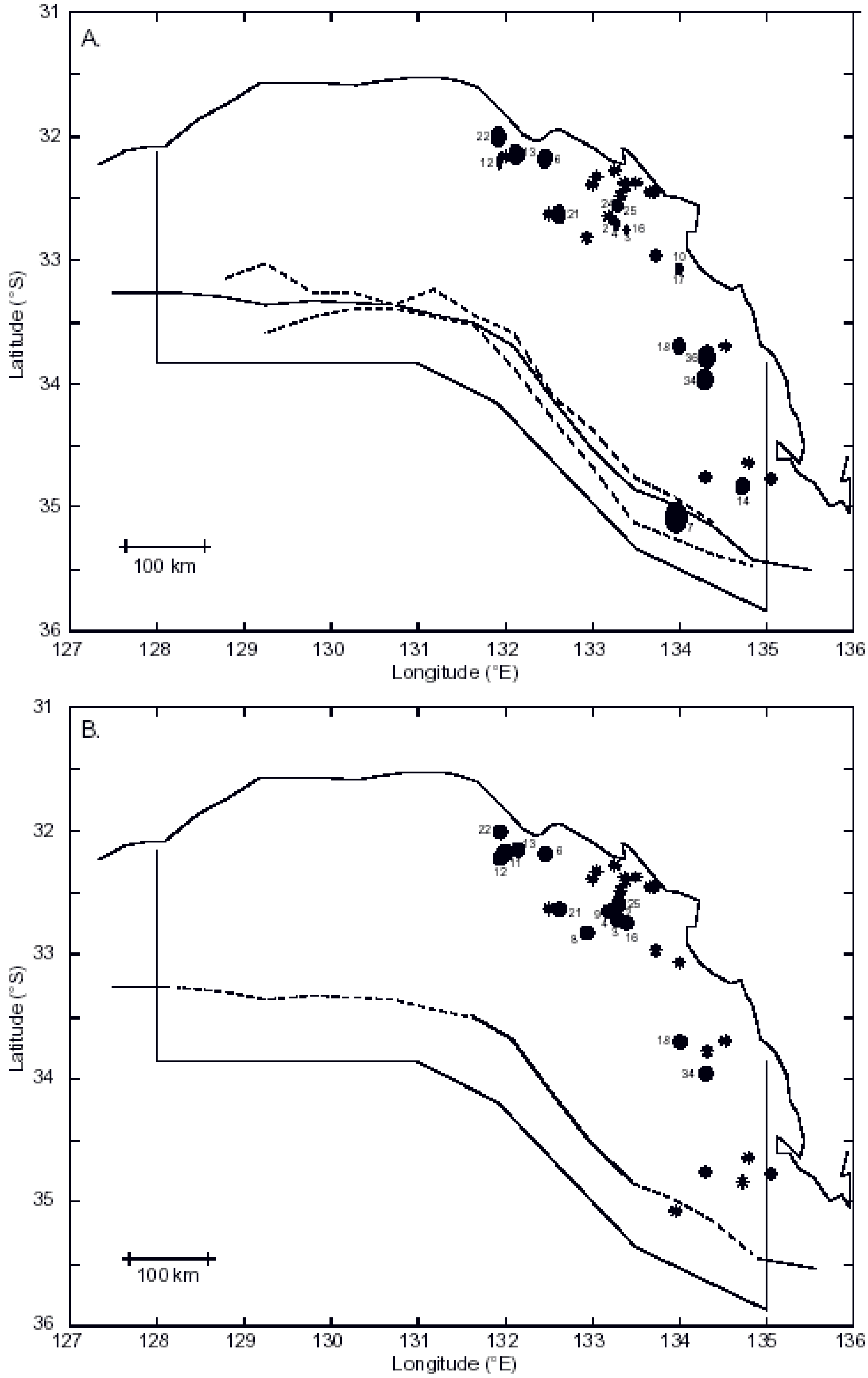


Figure 6. A. SBT attraction distances around the shelf break and lumps. The survey area, shelf break (solid line) and the location of the lumps are shown. A circle of the same radius as the attraction distance surrounds attractive lumps, which are also numbered. The broken line shows the attraction distance to the north and south of the shelf break. B. Lumps (filled circles) and shelf break segments (heavy line) with greater than average success (percentage of flights close to these features which detected SBT). Lump names are provided in Table 1; stars represent the location of unattractive lumps.

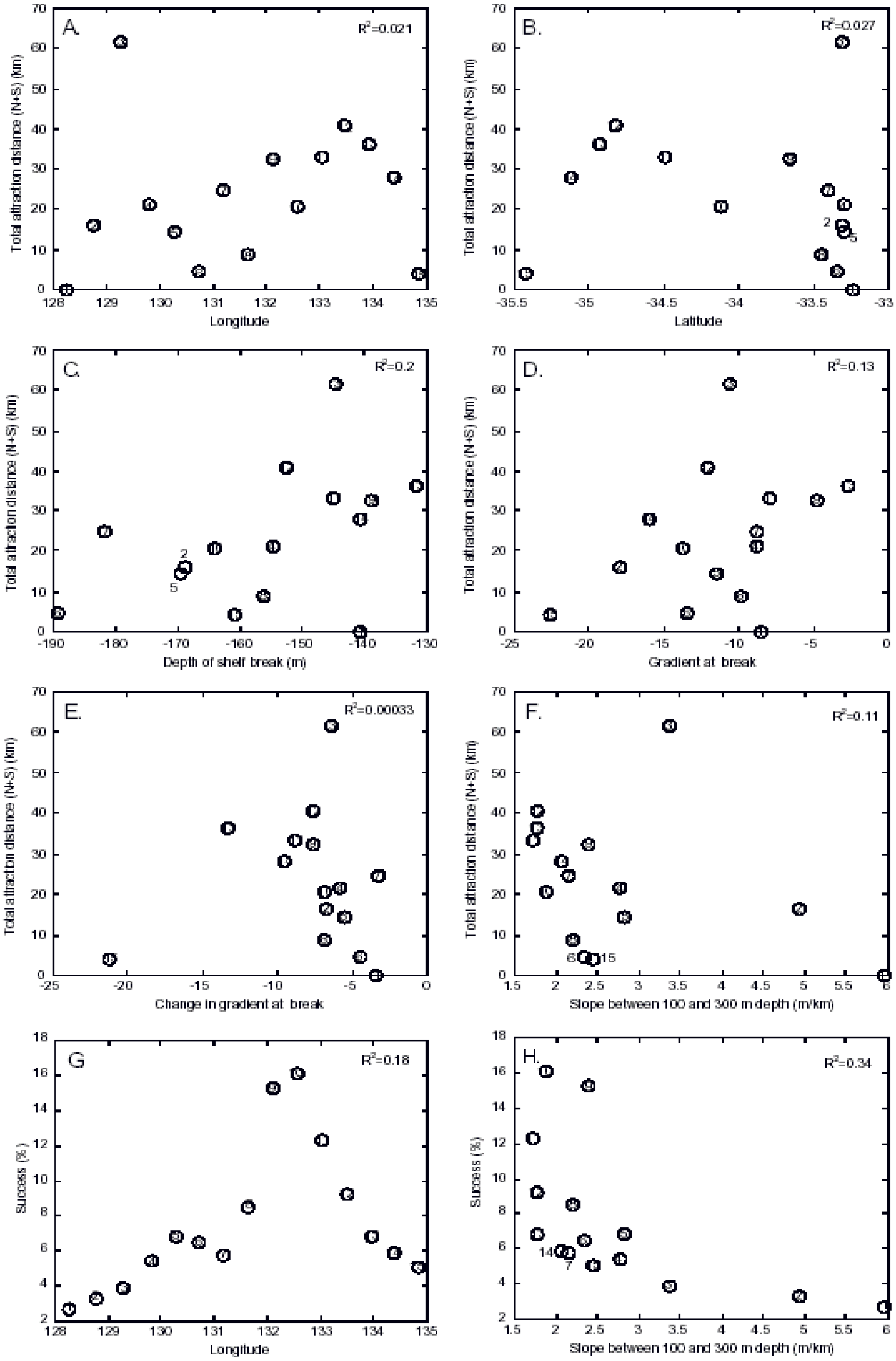


Figure 7. Relationships between the attraction measures and the topographic characters of the shelf break at each transect line in the GAB. The numbers for each transect line are shown on the symbols (transects are numbered from 1 in the west to 15 in the east). A-F. Relationships between the attraction distance and the topographic characters. G. Relationship between success (percentage of flights within 10 km of the shelf break which detected SBT) and longitude of the shelf break. H. Relationship between success and the slope between 100 and 300 m for the transect lines.

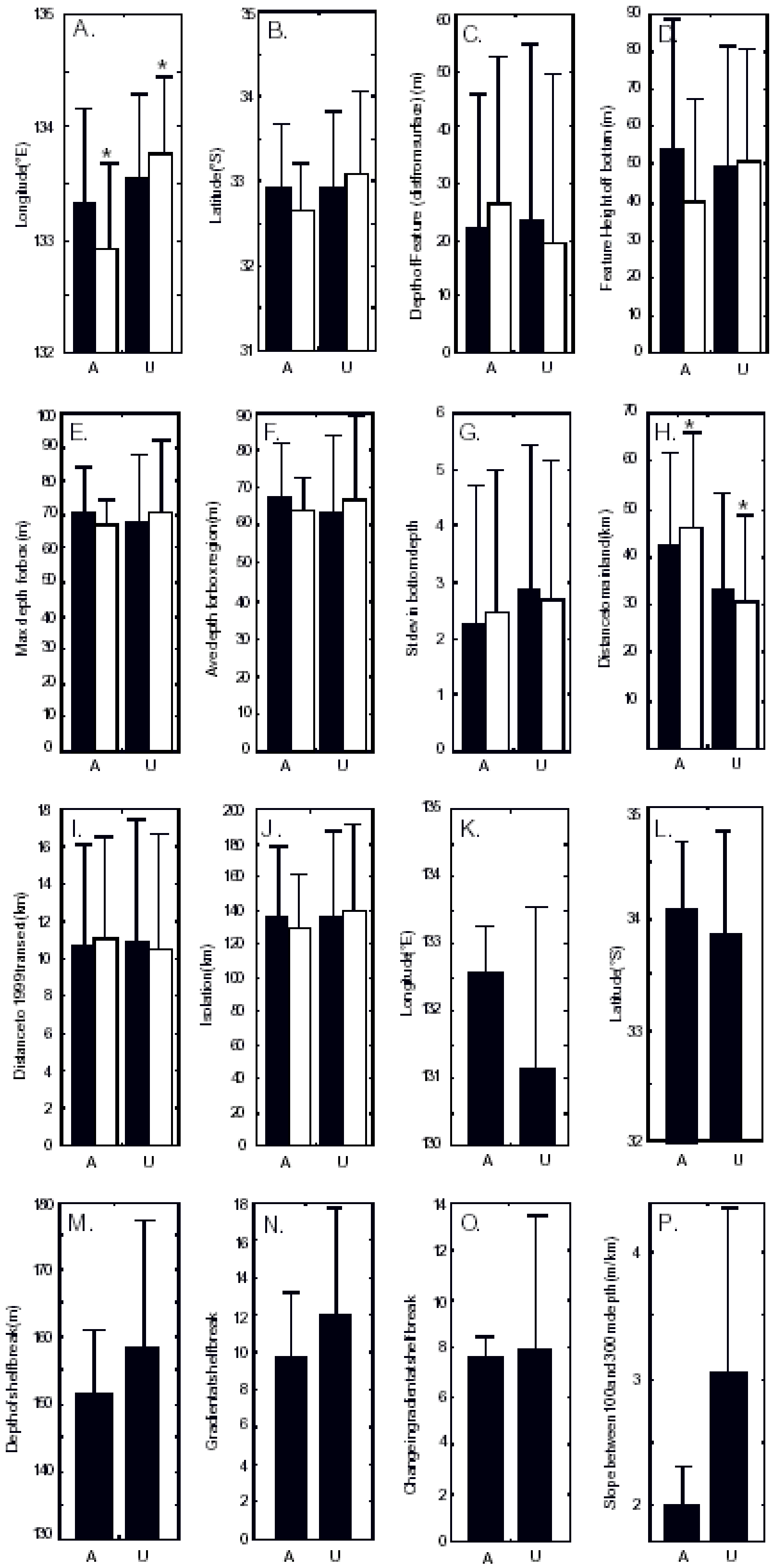


Figure 8. Mean (± 1 SD) topographic characters for attractive (A) and unattractive (U) lumps and shelf segments based on SBT sightings between 1991-2000. A-J. Lumps. Filled bars represent the attractive lumps, defined by having an attraction distance (n attractive = 17, n unattractive = 18). Open bars represent the attractive lumps defined using success>mean(success) (n attractive=15, n unattractive=20). Lump 7 was excluded from this comparison. K-P. Shelf characters, attractive segments defined by success>mean(success) (n attractive = 5, n unattractive = 10). Stars above bars indicate significant differences.

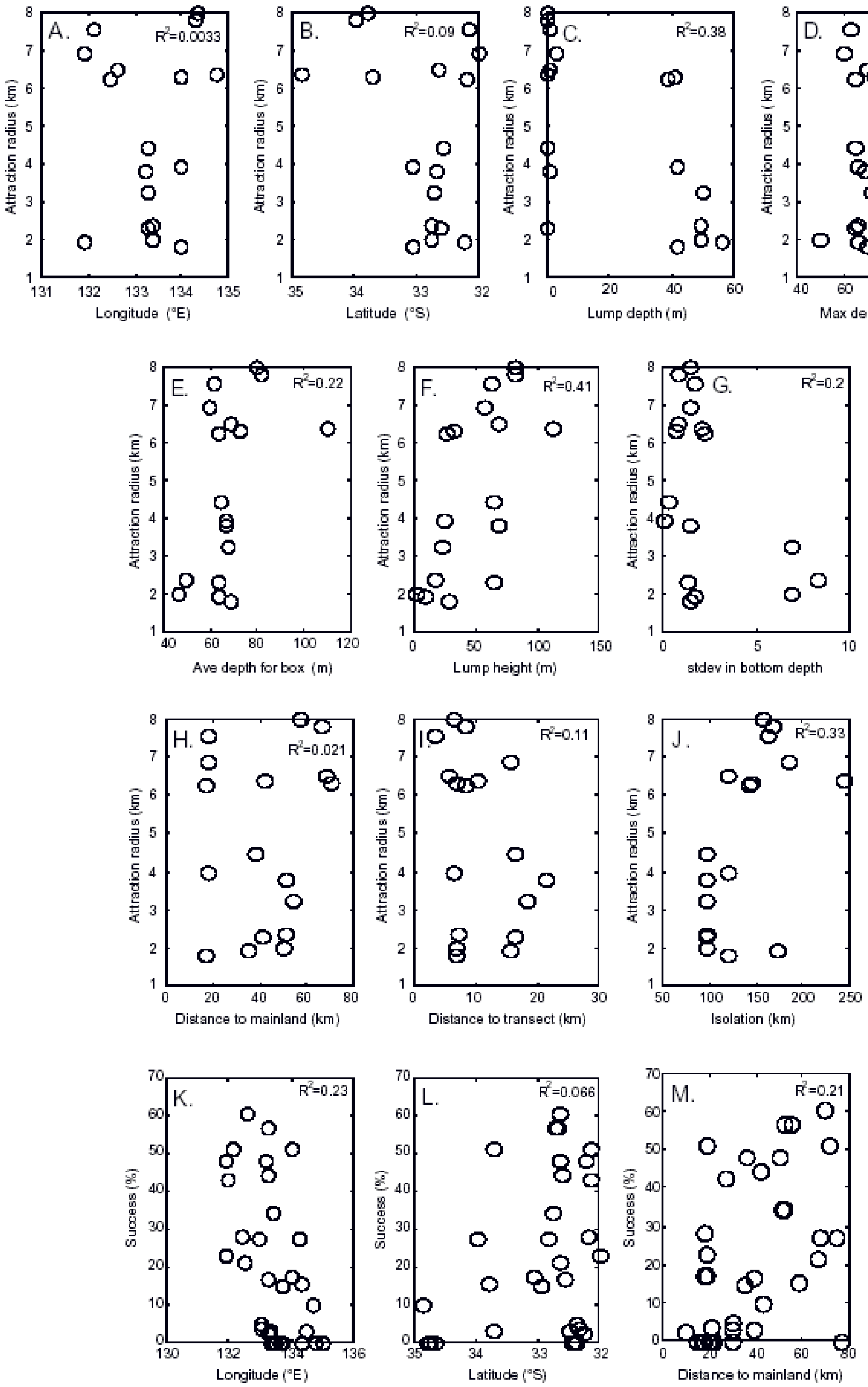


Figure 9. A-J. Relationships between the attraction distances and topographic characters of attractive lumps defined using attraction distance measure. K-M. Relationships between success, the percentage of close flights that detected fish, and topographic characters for all the lumps.

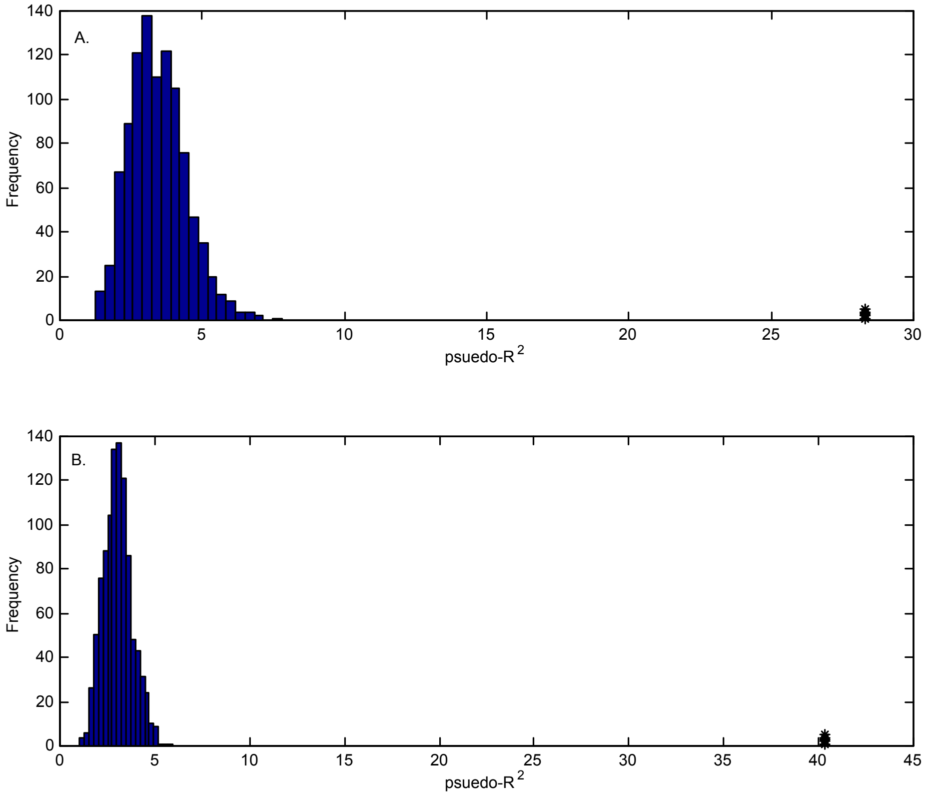


Figure 10. **Re-sampled GLM psuedo-R^2 values (n = 1000).** A. **Shelf.** B. **Lumps. The stars to the right of each distribution indicate the value of the actual model psuedo-R^2 value.**

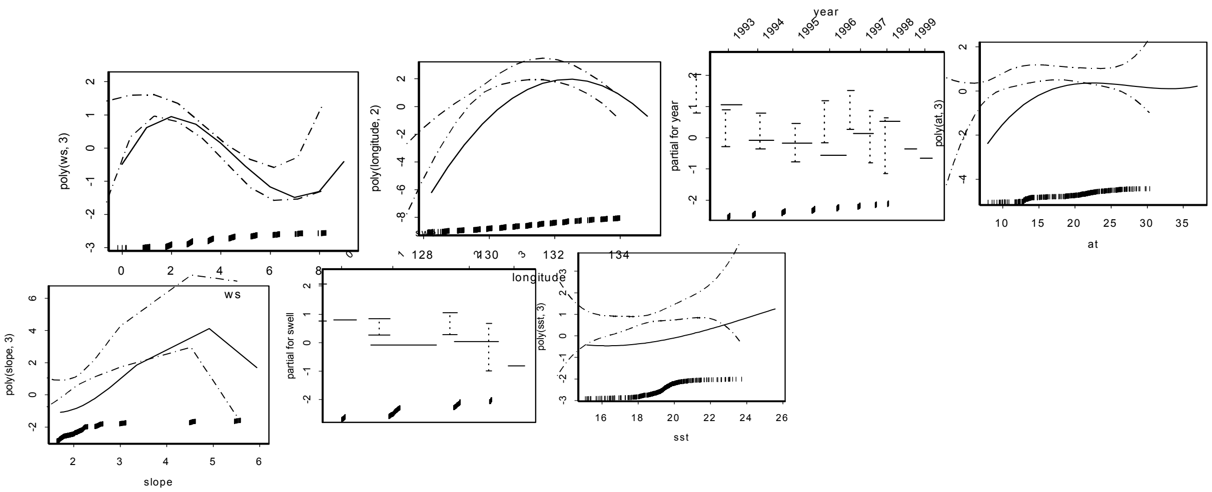


Figure 11. Results of GLM regression (solid lines) with the environmental and topographic predictor variables on the logit probability of detecting SBT on close flights past the shelf. One standard error boundary around the covariate effects is shown with dashed lines. Tick marks in the x-axis show the location of data points.

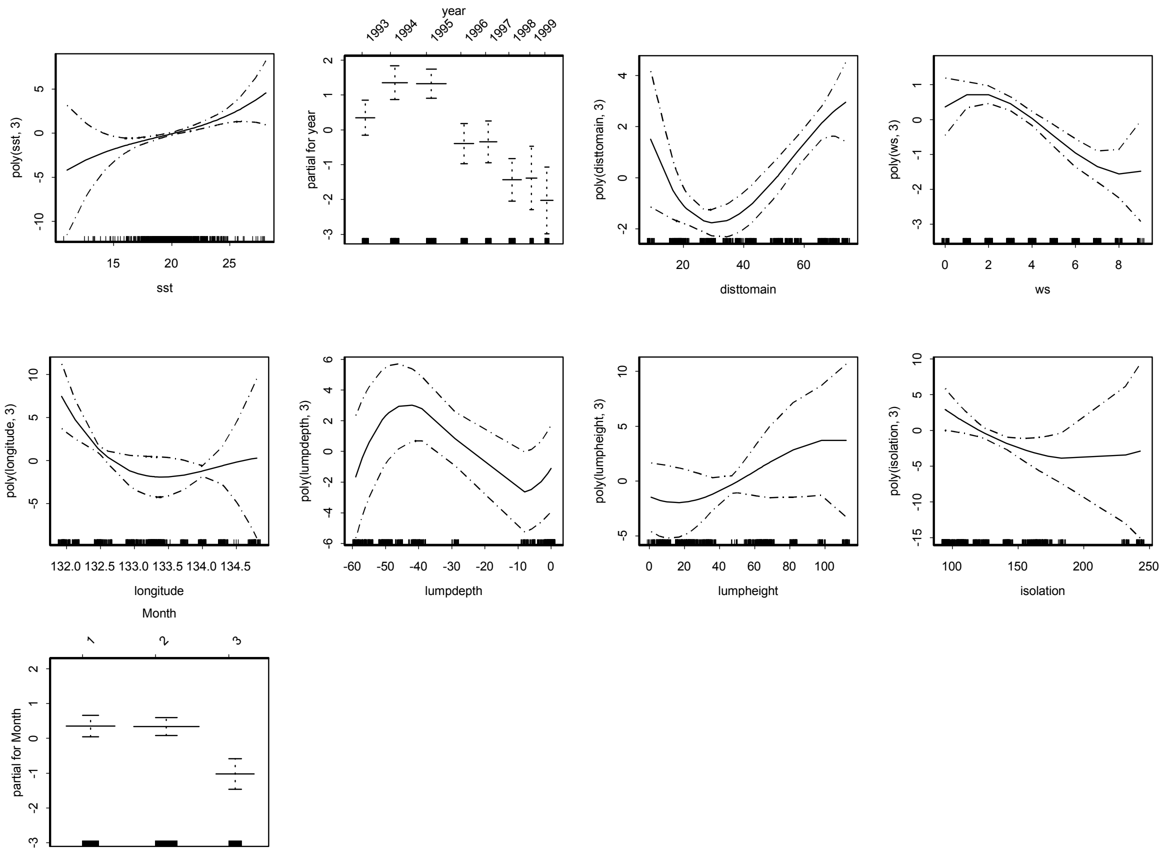


Figure 12. Results of GLM regression (solid lines) with the environmental and topographic predictor variables on the logit probability of detecting SBT on the close flights past the lumps. One standard error boundary around the covariate effects is shown with dashed lines. Tick marks in the x-axis show the location of data points.

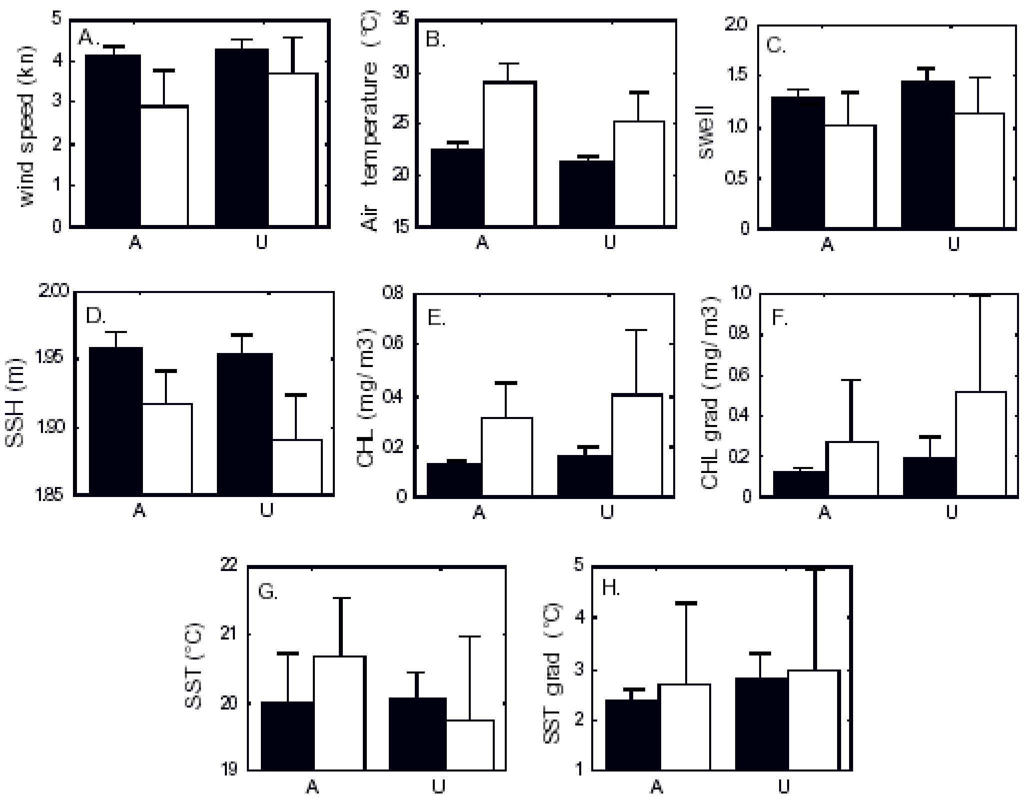


Figure 13. Mean (±1 SD) value of environmental variables at the attractive (A) and unattractive (U) shelf segments (solid bars) and lumps (open bars) in the GAB. The attractive features were defined using definition 2; those features where success>mean(success) for close flights between 1991-2000. The mean environmental variable for each feature was calculated before obtaining the mean for the group of features.

6.3 Estimating the position of juvenile SBT in the Great Australian Bight using light, water temperature and diving data collected by archival tags.

Jeremy O'Reilly, John Gunn and Alistair Hobday

Abstract

Information on southern bluefin tuna surfacing behaviour obtained from archival tags in different regions of the Great Australia Bight may reduce the errors associated with the estimation of an abundance index derived from the aerial survey. Before deriving these surfacing behaviours, it is necessary to improve the estimates of position within the GAB. In this process it was important to evaluate the performance of existing software with regard to position estimation. Subsequent improvements in the estimation of archival position are described in this section.

Introduction and Background

Since the development of archival tags in the early 1990's and their widespread use to study the movement and behaviour of tunas, the problem of how to estimate geoposition using data collected by tags, and quantification of the nature and extent of errors in those estimates, have been the subject of a number of research projects throughout the world (Gunn et al. 1994; Welch and Eveson, 1998; Hill and Braun, 2001; Musyl et al., 2001). Most of these validation projects have used data collected by stationary tags, attached to buoys for periods of months to years. Errors were estimated by comparing the calculated position relative to the known position of the buoys. These estimates present a best-case scenario for geolocation estimation, in which there is no vertical or horizontal movement of the tag to affect the levels of light recorded. Although the scales of errors varied among studies, in all cases the errors for longitude were significantly smaller than those for latitude. Musyl et al. (2001), for example, determined error for tags placed in the sub-tropical north Pacific and found the range of means to be 0.15-0.35° and 0.91-5.50° for longitude and latitude, respectively. In the only study conducted to date in which tags were carried by a moving fish, Gunn et al. (1994) found a similar magnitude of error to those for the stationary tag experiments. However, once again this was a best-case scenario, as the fish used in this experiment were being towed at a speed of 1 knot, in a cage that had a maximum depth of 20 metres, and thus vertical movement (and resultant degradation of light) was very limited.

In dealing with "real" data, collected by tags deployed on fish making regular and rapid dives in the open ocean, the challenges in estimating position are considerable, and the scale of errors much greater than those described in the controlled experiments discussed above. This is particularly true in the case of latitude estimation where even small errors in estimating the times of sunrise and sunset propagate into significant errors. Recent advances in methodology (Hill and Braun, 2001; Sibert et al., in press) suggest that it may be possible to use light information to estimate position if fish movement is restricted to upper layers of the water column. However, for species such as the bluefin tunas, bigeye tuna and swordfish, which all make regular vertical movements to depths greater than 100-200m, it seems likely that environmental data will be required to supplement the positions estimates made using light data.

In our studies of southern bluefin tuna (SBT) movement and behaviour in the Great Australian Bight (GAB) we are attempting to examine the relationship between behavioural patterns and a suite of environmental conditions. To allow this, high precision estimates of position are required, and this required evaluation of sea surface temperature (SST) and bathymetry as sources of data to assist the estimation of latitude. In this report, we describe the methodology developed and applied to improve the precision of geoposition estimates, and compare the latitude estimates produced using light, sea surface temperature and bathymetric data, either exclusively, or in combination.

Examination of existing non-light-based geoposition estimation methods

<u>Longitude estimation from archival tags</u>

The first step in estimating position of an archival tag is to estimate the longitude from the light record. Longitude is estimated by estimating the time of local midnight or midday and comparing this to the time of midday or midnight obtained from the tags' internal clock (usually GMT). Two different approaches have been taken in past studies for estimating longitude:

- 'Curve fitting': using a variety of statistical procedures (least squares, convex hull, quartile determination), this method estimates the time of midnight (or midday) directly by fitting a parametric curve to each dusk and dawn, and determines the time of dawn and dusk from the parameters of the fit. The midpoint of these time estimates provides an estimate of midnight or midday.
- 'Folding' methods, identify midnight and/or midday as the point of symmetry midway between dusk and dawn that produced maximum agreement between the dawn and dusk data. Two different means of undertaking folding have been proposed independently, one is available in ARCHTAG, the other from Wildlife Computers Inc (Seattle USA). The latter is based on the methods described by Hill and Braun (2001).

Juvenile SBT typically make deep dives around the time of dawn and dusk, and this presented particular problems for curve fitting methods, which depended upon accurately correcting for depth attenuation in the light levels. Folding methods were more robust to this behaviour, especially since these deep dives usually occurred symmetrically. The CSIRO proprietary software package ARCHTAG software was used to automatically detect each dawn and dusk event for the purposes of determining longitude by both the folding and curve fitting methods, however extremely deep dives near dawn and dusk, and bright nights (due to moonlight), frequently resulted in misidentification of dawn and dusk periods, and consequently, inaccurate estimates of longitude. Curve fitting methods were not robust to noise in the light data, and often gave poor estimates, even when dawn and dusk periods were identified correctly.

<u>Latitude estimation for archival tags: use of bathymetry and water temperature</u>

Sea surface temperature and bathymetry have been used to improve the estimation of the latitude of archival tags deployed on tunas (Gunn and Block, 2001) and sharks (West and Stevens, 2001) respectively. ARCHTAG includes modules that automate the estimation of latitude using both of these supplementary data sources. However, ARCHTAG was designed for research primarily focussed on the large-scale migration patterns of tunas and sharks, and it was unclear whether the temporal and spatial resolution of data sets being used in the automated latitude estimation routine were sufficient to provide the precise estimates required for fine-scale movement and behaviour analyses of SBT in the GAB. Thus, it was important to evaluate the performance of existing ARCHTAG software before attempting to estimate position for use deriving surfacing behaviours for use in the aerial survey index development.

(i) Bathymetric and maximum depth data

For any given longitude, ARCHTAG software allowed matching of a depth estimate from archival tag data against bathymetric data supplied by the Australian Geological Survey Office. In their study of shark movements in southern Australia, West and Stevens (2001) were able to use the depths at which the shark appeared to be either stationary or swimming along the bottom to conduct matching between a tag depth and bottom depth. This was relatively straightforward, as the shark depth data indicated long periods of bottom-orientated behaviour. Tuna rarely exhibit similar bottom-orientated behaviour. However, while in the GAB, each day they make regular deep dives and often the maximum depth of these dives is more-or-less constant over the period of a day. We hypothesised that these dives may be to close to the bottom. If this is the case, and as the continental shelf bathymetry of the GAB slopes gently from shallow coastal waters in the north to the continental shelf in the south, we could use the maximum depth in a twenty-four hour period to determine a northern bound on fish's position. In essence, if the fish dived to 50 m it couldn’t be north of the latitude of the 50m isobath. The ARCHTAG bathymetry module used sea floor bathymetry at 100 m isobath resolution, with the result that latitude estimates were pushed to lie along these contours. This gave the misleading impression that juvenile SBT preferred three regions: the shelf break (200 m); midway between the shelf break and the shoreline (100 m); and beyond the shelf break (>200 m). Thus, for finer scale analyses and more precise estimates higher resolution bathymetry data is required. This data exists in updated bathymetry files.

(ii) Sea Surface Temperature

The existing ARCHTAG SST modules derived latitude estimates at a specific longitude by matching SST estimated from tag data with SST estimated from satellite images. The module first calculated a daily mean SST from all the tag’s external water temperature readings that occurred at depths of less than 10m. The ARCHTAG module used weekly mean composite SST satellite images of the GAB at a 1° grid scale. Using these data, ARCHTAG produced a single latitude estimate corresponding to the mid-point of all latitudes where SST matched the estimated tag-derived SST. Again, it was apparent from the maps of the estimated positions that the resolution of the SST data was insufficient for the estimation of fine-scale positions within the GAB.

The ARCHTAG method had several additional limitations for fine-scale position determination. The principal limitation was the software for matching tag and satellite data allowed position estimates to be made even where there were large mismatches between the two data. Furthermore, such an estimate allowed matches between tag and satellite SST at positions that bathymetric-based northern bounds on latitude indicate should have been excluded. On the basis of this assessment, we concluded that the software needed to provide a basis for rejection of estimates where mismatches occurred, and that we should allow for a cross checking of the bathymetric and sea surface temperature to ensure that estimates were not mutually exclusive.

Methods

The improved position determination methodology was developed using Data from archival tags (113 Mk 7 Wildlife Computer) deployed on 3 and 4 year old SBT in the summer of 1998. The tags recorded light level, depth and the external and visceral temperature data every four minutes. Thirty of these tags have been returned to CSIRO and 24 contain data suitable for use in this analysis. The process of position estimation using the tag data and the supplementary data on sea surface temperature and bathymetry involved a series of steps.

Step 1. Position estimates based on light data

Wildlife Computer Inc (WC) proprietary software (Hill and Braun, 2001) was used to estimate the daily position based on depth-corrected estimates of light levels. Longitude was estimated by estimating the time of midnight based on a folding method. The WC software allows the neighbourhood of each dawn and dusk period to be identified interactively, so that sensible estimates of midnight were completed. This avoided problems caused by high levels of light due to moonlight during the full moon and noise in the light signal due to extreme diving behaviour, a problem in previous automatic position estimation methods. Furthermore, direct estimates of the accuracy and sensitivity of the longitude estimate are provided by the software. Latitudes were estimated by choosing the latitude that generated a theoretical light curve which best matched the corrected light levels observed around dawn and dusk events. The software used an initial known position (the release position) to 'learn' the optimal way to generate the latitude estimates. The software also allows for latitude estimates to be generated on any particular day based on either the observed dawn or dusk light curves, or both the dawn and dusk curves. For any particular estimate, the method chosen was that which reported the most accuracy in latitude (according to the rate of change of mean squared difference of curve fits for any given estimate).

Step 2. Determining the extent of error in the light-based position estimates

The quality of longitude estimates were examined by plotting the estimates produced by the WC software and checking for outliers according to a decision rule that any daily movements of more than 120 nm were incorrect. This rule was based on the movement rates observed by Davis and Stanley (2002) during their acoustic tracking studies of juvenile SBT in the GAB during the early 1990’s. To test the assumption that the WC software would estimate significantly higher errors for the outliers identified by the decision rule than for estimates within the bounds of acceptable movement distances, the distribution of estimated errors in the latitude and longitude estimates was examined.

An independent source of information on the error in WC software geolocation estimates came from data collected while tagged fish that were captured in the GAB were transported to Port Lincoln in towed cages. The cages containing captured fish are towed for distances of up to 400 nm at a speed of approximately one nautical mile per hour. The tag depth data indicates that date that fish were captured and it was possible to compare the estimated geolocation position of the towed fish along known tow-paths and calculate the error.

Step 3. Estimates of latitude using high resolution bathymetric and satellite SST data

Having recognised that the resolution of the bathymetric and SST data used in the existing ARCHTAG modules to estimate latitude was inadequate, higher resolution bathymetry data and daily satellite SST images were used. Software to use these data sets in combination, rather than treating bathymetric and SST-based latitude estimates separately, was also developed. Bathymetric latitude estimates provided a northern bound on the possible latitude of a fish in the GAB (at a given longitude), and so possible SST latitude matches were restricted to the subset of latitudes at or to the south of the northern bound. Furthermore, rather than characterize the SST reported by both tags and satellites by a single value, a collection of summary statistics was reported for each.

The maximum dive depth in any twenty-four hour period can be used in conjunction with the local sea floor bathymetry to determine which regions the fish could have plausibly visited on any given day. This assumes that the fish remains stationary during the day. In the GAB this maximum depth information can be used to restrict the latitude estimates for a given longitude. Furthermore, position estimates were made based on the assumption that when the maximum dive depth was less than 100 m the fish was diving to the seafloor. This allows for precise estimate of latitude, corresponding to positions on the continental shelf. When the depth was greater than 100 m, it was assumed that the fish was off the shelf and not necessarily diving to the sea floor, thus providing a northern bound on the latitude estimate. This assumption was also used in previous estimation methods, however, as mentioned previously, only course scale contour maps of the GAB bathymetry were used, causing position estimates to line up along a limited number of contours. In this study fine-scale bathymetric data collected at 0.5" resolution (in longitude and latitude) from the Australian Geographic Survey Organization was used (AGSO 98 dataset).

Attempts to match the observed tag-based SST directly with the observed satellite image SST's were also made. In this approach the range of observed tag-SST was matched to the range of observed satellite-SST in a specified longitude neighbourhood. In order to estimate the SST from a tag on a particular day it was necessary for the fish to have visited the surface during that day. Biases between satellite and tag-based SST will still result, as satellite-SST images are based on measurements of only the top few nanometres of the water column. This satellite SST will only be a good approximation of the temperature of the top few metres of the water column when the surface layer is well mixed. Obviously, the best agreement between satellite SST and tag temperatures will come when the tag is closest to the surface. Reducing the maximum depth for the surface definition of the fish (e.g. 0-2 m) resulted in fewer observations of temperatures in the interval each day. Relaxing the definition (e.g. 0 - 10 m) resulted in an increased number of observations for a tag-based SST estimate. It is desirable to choose a definition of 'surface' to use in defining tag-SST that is as shallow as practicable. Two definitions of surface were compared to assist selection of the final definition: when fish were in the upper 2m and upper 5m (**Table 1**). Since there was little difference in the number of missing values between using the 2 m and 5 m cutoff for the surface definition, the 2 m definition of surface depth is used in subsequent analyses, because of an expected closer match to the satellite measurements. Biases between the temperature measured by the tag and the satellite will likely still exist due to instrument measurement errors.

Table 1. Number of days of missing data for 22 tags using depth intervals of 0-2 m and 0-5 m as the definition of "surface" for tag-SST estimates.

Tag Number	97620	97618	97622	97632	97675	97707	97708	97711	97721	97731	97733	97734	97736	97743	97754	97755	97756	97757	97760	98007	98017	98024
Total days of suitable data	64	3	87	70	70	87	87	64	87	36	64	10	7	1	64	17	70	64	64	64	64	64
Number of days where there is no SST data (no observations in depth interval during the daytime)																						
Depth 0-2 m	0	0	5	4	0	0	7	1	29	0	23	0	2	0	0	1	3	0	8	1	4	2
Depth 0-5 m	0	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	1	0	1	0	2	2
Difference	0	0	4	4	0	0	6	0	28	0	23	0	2	0	0	1	2	0	7	1	2	0
Number of days where there is little or no SST data (five or fewer observations in depth interval during the daytime, including no observations at that depth)																						
Depth 0-2 m	2	0	15	12	11	9	17	9	62	2	37	5	3	0	4	4	16	4	28	2	21	7
Depth 0-5 m	1	0	5	2	0	0	6	6	1	2	1	0	0	0	3	1	6	2	16	1	5	5
Difference	1	0	10	10	11	9	11	3	61	0	36	5	3	0	1	3	10	2	12	1	16	2

Interpolation of tag SST

On days when the "surface" was not visited or was visited infrequently, it was not possible to calculate a sensible tag-SST estimate. Five or fewer data points were considered insufficient for the purposes of predicting the tag-SST. Therefore statistical models were developed which modeled the relationship between the minimum, median and maximum water temperatures recorded in the surface interval and the minimum, maximum and median water temperatures recorded at a depth of 10 - 20 m. These models were then used to predict the minimum, maximum, and median SST in those cases where there was insufficient data to calculate the surface values directly. Relationships between the mean, standard deviation, quartiles and inter-quartile range of water temperatures at the surface and 10-20m depth were also considered, however, these statistics did not exhibit as strong relationships as those for minimum, maximum and median temperatures.

Where there were fewer than five points, minimum and maximum predicted SST were compared with the minimum and maximum observed SST and adjusted to the observed values if the observed values were more extreme than the predicted values. The statistical models were generated using the temperature and depth data pooled from all tags. Separate models were not developed for each tag, since this would have restricted analysis to only those tags for which there was a substantial amount of information about surface temperature. Since the relationship relates two oceanographic variables we would not expect that there would be any systematic tag-to-tag variation.

Multivariate analysis of variance (MANOVA), generalised linear and generalised additive models were all investigated as tools to model and subsequently predict SST. MANOVA methods were investigated because it is expected a priori that the minimum, median and maximum SST will be correlated, however MANOVA models frequently predicted a minimum SST which exceeded the predicted maximum SST for new predictions, and so were not used.

Plots of observed and predicted SST against day-of-the-year for each tag were made in order to assess the quality of the predicted values. (Note that neither day-of-year nor tag were included in the models.) Checks were also made to ensure that predicted minima, maxima, and medians were ordered correctly for each set of observations. Based upon the fit of the models, the number of missing data points, and the quality of the predictions, a refined definition of 'surface' for calculating tag-SST was chosen. Only daytime (between 6am and 7pm local time) tag-SST was estimated since SST is often cooler at night in the GAB. In addition, it seems that juvenile SBT do not visit the surface as often at night. Therefore models were restricted to predicting the minimum and maximum daytime tag-SST. Accordingly, only daytime satellite images were used to perform our latitude matches.

Satellite Images of SST

The NOAA satellites that carry the SST sensors made between four and six passes over the GAB on most days (twenty four hours) between January 1 and March 31 1998. The images of SST, processed at CSIRO Marine Research, were at a spatial resolution of 0.01deg. The presence of thin, high cloud in the atmosphere can cause satellite SST to be underestimated by up to several degrees in the GAB (Chris Rathbone – CMR Remote Sensing pers. comm). Extreme contaminated values can be easily removed by rejecting all temperatures less than 8 °C, however, less extreme contamination remains problematic. Missing data in the satellite images can result from several other reasons, including incomplete spatial coverage of a single satellite pass and dense cloud cover.

Matching tag and satellite SST

To match tag-SST estimates with satellite-SST it was necessary to choose a matching criteria and select appropriate temporal and spatial scales at which to summarize the satellite SST. In order to calculate a minimum, maximum, and median satellite SST at a position it was also necessary to have sufficient data coverage. At any position, on a given date, these statistics were determined by selecting all of the available SST pixels in a square centred on the target location, from all satellite images within a specified number of days of that date. Only 'daytime' (between 6 am and 7 pm local time) satellite images were included. If satellite data were not available for the specified position and day, the closest day containing data for that position was used. The number of days considered for this process ranged from a single day up to three days either side of the target date. The area over which data were integrated ranged from 3 x 3 pixels (0.3 Nm side squares) up to 51 x 51 pixels (30 Nm side boxes).Clearly there was a trade-off between the resolution (in time and space) with which SST could be measured, and the risk of insufficient data to adequately summarize satellite SST at a given point.

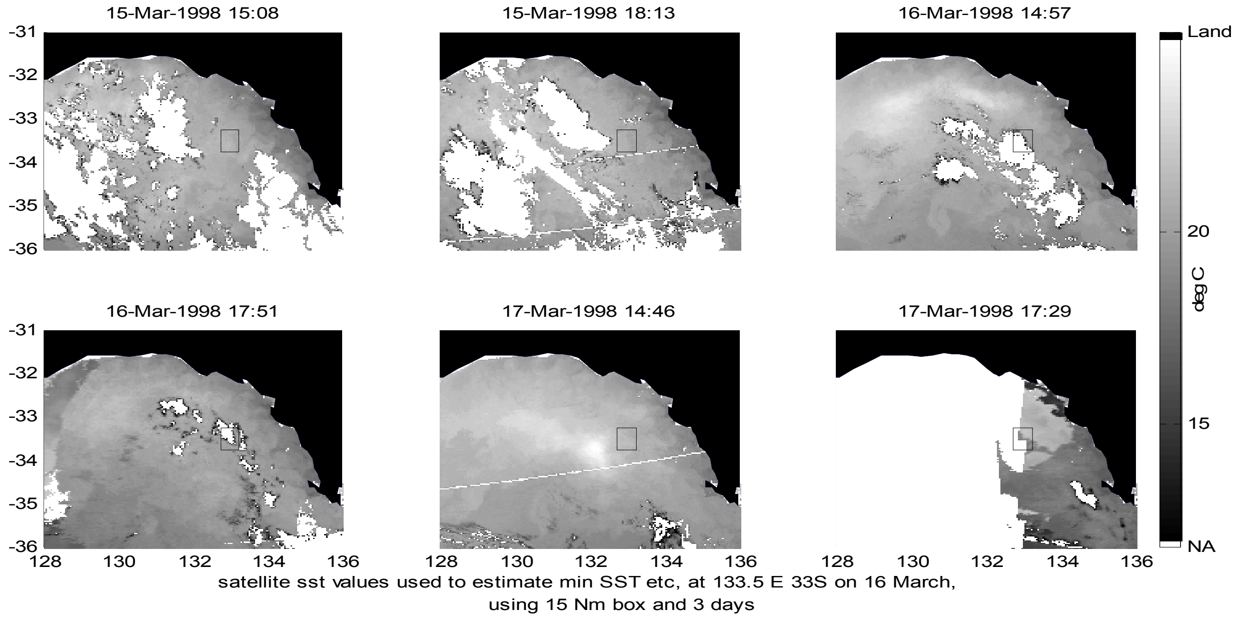


Figure 1. Example of the satellite data used to estimate the satellite SST for a location and date. The boxes mark the position of SST-values from all the images used to calculate the minimum, median and maximum SST at 133 deg E, 33.5 deg S on 16-March 1998, using a box size of 15 Nm (25 pixels) and 3 days of data.

Matching Criteria

Two criteria to match satellite and tag SST were considered:

1) Satellite and tag-based SST's were considered to "match" if the range of range of tag-SST was entirely contained within the range of satellite-SST.
   i.e. min(satellite-SST) <= min(tag-SST) & max(tag-SST) <= max(satellite-SST)
   This criterion was developed on the assumption that the satellite images approximated the true state of the environment, and that potential tag latitudes could be identified by determining where all the observed tag-SST were contained within the environmental range of satellite-SST.

2) Satellite [median, maximum] contained entirely within tag [median, maximum]
   This criterion was developed based on the knowledge that satellite images frequently report more extreme ranges in temperature variation than are actually present, due to contamination by cloud cover (satellite SST estimates are reduced by the presence of thin cloud). This criterion assumed that the observed tag-SST temperature data was much more accurate than that estimated by the satellites. However, it is also known that satellite SST temperatures may not be equivalent to the temperatures below the top few nanometres sensed by the satellite. Particularly on very calm days where mixing of the water column is low, the temperatures estimated by the satellite can be much higher than the temperature of water only a few centimetres below.

For each estimated longitude on a given day, matches were calculated for those latitudes that were plausible according to the bathymetric estimate of a northern bound on position. The northernmost and southernmost matching latitudes were returned as the 'range of possible latitudes' for that day. For each combination of spatial scale, time scale and matching criterion, the success (or reason for failure) of latitude estimates was determined and tabulated. Based on this information a matching criterion was chosen and used for all subsequent estimates. For each set of latitude range estimates for a given day for a given tag, the most precise (smallest range) estimate was used.

Choosing the final position estimate

Where no SST-based latitude matches were available, the bathymetric estimates of position were used. If the maximum daily dive was shallower than 100 m, the range of latitudes which corresponded to that maximum dive depth were reported as the latitude (usually this consisted of a single point estimate), when the depth was greater than 100 m the latitude range was reported as everywhere south of (and including) the northernmost match.

Once longitude and latitude-range estimates were determined, the optimal latitude estimate was selected by minimizing the distance travelled by each tag, subject to the latitude constraints for that day. This constrained optimisation was done using the solver package in Microsoft Excel-97. Latitude estimates that required the tag to move more than 120 Nm within twenty-four hours were considered outliers and removed, and the optimised trajectory determined once more. Missing positions (including removed outliers) were interpolated by selecting the latitude that lay on the estimated trajectory for that day. Missing positions were not interpolated outside the range of 128 to 136 deg E or south of 36 S, or where there were three or more missing neighbours (adjacent days with no position estimate).

Results and Discussion

A total of 30 tags released in the summer of 1998 in the GAB have been returned, of which we have included twenty-four in our analyses. All of the tags analysed using Wildlife Computers software had very large estimated errors in latitude (of the order of +/- 10 degrees). However, there was no obvious relationship between the size of reported errors between approximately correct (or even possible) latitude estimates and those of obvious outliers (where distance > 120 Nm from nearest position). Thus, there are no objective criteria identified by the software on which to base rejection of individual position estimates.

Predicting tag-SST when there was insufficient data to directly calculate it.

Throughout the data sets there were days when the fish did not spend sufficient time in the upper 2 m to allow calculation of a tag-SST. In all cases where there were insufficient data at the surface to directly calculate the tag-SST, there were sufficient data at greater depths (10-20m) to determine summary statistics about the temperatures at those depths.

Generalised linear models and generalised additive models both required univariate responses, and so each of the minimum SST, median SST, and maximum SST were modeled separately as functions of the three predictors (minimum, median and maximum temperature at depth). The quality of fit of both types of model was similar, however generalised linear models still occasionally predicted minima that were greater than the corresponding maximum SST. Generalised additive models did not suffer from this deficiency. In both cases, only the median and maximum temperatures at 10-20 m were significant predictors of SST. Where there were five or fewer data points for a given day and tag, the minimum, maximum and median tag-SST was predicted, and the predicted minimum and maximum compared with the observed data. If the observed extrema were more extreme than the predicted values the observed values were used, otherwise the predicted values were used. The predicted extreme was less than the observed extreme in less than one percent of predictions. The predicted median was always used in these circumstances. Plots of predicted values against each of the predictors and against day-of-year (which was not a predictor), were used to visually assess the quality of the predictions (**Figure 2**). It was noted that the predictions were often good at interpolating the minimum, maximum and median surface temperatures over time, despite the fact that day-of-year was not a predictor.

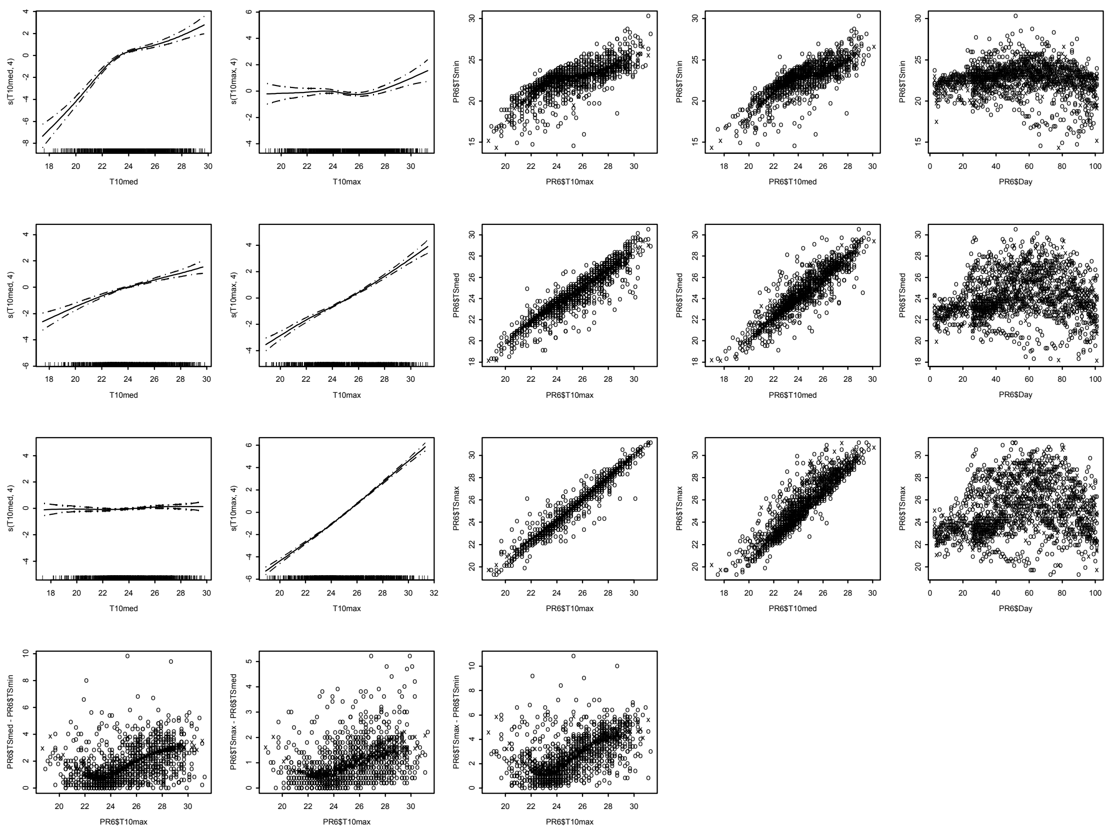


Figure 2. Predicted and observed minimum, maximum and median Tag SST. These are the GAM models used to predict SST at the surface based on water temperatures at 10-20 m (T10) -- All tags used in fitting model. In the first row the response variable is minimum SST, First column: predictor variable is median T10, Second column: Predictor variable is maximum T10. In the second row the response variable is median SST, 3rd, 4th column: Predicted (x) and observed (o) response vs. Median, Maximum. In the third row the response is maximum SST. 5th column: Predicted (x) and observed (o) response vs. Day-of-year. Final Row: check that no median< minimum, no maximum is < median, and no maximum < minimum. NB Day-of-year was not a predictor in the model, it is just plotted as a visual check.

Matching tag and satellite SSTs

Comparison between the two matching criteria revealed a greater success in matching satellite-SST with tag-SST using the first criterion: the range of observed tag-SSTs is contained entirely in the range of satellite-SSTs. The first criterion produced a match between tag and satellite SST for 66.5% of tag-days using at least one combination of days-of-images and box size. The second criterion resulted in only 11% of possible matches, in part due to a design fault. As more satellite data was included in the estimate of the "range" [median, maximum] of satellite SST estimates (i.e. by increasing box size or number of days of satellite images), the estimated range of satellite-SST at that location becomes larger, whilst the estimated range of tag-SST remained fixed. As a result, reducing the precision of the satellite-SST estimates reduced the chance of matching satellite and tag SST's. Final position estimates were therefore generated using the first matching criterion. Position estimation was carried out at every combination of spatial and temporal scale for each tag-day, and the most precise (smallest range) estimate from the collection of estimates at the various time/space scales was chosen as the final estimate for that tag-day. An example of the latitude bounding is shown in

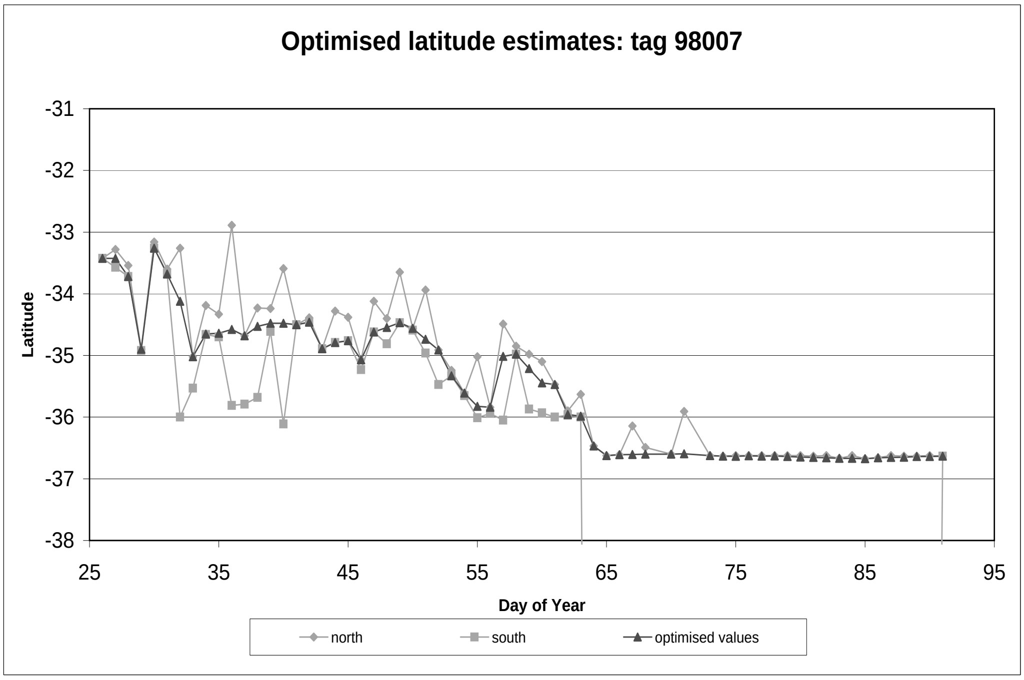


Figure 3. Optimised latitude estimates for an example tag (98007) with estimated northern and southern bounds on latitude. Note that latitude estimates from Day 63 onwards have no southern bound since these latitudes are outside the analysis region and coverage of the satellite images (128 - 136 E, 31-36 S) and have dives of greater than 100 m on those days. Those estimates were therefore discarded.

SST matching and position estimation can fail for a variety of reasons. **Table 3** summarizes the number of matches achieved for each tag according to the hierarchical matching criteria.

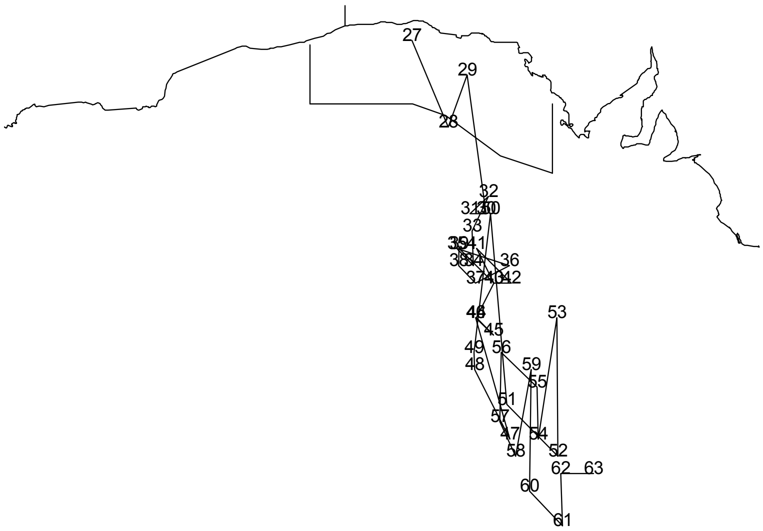
Table 3. Summary of the number of days of position estimation success before position interpolation. Match type 1: No longitude estimate or longitude outlier of >120 Nm. Match type 2: Position estimate (Latitude, Longitude) is an outlier by 120 Nm rule. Match type 3: No depth position (no dive >100m and no neighbours to interpolate position from or longitude is outside of the satellite image. Match type 4: Depth position places longitude estimate outside GAB and so no SST match possible. Match type 5: Depth position only (no match between satellite and tag SST information at that longitude). Match type 6: SST match possible.

Tag Number	97618	97620	97622	97632	97675	97707	97708	97711	97721	97731	97733	97734	97736	97741	97743	97754	97755	97756	97757	97760	98007	98017	98024
Match type	Number of matches (days)																						
1	0	1	0	0	3	0	0	1	7	0	1	0	0	1	0	0	0	0	0	3	2	0	1
2	0	0	0	2	5	11	10	4	6	2	1	0	0	3	0	7	0	4	8	0	0	0	5
3	4	19	17	1	0	0	0	0	5	0	0	0	0	14	0	14	0	1	0	13	25	18	28
4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
5	0	15	24	11	7	3	5	7	15	3	12	0	0	6	0	7	1	7	11	26	0	15	4
6	3	28	46	57	56	73	72	52	54	31	50	11	8	40	2	36	17	59	45	22	37	29	26
Total	7	64	87	71	71	87	87	64	87	36	64	11	8	64	2	64	18	71	64	64	64	64	64

Summary

Using the methods described here, positions were calculated for twenty-three tags at liberty between 1st January and 31st March 1998, for the period when the individual fish remained in the neighbourhood of the GAB survey region. An example of the final positions determined for a tag can be seen in **Figure 4**. In most cases failure to estimate a position on a particular day was due to the restriction of fine scale satellite data used in this analysis to the region of interest between 128 and 136 E, and not due to a failure of the method per se. In general most of those tags (fish) that left the survey region did so by moving first to the east and south-east. There is evidence from previous studies that some of those individuals subsequently moved west after leaving the survey region.

The improvement in estimated position is expected to allow fine-scale use of the associated behavioural information. In particular, these modified positions were used to determine the location of particular surfacing behaviours of the SBT, as in **Section 6.4**



Final position estimates for tag 98007

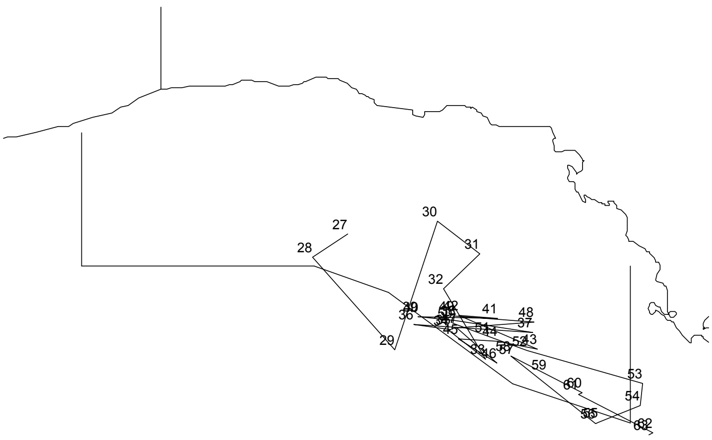


Figure 6. Initial (upper) and final (lower) position estimates for example tag 98007.

References

West, G.J. and Stevens, J.D. (*Editors*). 2001. Archival tagging of school shark, *Galeorhinus galeus*, in Australia: initial results. Environmental Biology of Fishes **60**.

Gunn, J. S. and Block, B. A. Archival, satellite and acoustic telemetry: where do tunas go in the wild and what do they do? In press.

Gunn, J.S., Polacheck, T., Davis, T.L.O., Sherlock, M., and Betlehem, A. 1994. The development and use of archival tags for studying the migration, behaviour and physiology of Southern Bluefin Tuna, with an assessment of the potential for transfer of the technology to groundfish research. Proc. I.C.E.S. Mini-Symposium on Fish Migration 23p.

Davis, T. L. O., Stanley, C. A., 2002. Vertical and horizontal movements of southern bluefin tuna (Thunnus maccoyii) in the Great Australian Bight observed with ultrasonic telemetry. Fishery Bulletin 100 (3), 448-465.

Hill, R. D. and Braun, M. J. 2001. Geolocation by light-level. The next step: Latitude. No. Proceedidngs of the Symposium on Tagging and Tracking Marine Fish with Electronic Devices, February 7-11, 2000, Honolulu.

Musyl, M. K., Brill, R. W., Curran, D. S., Gunn, J. S., Hartog, J. R., Hill, R. D., Welch, D. W., Eveson, J. P., Boggs, C. H., Brainard, R. E. (2001). Ability of archival tags to provide estimates of geographical position based on light intensity. In J. R. Sibert & J. L. Nielsen (Eds.), Electronic Tagging and Tracking in Marine Fisheries (pp. 343-367). London: Kluwer Academic Publishers.

Welch, D. W., Eveson, J. P., 1999. An assessment of light-based geoposition estimates from archival tags. Canadian Journal of Fisheries and Aquatic Sciences 56 (7), 1317-1327.

6.4 Surfacing behaviour of juvenile SBT in the Great Australian Bight

Jeremy O'Reilly, Alistair Hobday, John Gunn, Ann Cowling

Abstract

Depth and location information from archival tags deployed and recovered from juvenile southern bluefin tuna (SBT) in the Great Australian Bight (GAB) were used to investigate the relationship between surfacing, and hence detectibility in an aerial survey, and environmental conditions. Several methods for classifying depth information into behavioural definitions related to surfacing were considered for four time-scales in different areas of the GAB. Generalised linear models were developed to explore the relationship between the surfacing measure (response) and environmental (covariates) variables in each of three areas covered by the aerial survey. The surface-oriented behavioural definition and the whole-day time-scale were chosen for final model selection based on preliminary analysis of the data. The important environmental covariates in the final models included the atmospheric variables cloud cover, wind direction, barometric pressure and air temperature in Area 2, moonphase and chlorophyll in Area 3, and water temperature and chlorophyll in Area 4. The final models explained between 26.9% and 51.2% of the null deviance, although compared with bootstrapped models this was reduced to 19-25%. There was a significant difference in the SBT surfacing rates in the three areas examined; the range in the proportion of time at the surface was highest in Area 2 (55.3%) and lowest in Area 4 (35.9%).

Introduction

Juvenile southern bluefin tuna (SBT) spend the austral summer months in the Great Australian Bight (GAB) where they are targeted by a surface fishery. Fish are visible at the surface, and detection of schools is possible from light aircraft (Chen et al. 1995; Polacheck et al. 2000). Spotter planes are used to assist the fishing fleet, and to undertake scientific surveys of abundance. An aerial survey has been carried out since 1992, and an index of relative abundance has been developed (Chen et al. 1995; Cowling and Millar, 1998). This index of abundance is considered critical in detecting stock collapse and recovery of SBT.

The relative abundance index does not account for differences in surfacing rates in different parts of the GAB, or in different environmental conditions. If fish surface at different rates in different parts of the GAB or under different environmental conditions, then the relative abundance index might be influenced by interannual changes in SBT distribution or the GAB environment. If there is no difference in the surfacing rate, it will not influence the relative abundance index, as the rate will just be a multiplier.

Archival tags have been deployed on SBT to study vertical and horizontal movements. These tags record depth, internal and external temperature, and ambient light. The location of the fish, and the depth should allow estimates of the surfacing behaviour of the fish to be generated in a number of geographic regions. Thus, archival tags are a way of assessing the fraction of time that fish might be visible from the surface during an aerial survey. Corrections to sighting estimates can then be made based on the surfacing rates.

The goal of this study was to generate response variables linked to surfacing behaviour from archival tags on SBT within the GAB and develop models that explain relationships between the observed surfacing proportion and the environment.

Methods

The tag data analysed for surfacing behaviours were selected from a similar area within the GAB covered by the RMP aerial survey during January to March. Data from 21 tags at liberty in the GAB between January 1-March 31 in 1998 were used in this study. The archival tags recorded depth every four minutes; each four-minute recording is referred to as an observation. Each tag thus provided a time series of data; for this analysis depth, location, and external (water) temperature were used.

Fish location

Fish location was not recorded directly from the tags, but was calculated using light data recorded by the tags. This location estimation process is described in Section 6.3.

Fish location could not be estimated for all days; missing location estimates were linearly interpolated from known neighbouring locations. Interpolated locations were used to place observations into one of four spatial areas.

Spatial Areas

To allow incorporation of the results of this study into analyses based on aerial survey data, separate analyses of surfacing were planned for each of four areas used in the aerial survey analyses. The boundaries of the aerial survey regions were extended to include observations for tagged-SBT to 36°S. These observations were included the in regions to the north (**Figure 1)**. Location estimates for tagged-SBT to the east and west of the survey region were excluded from analyses. There were only four SBT location estimates from the tags in the western region (Area 1), and it was not considered further in this study. Thus, the analysis presented here considered surfacing behaviour in only Areas 2, 3, and 4.

Surfacing Behaviour Definitions

The response variables for models in this analysis were all based on the presence of fish at the surface, where they might be detected in an aerial survey. Four sets of behavioural classification for SBT surfacing were considered, although the final analysis was based on only one classification. All behaviours were based on the vertical position in the water column of tagged SBT and are discussed below:

1. Vertical behaviours,
2. Vertical behaviours visible at the surface (surface-oriented behaviours),
3. Sunbaking behaviour versus the other behaviours, and
4. Simple depth behaviour: Depth of observation less than 5 m

Vertical behaviours

Visual examination of depth time series from the archival tags led to a classification scheme with five distinct types of behaviour associated with the surface, and a further behaviour class not associated with the surface (**Table 1)**. An algorithm was developed to detect the presence of each behaviour within the tag time series. For each behaviour, an initial behaviour class was determined for each observation. Initial behaviour classification depended upon several local features of the observation in the depth time series. Most important of these was the depth of the observation and its neighbours. In addition, the number and duration of excursions between depth strata was used in some of the classification definitions. Once an initial behaviour type was identified, the series was 'smoothed' to ensure that the behaviours identified occurred in blocks, i.e. that singletons and short-term deviations from any behaviour did not change the overall behaviour in that temporal neighbourhood.

The time series section under consideration was converted to a binary time series according to the presence (1) or absence (0) of the particular behaviour at each observation. This binary time series was then smoothed using a binomial-weighted kernel smoother (see Hastie and Tibshirani, 1990). Smoothing bandwidth varied with each behaviour. The smoothed output was converted back to the binary form according to whether the smoothed value was less than or greater than 0.5 (< 0.5 => 0, > 0.5 => 1). Smoothing the binary time series sometimes resulted in poor estimation of the time when the behaviour changed. Therefore, for each continuous behaviour segment in the smoothed series, any observations before (after) the first (last) classification of that behaviour in the corresponding part of the unsmoothed series were re-classified as not exhibiting that behaviour (**Figure 2**).

Because behaviours were identified sequentially and smoothed from the original time series, it was possible that a single observation could be classified as belonging to more than one behaviour type. This does not imply ambiguity in the identification of behaviours, as some behaviours were nested (see descriptions below). Where multiple behaviours were identified for an observation, a single behaviour was chosen according to the following priority rule:

Periodic > Sunbaking > Midwater Surfacing > Midwater > Rare surfacing > Non-surfacing

Examples of the vertical behaviours used for this classification method are shown in **Figure 3,** and discussed below.

Sunbaking behaviour

Sunbaking behaviour was defined as a fish near the surface with only infrequent and short excursions to deeper water **(Figure 3 A)**. Sunbaking behaviour was classified by selecting all observations in the time series shallower than 5 meters and then smoothing the resultant time series using a bandwidth of five observations (20 minutes).

Midwater and Midwater Surfacing behaviours

High frequency changes in depth between 40 meters and the surface were termed midwater behaviours. They were defined as periods in which fish stayed predominantly above a depth of 40 m. Two types of midwater behaviours were classified, based on frequency of surface visitation.

Initially both midwater behaviours were identified by finding all observations shallower than 40 m. Each section of the time series was checked to ensure the behaviour was persistent, and not a transition between other behaviours. The boundaries of the midwater behaviours were defined by the turning points of the first (last) peak (trough) in each section. The peaks and troughs within the time series were identified by based on local behaviour (change in direction, duration of 'runs' in a particular depth direction, and a minimum time criterion for peaks and troughs). A window of nine observations was used to smooth the resultant segment.

Midwater Surfacing behaviour was then identified as the subset of Midwater behaviour where there were more than four visits to the surface (< 10 m) in each window of nine consecutive observations (**Figure 3 B)**. The boundaries of Midwater Surfacing sections were the first and last visit to the surface in the identified sections (**Figure 3 C)**.

Periodic surfacing behaviour

Periodic surfacing behaviour was defined by the fish periodically diving between some depth and the surface (**Figure 3 D)**. The time spent near and away from the surface was greater than observed in the high-frequency Midwater Surfacing behaviour. Periodic behaviour was identified by locating segments where observations within 10 meters of the surface were followed by an excursion to > 20 meters and back to 10 meters within four further observations (16 minutes), followed by another similar dive within three further observations (12 minutes). Periodic behaviour was smoothed using a window of seven observations (28 minutes).

Rare Surfacing and Non-Surfacing behaviours

Rare surfacing was those observations near to the surface (<5 meters) where Midwater or no other behaviour had been identified. The goal was to identify isolated visits to the surface. This behaviour was not smoothed, as a segment usually contained only a single surface observation. All remaining unclassified observations were defined as non-surfacing behaviour **(Figure 3 F).** An example of a depth time series classified according to these vertical behaviours is shown in **Figure 4.**

Vertical behaviours visible at the surface: surfacing-oriented behaviours

The first classification scheme could not easily be converted to a binary time series for the presence or absence of fish at the surface. Thus, a second time series for each tag, called surface-oriented behaviours, was developed based on the vertical behaviours described above. This set of time series for each tag was created by classifying each behaviour observation as either

- Surface-oriented (sunbaking, midwater surfacing, rare surfacing behaviours, and the portion of periodic behaviours where the depth was < 20 m), or
- Non-surface-oriented (midwater, the portion of periodic >20 m, and non-surfacing behaviours).

Periodic behaviour was divided into two depth strata to distinguish observations where the fish was near the surface from observations where it was not. This avoided inflating the proportion of time the fish spent near the surface. The numbers of surface-orientated and non-surface observations in the time interval being considered were counted to form a binomial response.

Sunbaking behaviour versus other behaviours

Sunbaking behaviour was the surfacing definition most strongly associated with the surface and least likely to suffer from misclassification. Thus, a third definition of surfacing was created, where only the sunbaking behaviour type was classified as surface-oriented and all other types classified as non-surfacing.

Simple depth behaviour

The fourth and most simple surfacing classification scheme was based simply on the vertical position of the fish at each observation. It was defined as all observations where depth was less than five m, and time series in binomial form for the presence/absence of the behaviour were generated for each tag.

Time scales for analyses

The relationship between the response variable (behaviour), and the predictor environmental variables, was considered at four temporal scales. The finest scale was approximately half an hour; the coarsest was the entire day. A third temporal scale was created by dividing the time between dawn and dusk into two even halves (AM and PM), while a fourth scale of response variables was created by dividing the day into five intervals. These intervals corresponded to periods of the day during which anecdotal accounts from spotters suggested differences in surface abundance of SBT:

1. Dawn: Two-hour period following sunrise (1/6th day length)
2. Morning: Two-hour period following dawn period (1/6th day length)
3. Midday: Four-hour period between Morning and Afternoon periods (1/3rd day length)
4. Afternoon: Two-hour period prior to dusk period (1/6th day length)
5. Dusk: Two-hour period before sunset (1/6th day length)

In all temporal scales an hour or half-hour was 1/12th or 1/24th of the time between sunrise and sunset. Thus, changes in day-length did not affect the number of intervals, but did affect the number of 4-minute observations per interval. Times of dawn (sunrise) and dusk (sunset) used to determine day-length and thus the duration of time intervals were calculated from the tag location using Clear Sky Institute's XEphem software (www.clearskyinstitute.com). As there is little difference in the time of dawn and dusk on a given day across all possible locations in the region of the GAB considered, these day length estimates are quite robust. The similarity of dawn and dusk times in this region is part of the problem in estimating latitude from the archival tags, as very small changes in dawn and dusk time correspond to large changes in latitude. When no location estimate was available for a day, the time of dawn and dusk was interpolated from surrounding days.

Environmental predictor variables (covariates)

Environmental variables were matched to each response variable, for each time interval considered, to generate a set of covariates. In addition, Julian day-of-the-year (between 1 January and 31 March) was also predictor variable. For days when fish location was unknown, the environmental variable was interpolated using values for known locations before and after that day, expect for moonphase which was the same for all locations each day and SST which was tag-based. Cubic spline interpolation was used to estimate missing values, except for chlorophyll concentration and cloud cover. Linear interpolation was used for these two variables, as other interpolation methods resulted in values unlike the neighbors from which they were interpolated. When data existed only before or after the day with unknown location, the nearest observed environmental values were used, as cubic spline and linear extrapolation usually produced unrealistic estimates. For intervals of less than 1 day, and where data existed at higher frequency (ie. DAR and tagSST data), interpolation was based on the value of the covariate in the same time interval on neighboring days. Variables from environmental data sets used as covariates are described below;

Sea surface height (SSH)

SSH data was produced from altimetry measurements made by the TOPEX/POISIDON satellite and corrected with tide-gauge data from the coast of the GAB. Data for the GAB were available for 9-day composites at a spatial resolution of 0.25°, and the closest value to the location estimate was selected.

Sea surface temperature (SST)

SST was calculated using data from the archival tags. The top two meters of the water column was defined as the surface and temperature measurements from observations within this depth range selected. The median SST for each time interval where the fish visited the surface at least once was calculated. If the fish did not visit the surface during the interval, the median SST was interpolated from neighbors to the interval mid-time.

Sea surface color (SSC)

Daily images of SeaWiFS satellite SSC (chlorophyll a) at 1-km resolution were available for the GAB. The natural log of chlorophyll-a concentration was used in analyses, since this was the scale at which errors in measurement were additive (this variable is actually called logchl in the remainder of the paper). Values for a 0.25° box centered on the location and for a period of seven days were matched to each observation.

CSIRO Division of Atmospheric Research (DAR) data

Five modeled atmospheric data at 6-hour intervals with a spatial resolution of 0.1° were used.

- Modeled air temperature at 1500 ft.
- Modeled total cloud cover.
- Modeled wind speed at 10m.
- Modeled wind direction at 10m.
- Modeled barometric pressure at 1m.

DAR data from the closest point were matched to the daily tag location estimates to get four daily values at 6-hour intervals. For analyses at finer intervals, these 6-hour values were interpolated to the mid-point of each time interval. When analysis was conducted at half-day intervals (am/pm), the 6-hour values from mid-morning and mid-afternoon were used. For the entire-day interval analysis, the mean of the mid-morning and mid-afternoon value was used. When interpolating wind direction, the orthogonal components of the wind vectors (wind speed and direction) were interpolated separately and then recombined.

Moon-phase

Moon-phase data (fraction of the moon surface illuminated) from the U.S. Naval Observatory Astronomical Applications Department website (http://aa.usno.navy.mil/data/docs/MoonFraction.html) was converted to an integer representation of moon-phase, for day 0 (full moon) through to day 28 and matched to each day of tag data.

Periodic Splines

The periodicity of two of the environmental covariates, wind speed and moon phase, was incorporated into the models. Periodic basis splines differ from sine and cosine terms in that the maximum amount of curvature is where there is most data, effectively allowing a data-driven choice of phase. Periodic basis splines with a total of 2 degrees of freedom were used in the final model selection process. (Generated using an adaptation of a splus code posted to S-news by Douglas Bates, bates@stat.wisc.edu).

Model-fitting

Binary response generalised linear models were initially used to identify significant covariates, and explore their relationship with the behavioural responses. However, preliminary plots of the fitted values vs. residuals at each time scale and for each behaviour response variable demonstrated a substantive lack of fit. Kernel estimates of the variation in residuals, even at the half-hour intervals (7-9 data points per interval), were an order of magnitude higher than what would be predicted by a binomial response. In was apparent that the assumption of an underlying binomial distribution and associated variance-mean relationship was invalid, and overdispersion was present. Model selection methods that did not account for overdispersion in the response thus produced extremely overfitted models that were not robust to minor changes in the data (such as removing a single outlier) (**Appendix 6.4A**).

Consequently binary response quasi-likelihood generalised linear models were fitted to the data since these allowed direct estimation of a scale parameter whilst maintaining the form of the variance-mean relationship. These models were fitted for each of the three different kinds of surfacing response at each of the four time scales in each of the areas, with the exception of the half-hour time interval. Fitting preliminary models using the half-hour interval showed over-dispersion remained problematic for each surfacing definition. This indicated overdispersion in this dataset could not be eliminated by subdividing the time series into smaller temporal intervals (**Appendix 6.4A**) and so the half-hour interval was not used further.

Final Model Selection

A single time scale and behaviour definition was chosen for final analysis in all areas, based on qualitative examination of the plots of fitted against response for each time and behaviour combination (**Figure 5**). The combination of surfacing response classification (surface-orientated behaviour) and time scale (whole-day) that produced the strongest relationships and most parsimonious fits was selected for final model analysis.

The final model selection method varied slightly according to the quantity of data in each area. Interactions were selected according to *a priori* hypotheses about potential relationships between covariates. Full models with covariates and interactions were reduced by backward selection to give an estimate of the number of parameters that would be necessary for a significant fit and eliminate terms from an over-specified model. The terms in the final models were chosen by applying the AIC in order to assess likely contribution to explanation of deviance by covariates. Where the degree of overdispersion was similar between models at different time scales, models that used fewer parameters to produce an equivalent quality of fit were considered superior.

For Areas 2 and 3, all covariates, covariate*tag interactions, covariate*day-of-the-year interactions, SSC*SSH interaction and wind speed*wind-direction interaction were included in an initial model, as were the third order interactions tag*SSC*SSH and tag*wind speed*wind-direction. Wind direction and moonphase were explicitly included as periodic covariates, each with two degrees of freedom, via the use of periodic basis splines. With the exception of the wind speed*wind-direction interaction, interactions involving wind-speed or moon-phase were treated like interactions involving non-periodic effects. Interaction between factor terms and periodic spline terms tended to introduce so much flexibility into the models that the main effects became meaningless. The wind speed*wind-direction interaction was modeled as bs.per(wind-direction, 2)*wind-speed. More flexible fits for the covariates were investigated by using natural splines of higher order for those covariates that remained in the model. For analyses at finer time scales than the whole day, interactions between covariate and time interval were also considered.

In Area 4 there was insufficient data to fit all second-order interaction effects simultaneously. Forward fitting was used to fit main effects (including periodic splines with two degrees of freedom for periodic covariates). Preliminary models found that tag was the only significant main effect, and so models of the form

tag + covariate + tag*covariate

were fitted for each covariate, and a new initial model built up using up to 3rd order interaction between those effects which were significant in individual models. Backward selection and higher order natural splines were then investigated as for Areas 2 and 3.

Preliminary plots of the tag interaction terms indicated that curvature effects in the interaction terms were perhaps fitting noise and identified tags with insufficient data. Those tags were removed from those analyses. Restricting interaction effects to a linear form in interactions between continuous and categorical variables produced models that were more robust to minor changes in the data (such as removing outliers).

Choosing the between alternate models: statistical tests

In typical analyses of binomial data, where overdispersion is not present, testing for significant differences between alternative, nested models is performed with chi-square (χ^2) test. However, when overdispersion is present, tests of the change in deviance between models need to account for a dispersion parameter. Explicit modeling of the variance-mean relationship in preliminary models showed a constant multiplier of the binomial variance-mean relationship best accounted for the form of overdispersion encountered and allowed the dispersion parameter to be estimated in subsequent analyses.
If the sample size is large and the observations are approximately independent the most appropriate test to use for testing change in deviance between models is an F-test, as both the estimate of the dispersion parameter ($\hat{\sigma}^2$) and the change in deviance are

expected to have approximately χ^2 distributions (on different degrees of freedom) (McCullagh and Nelder, 1989). While the time series nature of the observations invalidated independence of the responses, F-tests remained the best tool available to guide model selection.

F-tests were used to test scaled change in deviance between models at each step of model selection to assist selection of the most parsimonious model. In each case the dispersion parameter was estimated from the Pearson residuals of the larger model. Theoretically, if there are too few parameters in the model then the true dispersion parameter could be over-estimated, as apparent overdispersion could be due to lack-of fit. Conversely, a dispersion parameter estimated from an over-fitted model could underestimate the true dispersion by fitting noise. In practice, there was virtually no difference in the estimates of the dispersion parameter by either method. For comparisons of nested models it is necessary for the dispersion parameter used in statistical tests to remain constant (McCullagh and Nelder, 1989). Where backward model selection was possible, the best estimate of the dispersion parameter was obtained and used throughout model selection. If backwards selection was not possible (too little data), the dispersion parameter of the larger model from forward model selection was used.

Finally, models with interaction terms were compared graphically with models without interactions to examine whether the extra parameters in those models were justified by plotting the response (observed surfacing proportion) against fitted values. This graphical comparison was considered necessary as the statistical tests (F-test) used in model selection were only approximate. A subjective improvement of the fit was required for the inclusion of extra interaction terms to be justified.

In the final models, the scale of the response corresponds to the log odds (logit) of surfacing; numbers on the response scale (y-axis) can be converted to a surfacing probability;

$$p = \frac{e^{response}}{1 + e^{response}}$$

For example, a fitted value of –2 corresponds to a fitted probability of 0.12. Note, however, that the sum of the logit of probabilities due to each covariate determines the final fitted value for any particular observation.

2.8. Significance of the model fit: Bootstrapping and R^2 values

R^2 values for each model can be generated by calculating the proportion of deviance explained by the model to the total deviance in the data. Because significance tests for overdispersed binary response generalised linear models are approximate at best, bootstrapping was used to compare the observed model fit with the fit if the response variables were randomised. Randomizing the response variables associated with the set of predictor variables allowed models to be fit to covariates that had an identical correlation structure. The null deviance of these bootstrapped models is the same as that from the real data. A comparison of the observed R^2 with the distribution of 10,000 bootstrapped R^2 values for each model was used as a guide as to whether the fitted model was an improvement over what would be achieved if responses were randomly associated with the set of environmental covariates.

Results

A total of 19 tags was suitable for analysis (**Table 2**). Area 4 had the fewest tag days used for analysis (n = 152) and Area 3 the most (n = 663). Given the period of time considered, each tag could be present in the three areas for a maximum of 89 days; the most observed was 87 days (tag 97707) and the fewest was 17 (tag 97755). Each tag was present in an area for between 0 and 66 days, with the number of days suitable for analysis in each area between 0 and 64 days.

Final model selection

The best time scale and behavioural classification method was chosen by considering which combination produced the models with the strongest relationships. Fitted value plots for each surfacing definition and time interval combination are shown for Area 3 in **Figure 5**. Similar patterns were observed for Areas 2 and 4. At finer time scales, highly overfit models failed to produced fits of reasonable quality with little trend in residuals evident in the plots of observed vs. fitted values (data not shown). For the coarser time scales, model fits appeared to be better, so non-significant terms in the models were removed according to an AIC stepwise selection method, and plots of response versus residuals examined again. This approach allowed the time scale of response with the strongest relationship for the fewest parameters in each area to be identified.

In both Areas 2 and 3, the surface-oriented behavioural definition produced fits of similar quality and almost identical R^2 values at both the AM/PM and whole-day time scales compared to the other behaviour definitions. However, twice as many parameters were required to produce a fit of similar quality at the AM/PM time scale (twice as many observations as for the whole-day time interval). This suggested that the extra degrees of freedom were merely fitting noise. Examination of these plots also suggested the presence of an overdispersed response with respect to the binomial distribution (**Figure 5)**. The degree of overdispersion present in residuals for the 5-interval time scale (**Figure 6**) was so high that there was no clear variance-mean relationship evident, and this time-scale was not used further.

The comparison between the surface-oriented behaviour, sunbaking, and simple depth behaviour classification schemes (**Figure 5)** led to the decision to use the surface-oriented behaviour classification calculated over the whole-day interval in the final models for each area.

Final Model Area 2

Backward selection of main effects and interactions, and subsequent investigation of curvature effects in model terms by forward fitting, resulted in a final model for Area 2 of the form:

Surfacing ~ tag + bs.per(dar.winddir, period = 360, degree = 2) + dar.cloud + dar.air + dar.bp

This model explained 26.90% of the null deviance (**Table 3**). Although cloud cover appeared to be the most non-significant covariate in the model, ($p<0.1129$), a model with it removed explained significantly less deviance (change in deviance F test,

$p<0.005$), and so it was retained. Because the cloud cover (dar.cloud) was significant, it follows that air temperature (dar.air), which explained more deviance for the same degrees of freedom, was also significant. Hence, both terms were retained in the final model presented above.

Plots of the fitted terms for this model**,** indicated that most of the explained deviance was due to the tag effect (**Figure 7 A**). There was a slight increase in surfacing when the wind direction was from 200° (~SSE) **(Figure 7 B).** A decrease in surfacing was associated with increased cloud cover (**Figure 7 C)**, decreased air temperature (**Figure 7 D)**, and decreased barometric pressure (**Figure 7 E)**. The residuals about the fits also demonstrated most of the variation in the data remained unexplained by the fitted model. Diagnostic plots of the standardised residuals showed no systematic trend (**Figure 8 A)** or heteroscedacity (**Figure 8 A B)**. The scatter of the points on the plot of observed versus fitted values showed a trend, and indicated the strength of the fitted model (R^2) (**Figure 8 C**). Furthermore, the quantiles of the standardised residuals were surprisingly normal for a binomial model (**Figure 8 D)**.

Final Model Area 3

In a preliminary model, there was a weak but significant response to SSC (logchl) and the tag*logchl interaction, which may have been due to influential SSC values. Therefore, 12 records with extreme logchl values [logchl < 0.01 or logchl > 0.3] were removed from the data set, and it was re-analyzed. Both logchl and tag*logchl remained significant in the final model (**Figure 9, Table 4).**

Surfacing ~ tag + bs.per(moon, 2, period = 29) + ns(logchl, 2) + ns(Day, 2) + tag*Day + tag*logchl

In Area 3, most of the variation remained unexplained by the fitted model (**Table 4**, R^2 = 30.06%), and inter-tag variability accounted for a large portion of the explained variation, as tag was the most variable covariate (**Figure 9 A)**. There was a clear decrease in surfacing in Area 3 as the year progressed **(Figure 9 D)**. A slight increase in surfacing was observed during the new moon (**Figure 9 B)** and when logchl was moderate (**Figure 9 C,** logchl ~ 0.05 - 0.15). The tag*day interaction showed differences between tags **(Figure 10)**. Two tags (**Figure 10,** subplot 7 (tag 97711) and 10 (tag 97733), from left to right) showed a reduced decrease in surfacing with day-of-the-year compared to other tags, with a slope in the opposite direction to the covariate "day". Two other tags (**Figure 10,** subplot 16 (tag 97760) and subplot 19 (98024), from left to right) showed a stronger decrease in surfacing with day-of-the-year than the covariate "day", as the slope was in the same direction. Tags with a slope close to zero approximate the same relationship as for the covariate alone. The second interaction, tag*logchl, also showed some tags with different effects than the chl covariate alone, as indicated by slopes different from zero **(Figure 11,** subplot 5 tag 97707, subplot 15, tag 97757).

Again, plots of residuals indicated normality of the standardised residuals (**Figure 12 D)**, with no substantial trend or heteroscedacity (**Figure 12 A and B).** There was a moderately strong relationship between observed and fitted values across most of the data range **(Figure 12 C**).

Final Model Area 4

There was limited data in Area 4 (**Table 5)**, therefore it was not possible to backward select models starting with a model which fit all covariates and interactions of interest.

Backward selection from a model with only linear main effects resulted in a final model in which tag was the only significant predictor. To generate more complete models, forward fitting of individual covariates, plus the associated interaction with tag, was undertaken for each covariate separately. Only two of these models had significant terms other than tag; those terms were tagsst and logchl. A model with both of these terms and all possible interactions up to third order was used as an 'initial' model. Backward selection removed the third order interactions, but all other interactions remained significant, to produce the final model;

surfacing ~ tag + tagsst + ns(logchl,2)+ tag*tagsst + tag*logchl + tagsst*logchl

Forward fitting of curvature effects (via natural splines, ns) of the complex model 4.2 indicated that slight curvature in the logchl effect was significant. More deviance in the data was explained by the final model than for the other areas (**Table 5,** R^2 = 51.21 %). As in other areas, the bulk of the explained variance was due to inter-tag variation (**Figure 13 A)**. In particular, some tags had very few observations in Area 4, and might have undue influence on the interaction fits (**Figure 14,** subplot 8 tag 97741; **Figure 15,** subplot 4, tag 97711). Surfacing showed a slight decrease with increased SST (**Figure 13 B**), however, there was much variation between tags, as indicated by the range and magnitude of slopes in the tag*tagsst interaction plots (**Figure 14).** Surfacing was reduced at intermediate values of logchl (**Figure 13 C).** This pattern was found in most tags, as indicated by slopes close to zero in the tag*logchl interactions **(Figure 15).** The final interaction term in the model was between tagsst*logchl (**Figure 13 D).** To display the interaction effect, only a single plot is required if the relationship between the interaction effect and the product of the mean corrected covariates is shown. This interaction shows a decrease in surfacing as the product of the mean corrected covariates decreases. This effect is probably due to influential values of logchl, although the most influential values were already removed from the final model.

Plots of residuals indicate no substantial trend, but slight heteroscedacity (**Figure 16 A and B).** The relationship between observed and fitted values across most of the data range was strongest for highest predicted surfacing **(Figure 16 C**). The normality of the standardised residuals is reasonable for a binomial model for this sample size (**Figure 16 D**).

All models

The terms in the final models are summarised in **Table 6.** No single environmental variable was important in all areas. The atmospheric variables were important only in the nearshore Area 2, while in Areas 3 and 4 chlorophyll was significant.

The final models were used to generate surfacing rates for each area, based on the median values of the significant environmental variables in each area (**Table 7**). The surfacing rate for an individual tag in an area ranged from 4% to 86 %, with an overall average of 44% (**Table 8).** The average rate for each tag for all the areas ranged from 14% to 66 % (average 43%). The surfacing rates in each area differed, with surfacing being highest overall in the inshore Area 2, followed by the offshore central portion, and were lowest in the eastern Area 4 (**Table 8).** The difference in surfacing rates between areas was significant (ANOVA on arcsin $\sqrt{p}$ transformed values, $F_{2,40} = 4.022$, $p < 0.03$).

Bootstrapping

Examination of the plots of fitted values for effects against residuals, demonstrates that there the variability in the residuals is much larger than the amount of variation explained by the individual terms in the model. This is reflected in the relatively low R^2 statistics for each of the models (**Table 9).** The distribution of R^2 statistics generated via bootstrapping for models with the same structure as the final model, but with random responses, showed strong evidence that each model was describing a relationship between surfacing and the environmental covariates (**Table 9).** The bootstrap estimates were approximately normally distributed with only a slight skew to the right. In each area, the observed R^2 statistic was larger than the largest of the 10,000 bootstrap estimates of R^2 made under the hypothesis of no relationship between the response (surfacing) and environmental covariates. Thus, a the hypothesis is rejected at a significance of $p < 0.0001$ in each case. The difference between the observed R^2 and the mean (or median) bootstrapped estimate is an estimate of how much extra variation had been explained by the models than expected if there had been no relationship. The adjusted R^2 values (observed – mean) are lower, and indicate the models explained between 19 and 25% of the deviance.

Discussion

This analysis showed that surfacing behaviour could be explained by models that incorporated environmental covariates. The final models chosen were different between areas. Because the goal of this study was to find the best model of surfacing response for each area, no test of the ability of a model from one area to explain deviance in another area was undertaken. This would be a useful exercise in the future, as it would provide some indication of the generality of the model results.

Final Models; Significant terms

<u>Tag</u>

A significant tag term in the final models indicated that tagged SBT spent different amounts of time at the surface within each area.

Covariate terms

The significant covariate terms in each model varied between areas. In general, atmospheric variables were important in Area 2. Without using information from tags, these variables might be considered more likely to influence the detection of fish at the surface, rather than the behaviour of the fish. This analysis shows that the behaviour of the fish appears to be related directly to these atmospheric variables. The relationship between surfacing and covariates was weakest in this area, suggesting that the environmental variables are themselves proxies for oceanic variables that the SBT could directly sense. In Areas 2 and 3 some oceanographic variables (chlorophyll and sea surface temperature) were significant. The relationship between these variables and SBT behaviour in under current investigation, as they appear to indicate preferences for feeding and thermal environments.

Only in Area 3 was day of the year significant

*Tag*covariate terms*

Lack of significant tag*covariate terms in the models indicated that the individual tags (SBT) responded to the environment in the same way. Some significant differences between how the fish responded to the environment were found, as outlined in the Results. Such differences might be due to differences in the size of the fish, or recent feeding history. These differences should be further investigated.

Lack of fit

All of the fitted models indicated lack of fit, and had relatively low amounts of null deviance explained after correction with the bootstrapping method. Collectively the results suggest that the relationship between surfacing and environment is unlikely to be useful in a predictive sense. This lack of fit may be due to a number of factors, including

- a mismatch in the scale at which fish experience their environment and the scale at which it was included in this study.
- imprecision in location estimates upon which these models depend.

Individual to school

These surfacing behaviours and relationships with environmental covariates are generated from individually tagged fish. To extend the results from an individual fish to a school, an assumption must be made about how individual behaviour is related to school behaviour. Without additional information, that assumption is that when an individual was close to the surface, or exhibiting a surface-orientated behaviour, some fish from the school would have been visible at the surface throughout the period of that behaviour.

Application to Aerial Survey Abundance Index

These surfacing rates can now be applied to analysis of abundance estimates from the RMP aerial survey data. The difference between spatial areas used in this analysis indicates that the abundance index should be derived for the same areas and then corrected for surfacing rates, to produce an overall abundance index.

Future Analyses

Further investigation of SBT surfacing behaviour might be advanced by applying generalised estimation equation and ordinary least squares approaches to the analysis of this data to take into account both the correlation structure between the covariates and any autoregressive structure within the response variables. While this approach might lead to small improvements in the explanatory power of the models, uncertainty in the location of the tagged fish will remain an issue, as uncertainty prevents a better spatial match between the environmental covariates and the tagged fish. Behavioural analysis at larger spatial scales, such as comparisons between the GAB and the Indian Ocean may be more instructive with regard to developing an understanding of how SBT surface in different environments.

The influence of surfacing on the fishery-independent SBT abundance index remains an important topic for research; these results indicate the potential importance of including surfacing and the relationship between surfacing and the environment in future analyses.

Acknowledgements

Madeline Cahill and David Griffin (CSIRO, Marine Research) provided altimetry data. Satellite CHL data were provided by Chris Rathbone (CSIRO, Marine Research). Mark Palmer, Paige Eveson and Tom Polacheck provided discussion and analysis suggestions. This project was supported by FRDC grant (1999/105) to Ann Cowling and John Gunn, and by JAMARC, CSIRO Marine Research, and the Australian Fisheries Management Agency as part of the Japan/Australia SBT Recruitment Monitoring Program.

References

Chen, S., Cowling, A., Polacheck, T. 1995. Data analysis of the aerial surveys (1991-1995) for juvenile southern bluefin tuna in the Great Australian Bight. 1995 Southern Bluefin Tuna Recruitment Monitoring Workshop Report. RMWS/95/6, 57 pp. CSIRO Marine Laboratories Hobart

Cowling, A., Millar, C. (1998). Data analysis of the aerial surveys (1991-1998) for juvenile southern bluefin tuna in the Great Australian Bight. Southern bluefin tuna recruitment monitoring and tagging program. Report of the tenth workshop. 1998; RMWS/98/3.

Hastie, T. J., Tibshirani, R. J. (1990). Generalised Additive Models. London: Chapman and Hall.

O'Reilly, J., and Gunn, J. 2001. Estimating the position of juvenile southern bluefin tuna in the Great Australia Bight using light, water temperature and diving data collected by archival tags. 2001 Southern Bluefin Tuna Recruitment Monitoring Workshop Report. RMWS/01/X. CSIRO Marine Laboratories Hobart

McCullagh, P. and Nelder, J. A. 1989. Generalised Linear Models. Second edition. Monographs on Statistics and applied probability 37. London: Chapman and Hall.

Polacheck, T., Gunn, J., Davis, T., Kalish, J., Hausknecht, C. 2000. An Overview of the Fishery for Southern Bluefin Tuna. Background Document for the Indonesia Australia Workshop on Shark and Tuna: future collaboration including research and fisheries management cooperation Bali, Indonesia Workshop, 2-3 March 2000. CSIRO.

Table 1. Summary of the behaviour descriptions used for classifying the vertical position of juvenile SBT from archival tag data.

Behaviour	Description
Sunbaking	Fish stays near the surface of the water (< 5 m).
Midwater Surfacing	Fish explores the top layer of the water column
Midwater	Fish is in top layer but rarely goes to the surface.
Periodic	Fish from depth >20 m to surface, in a periodic fashion
Rarely surfacing	Isolated events where fish at the surface and then returns to deeper water
Not surfacing	Fish not exhibiting any surface-oriented behaviour

Table 2. Number of days of location information in each area for each tag used in the final analyses. In parentheses is the total number of days in the area. Tags with fewer than six observations in an area were excluded from the final model for that area. Observations with influential values (extreme outliers) among the covariates used in the final model were also excluded (see text). Tags 97734 and 97736 were excluded from all analyses since they were at liberty for 10 or fewer days, and only in a single area.

Tag	Area 1	Area 2	Area 3	Area 4	Total
97620	0 (0)	0 (0)	28 (28)	12 (12)	40 (40)
97622	0 (0)	0 (5)	31 (31)	26 (26)	57 (62)
97632	0 (0)	9 (9)	41 (41)	20 (20)	70 (70)
97675	0 (0)	15 (16)	50 (54)	0 (0)	65 (70)
97707	0 (1)	23 (23)	63 (63)	0 (0)	86 (87)
97708	0 (0)	20 (20)	64 (66)	0 (1)	84 (87)
97711	0 (0)	22 (22)	33 (33)	8 (8)	63 (63)
97721	0 (0)	20 (21)	41 (41)	13 (13)	74 (75)
97731	0 (0)	0 (1)	25 (25)	10 (10)	35 (36)
97733	0 (0)	16 (16)	38 (38)	10 (10)	64 (64)
97734	0 (0)	0 (0)	0 (10)	0 (0)	0 (10)
97736	0 (0)	0 (0)	0 (7)	0 (0)	0 (7)
97741	0 (0)	16 (18)	25 (27)	5 (5)	46 (50)
97754	0 (0)	20 (20)	29 (29)	0 (1)	49 (50)
97755	0 (0)	0 (0)	17 (17)	0 (0)	17 (17)
97756	0 (0)	15 (17)	47 (48)	5 (5)	67 (70)
97757	0 (1)	12 (16)	43 (47)	0 (0)	55 (64)
97760	0 (0)	0 (0)	20 (20)	10 (10)	30 (30)
98007	0 (0)	0 (1)	22 (22)	10 (10)	32 (33)
98017	0 (0)	0 (0)	20 (20)	17 (17)	37 (37)
98024	0 (2)	0 (0)	26 (26)	6 (6)	32 (34)
Total	0 (4)	188 (205)	663 (693)	152 (154)	1003 (1056)

Table 3. Area 2 sequential analysis of deviance for significant covariates in the final model. Terms in the model were added sequentially (first to last).

Term	Df	Deviance	Resid. Df	Resid.Dev	F Value	Pr(F)
NULL			187	10754.53		
tag	10	1491.460	177	9263.07	3.53480	0.0002854
bs.per(dar.winddir, period = 360, degree = 2)	2	447.400	175	8815.67	5.30175	0.0058300
dar.cloud	1	107.130	174	8708.54	2.53900	0.1129009
dar.air	1	120.212	173	8588.33	2.84905	0.0932405
dar.bp	1	727.461	172	7860.87	17.24102	0.0000517

Table 4. Area 3 sequential analysis of deviance for significant covariates and interaction terms in the final model. Terms in the model were added sequentially (first to last).

Term	df	Deviance	Resid. Df	Resid. Dev	F Value	Pr(F)
NULL			662	40529.71		
tag	18	4712.264	644	35817.44	6.06561	0.00000000
bs.per(moon, 2, period = 29)	2	1118.134	642	34699.31	12.95333	0.00000311
ns(logchl, 2)	2	1176.182	640	33523.13	13.6258	0.00000163
ns(Day, 2)	2	1423.409	638	32099.72	16.48987	0.00000011
Tag*Day	18	2455.88	620	29643.84	3.1612	0.00001241
Tag*logchl	18	1299.086	602	28344.75	1.67218	0.03986043

Table 5. Area 4 sequential analysis of deviance for significant covariates and interaction terms in the final model. Terms in the model were added sequentially (first to last). Note that two of the covariates (tagsst , and logchl) were not significant, however, the interaction between tag and these covariates was significant, therefore they were retained in the model. An F-test on the change of deviance between this model and a simpler model which replaced ns(logchl,2) with logchl, indicated that the extra curvature in modeling logchl was significant ($p<0.05$).

Term	Df	Deviance	Resid.Df	Resid.Dev	F Value	Pr(F)
NULL			151	7027.977		
tag	12	1599.287	139	5428.69	4.519323	0.0000068
tagsst	1	2.614	138	5426.076	0.088658	0.766447
ns(logchl, 2)	2	2.02	136	5424.055	0.034255	0.966336
Tag*tagsst	12	974.184	124	4449.871	2.752886	0.002624
Tag*logchl	12	959.049	112	3490.822	2.710116	0.003028
Tagsst*logchl	1	61.631	111	3429.191	2.089909	0.049092

Table 6. Significant covariates and interaction terms in the final models for each area in the GAB.

Model Terms	Area 2	Area 3	Area 4
Tag covariate	tag	tag	tag
Environmental covariates	bs.per(dar.winddir, period=360, degre=2) dar.cloud dar.air dar.bp	bs.per(moon, 2, period = 29) ns(logchl, 2) ns(Day, 2)	tagsst ns(logchl,2)
Interaction Terms	tag*wind direction wind_direction*wind speed.	tag*day tag*logchl	tag*tagsst tag*logchl tagsst*logchl

Table 7. Median values of the significant environmental variables in each area. These values were used to predict surfacing rates in each area.

Area	Median values of model covariates in that area			
2	dar.winddir	dar.cloud	dar.air	dar.bp
	218.727377	0.13	13.075	1018.825
3	Day	moon	logchl	
	36	14	0.119000073	
4	Day	tagsst	logchl	
	218.675	133.928	-3.60252765	

Table 8. Predicted surfacing rates for the 19 tags in each GAB area using the median values of the (significant) environment variables for that area. The average surfacing rate for each tag and each area is also shown.

Tag	Area 2	Area 3	Area 4	Mean
97620		0.390	0.408	0.399
97622		0.526	0.315	0.420
97632	0.587	0.441	0.293	0.440
97675	0.572	0.640		0.606
97707	0.541	0.562		0.552
97708	0.692	0.629		0.660
97711	0.409	0.390	0.136	0.312
97721	0.553	0.518	0.769	0.613
97731		0.395	0.452	0.424
97733	0.647	0.400	0.409	0.485
97741	0.656	0.289	0.859	0.601
97754	0.602	0.350		0.476
97755		0.325		0.325
97756	0.503	0.444	0.248	0.398
97757	0.324	0.486		0.405
97760		0.138	0.134	0.136
98007		0.529	0.325	0.427
98017		0.249	0.279	0.264
98024		0.482	0.041	0.262
Mean	0.553	0.431	0.359	
se. of mean	0.109	0.128	0.234	

Table 9. Comparison of values for 10,000 bootstrapped R^2 statistics under the null hypothesis of no relationship between response and covariates, and observed R^2 statistics for the surfacing models in each area.

Area	Observed Model R^2	Bootstrapped R^2				Observed - bootstrap mean	Significance
		Min	Max	Median	Mean		
Area 2	26.90%	1.4%	19.3%	6.9%	7.2%	19.70%	$p < 0.0001$
Area 3	30.06%	3.9%	14.7%	8.1%	8.2%	21.88%	$p < 0.0001$
Area 4	51.21%	10.9%	45.1%	25.4%	25.7%	25.53%	$p < 0.0001$

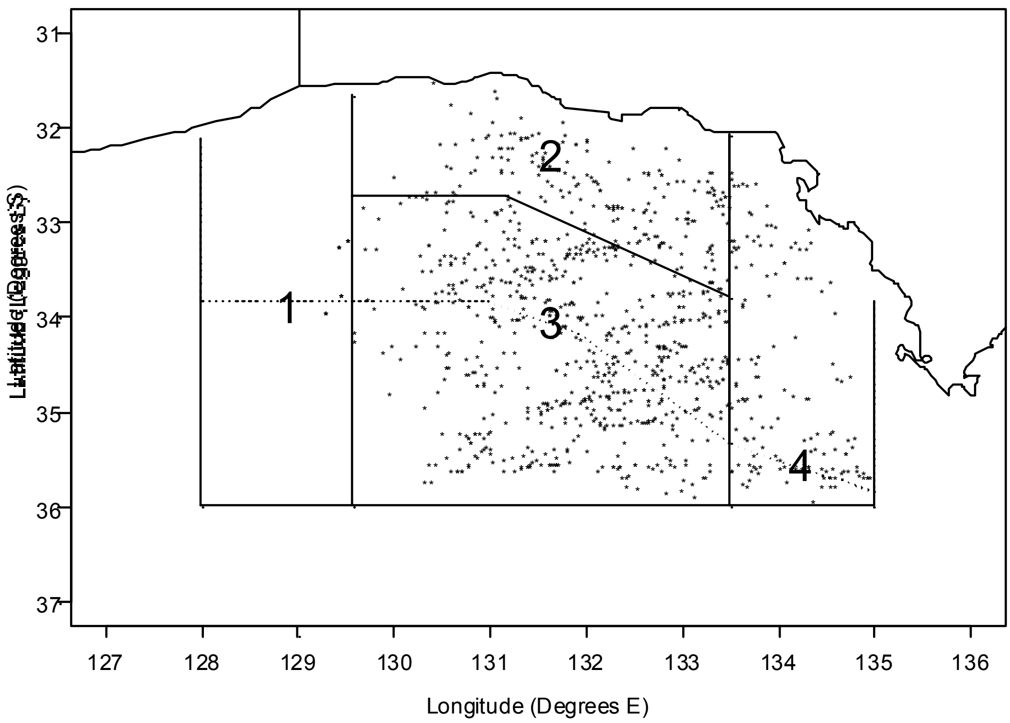


Figure 1. Spatial areas for analysis. Dotted lines represent the southern boundary of the aerial survey region. The location of each observation from all tags is indicated with dots. Area 1 was not considered as it had only 4 observations.

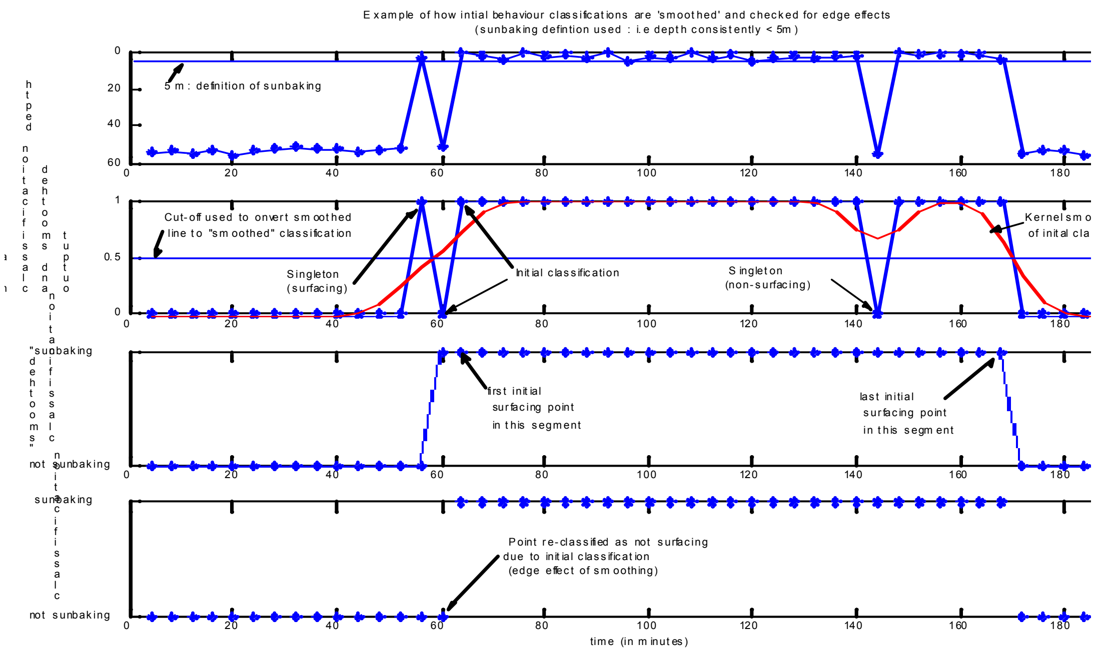


Figure 2. Example of the smoothing process for the vertical behaviour sunbaking classification. A. Initial depth time series, with a line at 5 m indicating cutoff for sunbaking behaviour. B. Initial classification of behaviour (behaviour present = 1, behaviour absent=0). A kernel smooth is fitted to the 0/1 series, curved line. C. The smoothed signal produced in panel B is converted back to a presence/absence time series based upon a cutoff value of 0.5. D. The edges of the portions of the time series classified as "surfacing" are compared to the initial classification. The first and last points in this section which were initially classified as surfacing become the edge points of the new classification. Any points in the section that lie outside that region are reclassified as 'not exhibiting the behaviour'.

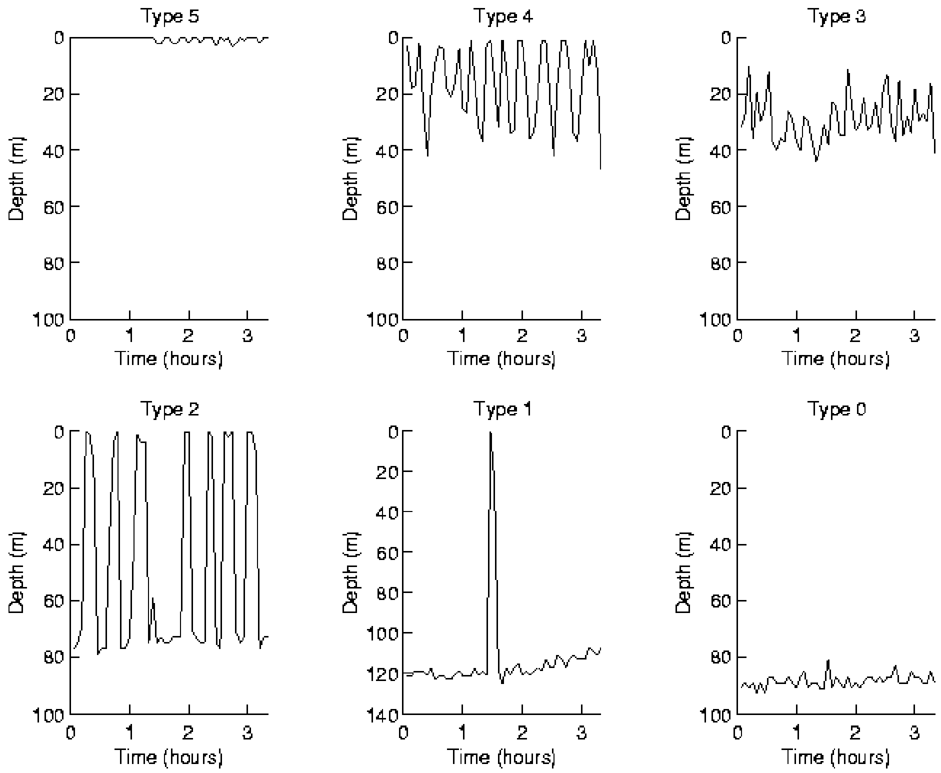


Figure 3. Example of each of six southern bluefin tuna vertical behaviours derived from archival tag data described in the text. Type 5: Sunbaking, Type 4: Midwater Surfacing, Type 3: Midwater, Type 2: Periodic, Type 1: Rare surfacing, Type 0: Non-surfacing.

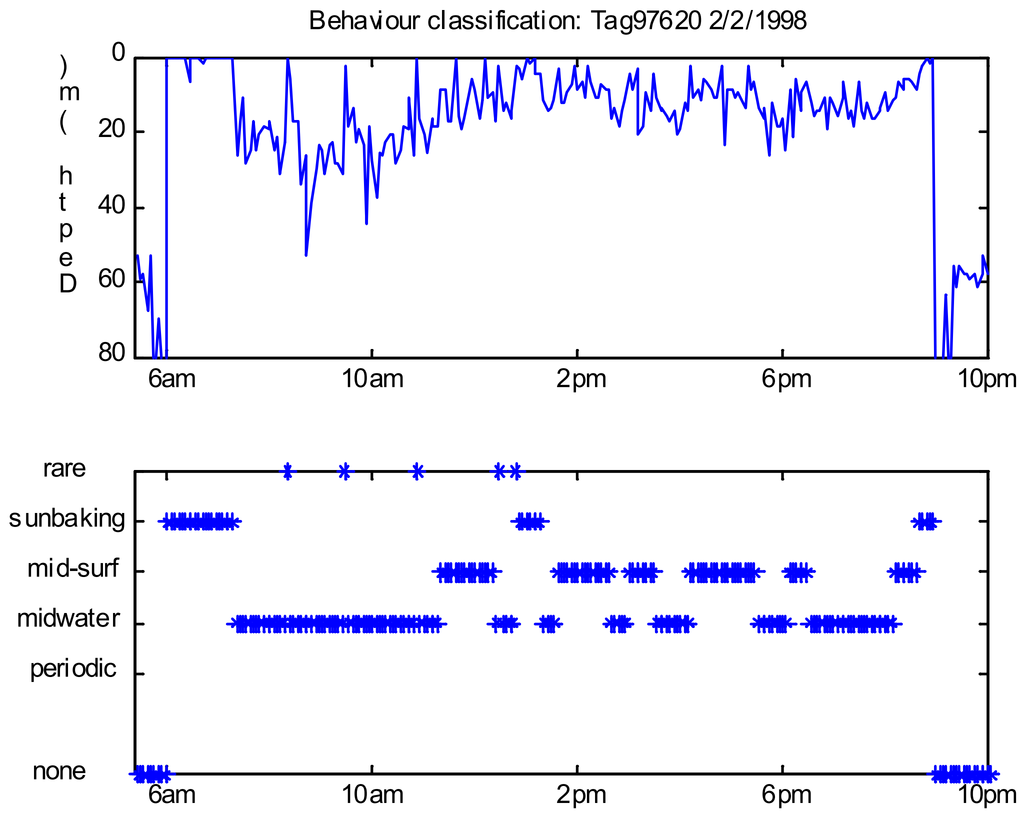


Figure 4. An example of final vertical behaviour classification for a juvenile southern bluefin tuna for a single day (360 4 minute observations) for Tag 97620 on February, 2 1998.

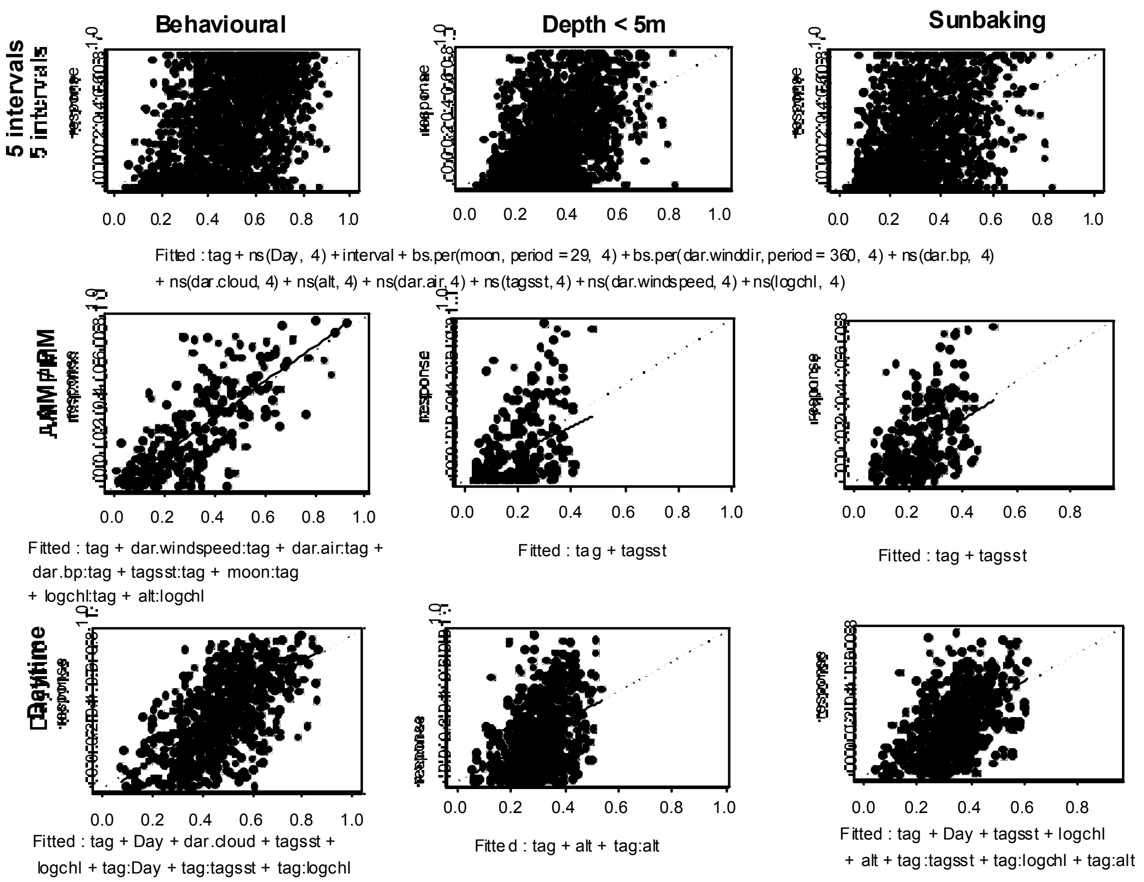


Figure 5. Plots of response vs fitted values for preliminary Area 3 models. Top Row. GLM fits of an over-specified model of surfacing defined on five intervals per day. Second row GLM fits for models of surfacing defined on two intervals per day (am/pm). Third row. GLM fits for one interval per day. Column 1: GLM of vertical behaviour definitions of surfacing, Column 2: Depth-based definitions of surfacing. Column 3. Sunbaking-only definition of surfacing.

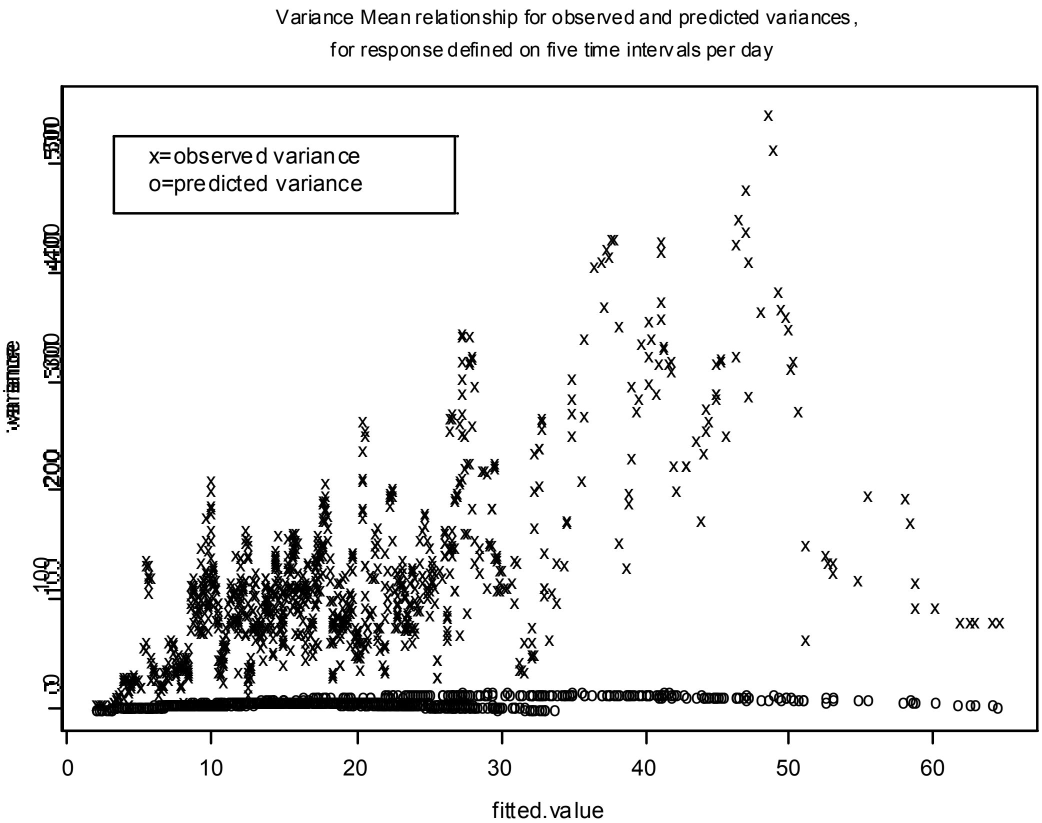


Figure 6. Observed and predicted variances of residuals for the 5-interval time scale. Predicted variances were based on a binomial assumption and observed variances were estimated by a kernel smooth. Note that the predicted and observed variances are for a binomial response: ie the <u>number</u> of surfacing events in an interval, and not the <u>proportion</u> of surfacing events in the interval.

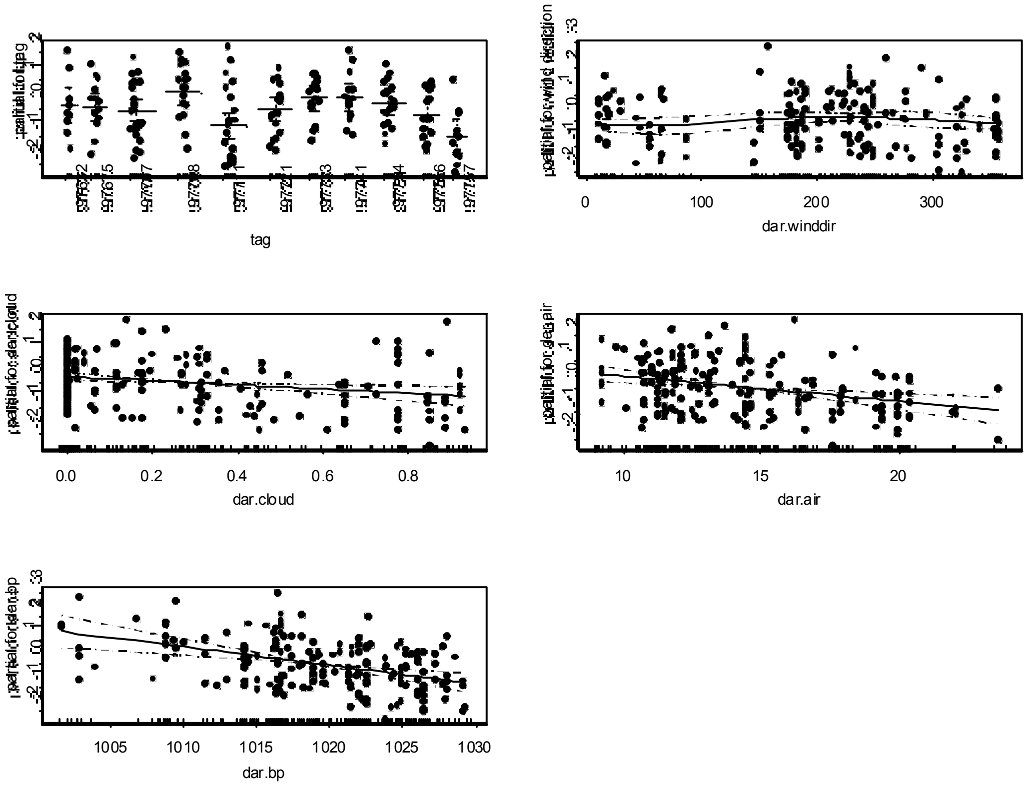


Figure 7. Final model for Area 2, significant covariates with residuals.

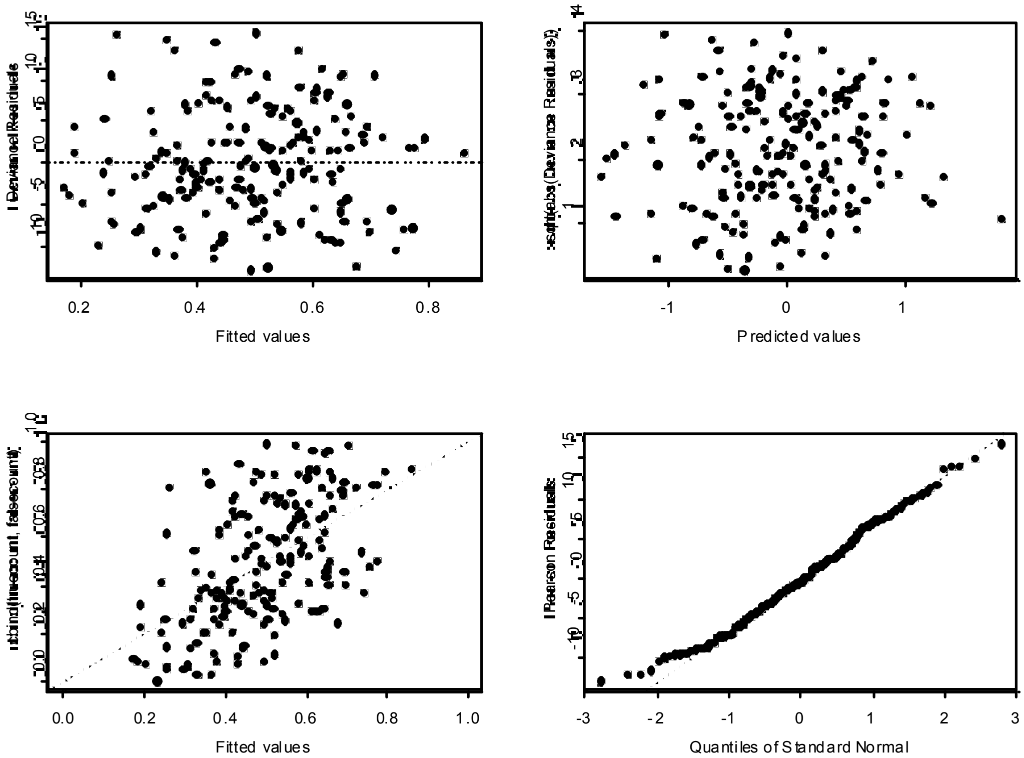


Figure 8. Area 2 final model diagnostic plots. A. Relationship between fitted model values and residual deviance B. Relationship between the fitted model and the square root of the residual deviance C. Relationship between observed and fitted values, with 1:1 line. D. Relationship between the standardised normal scores and the Pearson residuals. The diagonal line indicates the position along which normally distributed standardised residuals would lie.

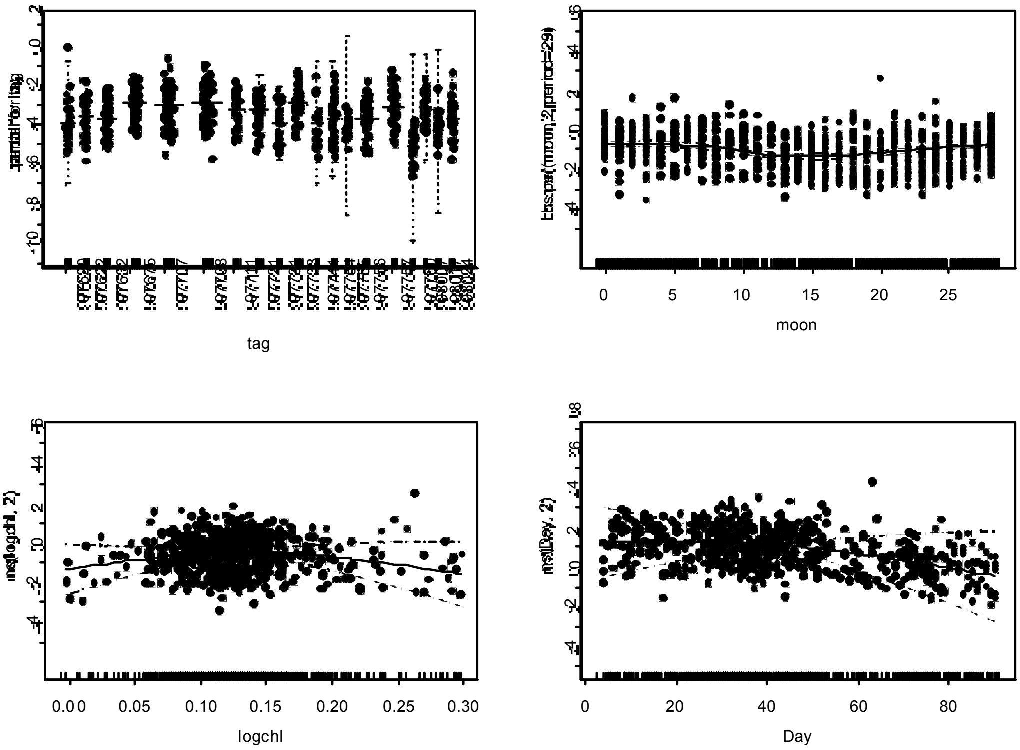


Figure 9. Final model for Area 3, covariates with residuals.

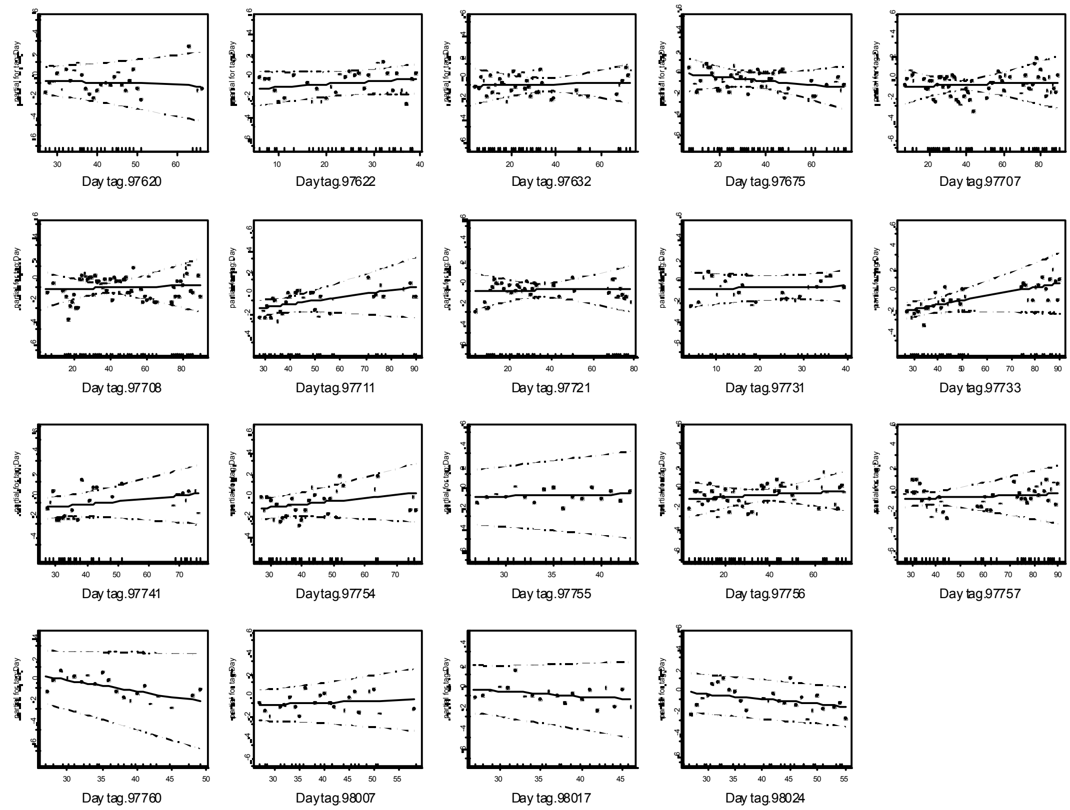


Figure 10. Area 3, final model interaction term one: Tag*day

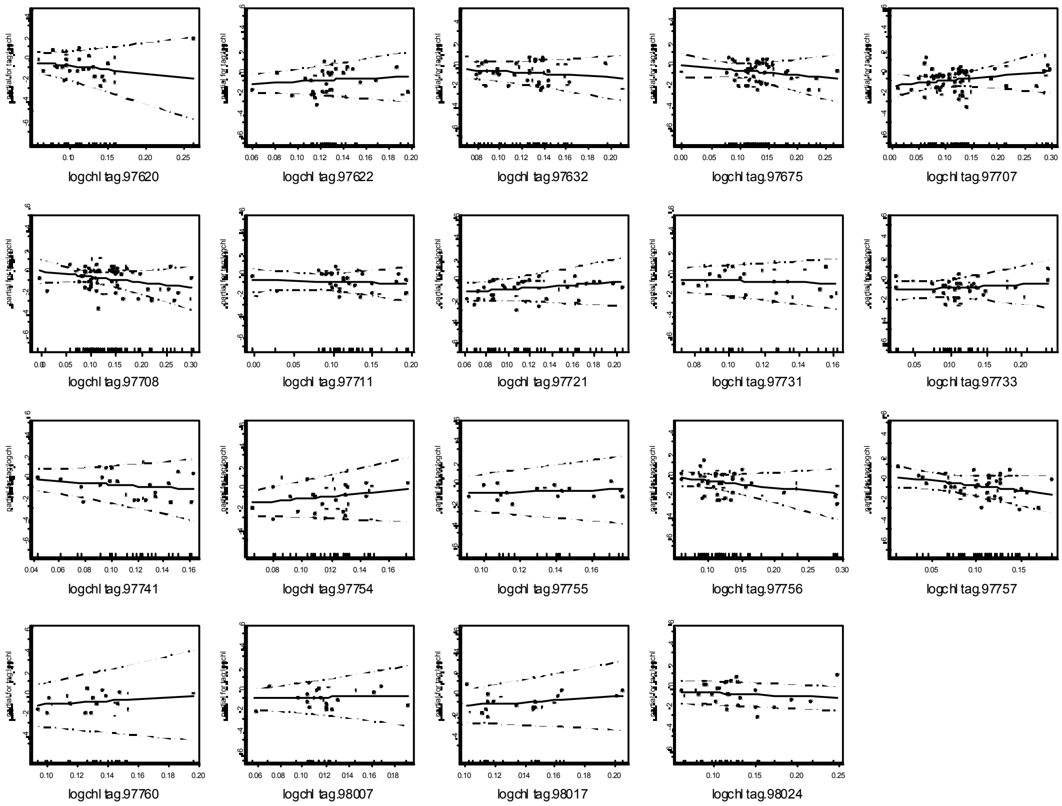


Figure 11. Area 3, final model interaction term two: Tag*logchl

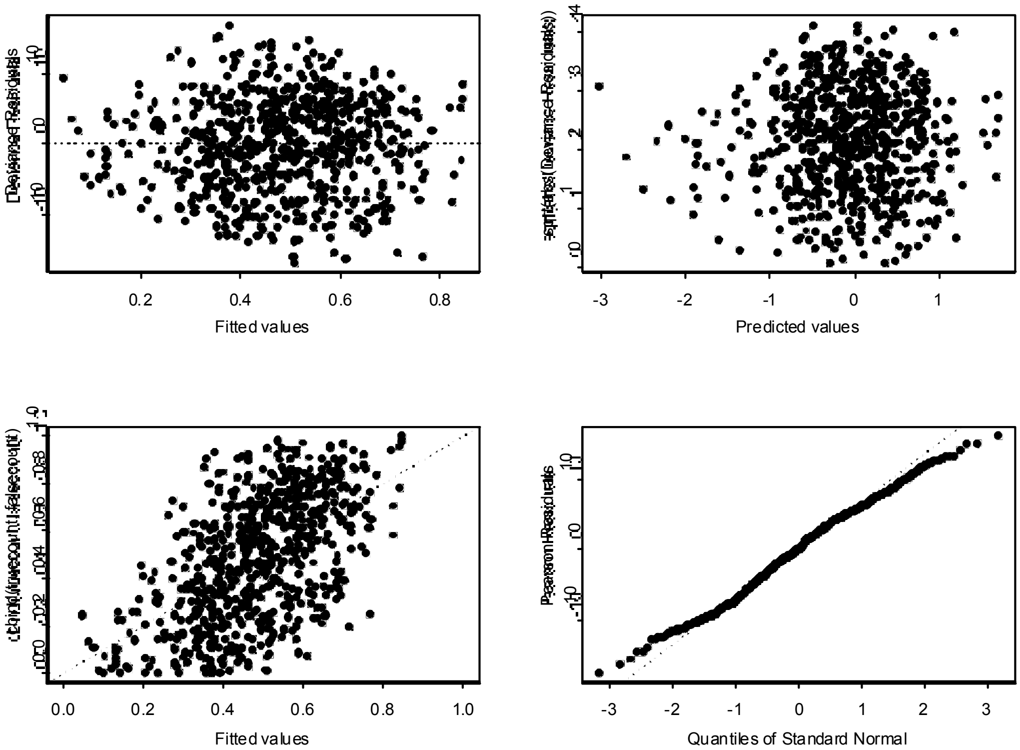


Figure 12. Diagnostic plots for Area 3. A. Relationship between fitted model values and residual deviance B. Relationship between the fitted model and the square root of the residual deviance C. Relationship between observed and fitted values, with 1:1 line. D. Relationship between the standardised normal scores and the Pearson residuals. The diagonal line indicates the position along which normally distributed standardised residuals would lie.

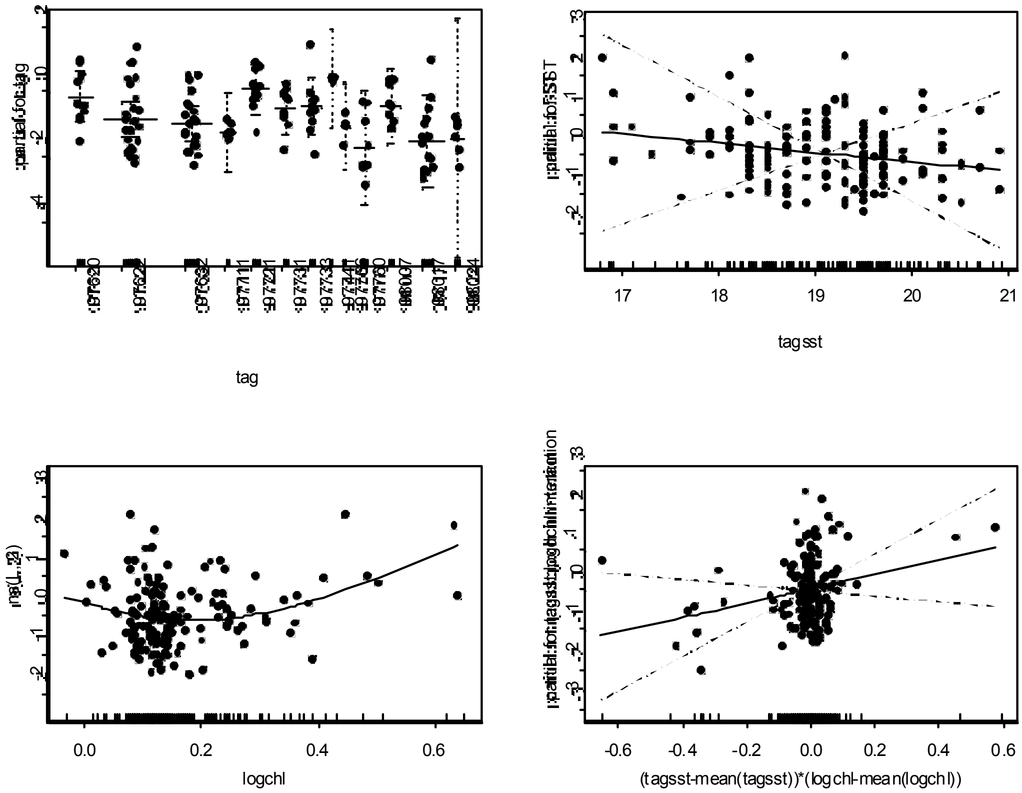


Figure 13. Area 4 final model, covariates with residuals. The standard errors of the logchl effect are not shown because they reduced the clarity of the plot.

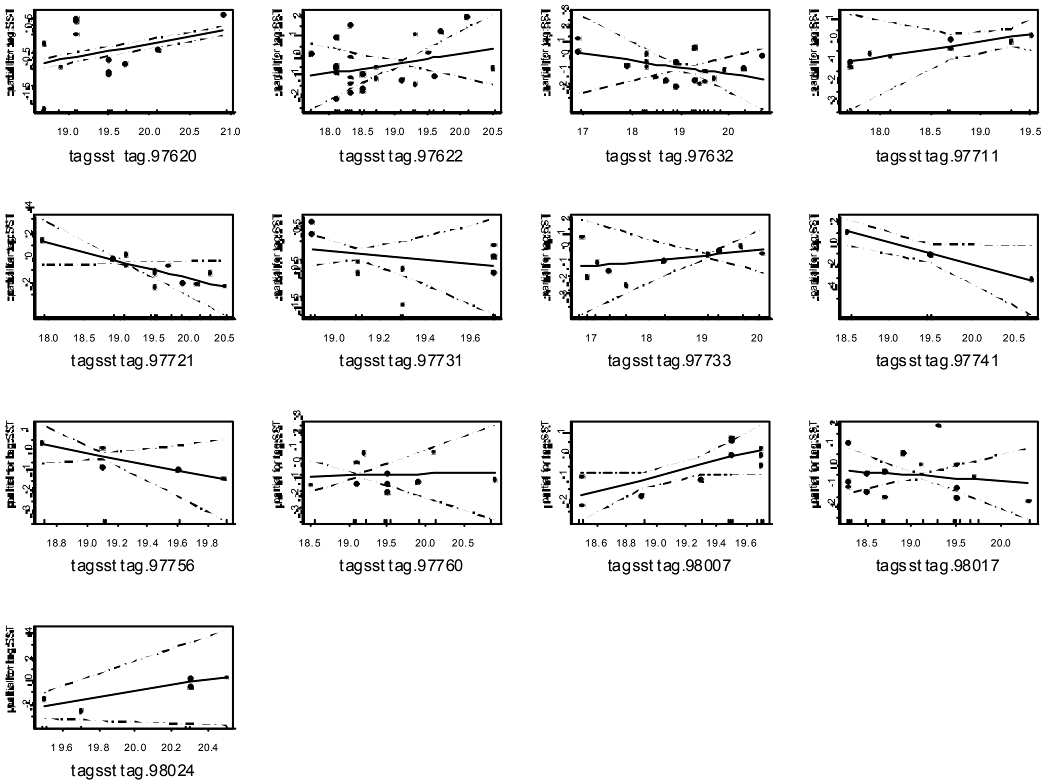


Figure 14. Area 4 final model interaction term one: tag*tagsst

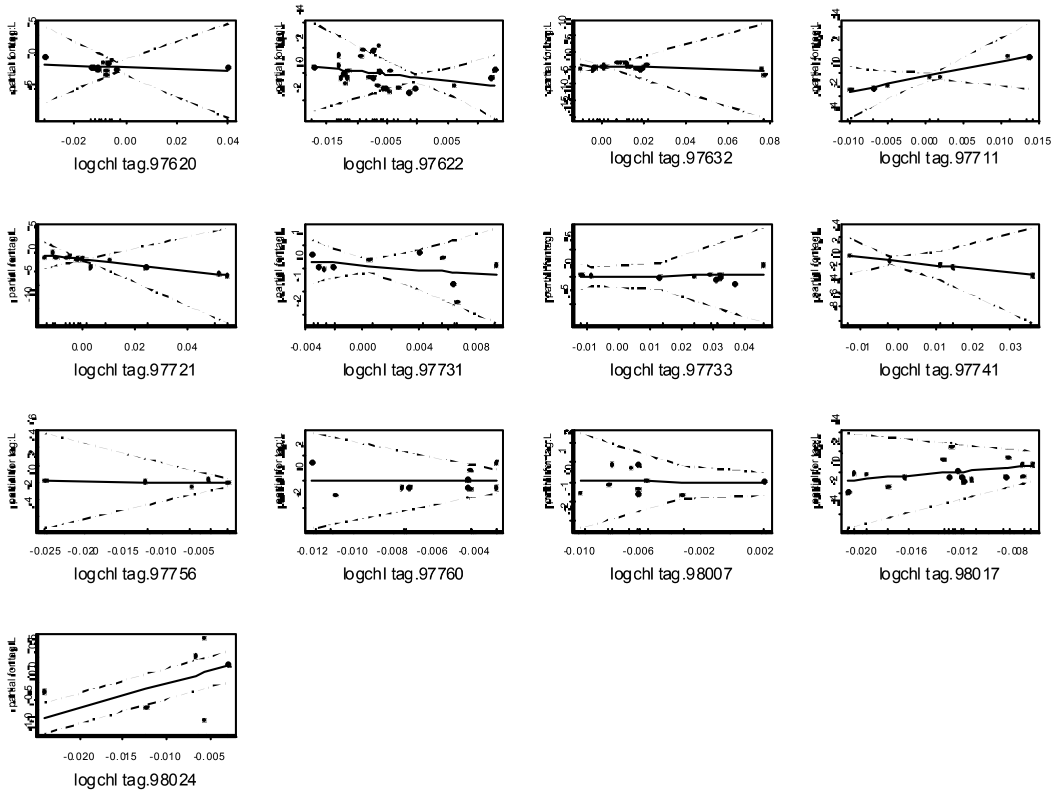


Figure 15. Area 4 final model interaction term two: tag*logchl.

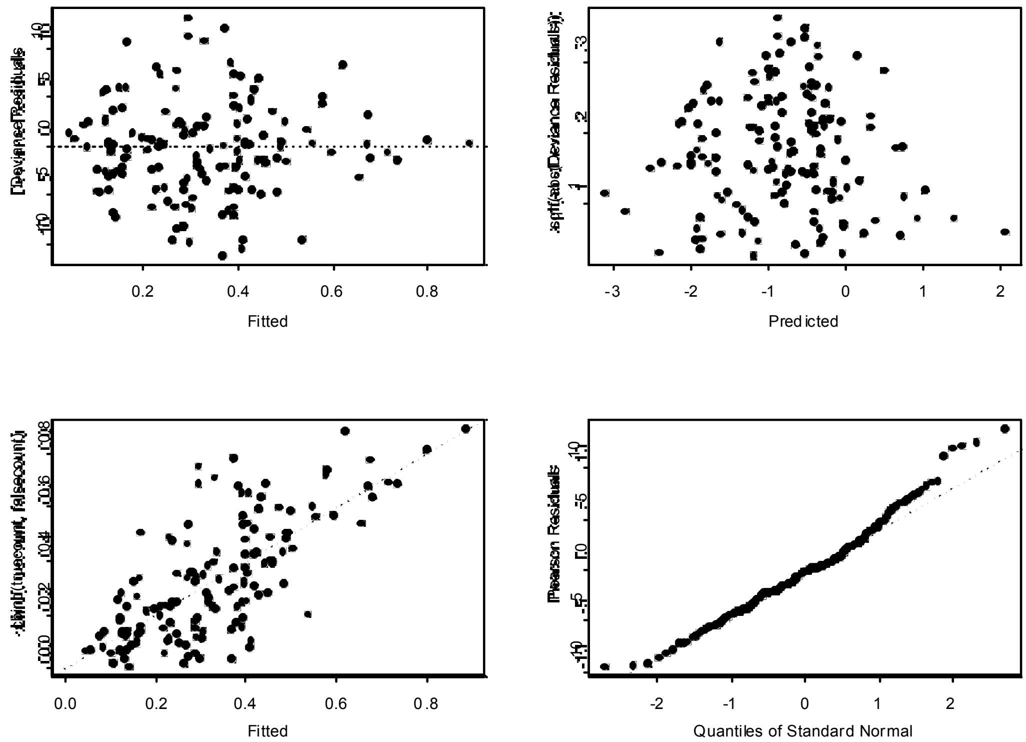


Figure 16. Area 4 final model diagnostic plots. A. Relationship between fitted model values and residual deviance B. Relationship between the fitted model and the square root of the residual deviance C. Relationship between observed and fitted values, with 1:1 line. D. Relationship between the standardised normal scores and the Pearson residuals. The diagonal line indicates the position along which normally distributed standardised residuals would lie.

Appendix 6.4A. Overdispersion

Overdispersion can arise when several sub-populations with different means are sampled and the sample is treated as if it came from a single homogenous population. In the context of analyzing the surfacing data, reducing the length of the interval could result in better separation of populations, and reduce the overdispersion. As the interval size decreases, however, the number of samples within each interval decreases, which reduces the precision of the estimated mean for that interval. Thus, Pearson's statistic and the deviance are no longer suitable statistics for measuring model goodness-of-fit. In the extreme case of treating each observation as a response, Pearson's statistic for the data set will reduce to the number of records, and the deviance will be independent of the fitted values. For less extreme cases, Pearson's statistic and the deviance can be used to examine the goodness of fit but require the further calculation of conditional means and variances in order to determine statistical significance (McCullagh and Nelder, 1989 p122).

Analysis of overdispersed data requires estimation of the dispersion parameter. This estimate is approximately chi-squared, and so the estimate of change in deviance between models, divided by the estimated dispersion parameter should be approximately F-distributed (McCullagh and Nelder, 1989). Estimation of the dispersion parameter is performed by fitting quasi-likelihood models to the binomial data.

6.5 Further data analyses of the aerial surveys for juvenile SBT in the Great Australian Bight

Ann Cowling

Abstract

A number of indices of abundance of juvenile southern bluefin tuna (SBT) in the Great Australian Bight result from further development of a model of SBT abundance based on analysis of aerial survey data, through the addition of new oceanographic data and inclusion of SBT surfacing rates. All the indices are robust to the period of time used to develop the underlying model. Comparison of all the abundance indices developed to date leads to the recommendation of a presence/absence index for use as a long-term monitoring index due to its low CV's, relatively low cost and relative robustness to changes in staff over time. The indices developed in this section show evidence of a decline in juvenile SBT abundance in the survey area over the period of the surveys (1993-2000).

Introduction

The aerial survey was proposed and developed as one of the projects in the Recruitment Monitoring Program (RMP) as a potential tool for monitoring the abundance of juvenile southern bluefin tuna (SBT) in the Great Australian Bight (GAB). Under this program, aerial surveys of the GAB were conducted each summer from 1993 to 2000. However, there are some doubts as to whether the information from the aerial survey is adequate for monitoring juvenile abundance. The main issue is whether the coefficients of variation (CVs) associated with the abundance estimates are so wide that the index is unable to detect even a major increase or decrease in abundance.

Two steps have been taken to address this problem: a new project, the integrated analysis project, was initiated under the RMP in 1999, and a workshop was held in Port Lincoln in 2000 to discuss and agree on alternative methods for analysing the data. The integrated analysis project has led to new environmental and SBT behavioural variables being identified and extracted for inclusion in the indices. Predicted surfacing rates in different areas of the GAB have been derived from models of surfacing behaviour developed from archival tag data. The 2000 Port Lincoln aerial survey workshop agreed on the proposed change in analysis methodology to a linear modeling approach. This approach allowed environmental variables to be incorporated into the models thereby allowing the estimates to be standardised for differences in environmental conditions between years.

This section of the report firstly further develops the models and analysis of the aerial survey data presented in **Section 6.1**. The abundance indices developed to date have been based on a strip-transect method, after conventional line transect methods were found to be inappropriate, or a modeling-based approach. The modeling-based approach used both a presence/absence and a conditional biomass (biomass given presence) form. These models are discussed and developed more fully in the materials references in **Section 6.1.**

In this section, the appropriate scale of analysis is determined, selection of basic environmental variables explained, and the importance of different temporal periods explored. Then, additional oceanographic data identified as having an effect on SBT abundance in a separate analysis (**Section 6.2)** are incorporated into the models. Finally, estimates of surfacing behaviour (**Section 6.4)** are included in the indices. For each stage in the development of the models, assessment whether the development of the model has reduced the CVs of the annual estimates is included. Finally, the temporal trends in the final indices are presented.

Methods and Results

Appropriate units for analysis

In analysis of the aerial survey data it is important to consider the appropriate unit for modeling as well as the appropriate unit for the data. The two scales may be different: the appropriate unit for modeling depends on the spatial scale at which we wish to develop and interpret results. The data unit depends on the unit of data collection and the appropriate level of subdivision of that unit to best capture spatial and environmental variation during the surveys. To resolve this problem, models are fitted to data using three data units (whole line, half line, and quarter line), three transect line widths (4 nm, 5 nm and 6 nm) and five modeling units. The five modeling units are;

1. inshore and offshore halves of the 15 transect lines used in the aerial survey
2. inshore and offshore halves of each block (3 transect lines per block)
3. blocks
4. inshore and offshore components of larger spatial units than blocks
5. no spatial units

The results of these analyses (see **Appendix 6.5A**) show that

1. the most appropriate data unit in the survey analysis is the quarter line truncated at 6 nm each side of the transect line
2. the most appropriate modeling units are the inshore/offshore (IO) components of each of the five survey blocks. The five survey blocks are created from adjacent sets of three north-south transects lines, and divide the survey area into five equal east-west segments.

Accordingly, these data and modeling units are used in the remainder of this section.

Comparison of plane SST, plane AT and satellite SST

The environment has the potential to influence the detectibility and/or presence of SBT in the survey area. Early attempts to produce an abundance index for juvenile SBT did try to include environmental variables that were measured when the survey was undertaken. In particular, water and air temperature appeared to be related to the detection of SBT in the GAB. In this sub-section models using plane SST, plane AT and satellite SST are compared to determine which of these environmental variables is most suitable for use in the development of an abundance index.

Previous analysis (Cowling, 2000) showed very different year-to-year trends in the abundance indices depending on whether plane-based air temperature (AT) or plane-based sea surface temperature (SST) had been included in the model from which the index was derived. The view was taken that because AT is highly correlated with SST, only one of the two variables should be included in the models.

Both of the plane-based environmental variables are measured with considerable errors due to the quality of the instruments used in the field. The infrared sensors for measuring SST from the survey plane were changed after both the 1993 and the 1998 field seasons. Both changes were intended to improve the technology, but the SST sensors used from 1994 to 1998 were unreliable, and were hence upgraded again in 1999. Calibration procedures to adjust for differences between planes were used from 1997 onwards. Unfortunately, these SST comparisons showed that the 1994-98 instruments gave such highly variable readings that meaningful adjustments could not be made. Calibration of the 1999 and 2000 SST instruments showed they are very reliable (accurate and precise) and no adjustments are necessary. Because of differences in the quality of the plane-derived SST, satellite SST was extracted to provide a complete and consistent set of SST data for use in the models. Values of satellite SST outside the range 10-30^{o}C were excluded from the dataset used for modeling.

Air temperature was measured using the planes' thermometers, which are intended to indicate when there is a possibility of ice formation that would compromise flying safety. This primary use suggests these instruments are intended to be accurate around 0^{o}C. In some years, when the AT measurements should be the same, such as when the two survey planes are stationary on the ground and facing the same direction, differences of up to 8^{o}C have been recorded. In 2000, thermisters were added to the CSIRO instrumentation on the planes to measure AT, and these data showed close agreement between planes in calibration tests.

To compare the explanatory power of plane SST, plane AT and satellite SST in the models, the "presence/absence" and "log(biomass) given presence" (conditional biomass) models were fitted including satellite SST, plane SST and AT in turn. Other variables included in the models were wind speed and line length (the most highly significant variables from other analyses) and terms for block and year effects. The proportion of deviance explained by plane SST, plane AT or satellite SST in each model was compared.

The environmental variables varied with regard to the amount of missing data (**Table 1**). This table shows 22.8% of the plane SST data is missing, compared to 9.2% of the satellite SST data and none of the plane AT data.

Table 1. The proportion of missing values for plane SST, plane AT and satellite SST in the restricted quarter line data.

Data	Plane SST	Plane AT	Satellite SST
Number of quarter lines	2393	2393	2393
Number of quarter lines missing data	546 (22.8%)	0 (0%)	221 (9.2%)

The missing values of plane SST occurred due to malfunctions in the recording equipment during the flights. Also, in 2000, one plane did not have SST recording equipment due to delays in approval from CASA to make holes in the plane to insert the instruments. The differing numbers of data in each analysis invalidates direct comparison of the fits of the models using analysis of variance/deviance. The analysis of deviance (presence/absence models) and variance (conditional biomass models) tables for the three model versions (plane AT, plane SST, and satellite SST) are given in **Appendix 6.5B**. The two models forms are compared using different criteria. The criteria used for determining which of the three explanatory variables provided the best fit for the presence/absence and conditional biomass models are summarised in **Table 2**.

Table 2. Criteria used to compare between three environmental variables for use in the SBT abundance indices.

Presence/absence models	Plane AT	Plane SST	Satellite SST
Deviance (3df)	30.61	27.81	21.48
Number of missing data	0	546	221
Proportion of deviance explained by model	19.3	18.6	19.2
Conditional biomass models	Plane AT	Plane SST	Satellite SST
Sum of squares (3df)	48.6746	28.2784	34.7553
Number of missing data	0	546	221
Multiple R^2	16.26	19.69	13.72

These analyses suggest that satellite SST may perhaps be nearly as useful at plane AT in predicting presence/absence (proportion of deviance explained by model). On the basis of this table, plane SST would be the best variable to use in a combined index. However, the plane SST data must be able to be obtained reliably and this has not been the case in the past surveys. Accordingly in the rest of this analysis, satellite SST is used. It provides an acceptable compromise between reliable data, comparability between years, good fit and minimum numbers of missing data.

The presence/absence, conditional biomass and the combined indices calculated from models including plane AT, plane SST and satellite SST for the offshore section of Blocks 2, 3 and 4 are shown in **Figure 1.**

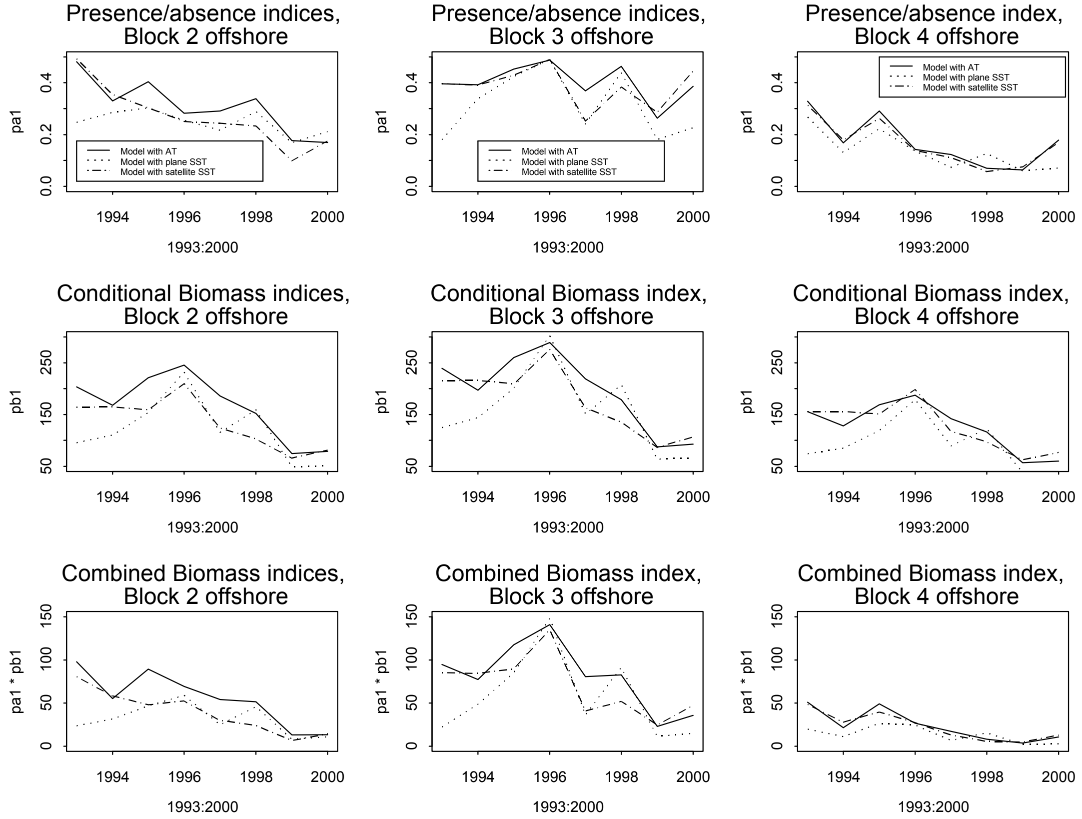


Figure 1. Comparison of the presence/absence, conditional biomass, and the combined indices calculated from models including plane AT, plane SST and satellite SST for the offshore section of Blocks 2, 3 and 4.

The presence/absence indices including plane AT and satellite SST are very similar in the offshore sections of Blocks 2, 3 and 4, however, the presence/absence index including plane SST is different. The conditional biomass indices are all different, but the most irregular fluctuations, suggestive of noise, occur in the plane SST based index. The overall trend in the combined index is most similar for plane AT and satellite SST. These plots suggest that there may be most noise in the plane SST data. As the satellite SST data is known to be consistent from year to year, it provides the preferred index. The plots confirm the decision based on **Table 10** that satellite SST is preferred for use in the indices.

Effect of different environment standardizations

Preliminary analyses have showed slight differences in the predicted abundance trends with inclusion of different environmental conditions and standardization of the conditions. In this sub-section, the indices given by the predicted values of the models in two sets of conditions, “average conditions” and “sighting conditions” are compared (**Table 3**) and the “regional average conditions” and “regional sighting conditions” are presented. Average conditions are the median of each environmental condition over all included quarter lines. Sighting conditions are the median over all included quarter lines *on which sightings were made* of each environmental condition. The conditions in which sightings are made are warmer (higher AT and SST), calmer (lower wind speed) and less cloudy (less high cloud) (**Table 3)**. There is no discernable difference in swell.

A line length of 33 nm is used for prediction as this is the median length of the quarter lines.

Table 3. A selection of average and sighting conditions for the aerial survey. Average conditions are the median of each environmental condition over all included quarter lines. Sighting conditions are the median over all included quarter lines *on which sightings were made* of each environmental condition.

	AT	Plane SST	Satellite SST	Wind speed	Swell	High cloud	Wind direction
Average conditions	23.6	19.2	20.3	3.9	1.0	1.5	ESE
Sighting conditions	26.3	19.8	20.6	3.2	1.0	0.9	ESE

The "regional" sets of conditions allow for spatial variability in environmental conditions across the survey area by taking the medians of the conditions in the inshore/offshore sections of each aerial survey block. Regional standard conditions" for the key environmental conditions are shown in **Table 4**, and **"**regional sighting conditions" for the key environmental conditions are shown in **Table 5.**

Table 4. Regional standard conditions: the medians of the environmental conditions in the inshore (I) and offshore (O) sections of the aerial survey blocks

Block	1		2		3		4		5	
	I	O	I	O	I	O	I	O	I	O
Wind speed	4.6	4.6	4.0	4.0	4.0	4.0	3.6	3.6	3.3	3.4
Swell	1.5	1.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.3
Satellite SST	21.1	20.6	20.8	20.8	20.4	20.3	20.2	20.1	19.2	19.3
Wind direction	ESE	ESE	ESE	ESE	ESE	ESE	ESE	ESE	ESE	N

Table 5. Regional sighting conditions: the medians of the environmental conditions in the inshore (I) and offshore (O) sections of the aerial survey blocks.

Block	1		2		3		4		5	
	I	O	I	O	I	O	I	O	I	O
Wind speed	3.0	3.2	3.2	3.3	3.4	3.4	3.3	3.1	2.9	2.1
Swell	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Satellite SST	21.9	20.8	21.4	21.2	20.6	20.4	20.4	20.4	20.0	20.8
Wind direction	ENE	ENE	N	ESE	ESE	ESE	ESE	ESE	ESE	N

There are larger differences in the regional standard conditions (**Table 4)** across the survey area than in the regional sighting conditions (**Table 5**) which are much more stable. The similarity in the values of wind speed and satellite SST across the GAB suggests that indeed there may be specific conditions in which detection of surfacing SBT is more likely. Differences between blocks are generally larger than differences between inshore and offshore halves of the same block. Indices based on the average conditions in each block could be derived and would allow for the slight differences in prevailing conditions between each block, however, this analysis has not been undertaken at this time.

The variables having a significant association with SBT in the presence/absence model are wind speed, line length, SST, swell, pilot, wind direction, block*IO and block*year (**Table 6)**. It should be noted that the block*year interaction is only marginally significant. The variables having a significant association with SBT based on the conditional biomass model are wind speed, SST, pilot and block (**Table 7)**. It should be noted that the year-to-year differences in conditional biomass are not statistically significantly different.

Comparing the average conditions with sighting conditions within the offshore sections of Block 3 (B3O) or Block 4 (B4O) (**Figure 2)**, shows the indices are a little higher in sighting conditions than in average conditions. A comparison of the final combined indices using "average sighting conditions" and "average conditions" would show that the indices using average sighting conditions would be slightly higher. This prediction has not been confirmed analytically. In the remainder of this section, predictions are made using average sighting conditions.

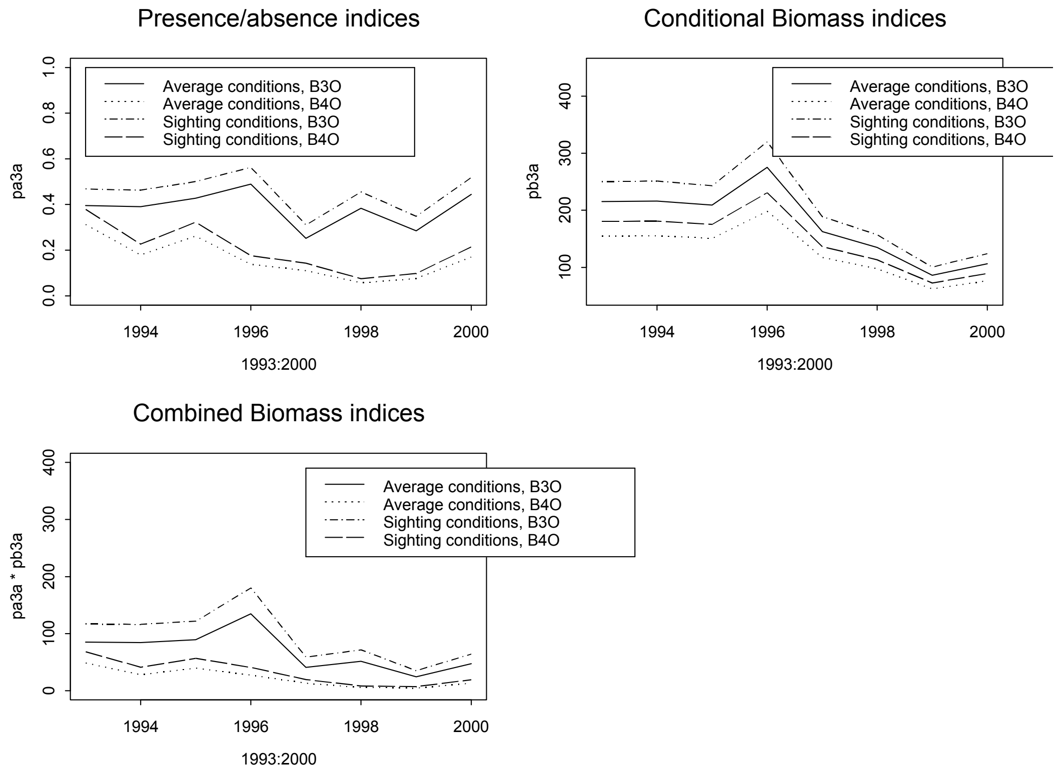


Figure 2. The presence/absence indices, conditional biomass indices and combined indices in average and sighting conditions for the offshore section of Blocks 3 and 4.

Table 6. The analysis of deviance and variance results for the (reduced length) presence/absence index.

	df	Deviance	Residual df	Residual Deviance	Pr(Chi square)
NULL			2165	2183.728	
poly(avws, 2)	2	145.606	2163	2038.122	0.00000000
poly(len, 3)	3	70.1361	2160	1967.985	0.00000000
poly(SST, 3)	3	21.3429	2157	1946.643	0.00008930
avsw	1	18.5684	2156	1928.074	0.00001640
pilot	3	18.6492	2153	1909.425	0.00032310
avwdir	5	12.7481	2148	1896.677	0.02585770
block	4	67.7534	2144	1828.923	0.00000000
IO	1	1.6876	2143	1827.236	0.19391380
year	7	11.1121	2136	1816.124	0.13380360
block:IO	4	33.8208	2132	1782.303	0.00000080
block:year	28	41.2412	2104	1741.062	0.05103100

Table 7. The analysis of variance results for the conditional biomass model.

	df	Sum of Sq	Mean Sq	F Value	Pr(F)
block	4	30.5477	7.63693	3.53062	0.0075458
poly(SST, 2)	2	50.7358	25.36788	11.72780	0.0000111
pilot	3	47.1698	15.72326	7.26900	0.0000919
poly(avws, 3)	3	18.8578	6.28593	2.90604	0.0344973
year	7	20.3003	2.90005	1.34072	0.2293821
Residuals	419	906.3203	2.16306		

Demonstration that the analysis is robust to years

It has been suggested that the models describing juvenile SBT abundance in the GAB may be driven by particular years: for example, abundance appears to have been particularly high in 1993, and in the last two years of the survey, trainee spotters were used, perhaps resulting in lower number of observations. There is a possibility that these years may be causing the apparent trends in the indices. This is extremely unlikely, however, as analyses do not reveal high influence points related to particular years. In addition, the strip-transect analyses show a similar general trend to that observed in the overall index. If the modeled analysis were being driven by particular years, large differences between the strip-transect analysis and the modeled analysis in those years would be expected. This is because in the strip-transect analysis, each year is analysed independently and so results in one year do not affect those in another.

Nevertheless, to demonstrate that the model is robust to the particular years of data included in the analysis, additional analyses are performed including only the years

1. 1993-1996
2. 1993-1998
3. 1994-1996
4. 1994-1998

All these analyses remove 1999 and 2000 as trainee spotters were used in these years. Predicted trends and standard errors were obtained for each using standard methodology

The predictions from the three models fitted to the data for various sets of years are shown in **Figure 3.** The presence/absence indices for 1993-1996 and 1994-1996 are somewhat higher than those for the other groups of years. However, the differences lie within the 95% confidence intervals for the 1993:2000 presence/absence indices (**Figure 4)** and so are not statistically significant. The set of years used is thus considered not to strongly influence the model used to calculate the abundance index. The differences are likely due to random variability.

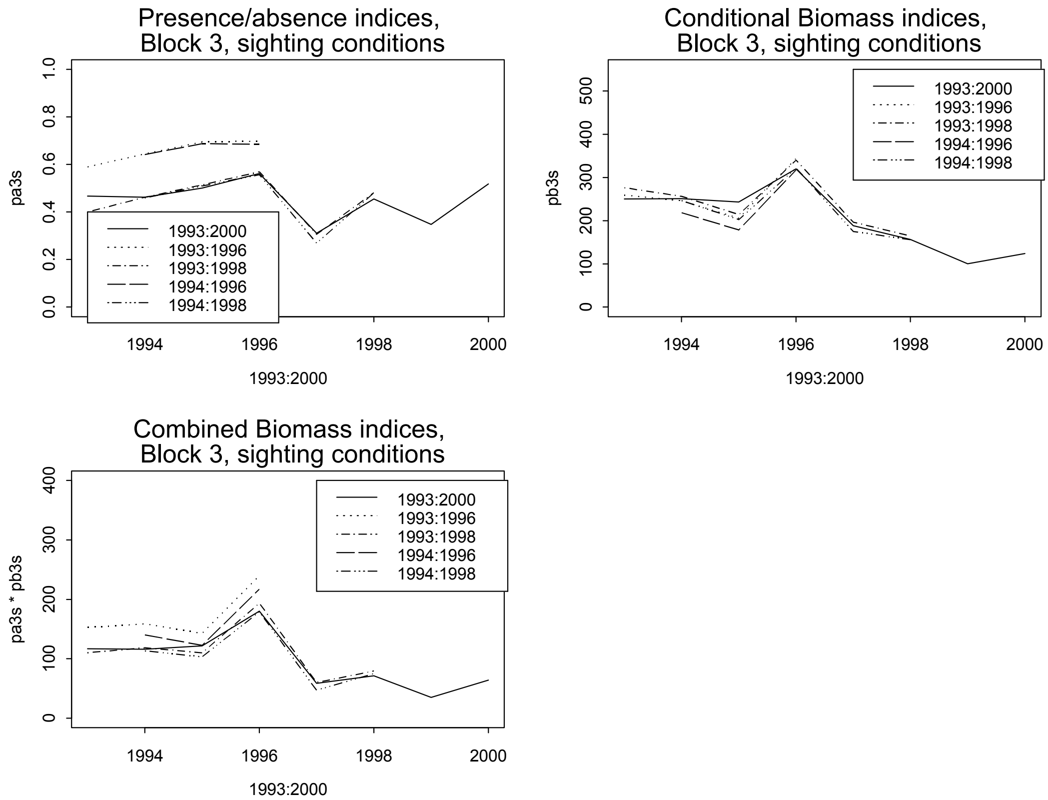


Figure 3. Predictions of the abundance indices from the three models (presence/absence, conditional biomass and combined) fitted to the aerial survey data for various sets of years.

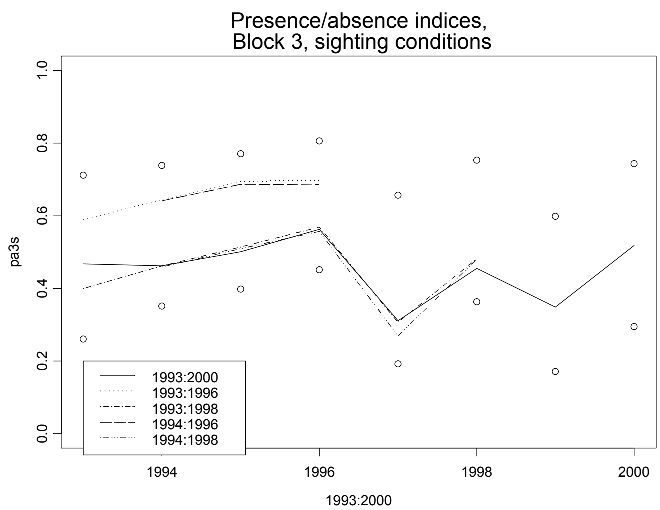


Figure 4. Presence/absence model-based indices for Block 3 for different sets of years. The 95% confidence interval limits around the point estimates are shown by small circles.

Effect of model form on CVs

Basic strip-transect estimates: Whole line data vs quarter line data

Strip-transect estimates (both weighted by length and unweighted) were obtained using the whole line data and the quarter line data. These two strip-transect estimates of biomass density based on the (length restricted) whole line data truncated at 6 nm are given in **Table 8** and shown in **Figure 5**. The table shows that the analysis weighted according to length has slightly lower coefficients of variation (CVs) and that CVs are minimum in 1994-1996 when the survey was replicated 7 or 8 times. In other years when fewer replicates, and hence fewer whole lines, were surveyed, the CVs are higher. Clearly there is very little difference between the weighted and unweighted estimates (**Figure 5)**. The CVs are marginally lower in the unweighted analysis, but are marginally narrower for the weighted analysis. There are wide confidence intervals around the index between 1993 and 1998 (**Figure 5)**. An increase in surface abundance of 50% during that period, a decrease of 50% during the period, and no change are all possible interpretations of the index based on whole line data.

Table 8. Strip-transect based estimates of biomass density (BD) using the (length restricted) whole line data truncated at 6 nm each side of the transect line. The first set of estimates (BD 1) gives each line equal weight, and the second set (BD 2) weights the biomass on the lines according to the length of the line.

Whole Line Data	1993	1994	1995	1996	1997	1998	1999	2000
BD 1	134.2	83.7	113.2	97.5	80.4	88.8	29.1	66.8
SE 1	40.6	16.7	24.7	24.2	20.1	33.5	12.1	29.0
CV 1 (%)	30	20	22	25	25	38	42	43
BD 2	144.8	91.0	119.7	99.5	88.3	85.8	29.7	55.1
SE 2	42.6	17.4	25.0	24.5	20.4	32.7	12.2	24.6
CV 2 (%)	29	19	21	25	23	38	41	45
Number of replicates	4	8	8	7	5	5	4	4

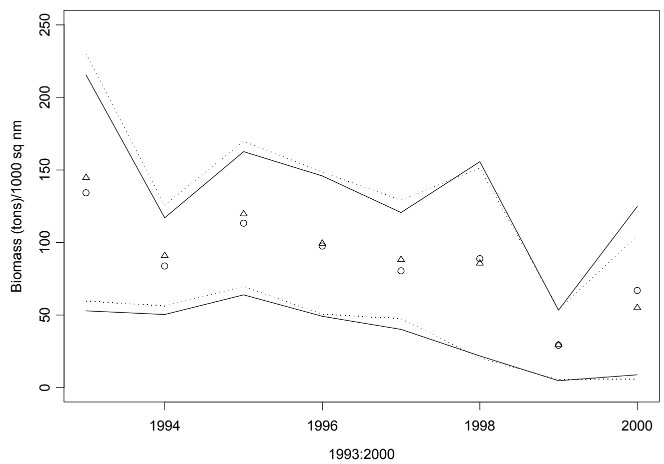


Figure 5. Comparison of weighted and unweighted estimates of biomass density, using whole lines and 6 nm truncation. Circles and solid lines show the estimates and 95% confidence intervals for the unweighted analysis, diamonds and dotted lines show the estimates and 95% confidence intervals for the weighted analysis.

Comparison of the unweighted quarter line estimates (BD 3) with the unweighted whole line estimates (BD 1) (**Table 9 and Figure 6)** shows that increasing the number of units has a noticeable effect on decreasing the CVs and narrowing the confidence intervals. This is to be expected from basic sample size considerations.

Table 9. Comparison of the unweighted quarter line estimates (BD 3) with the unweighted whole line estimates (BD 1), which is the same data in Table 8.

	1993	1994	1995	1996	1997	1998	1999	2000
BD 1	134.2	83.7	113.2	97.5	80.4	88.8	29.1	66.8
SE 1	40.6	16.7	24.7	24.2	20.1	33.5	12.1	29.0
CV1 (%)	30	20	22	25	25	38	42	43
BD 3	127.5	75.2	105.5	95.4	82.0	86.1	30.7	54.1
SE 3	31.1	13.6	21.3	20.7	19.4	29.2	12.5	19.3
CV3 (%)	24	18	20	22	24	34	41	36

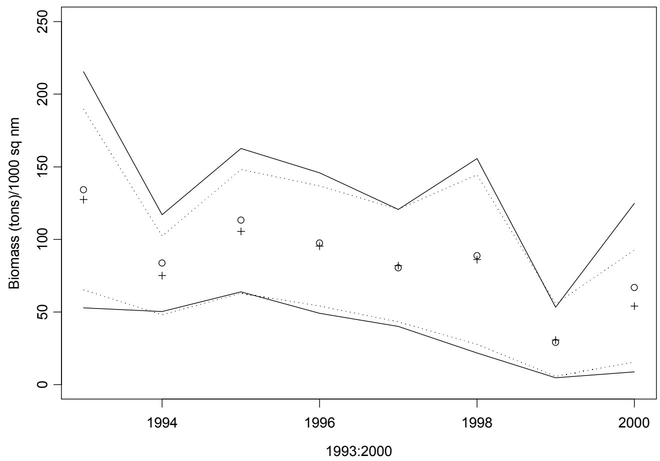


Figure 6. Comparison of weighted estimates of biomass density for whole and quarter lines, with 6 nm truncation. Circles and solid lines show the estimates and 95% confidence intervals for the whole line analysis, crosses and dotted lines show the estimates and 95% confidence intervals for the quarter line analysis.

Comparison of CVs for strip-transect and model-based indices

In this sub-section the CVs of the strip-transect and model-based indices are compared. For comparative purposes, quarter line strip-transect presence/absence indices are also given. The various indices are provided in **Table 10 and Table 11.** Comparison of presence/absence indices with biomass indices indicates that the presence/absence indices have much lower CVs than the biomass indices (**Table 10)**. This is because of high variability in the biomass detected on a quarter line (**Figure 7, Table 12)**, and because of the large variation between and within spotters in their estimates of biomass (**Section 6.1**).

Table 10. The bootstrap model-based point estimates, standard errors and CVs for the presence/absence, conditional biomass and combined biomass indices for the years 1993-2000.

	1993	1994	1995	1996	1997	1998	1999	2000
Presence/absence index	0.363	0.337	0.342	0.278	0.233	0.241	0.153	0.200
Estimated SE	0.067	0.060	0.062	0.041	0.046	0.040	0.047	0.043
Estimated CV (%)	18	18	18	15	20	17	30	21
Conditional biomass index	223.7	205.7	168.5	251.6	167.2	134.7	111.2	123.0
Estimated SE	74.7	65.8	54.0	64.1	56.5	33.5	43.6	37.7
Estimated CV (%)	33	32	32	25	34	25	39	31
Combined biomass index	104.8	93.3	78.2	107.5	59.6	51.7	31.3	41.7
Estimated SE	37.7	32.3	28.0	29.0	20.9	14.0	12.8	13.1
Estimated CV (%)	36	35	36	27	35	27	41	31

Table 11. The unweighted strip-transect estimates, standard errors and CVs for the whole line biomass, quarter line biomass, and quarter line presence/absence indices for the years 1993-2000.

	1993	1994	1995	1996	1997	1998	1999	2000
Whole line biomass index	134.2	83.7	113.2	97.5	80.4	88.8	29.1	66.8
SE	40.6	16.7	24.7	24.2	20.1	33.5	12.1	29.0
CV (%)	30	20	22	25	25	38	42	43
Quarter line biomass index	127.5	75.2	105.5	95.4	82.0	86.1	30.7	54.1
SE	31.1	13.6	21.3	20.7	19.4	29.2	12.5	19.3
CV (%)	24	18	20	22	24	34	41	36
Quarter line presence/absence	0.775	0.491	0.494	0.400	0.446	0.477	0.321	0.465
SE	0.088	0.047	0.049	0.048	0.059	0.065	0.058	0.070
CV (%)	11	10	10	12	13	14	18	15

Table 12. Quantiles of the conditional biomass model.

Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum
0	1.398	1.806	1.806	2.296	3.395

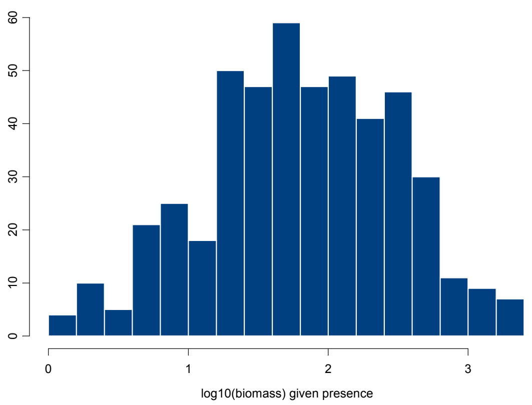


Figure 7. Histogram of conditional biomass for the quarter lines. Note that the conditional biomass varies by over 3 orders of magnitude.

Patches typically do not occur in isolation: they often form groups or clusters of 2-5 patches, usually within a diameter of 1-5nm**.** Between 1993 and 1998, 713 sightings of SBT were made. Of these, 52% were in clusters of more than one patch, only 15% occurred in single patches and 29% in clusters of more than 10 patches (**Table 13)**. While there is uncertainty in the exact number of patches in a sighting, the data is sufficiently reliable in grouped form to be indicative.

Table 13. Number of patches per sightings during the aerial survey in the GAB between 1993 and 1998.

	No of patches in sighting			
	1	2-5	6-10	>10
Number of sightings	345	274	58	36
Percentage of sightings	48	38	8	5
Total patches detected	345	842	435	646
Percentage of patches detected	15	37	19	29

Comparison of strip-transect and model-based estimates shows that the main difference between the strip-transect indices and the model-based indices is that the strip-transect indices are not adjusted for between-year and within-year differences in environmental conditions, whereas the model-based indices are adjusted for a set of known and fixed conditions that are the same from year to year. One way of judging the effect of the environmental data is to compare the (i) strip-transect estimates, (ii) model-based index which included only the terms block, IO and year, and (iii) model-based index that include the environmental data (**Figure 8)**. To aid the comparison between these three types of presence/absence indices all were rescaled by setting the 1995 estimate to unity.

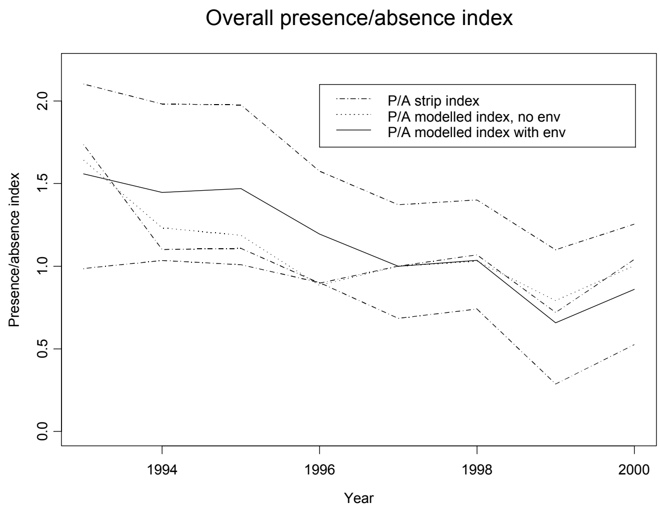


Figure 8. Comparison of the three presence/absence indices, (i) strip-transect, (ii) modeled and no environmental variables, and (iii) modeled and environmental variables. Only the 95% bootstrap confidence intervals from the modeled index with environmental data have been added to the plot.

The comparison shows that differences between the models are not statistically significant. In particular, the modeled index fluctuates less than the strip-transect estimate (**Figure 8)**. The inclusion of the environmental data increases the indices for the years 1994-1996 and reduces it in the year 2000. This appears to be largely due to adjustments due to inclusion of satellite SST – 1994-1996 were cooler than other years (see **Appendix 6.5C**). The fact that the differences between the models are not statistically significant, based on confidence intervals from the modeled index with environmental data, raises questions about the quality of the environmental data that have been used in the models.

Indices based on models including aerial survey variables

The statistical method used to analyse the data is outlined in Cowling (2000). Predictions are made for the inshore and offshore components of each block in sighting conditions. Final estimates could also be obtained by weighting each component of each block by its area, however, sighting conditions were averaged in the final combined indices.

The presence/absence and conditional biomass indices and confidence intervals are calculated using both the bootstrap and predictions from the respective models. There is extremely good agreement between both methods of calculating the indices and the confidence intervals (**Figure 9)**. The point estimates and confidence intervals for the overall index are calculated using the bootstrap. Based on these results, overall indices were created.

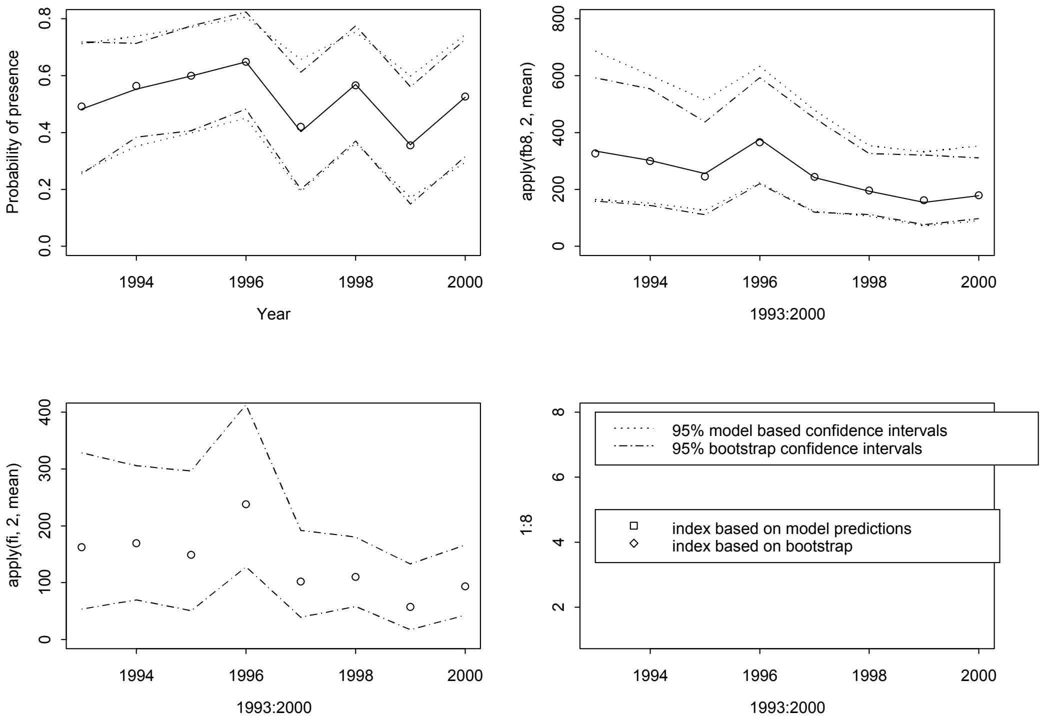


Figure 9. The presence/absence, conditional biomass, and overall combined indices for Block 3 offshore.

Overall abundance indices for each inshore and offshore segment in each block are shown in **Figure 10**. The overall abundance index is highest in Block 3 Offshore (**Figure 10**). In fact, the domestic fishing ground for SBT was located in this area from 1996-2000. There is an apparent decline in abundance over time in Blocks 1, 2, 4 and 5, with less of a decline in Block 3. The final indices presented are the overall GAB indices obtained by averaging the indices over the sections of each block (**Figure 11)**. There may be minor changes to the indices if the sections of each block are correctly weighted by area rather than averaged.

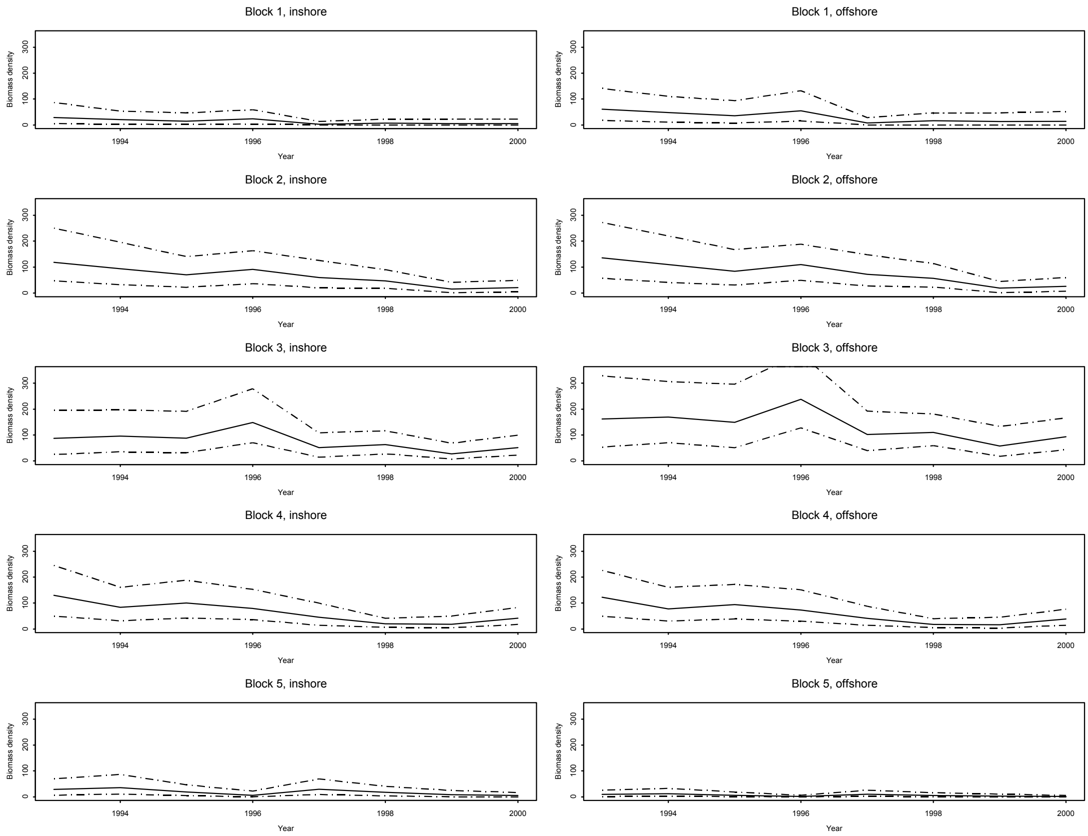


Figure 10. The overall indices for all the survey blocks together with bootstrap confidence intervals.

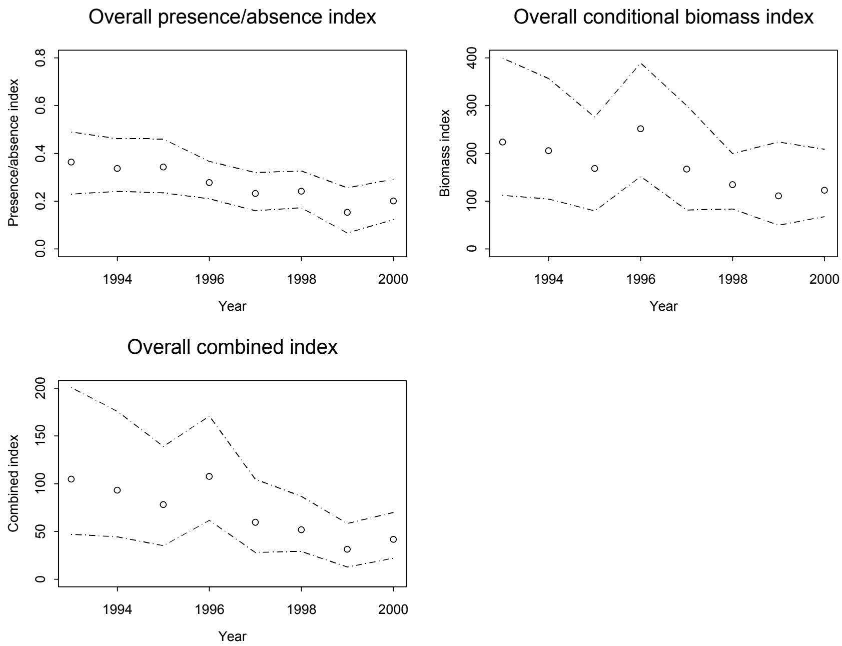


Figure 11. The final indices (presence absence, conditional biomass, and combined) for the entire GAB obtained by averaging the indices over the inshore/offshore sections of each block (Block*IO). Bootstrap confidence intervals are shown for each index.

The presence/absence index is relatively linear with the narrowest confidence interval (**Figure 11**). The conditional biomass index has the widest confidence interval. This is due to the high variability in the biomasses detected on the quarter lines. The overall combined index has wide confidence intervals due to the high variation in the conditional biomass indices.

Addition of additional important environment variables

During earlier portions of the project (**Section 6.2**), several additional important remotely-sensed oceanographic variables were identified. These variables are chlorophyll (CHL), chlorophyll gradient and SST gradient. Moon phase was also included in these analyses. The average value for each variable was extracted for each quarter line. Unfortunately, these variables differ with regard to the years they are available. CHL has only been measured from satellites since 1998 and so models including CHL only cover the years 1998 to 2000. Data for two important variables, SST-gradient and moon-phase, were available for each year of the aerial survey. The effect of these variables on two of the indices (presence/absence and conditional biomass) is shown in the analysis of deviance and variance tables in **Appendix 6.5D**. These analyses show that the variable SST-gradient does not have a significant association with presence/absence or conditional biomass, whereas moon-phase does have a significant association with conditional biomass but not with presence/absence. Moon-phase is therefore retained in the conditional biomass models. SST-gradient is therefore removed from both models. The median value of moon-phase is 0.06, and the median value in sighting conditions is 0.07; both these values are close to zero. Prediction at a value of 0 of moon-phase is equivalent to not including moon-phase at all in the model, thus, models which do not include moon-phase result in almost identical indices as models that do. The modeled relationship between conditional biomass and moon-phase shows that if SBT are observed, conditional biomass is at a peak two days after the full moon (**Figure 12)**. This relationship shows that biomass declines as the full moon wanes until two days past the new moon then starts increasing again, reaching a peak two days after the full moon.

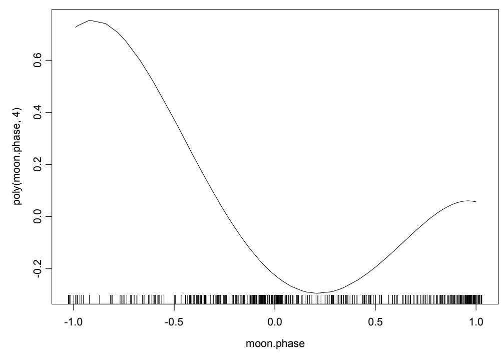


Figure 12. Moon-phase fitted as a 4th order polynomial in the conditional biomass presence model. (Note that while moon-phase was fitted as a polynomial, the ends are very close. A cyclic function will make the two ends of the curve meet. Models were also fitted using sine and cosine terms and results were very similar.)

Chlorophyll was found to be an important predictive variable for presence/absence of SBT in **Section 6.2.** However, chlorophyll can be examined only for 1998 onwards, as the satellite was only launched in September, 1997. It is therefore not useful in an index starting in 1993 and was not considered further for inclusion. Chlorophyll does appear to have some effect on presence/absence though not on abundance: as chlorophyll increases, the probability of SBT presence declines (**Appendix 6.5D**).

Inclusion of surfacing behaviour: Linking the aerial survey indices with SBT surfacing from archival tags

If SBT surfacing rates are the same all over the GAB in average or sighting conditions, there is no need to adjust the aerial survey indices for differences in surfacing rates: the same multiplier would be used everywhere. As a first step to including spatial differences in surfacing rates in an SBT abundance index, appropriate areas of the GAB in which to test for differences must be identified. The spatial accuracy of the daily position estimates from the archival tag data in the GAB is not sufficient to allow separate models to be developed for the inshore and offshore sections of each block. There is therefore a need to develop a meaningful but coarser spatial basis for which to model the surfacing data and develop indices from the aerial survey data. The average number of patches detected per quarter line in each block and the average biomass detected per quarter line in each block in the inshore and offshore halves of each block is given in **Table 14**.

Table 14. Mean number and biomass of SBT patches detected per quarter line by block for the years 1993-2000.

Area	Mean	Block 1	Block 2	Block 3	Block 4	Block 5
Inshore	n	0.18	1.70	2.21	2.91	0.69
Offshore	n	0.22	0.97	1.53	1.32	0.37
Inshore	biomass	4.27	65.65	62.31	84.43	14.17
Offshore	biomass	5.08	32.00	49.69	34.09	4.32

Blocks 2, 3 and 4 have higher abundances of observed SBT than the outer two blocks where few SBT are detected (**Table 14)**. For the purpose of the development of the indices including environmental and behavioural (surfacing rate) covariates, the survey area is divided into 4 areas, all of Block 1, the inshore of Blocks 2-4, the offshore of Blocks 2-4 and Block 5. These are subsequently denoted as Areas 1-4 respectively (**Figure 13)**. The number of patches and biomass in each of these Areas is shown in **Table 15.**

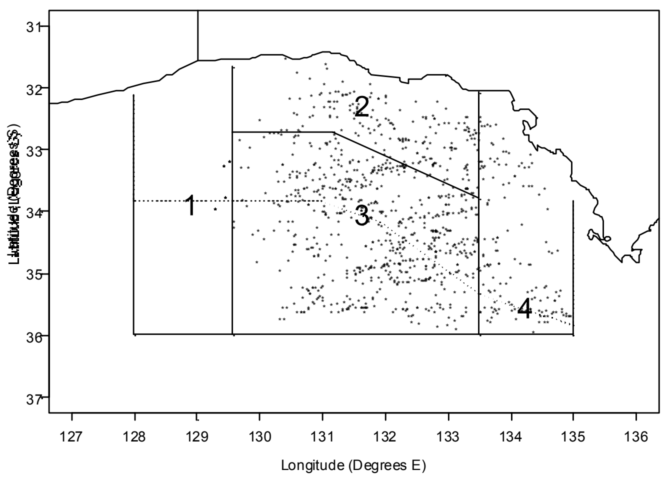


Figure 13. The four spatial areas used for development of an index incorporating behavioural information on surfacing rate of SBT derived from archival tag analysis. Dotted lines represent the southern boundary of the aerial survey region. All SBT sightings for 1993-2000 are shown as dots.

Table 15. Mean number and average biomass of patches detected per quarter line in each Area used to develop an index incorporating information on surfacing rate and environmental covariates.

	Area 1	Area 2	Area 3	Area 4
Mean number	0.20	2.27	1.27	0.53
Mean biomass	4.67	70.66	38.59	9.22

Separate estimates of surfacing behaviour for each area were derived using the standard and sighting conditions (**Section 6.4**). Different surfacing rates would allow calculation of presence/absence and biomass indices for each area. The indices can then be adjusted for the differences in surfacing rates to produce the final indices.

The models for surfacing rate in each area (**Section 6.4**) were used to give predicted surfacing rates in standard conditions over the GAB for each tagged SBT and each area. As with the aerial survey predictions, standard conditions are defined as the median of the conditions over the region observed during the study. The significant explanatory variables were different in the models for each area (**Table 16**).

Table 16. Standard conditions for each Area: DAR is Division of Atmospheric Research, the source for some of the environmental data, as noted in the table. Note Area 1 was not included due to insufficient surfacing observations to allow construction of a surfacing behaviour model. Where standard conditions are not included for an area, it was because the conditions were not significant terms in a model describing surfacing for that area.

Area	DAR wind direction	DAR wind speed	DAR cloud	DAR AT	DAR barometric pressure	Log(chl)	Day-of-year	Moon-phase	tag-SST
2	227.76	11.6035	0.165	12.665	1019.54	0.126			
3						0.126	44	15	
4						0.126			20.1

Table 17. Predicted surfacing probability for each tag within each area. Note Area 1 was not included due to insufficient surfacing observations to allow construction of a surfacing behaviour model. Not all tagged SBT were found in all areas.

Tag	Area 2		Area 3		Area 4	
	Surfacing	SE	Surfacing	SE	Surfacing	SE
97620			0.322106	0.052303	0.436868	0.113057
97622			0.556031	0.096883	0.349519	0.081992
97632	0.612971	0.078675	0.406700	0.04986	0.177274	0.06318
97675	0.607584	0.066238	0.565126	0.043738		
97707	0.581719	0.056886	0.548172	0.040635		
97708	0.716624	0.051798	0.598656	0.039244		
97711	0.439109	0.058858	0.416085	0.047651	0.190022	0.204819
97721	0.580901	0.054282	0.501029	0.050691	0.425374	0.080296
97731			0.371886	0.109611	0.273183	0.13622
97733	0.677949	0.056094	0.452263	0.048162	0.413983	0.131808
97741	0.692125	0.057366	0.293537	0.049033	0.450696	0.139759
97754	0.641319	0.056203	0.396965	0.048329		
97755			0.329121	0.109731		
97756	0.529174	0.065377	0.422868	0.045500	0.063728	0.06952
97757	0.354757	0.074603	0.451346	0.049196		
97760			0.060585	0.028799	0.115587	0.053673
98007			0.534037	0.063425	0.626629	0.129668
98017			0.171532	0.05523	0.185236	0.088598
98024			0.352923	0.051112	0.214402	0.083025

To test for differences between surfacing probability (**Table 17)** between areas, a mixed model was fit to the predicted values using tag as a random effect and area as a fixed effect. The standard errors for each fish and area were ignored as they are relatively similar in magnitude and so would make little difference in an analysis (**Table 17)**. Because of the lack of balance in the data, the model was fit using the "reml" directive in Genstat. The differences in surfacing rates between regions were found to be significant ($p < 0.001$). The predicted mean surfacing rates and standard errors are shown in **Table 18**. The significant differences between areas under standard weather conditions indicates that indices should be corrected for surfacing rates in these different regions of the GAB.

Table 18. Mean predicted surfacing rates for juvenile SBT in three regions of the GAB.

	Area 2	Area 3	Area 4
Mean	0.5679	0.4079	0.3131
SE	0. 0400	0.0324	0.0374

The "presence/absence" and "conditional biomass" models developed in an earlier sub-section were recalculated, replacing block by Area, (the four areas defined earlier in this section). One thousand bootstrap samples, models and new predictions were made for each of the four areas. This resulted in 1000 values of predicted surface abundance in sighting conditions for each of the four areas. When combining Areas 2, 3 and 4, they are weighted in the ratio 3:3:2; an approximation to their relative areas.

These abundance indices are adjusted for surfacing rates in average conditions by dividing them by the predicted surfacing rates from the archival tag models in the corresponding Area of the GAB. It is arguable that surfacing rates for "average sighting conditions" are more appropriate, however, only the "average conditions" result is reported here. Variation in surfacing rates for each Area was incorporated in the adjustment by dividing by normal random variables with mean and standard deviation equal to the mean and standard deviation for that Area. In making these corrections, there is an assumption that surfacing rates are the same in each year. As surfacing behaviour was only derived from archival tags in 1998, this assumption cannot be validated at this time.

When surfacing behaviour is included, there is very little difference between the Area-based indices (**Figure 14, Table 19)** and the block*IO indices (**Figure 11)**.

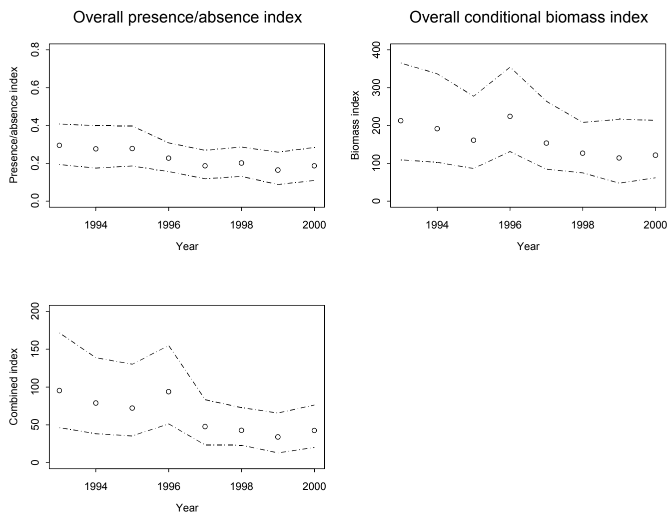


Figure 14. The final indices (presence absence, conditional biomass, and combined) given by the models using the four Areas rather than block*IO indices shown in Figure 11.

Table 19. The bootstrap model-based point estimates, standard errors and CVs of the surface adjusted overall index based on Area rather than Block*IO.

	1993	1994	1995	1996	1997	1998	1999	2000
Combined biomass index (Area)	95.3	78.7	72.0	93.8	47.7	42.6	33.9	42.5
SE	33.4	26.7	24.4	26.8	15.1	12.8	13.9	15.0
CV (%)	35	34	34	29	32	30	41	35

The effect of removing Area 1 (for which there are no estimates of surfacing rate) and dividing by the simulated surfacing rates to get surfacing-adjusted indices is shown in **Figure 15.** Including the surfacing probabilities has the effect of shifting all the indices upwards. There is almost no effect at all on the variability in the indices apart from a slight increase, as expected when surfacing behaviour is included.

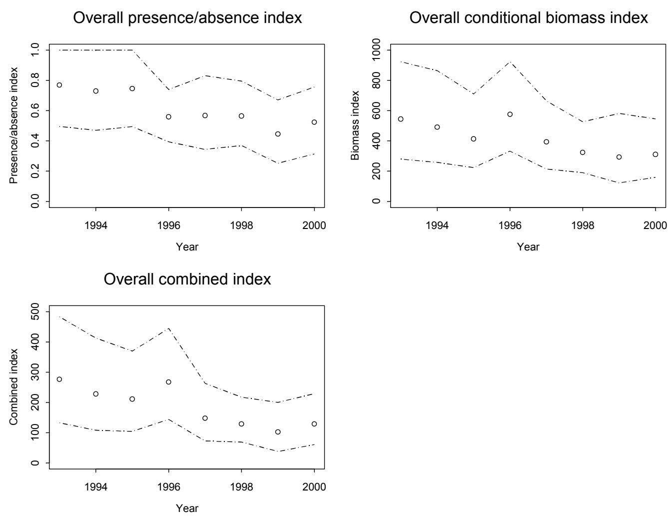


Figure 15. Indices resulting after removing Area 1 (for which there are no estimates of surfacing rate) and dividing by the simulated surfacing rates to get surfacing-adjusted indices.

The bootstrap model-based point estimates, standard errors and CVs of the surface-adjusted overall and presence/absence indices are shown in **Table 20**. Note that the CVs of the presence/absence index have increased a little more than those of the biomass index. These estimates are the final set produced in this section of the report.

Table 20. The bootstrap model-based point estimates, standard errors and CVs of the surfacing-adjusted overall and presence/absence indices.

	1993	1994	1995	1996	1997	1998	1999	2000
Surfacing-adjusted combined biomass index	276.5	228.0	211.1	267.7	148.1	128.8	102.5	128.9
SE	96.7	78.5	72.9	78.9	48.1	39.4	43.1	45.6
CV (%)	35	34	35	29	32	31	42	35
Surfacing-adjusted combined presence/absence index	0.77	0.73	0.75	0.56	0.57	0.57	0.45	0.53
SE	0.15	0.15	0.14	0.09	0.13	0.10	0.11	0.11
CV (%)	19	20	19	17	22	18	25	22

Trends in SBT abundance

There appears to be changes in abundance over time in all these indices. Fitting year in the presence/absence model as a linear term rather than as a factor, results in the analysis of deviance tables shown in **Appendix 6.5E**. After allowing for all the preceding terms in the model, there is a strongly significant effect of year ($p = 0.009$) (**Appendix 6.5E, Table E1**). The estimated coefficient is –0.188 with standard error 0.043, which is evidence of a decline in abundance. If the last 2 years of the data (1999, 2000) are omitted from the analysis, this effect is still observed (**Appendix 6.5E, Table E2)**. The p-value for year is nearly the same (0.009). The estimated coefficient is -0.141 with a standard error of 0.053.

With regard to the model for the conditional biomass index, there is also a significant effect of year ($p = 0.021$) (**Appendix 6.5E, Table E3)**, with a negative coefficient for the slope (-0.114 with standard error of 0.049), which corresponds to a decline through time. When the years 1999 and 2000 are removed, the effect of year is only marginally non-significant ($p=0.064$) (**Appendix 6.5E, Table E4)**. The slope is slightly reduced -0.101 with standard error 0.054. This result should be interpreted in conjuction with the observation that the conditional biomass index has a high point estimate for 1996. In the strip-transect estimates, 1996 is not a particularly high year. Therefore, the 1996 value must have been increased by the inclusion and correction for environmental conditions. In fact, in 1996 the wind speeds were higher than in any other year and the satellite SST are lower than any other year (**Appendix 6.5D**). Adjustment for each of these factors will increase the predictions for 1996 more than for other years. As this point is so unusual compared to the other points in the sequence, the reliability of the wind speed and SST data should be investigated further. If that point were removed, the significance of a decline in abundance given presence would be much higher, and so the current estimate of a decline may be a conservative estimate.

Conclusion

This section of the project has addressed the issue of the robustness of the model-based methodology, showing that it is robust against the particular years of data used in the analysis, the conditions used for making predictions and the spatial units used for making predictions (inshore/offshore block segments or Areas).

Which index is preferred?

A major issue addressed by this section is the choice of an index. A choice must be made between a presence/absence index and a biomass index. The presence/absence index is recommended for the following reasons:

1. It has a lower CV,
2. It does not use the spotters estimates of biomass, which are unreliable (**Section 6.1**),
3. If surveys are to continue, a presence/absence survey and index will be quickest, and hence cheapest, to conduct. It has the simplest protocols and will be least affected by changes in survey staff over time.

An interesting result of the surfacing analysis (**Section 6.4)** is that there are differences in surfacing behaviour in different areas of the GAB. However, it may be preferable to see one model applied to all Areas, including interactions between all environmental variables and Area to account for variation between Areas. There is a need for the surfacing model to be further developed to allow this improvement. This will also allow for statistically rigorous testing of differences in SBT abundance between Areas.

Despite the differences in surfacing rates in the different Areas, the effect of the adjustments on the indices was negligible – no noticeable effect on overall shape, a general increase in level and a very slight increase in CV. If the surfacing model used the same environmental variables as the surface abundance models, predictions of both surface abundance and surfacing probability in the same conditions could be made and integrated over a set of standard environmental conditions to get a final index.

The final issue is whether to include environmental variables in the analysis and production of a fishery-independent index of abundance for juvenile SBT. While the inclusion of the environmental variables did not result in statistically different indices, it did reduce year-to year fluctuations in the indices. One interpretation is that much of the observed interannual variation in abundance might be attributed to interannual changes in the environmental conditions. To be sure that such effects are accounted for, future indices should include environmental variables. In conclusion, of all the indices, the presence/absence model-based indices including surfacing probabilities appear to be most useful. In future it would be useful to compare indices based on “regional sighting conditions”, and develop at least one other method of including surfacing probabilities in the models (**see Section 6.6**).

Future considerations

While a set of fishery-independent indices for the abundance of juvenile SBT in the GAB have been developed and extensively compared, there are a number of additional issues that are not addressed in this section. These include

1. Differences between trained spotters
2. Effect of the trainee spotters in the last two years (1999-2000)
3. Differences in the spotters size estimates (important for biomass indices only).
4. Lack of independence of estimates of pilot and spotter within a spotting team (artificially reduce the uncertainty in size and detection estimates)

These issues are explored further in the final section (**Section 6.6**). The indices are also developed in this next section, and should replace those presented here.

References

Cowling, Ann (2000). Data analysis of the aerial surveys (1993-2000) for juvenile Southern Bluefin Tuna in the Great Australian Bight. RMWS/00/3. 2000 SBT Recruitment Monitoring Workshop. Hobart.

Appendix 6.5A. Appropriate units for analysis

In this appendix the justification for the selection of the following units are described:

1. Selection of length of units for analysis
2. Choice of strip width for strip-transect analysis
3. Choice of strip length for analysis
4. Choice of spatio-temporal units for modeling

Selection of maximum and minimum length of units for analysis

The goal in this first sub-section of this appendix is to determine minimum and maximum lengths for the whole, half and quarter lines to be included in the analyses. An aim is to exclude as few data as possible, but as the current model-based analysis is not weighted by survey effort, giving equal weight to very incomplete units should be avoided. Variation in line length can result in the survey because sometimes a transect line was not completely surveyed and rarely 10 nm or less of a line was surveyed. This occurred when the weather deteriorated during the course of a flying day or when, towards the end of a day, fuel was running too low to continue surveying. In addition, when a partially completed line is divided into half and quarter lines for the analysis short fragments may result. The results of the analysis are provided in **Table A1**. In the remainder of this sub-section additional details of how these results were obtained are provided.

Table A1. The length of lines excluded for each scale of analysis. This is the summary of the analysis in this appendix.

Scale	Excluded lines
Whole lines	Length < 40 nm
Half lines	Length < 30 nm & Length > 100 nm
Quarter lines	Length < 10nm & Length > 65 nm

Whole lines

A few short and long whole lines were present in the data (**Figure A2).** The goal is to remove the few very short lines in the data from the analysis (**Table A2)**. Accordingly, in subsequent analyses, the <0.5% of the lines with length < 40nm (corresponding to < 20 minutes of search effort at 120 knots) were removed.

Table A2. Extreme quantiles of the whole, half and quarter line length (nm) distributions.

Quantile (%)	0	2	6.67	99.50	99.75	100
Length: Whole line	0.3	30.9	42.4	188.3	188.6	206.0
Length: Half line	0.3	15.1	33.2	93.9	96.5	147.4
Length: Quarter line	0.3	10.6	20.4	59.0	61.2	81.5

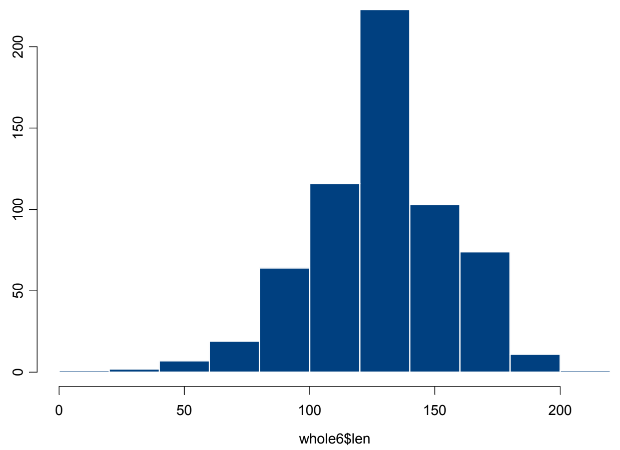


Figure A1. Frequency distribution of lengths of whole transect lines in the aerial survey data.

Half lines

Breaking whole lines into half lines has resulted in an increased proportion of line fragments, and there is one very long line (**Figure A2**). Extreme quantiles of the half line length distribution (**Table A2)**, and were used to select the lengths for exclusion. The few very short lines in the data and the abnormally long lines were removed from further analysis. Thus, in subsequent analyses, the <2% of the lines <30nm in length (corresponding to <15 mins search effort at 120 knots), and the one line over 100 nm in length[1] were removed.

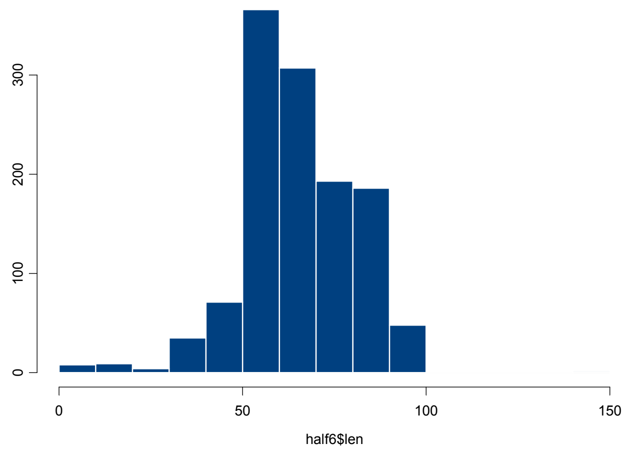


Figure A2. Frequency distribution of lengths of half lines created from the aerial survey data.

Quarter lines

Breaking the transect lines into quarter-line segments also resulted in an increased proportion of line fragments (**Figure A3**). It was considered desirable to remove the few

[1] The long line occurred as a result of the method of numbering lines from the early years of the survey when there was a random element in the placement of the lines. Two lines that fell close together in one replicate of the survey have been assigned the same line number as they both fell in the region assigned to that line number

very short lines and any abnormally long lines from the analysis. Extreme quantiles of the quarter line length distribution (**Table A2**) were used to select the lengths for exclusion. In subsequent analyses, the lines <10 nm in length (corresponding to 5 minutes of search effort at 120 knots, n < 2%), and one line over 65 nm in length that may represent an error in line division, were removed.

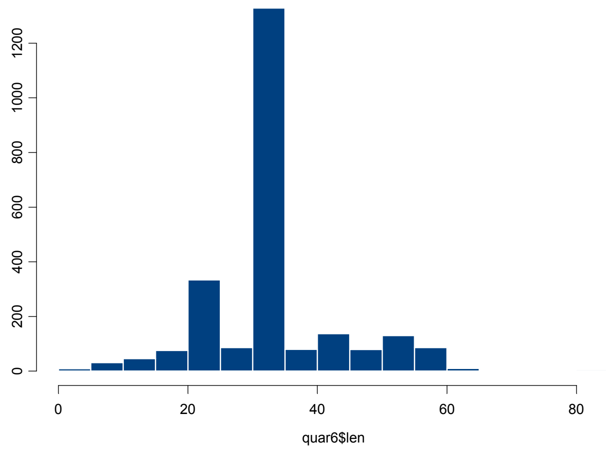


Figure A3. Frequency distribution of the lengths of the quarter lines created from the aerial survey data.

Choice of strip width

The aerial survey observations along each transect line occur at a range of distances from the track line. In conventional line transect surveys, the number of sightings decreases with distance from the transect line, due to a reduction in detection at greater distances from the observer. In the current analyses, the observations of SBT have been truncated at three different strip half-widths: 4, 5 and 6 nm. There are more sightings and higher biomass in the widest (6 nm) width strip and least in the narrowest strip (4 nm). Somewhat surprisingly, there is no decline in detectability of SBT in the range of 4-6 nm from the trackline. Thus, analyses that do not model any decline in detectability with distance from the track line are appropriate. The analysis is also based on the assumption that detectability is equal throughout the surveys.

Strip-transect analyses based on three strip widths using the whole line as the unit of analysis show that there are no consistent differences between the estimates including data truncated at different widths (i.e. no consistent declines in detectability). The CVs increase as truncation width decreases and there is no evidence of a decline in detectability with strip width (**Table A3, Figure A4)**. Based on these results, a truncation distance of 6 nm each side of the transect line is used for all strip-width analyses.

Table A3. Biomass density (BD, tonnes/1000 sq nm), standard error of the biomass density (SE) and coefficient of variation (CV) for the strip-transect analyses based on three strip widths using whole line as the unit of analysis.

	1993	1994	1995	1996	1997	1998	1999	2000
6 nm truncation								
BD	134.2	83.7	113.2	97.5	80.4	88.8	29.1	66.8
SE	40.6	16.7	24.7	24.2	20.1	33.5	12.1	29.0
CV (%)	30	20	22	25	25	38	42	43
5 nm truncation								
BD	128.3	87.5	117.4	92.5	85.8	88.7	32.0	64.6
SE	33.1	17.6	25.7	22.6	23.2	37.5	13.4	32.2
CV (%)	26	20	22	24	27	42	42	50
4 nm truncation								
BD	146.8	84.9	111.4	86.5	92.3	88.5	37.1	74.2
SE	40.6	19.1	24.8	22.8	27.5	42.6	15.6	39.3
CV (%)	28	23	22	26	30	48	42	53

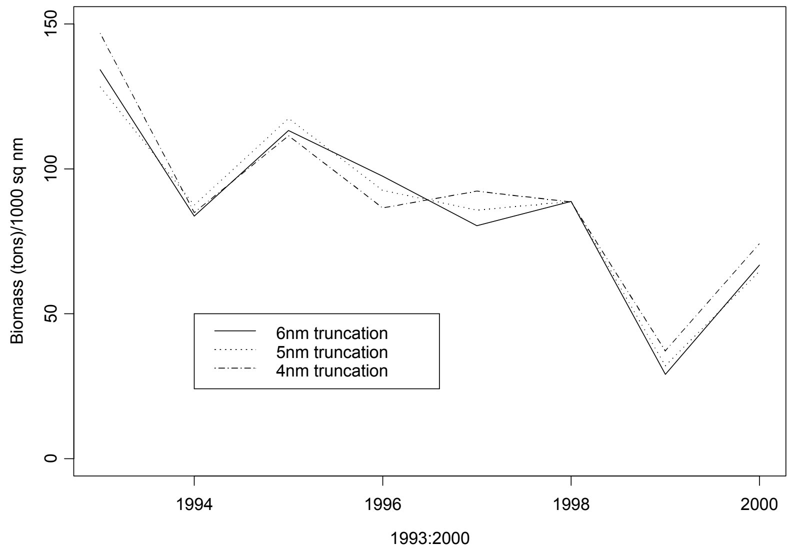


Figure A4. Biomass density of SBT in the GAB 1993-2000 using three potential strip widths.

Choice of strip length: whole, half and quarter lines

In an ecological study, the best quadrat size is related both to the scales of pattern that are of interest and to the overall scale and purpose of the study. The quadrat size should also have some kind of environmental interpretation: in this case, be linked to oceanographic scales of variation. SBT patches usually occur in groups (clusters) rather than in isolation. When clusters are well spaced, the spatial unit should have a high chance of covering most of a cluster, and so should be a little larger than the size of the cluster. In the example below (**Figure A5)**, the smallest quadrat is too small to cover a whole cluster, and the largest cluster is too large – it is capable of covering a number of distinct clusters. The medium sized cluster is the best of these. It is a little larger than the average cluster size. However, in studies in which there is poor separation of clusters, or if clusters differ markedly in size, there is no ideal quadrat size.

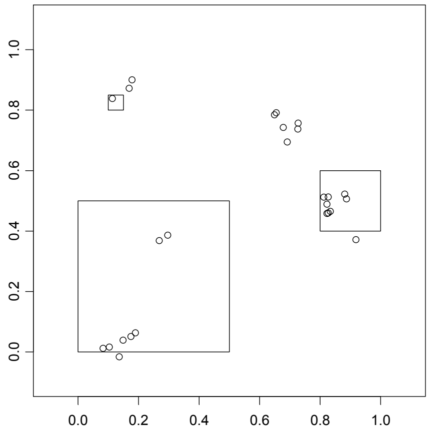


Figure A5. Example of how different size sampling or analysis units (quadrats) would cover a set of SBT sightings composed of differing numbers of patches.

In additional aerial survey experiments conducted in 1998 in which only patches within 10 nm from the transect line were recorded, 118 patches were detected on just one side of a 60 nm transect. The variability in size of the groupings means that there is no single best quadrat size. However, a large proportion of clusters tend to occur inshore associated with the "lumps" and offshore associated with the shelf break (see **Section 6.2)**. A division of the lines into inshore and offshore halves separates the data naturally into these two components. In some years the offshore clusters are inside the shelf break and in other years outside the shelf break. Thus, dividing the offshore half into two parts along the shelf break makes another natural division. The shelf break is approximately ½ way along the offshore half lines. Thus the quarter line unit is another appropriate unit: the outer quarter extends between the shelf break and the southernmost boundary of the survey area.

Empirical evidence that this is a reasonable decision is obtained by comparing the number of detections of presence for three line lengths (**Table A3)**. In splitting lines it is preferable not to split groups detected as a unit. Using smaller lines as the unit of analysis does not double the number of presences, but increases it by a relatively small proportion (**Table A3)**. The number of presences increases by 32% when the whole lines are split, and by 17% when the half lines are split again to make quarter lines. These are acceptable increases, however, further division of the lines is undesirable as SBT are only detected on 20% of the quarter lines. Finer division of the lines will further reduce this percentage, making the SBT presence signal weaker.

Table A3. Frequency of detection of SBT per total unit for each of three possible line lengths for the analysis of SBT abundance.

Data Unit	Absence (n)	Presence (n)	Total units	Presence (%)
Whole lines	311	310	621	50%
Half lines	819	409	1228	33%
Quarter lines	1955	479	2434	20%

In the modeling approach used to derive abundance indices, environmental variables are matched with each line segment. These environmental variables, such as wind speed or temperature can vary considerably over the course of a 150 nm transect. The average of the values recorded on the line segment (weighted by the line length associated with that value) are matched with each segment. Thus, finer subdivision of the lines provides more environmental resolution and increases the likelihood of detecting any environmental relationships present in the data. The observed inshore/offshore distribution of SBT, empirical evidence for an appropriate scale (**Table A3)**, and increased environmental resolution, all support the choice of the quarter line as the main unit for the analysis. Accordingly, environmental conditions used in analyses are also averaged over the quarter line.

Choice of spatio-temporal units for modeling

Previous analyses have shown that there are significant differences between the abundance of SBT in the quarter lines in the survey area. However, broader scale areas, such as blocks (sets of transect lines), or groups of blocks (separated into inshore and offshore areas) are of more interest for analysis and prediction than particular lines. In addition, a modeling approach can allow for different temporal trends in different areas. However, the quarter line is too small a unit for this interaction modeling as the quarter line-by-year interaction has 413 degrees of freedom – over 17% of the available degrees of freedom in the presence/absence analysis. To guard against finding meaningless (noise) terms as significant, a rule of thumb is that no more than 10% of the degrees of freedom be used in observational studies such as this one.

Therefore the following five versions, each using different spatial-temporal terms, are compared using two models; (i) presence/absence, and (ii) conditional biomass.

1. inshore and offshore halves of the 15 transect lines used in the aerial survey (term: line*IO*year)
2. inshore and offshore halves of each block (3 transect lines per block) (term: block*IO*year)
3. blocks (term: block*year)
4. inshore and offshore components of larger spatial units than blocks (denoted BLOCK) with three levels: the high abundance Blocks (2, 3 and 4) are in one level, the consistently low abundance Block 1 in another level, and the variable/low abundance Block 5 in another level[2]) (term: BLOCK*IO*year)
5. no spatial units in model (term: year)

The term IO is a factor with 2 levels: inshore (I) and offshore (O). The full models are nested and the analysis of deviance and variance results are shown in **Table A7 and Table A8**. The significance of the spatial aggregation in the presence/absence models can be assessed using the change in deviance in the analysis (**Table A5)**, while for the conditional biomass models F-tests can be used to assess the significance (**Table A6)**.

[2] The basis for this grouping is that the mean biomass detected per year is consistently high in the 3 central blocks (2,3 and 4) and low in Block 1 and variable/low in Block 5.

	1993	1994	1995	1996	1997	1998	1999	2000
Block 1	5.9	1.2	7.8	5.6	1.8	7.5	5.6	1.2
Block 2	30.9	19.7	82.7	57.7	48.2	110.5	2.4	8.9
Block 3	85.1	56.4	53	90.7	32.9	37.7	42.4	40.1
Block 4	125.7	87.1	86.8	37.4	53	13.6	6.8	41.7
Block 5	22	15.1	2.9	0.3	32.7	1.6	2.5	1.2

Table A5. Comparison between the spatio-temporal units in the versions using the presence/absence model.

Models versions compared	Change in deviance	Change in df	P(Chi-square)
1 and 2	241.1172	160	0.00004
1 and 3	297.7126	200	0.000009
1 and 4	303.8922	192	0.0000005
1 and 5	467.0645	232	0

Table A6. Comparison between the spatio-temporal units in the versions using the biomass conditional on presence model.

Models versions compared	F-ratio	P(F)
1 and 2	1.1371	0.2066
1 and 3	1.1855	0.1172
1 and 4	1.1352	0.1888
1 and 5	1.3182	0.0198

There are significant differences in the deviance according to the spatial scale of the model (**Table A5),** but little difference in the F-ratios (there is a difference only when no spatial units are used), (**Table A6).** These results indicate that the presence/absence models are much more sensitive to the spatial scale than the conditional biomass models. There is a better fit to the presence/absence models when there is finer spatial detail in the model. There is more spatial variation in presence/absence than in conditional biomass. In the rest of **Section 6.5** blocks are used as the spatial unit.

The presence/absence model can be slightly simplified: the block*IO*year and the IO*year interaction terms are not significant (**Table A7**). However, the block*IO interaction and the block*year interactions are significant, ie there are inshore/offshore differences in blocks, and there are year to year differences in blocks in the probability of detecting SBT. In the rest of **Section 6.5**, the presence/absence models include the terms block*IO and block*year as the spatio-temporal units, and in the conditional biomass models block and year are the spatio-temporal units.

Table A7. Analysis of deviance tables for presence/absence models using different spatial temporal variables in five model versions.

Model version 1: term: line*IO*year	df	Deviance	Resid. df	Resid. Dev	Pr(Chi)
NULL			2392	2396.035	
Line	14	153.1556	2378	2242.879	0
IO	1	0.0858	2377	2242.793	0.769623
Year	7	22.4005	2370	2220.393	0.002166
Line*IO	14	32.7999	2356	2187.593	0.003079
Line*year	98	158.3694	2258	2029.223	0.000109
IO*year	7	13.8479	2251	2015.376	0.053954
Line*IO*year	98	104.3126	2153	1911.063	0.312437
Model version 2: term: block*IO*year	df	Deviance	Resid. df	Resid. Dev	Pr(Chi)
NULL			2392	2396.035	
Block	4	108.914	2388	2287.121	0
IO	1	0.118	2387	2287.003	0.731214
Year	7	22.031	2380	2264.972	0.002509
Block*IO	4	9.5312	2376	2255.44	0.04911
Block*year	28	56.5227	2348	2198.918	0.00111
IO*year	7	12.2779	2341	2186.64	0.091782
Block*IO*year	28	34.4595	2313	2152.18	0.18615
Model version 3: term: block*year	df	Deviance	Resid. df	Resid. Dev	Pr(Chi)
NULL			2392	2396.035	
Block	4	108.914	2388	2287.121	0
Year	7	22.0369	2381	2265.084	0.002503
Block*year	28	56.308	2353	2208.776	0.00118
Model version 4: term: BLOCK*IO*year	df	Deviance	Resid. df	Resid. Dev	Pr(Chi)
NULL			2392	2396.035	
BLOCK	2	97.46805	2390	2298.567	0
IO	1	0.12365	2389	2298.443	0.725108
Year	7	21.94364	2382	2276.499	0.002598
BLOCK*IO	2	7.19915	2380	2269.3	0.027335
BLOCK*year	14	29.91773	2366	2239.382	0.007833
IO*year	7	11.52281	2359	2227.86	0.117385
BLOCK*IO*year	14	12.90435	2345	2214.955	0.534069
Model version 5: term: Year	df	Deviance	Resid. df	Resid. Dev	Pr(Chi)
NULL			2392	2396.035	
Year	7	17.90708	2385	2378.128	0.012396

Table A8. Analysis of deviance tables for conditional biomass models using different spatial temporal variables in five model versions.

Model version 1: term: line*IO*year	df	Sum of Sq	Mean Sq	F Value	Pr(F)
Line	14	89.9832	6.42737	3.056345	0.00019
IO	1	19.6363	19.63629	9.337456	0.002443
Year	7	44.0038	6.28626	2.989243	0.004759
Line*IO	14	27.7508	1.9822	0.942578	0.513119
Line*year	80	212.561	2.65701	1.263463	0.083933
IO*year	7	14.4092	2.05846	0.97884	0.446642
Line*IO*year	49	103.7465	2.11728	1.006808	0.466638
Residuals	306	643.5057	2.10296		
Model version 2: term: block*IO*year	df	Sum of Sq	Mean Sq	F Value	Pr(F)
Block	4	47.4585	11.86462	5.46029	0.000273
IO	1	29.233	29.23297	13.45349	0.000277
Year	7	46.7849	6.68355	3.07588	0.003616
Block*IO	4	14.8186	3.70466	1.70494	0.148007
Block*year	28	75.8116	2.70756	1.24606	0.184039
IO*year	7	17.6282	2.51831	1.15897	0.325425
Block*IO*year	23	46.0134	2.00058	0.9207	0.57063
Residuals	404	877.8485	2.17289		
Model version 3: term: block*year	df	Sum of Sq	Mean Sq	F Value	Pr(F)
Block	4	47.4585	11.86462	5.341614	0.00033
Year	7	53.4854	7.64078	3.439983	0.001353
Block*year	28	79.5603	2.84144	1.279255	0.157416
Residuals	439	975.0924	2.22117		
Model version 4: term: BLOCK*IO*year	df	Sum of Sq	Mean Sq	F Value	Pr(F)
BLOCK	2	46.205	23.1025	10.56012	3.32E-05
IO	1	28.6644	28.66439	13.10246	0.00033
Year	7	45.7788	6.53983	2.98935	0.004487
BLOCK*IO	2	0.5409	0.27045	0.12362	0.883745
BLOCK*year	14	42.0831	3.00594	1.37401	0.161869
IO*year	7	23.7196	3.38851	1.54889	0.148936
BLOCK*IO*year	9	14.763	1.64033	0.74979	0.663128
Residuals	436	953.8419	2.18771		
Model version 5: term: Year	df	Sum of Sq	Mean Sq	F Value	Pr(F)
Year	7	54.707	7.815251	3.343643	0.00173
Residuals	471	1100.89	2.337346		

Appendix 6.5B. Comparison of plane SST, plane AT and satellite SST in the models.

Inclusion of either SST or AT in the models has lead to differing trends in the aerial survey indices in past analyses. They are comprehensively compared in this Appendix, and the results are discussed in the text.

Table B1. Presence/absence model analysis of deviance results from including three different environmental variables, plane air temperature (AT), plane SST and satellite SST. The scale of the analysis unit is the quarter line. The models included the following terms; average wind speed for the unit (avws), the length of the unit (len), average AT, average plane SST, or average satellite SST (avat, avsst1, or avsst2), block, inshore/offshore (IO), year, and two interaction terms, block*IO and block*year.

1. Model with AT	df	Deviance	Resid. df	Resid. Dev	Pr(Chi)
NULL			2392	2396.035	
poly(avws, 3)	3	164.5316	2389	2231.503	0.00000
poly(len, 3)	3	78.2715	2386	2153.232	0.00000
poly(avat, 3)	3	30.6058	2383	2122.626	0.00000
block	4	84.9345	2379	2037.691	0.00000
IO	1	1.0003	2378	2036.691	0.31725
year	7	27.4902	2371	2009.201	0.00027
Block*IO	4	32.2964	2367	1976.904	0.00000
Block*year	28	43.6718	2339	1933.233	0.02994
2. Model with plane SST	df	Deviance	Resid. df	Resid. Dev	Pr(Chi)
NULL			1846	1802.096	
poly(avws, 3)	3	145.7744	1843	1656.321	0.00000
poly(len, 3)	3	33.1182	1840	1623.203	0.00000
poly(avsst1, 3)	3	27.8134	1837	1595.39	0.00000
block	4	69.3878	1833	1526.002	0.00000
IO	1	4.667	1832	1521.335	0.03075
year	7	19.0831	1825	1502.252	0.00793
Block*IO	4	12.1732	1821	1490.078	0.01611
Block*year	28	23.9991	1793	1466.079	0.68158
3. Model with satellite SST	df	Deviance	Resid. df	Resid. Dev	Pr(Chi)
NULL			2169	2185.538	
poly(avws, 3)	3	146.3841	2166	2039.154	0.00000
poly(len, 3)	3	69.9997	2163	1969.154	0.00000
poly(SST, 3)	3	21.4808	2160	1947.673	0.00008
block	4	78.0134	2156	1869.66	0.00000
IO	1	1.367	2155	1868.293	0.24233
year	7	29.7198	2148	1838.573	0.00011
Block*IO	4	34.4019	2144	1804.171	0.00000
Block*year	28	38.7594	2116	1765.412	0.08487

Table B2. Conditional biomass (log(biomass) given presence) analysis of variance results from model versions including one of three different environmental variables, plane air temperature (AT), plane SST and satellite SST. Terms were added sequentially (first to last). The scale of the analysis unit is the quarter line. The models included the following terms; average wind speed for the unit (avws), the length of the unit (len), average AT, average plane SST, or average satellite SST (avat, avsst1, or avsst2), block, and year.

1. Model with AT	df	Sum of Sq	Mean Sq	F Value	Pr(F)
poly(avws, 3)	3	12.6056	4.20187	1.988666	0.114869
poly(len, 3)	3	4.4362	1.47872	0.69985	0.552513
poly(avat, 3)	3	48.6746	16.22488	7.678931	5.13E-05
block	4	66.4681	16.61703	7.864532	3.9E-06
year	7	55.7002	7.95717	3.76598	0.000554
Residuals	458	967.7119	2.11291		
2. Model with plane SST	df		Mean Sq		Pr(F)
poly(avws, 3)	3	20.9661	6.98869	3.330433	0.019806
poly(len, 3)	3	16.388	5.46267	2.603214	0.051916
poly(avsst1, 3)	3	28.2784	9.42614	4.491988	0.004166
block	4	49.8749	12.46872	5.941918	0.000125
year	7	55.2964	7.89949	3.764469	0.000607
Residuals	332	696.6799	2.09843		
3. Model with satellite SST	df	Sum of Sq		F Value	Pr(F)
poly(avws, 3)	3	13.5319	4.51063	2.034763	0.108378
poly(len, 3)	3	5.203	1.73433	0.782362	0.504274
poly(SST, 3)	3	34.7553	11.5851	5.226082	0.001498
block	4	52.8534	13.21334	5.960589	0.000113
year	7	40.972	5.85315	2.640378	0.011128
Residuals	418	926.616	2.21678		

Appendix 6.5C: Variation in weather conditions/survey operations from year to year

In all long-term environmental surveys there are year-to-year differences in weather conditions. In this survey for example, year-to-year differences in SST, wind-speed, air-temperature and so on are expected. A number of the key environmental variables in the aerial survey (wind speed, swell, high cloud) are subjective estimates made by the spotting crew. It is therefore not clear whether differences between planes in any year are due to differences in the crew, differences in the areas flown, or true differences in the survey conditions experienced. Year-to-year differences in instrumentation also mean that differences between instruments may be the cause behind apparent variation. Average conditions in the key weather conditions for each year support the view that interannual environmental variation occurs in the GAB.

Table C1. Average environmental conditions for the aerial survey for the years 1993-2000. The median conditions when SBT were detected in all years is also provided. The environmental conditions are matched for each quarter line in the survey.

	Sighting conditions	1993	1994	1995	1996	1997	1998	1999	2000
Estimated by the pilot									
Wind-speed	3.2	3.8	3.8	4.0	4.4	4.3	3.7	3.9	4.3
Swell	1.0	1.0	1.0	1.7	2.0	1.6	2.0	1.0	0.1
High cloud	0.9	1.0	1.7	2.7	1.4	0.7	0.4	2.4	2.5
Measured on the plane									
Air Temperature	26.3	25.0	26.0	16.1	24.3	27.6	23.1	21.9	28.8
Plane SST	19.8	20.8	19.3	18.0	18.3	20.6	18.5	20.0	21.4
Measured remotely									
Satellite SST	20.6	20.4	20.0	20.0	19.8	21.1	21.1	20.3	20.6

Appendix 6.5D: analysis of variance and deviance tables including oceanographic variables

Effect of including SST gradient and moon-phase terms in models

The SST gradient (SSTgrad, based on satellite SST) and moon-phase were not significantly associated with the presence/absence of SBT (**Table D1)**. In the alternative model, SST gradient is not significantly associated with conditional biomass, however, moon-phase is significantly associated with conditional biomass (**Table D2)**. Accordingly, SST gradient was dropped in the subsequent models.

Effect of including CHL and CHL gradient in the models

To include the variables CHL and CHL gradient in the models requires restricting the data to that from the years 1998-2000. It is thus important to compare the results of the same model for the period 1993-2000 with the abbreviated time period. This comparison is seen in **Table D3 and Table D4,** which are the baseline models including moon-phase, satellite SST, average wind speed (avws), line length (len), average swell (avsw) average wind direction (avwd).

When CHL and CHL gradient are included in the models (**Table D5 and Table D6)** the majority of the signal is seen in the presence/absence model, with little change in the biomass conditional on presence model, which suggests that there is density dependence operating at a large scale.

Table D1. Presence/absence analysis of deviance results based on data for 1998-2000 and including the terms satellite SST (SST), SST gradient, moon-phase, average wind speed (avws), line length (len), average swell (avsw), pilot, average wind direction (avwdir), block, inshore/offshore, year, and two interaction terms.

	df	Deviance	Resid. df	Resid. Dev	Pr(Chi)
NULL			2165	2183.728	
poly(avws, 2)	2	145.606	2163	2038.122	0.00000
poly(len, 3)	3	70.1361	2160	1967.985	0.00000
poly(SST, 3)	3	21.3429	2157	1946.643	0.00009
poly(SSTgrad, 3)	3	5.3321	2154	1941.31	0.14903
poly(moon.phase, 4)	4	2.1378	2150	1939.173	0.71043
avsw	1	17.6889	2149	1921.484	0.00003
pilot	3	18.7352	2146	1902.748	0.00031
avwdir	5	12.231	2141	1890.518	0.03176
block	4	67.1758	2137	1823.342	0.00000
IO	1	1.4999	2136	1821.842	0.22069
year	7	12.8173	2129	1809.025	0.07669
Block*IO	4	34.9751	2125	1774.05	0.00000
Block*year	28	41.3469	2097	1732.703	0.04990

Table D2. Conditional biomass analysis of variance results for a model based on data for 1998-2000 and including the terms satellite SST (SST), SST gradient, moon-phase, pilot, average wind speed (avws), and year.

	df	Sum of Sq	Mean Sq	F Value	Pr(F)
block	4	30.5477	7.63693	3.61315	0.00657
poly(SST, 2)	2	50.7358	25.36788	12.00195	0.00001
pilot	3	47.1698	15.72326	7.43893	0.00007
poly(SSTgrad, 3)	3	10.1974	3.39914	1.60819	0.18686
poly(moon.phase, 4)	4	22.2505	5.56263	2.63177	0.03392
poly(avws, 3)	3	22.4514	7.4838	3.5407	0.01476
year	7	19.7569	2.82241	1.33533	0.23192
Residuals	412	870.8222	2.11365		

Table D3. Presence/absence model analysis of deviance results based on data for 1998-2000 and including the terms satellite SST (SST), moon-phase, average wind speed (avws), line length (len), average swell (avsw), pilot, average wind direction (avwdir), block, inshore/offshore, year, and two interaction terms.

	df	Deviance	Resid. df	Resid. Dev	Pr(Chi)
NULL			577	550.8301	
poly(avws, 2)	2	29.26994	575	521.5602	4E-07
poly(len, 3)	3	6.38235	572	515.1779	0.09442
poly(SST, 3)	3	15.09904	569	500.0788	0.001734
poly(moon.phase, 4)	4	3.59958	565	496.4792	0.4629
avsw	1	0.69816	564	495.7811	0.403402
pilot	2	1.61933	562	494.1617	0.445008
avwdir	4	6.55936	558	487.6024	0.161088
block	4	18.21191	554	469.3905	0.001122
IO	1	3.35929	553	466.0312	0.066827
year	2	0.52559	551	465.5056	0.768898
Block*IO	4	16.80594	547	448.6996	0.002108
Block*year	8	6.46148	539	442.2382	0.595684

Table D4. Conditional biomass analysis of variance results for a model based on data for 1998-2000 and including the terms satellite SST (SST), moon-phase, pilot, average wind speed (avws), block and year.

	df	Sum of Sq	Mean Sq	F Value	Pr(F)
block	4	43.4642	10.86604	4.721341	0.001691
poly(SST, 2)	2	11.6413	5.82063	2.529089	0.085514
pilot	2	11.0011	5.50057	2.390022	0.097552
poly(moon.phase, 4)	4	6.9557	1.73893	0.755573	0.556983
poly(avws, 3)	3	8.7735	2.92451	1.270714	0.289411
year	2	4.1687	2.08435	0.905657	0.40801
Residuals	88	202.5296	2.30147		

Table D5. Presence/absence analysis of deviance results for the same model shown in Table D3 but including the variables CHL and CHL gradient.

	df	Deviance	Resid. df	Resid. Dev	Pr(Chi)
NULL			324	341.22	
poly(avws, 2)	2	13.68429	322	327.5357	0.001068
poly(len, 3)	3	5.96816	319	321.5676	0.11317
poly(SST, 3)	3	4.02544	316	317.5421	0.25873
poly(moon-phase, 4)	4	8.05099	312	309.4911	0.089728
poly(CHL, 3)	3	11.51332	309	297.9778	0.009251
poly(CHLgrad, 3)	3	2.0651	306	295.9127	0.559006
avsw	1	1.48382	305	294.4289	0.223177
pilot	2	3.40315	303	291.0257	0.182396
avwdir	4	5.74069	299	285.2851	0.219369
block	4	12.0113	295	273.2738	0.017267
IO	1	0.14731	294	273.1265	0.701123
year	2	3.46176	292	269.6647	0.177129
Block*IO	4	13.25534	288	256.4093	0.010093
Block*year	8	7.66215	280	248.7472	0.467148

Table D6. Conditional biomass analysis of variance results for the same model shown in Table D4, but including the variables CHL and CHL gradient.

	df	Sum of Sq	Mean Sq	F Value	Pr(F)
block	4	38.28796	9.571989	5.047765	0.001815
poly(SST, 2)	2	12.13787	6.068937	3.200439	0.049764
pilot	2	9.4476	4.723801	2.491085	0.093696
poly(moon.phase, 4)	4	11.40665	2.851662	1.503817	0.216329
poly(CHL, 3)	3	6.07466	2.024886	1.067818	0.371791
poly(CHLgrad, 3)	3	2.40027	0.800091	0.421926	0.738134
poly(avws, 3)	3	16.244	5.414668	2.855412	0.047037
year	2	7.26255	3.631277	1.914945	0.158672
Residuals	47	89.12529	1.896283		

Appendix 6.5E: Analysis of trends in abundance of juvenile SBT in the GAB.

The final presence/absence and conditional biomass models were tested for a linear trend in years. The trends were considered using the full period of the survey (1993-2000) (**Table E1 and Table E3)** and for the period excluding 1999-2000 when trainee spotters were used (**Table E2 and Table E4)**. The conclusions of this analysis are discussed in the main body of **Section 6.5**.

Table E1. Analysis of deviance results for the presence/absence model with year fitted as a linear term and not a factor.

	df	Deviance	Resid. df	Resid. Dev	Pr(Chi)
NULL			2165	2183.728	
poly(avws, 2)	2	145.606	2163	2038.122	0.00000
poly(len, 3)	3	70.1361	2160	1967.985	0.00000
poly(SST, 3)	3	21.3429	2157	1946.643	0.00009
avsw	1	18.5684	2156	1928.074	0.00002
pilot	3	18.6492	2153	1909.425	0.00032
avwdir	5	12.7481	2148	1896.677	0.02586
block	4	67.7534	2144	1828.923	0.00000
IO	1	1.6876	2143	1827.236	0.19391
year	1	6.8129	2142	1820.423	0.00905
Block*IO	4	33.8851	2138	1786.538	0.00000
Block*year	4	7.0642	2134	1779.474	0.13253

Table E2. Analysis of deviance table for presence/absence model with year fitted as a linear term and not as a factor using data from 1993-1998 only

	df	Deviance	Resid. df	Resid. Dev	Pr(Chi)
NULL			1873	1887.094	
poly(avws, 2)	2	127.7488	1871	1759.346	0.00000
poly(len, 3)	3	83.967	1868	1675.379	0.00000
poly(SST, 3)	3	12.6168	1865	1662.762	0.00554
avsw	1	25.3831	1864	1637.379	0.00000
pilot	2	6.9799	1862	1630.399	0.03050
avwdir	5	13.1127	1857	1617.286	0.02234
block	4	62.4881	1853	1554.798	0.00000
IO	1	0.971	1852	1553.827	0.32442
year	1	6.8667	1851	1546.96	0.00878
Block*IO	4	34.2257	1847	1512.735	0.00000
Block*year	4	12.3757	1843	1500.359	0.01477

Table E3. Analysis of variance results for the conditional biomass model using data from all years.

	df	Sum of Sq	Mean Sq	F Value	Pr(F)
block	4	30.5477	7.63693	3.54699	0.00733
poly(SST, 2)	2	50.7358	25.36788	11.78216	0.00001
pilot	3	47.1698	15.72326	7.30269	0.00009
poly(avws, 3)	3	18.8578	6.28593	2.91951	0.03387
year	1	11.5633	11.56334	5.37062	0.02095
Residuals	425	915.0572	2.15308		

Table E4. Analysis of variance table for the conditional biomass model using data from 1993-1998

	df		Mean Sq	F Value	Pr(F)
block	4	29.2321	7.30803	3.50789	0.00794
poly(SST, 2)	2	49.7307	24.86537	11.93551	0.00001
pilot	2	19.8804	9.94022	4.77136	0.00900
poly(avws, 3)	3	21.2835	7.09451	3.4054	0.01782
year	1	7.1658	7.16582	3.43963	0.06445
Residuals	366	762.4912	2.08331		

6.6 Further considerations on the analysis and design of aerial surveys for juvenile SBT in the Great Australian Bight

Mark V. Bravington

Abstract

The SBT aerial survey from 1993-2000 is re-analyzed, and the potential precision as an index of relative abundance is investigated. The most important limiting factor in analysis, is differences between the observers. However, provided sufficient attention is given to observer intercalibration, through protocols such as using pairs of observers and through appropriate analyses, then it should still be possible to construct a usefully precise index of GAB abundance without much further work. As time goes by and more data are collected, it is in principle possible to retrospectively improve the precision of past estimates by at least 5 percentage points (on a CV scale), reducing annual CVs to under 30%. Issues relating to the design and analysis of future surveys are discussed.

Introduction

Standardised aerial surveys for SBT in the GAB took place annually from 1993 to 2000, and a partial survey resumed in 2002. A number of specific experiments have also been conducted during the course of the surveys, to study particular operational aspects. The ultimate goal of this work has been the construction of an index that is proportional to the abundance of juvenile SBT in any year. The initial hope was to produce an age-disaggregated index, but this has proven infeasible and attention has shifted to the feasibility of producing a simpler composite index across age groups 2, 3, and 4.

Over this long period, a number of issues have arisen concerning how best to analyse these particular data, and over the scope and role of any future SBT aerial surveys. While many questions have been answered through experiment and detailed analysis, there are still a number of unresolved points (e.g. those in Cowling, 2001a). Even though many years have now passed since a stable design has been implemented, there is still no finally-agreed way to produce a numerical index for use in assessment. Several approaches have been suggested (Research, 2000; Cowling & Laslett, 2000; Cowling, 2001b), but it remains unclear how much accuracy we could realistically expect from any index (in the sense of estimation uncertainty year-to-year). The main point of this report is to reconsider the feasibility and, assuming that a reliable index can be produced from the historical, to consider how best to proceed in future. The historical survey design is expensive, and relies on having available two planes and several observers/pilots over a wide area across the full extent of the fishing season. Logistical constraints may preclude running future surveys in the same way as past surveys. The questions here are whether the design can be made more efficient, what use can be made of other data to assist interpretation, and whether indices of abundance based on modified designs can be made compatible with indices based on the historical data.

The bulk of this report describes the results of analyzing separately three components of biomass (Biomass per Patch, BpP; Patches per Sighting, PpS; Sightings per Mile, SpM), examining the current and potential precision of each. Imprecision (aka estimation uncertainty), usually measured as a coefficient of variation CV, arises from two sources: (i) intrinsic variability in the data (e.g. that the mean value of some quantity is estimated from a limited number of observations each year); and (ii) the possibility that "nuisance" covariates such as SST rather than "interesting" covariates such as year, may be responsible for an observed effect. In any given year, the former can only be addressed by increasing the sample size. On the other hand, uncertainty due to nuisance covariates can potentially be reduced over time in two ways: using data from many years to pin down the real effect of that covariate, and using data separate from the main survey design, such as calibration experiments on observers, and archival tagging of fish. In this report, I have tried to show roughly how much uncertainty comes from each of the two sources, and therefore how much potential for improvement in CV there is without changing historical levels of effort. The discussion accompanying this section raises several issues pertinent to any continued or modified aerial survey in the GAB, and to the role of recruitment indices in stock assessment and management.

This report inevitably draws upon, and covers much of the same ground as, earlier aerial survey reports (see References), because the underlying data is the same even though the focus is slightly different. Many of the conclusions reached already on the basis of extensive experiments and analysis, are taken as implicit in this report: for example, that fish size estimates are simply not useful, and that satellite SST is the most appropriate temperature covariate to include.

General issues of SBT aerial survey analysis

Aerial and ship-based surveys are a well-accepted method of estimating abundance for other marine species such as whales and dolphins. There is an extensive set of statistical methods ("line transect theory") that has been developed specifically to handle aerial-survey-type data, dealing with such issues as how sighting rates change with distance from the trackline. In the case of the SBT aerial survey, though, there are three main complications which prevent the straightforward application of textbook methods, and which make analysis generally difficult.

1. SBT aren't always at the surface, and the proportion of time for which they are visible from an airplane appears to vary widely depending on environmental conditions.

2. The notion of "a sighting" of SBT is imprecise. A visible patch of SBT is easy to define, but a loose aggregation of patches may be classed as one sighting or as several, depending on how far it is from the flight path. This plays havoc with conventional line transect analysis.

3. Observers may differ widely in their ability to see patches under different conditions, and/or in their estimates of biomass. Unless the exact same set of observers and pilots is used over time (even then, assuming that individuals don't change their behaviour), then there is a risk that estimates of abundance will be driven by who is doing the spotting, rather than by how many fish are really out there.

All three concerns can also be problematic for ship and aircraft surveys of whales and dolphins. However, the surfacing behaviour of whales and dolphins is much more predictable (unlike tuna, they need to breathe) and is usually easier to account for in analysis. Dealing with point 3 is generally hard, but part of the usual protocol of such surveys is to rotate numerous observers operating in pairs or larger teams, to minimise the risk (and perhaps to allow calibration of observer effects based on after-the-fact analysis of number of animals seen). Although efforts were made to do this as far as possible in the GAB SBT survey, logistics prevented a fully robust solution. For example, in 1999 and 2000, only new trainee spotters were used, and half the flights used a new-to-the-survey pilot (albeit commercially experienced). Sighting rates from the trainee spotters turned out to be much lower, and there was too little data to allow reliable calibration of individual effects. Cowling, 2000 shows that survey indices from those two years depend substantially on what assumptions are made about sighting rates where the trainees were involved.

NOTE: in this report, “observer” means either spotter or pilot. All names have been changed to codes to protect anonymity.

Design- vs. model-based analysis

The Port Lincoln workshop in 2000 (Research, 2000) considered these points, and made several general recommendations about how to proceed analytically. The basis of line transect theory is that sighting rates tend to decline with distance from the transect (i.e. flight path), but in fact there is no evidence of this for aerial survey, at least for distances up to 6 nm. To avoid complications associated with point 2 above, the workshop therefore suggested a strip-transect approach instead; this means, in effect, that the index should simply “add up” all observed biomass within 6 nm either side of the flight path. To deal with points 1 and 3, the workshop recommended developing model-based rather than design-based estimators (see below), which would take into account environmental and observer factors, estimate their effects on observed biomass, and produce corrected indices that make allowance for the conditions in particular years. The workshop did not make detailed recommendations about exactly what model-based analysis to try.

Design-based analyses attempt to deal with environmental and observer effects by ensuring balanced sampling in space and time. The effects of position and time of year on distribution and sightability are not explicitly considered, since they will be matched between different years and won’t distort overall year-to-year comparisons. However, in practice it is impossible to fix environmental covariates such as sea surface temperature, so most design-based analyses need to include a certain model-based element.

Model-based analyses attempt to describe abundance and sightability specifically in terms of space and time factors, as well as observer and environmental effects. The approach taken in this paper is to decompose "biomass per length of trackline" into three components: biomass per patch (BpP), number of patches per sightings (PpS), and number of sightings per nm (SpM) within a 6 nm-wide strip. In terms of the three-stage model in this paper

$$\text{true local biomass per } 12 \text{ nm}^2 = \text{try local BpP} \times \text{true local PsS} \times \text{true local SpM} \qquad (1)$$

at any given moment, and the total biomass is just the sum of local biomasses across the whole area of interest.

There are various different ways that the data could be decomposed: for example, Cowling, 2000 uses biomass conditional on fish presence per trackline fragment, together with fish presence. One advantage of the three-stage decomposition used here, is that it may be easier to link each component to specific aspects of tuna biology or observer behaviour. Also, this decomposition can be applied at any spatial scale, without requiring data to be aggregated or averaged over different fractions of a transect line; this may also help links to biological data, e.g. on surfacing behaviour.

Statistically, the three components BpP, PsS and SpM can be analyzed independently. Any or all can be affected in different ways by position, time (day-of-year as well as year itself), and other covariates (observer, temperature, wind speed, etc.) as appropriate. Some covariates will genuinely affect local abundance, others will affect sightability (the probability of a particular tuna being seen if it is within the visibility range of the plane), which in turn reflects both tuna behaviour (at surface or not) and observational limitations. The general aim is to estimate *true* local BpP (or PsS or SpM) based on local observed BpP, taking into account those environmental and observer covariates that will affect sightability. The simplest approach is to predict what local BpP etc. would have been if all sightability covariates had been set to some standardised values: no cloud cover, no glare, one observer of average skill, etc. Finally, to form an overall index for a particular time-of-year, estimates of the three components of biomass at a particular place and time are first multiplied together, and then added up across all places and times according to the above formula. To keep down complexity in this report, I have restricted attention to models which enforce the same spatial pattern or movement pattern across different years. It is certainly possible to fit more complex models e.g. with year-dependent migration timing, but issues of model selection and appropriate levels of complexity become quite intricate.

Model-based analyses can be used to investigate how the overall uncertainty in an index is affected by uncertainty in the individual components. For example, it may turn out that the main reason for high uncertainty in a particular year's index, is that a new spotter was used in that year only; or that sea surface temperatures were particularly low, and that only limited data from other years were available to show how this could affect an index. It is also possible to check which factors are limiting precision: is the main limitation simply the number of sightings, or is it uncertainty about the average biomass of a patch? All this information can be useful in refining survey design, figuring out how much survey effort is necessary, and designing other experiments (such as archival tagging) which may shed light on the role of particular covariates.

It is important to note that, in model-based analyses, estimates of "interesting" quantities such as abundance in a particular year, can be partly confounded with estimates of "nuisance" quantities such as the effect of depth on tuna surfacing behaviour. Because the data can to some extent be explained equally well by varying either the estimated depth effect or the abundance estimate, any uncertainty in the estimated depth effect will spill over to create uncertainty about abundance. However, there is a positive aspect to this too: as time goes by, more and more data will accumulate allowing the depth effect to be pinned down increasingly tightly (assuming this is a consistent property of tuna behaviour), and this should allow *retrospective* improvement of historical abundance estimates. In this report, I have used this idea to try to set limits on what precision could ultimately be obtained from an aerial survey if a really long time-series, or additional non-survey information on environmental/observer effects, became available.

Model-based analyses build up a picture of spatial distribution that automatically takes account of variations in the underlying effective survey effort. Therefore, provided that fundamental protocols do not change, it is in principle possible to change the design of the survey and still obtain indices that can be meaningfully compared with historic indices. However, the role of good survey design cannot be overemphasised, even if model-based analyses are ultimately used. A poorly-balanced design — e.g. where one spotter only observes in the west, while the other only observers in the east — makes it impossible to carry out useful model-based analysis, because there is no way to distinguish genuine distributional differences from observer differences. Generally speaking, the more unbalanced the design, the more uncertainty there will be in the estimated indices. Exactly the same problem is encountered when trying to standardize CPUE indices for use in stock assessment, for example, where badly-unbalanced data is commonplace. The ensuing difficulties are familiar and often unfixable.

Although model-based analyses of the SBT aerial survey have numerous benefits, and in fact are probably unavoidable at least to some extent, it is important not to gloss over the difficulties. Perhaps the biggest is *model selection*. It is not always obvious *a priori* which covariates ought to be included in models, and the resulting index can be sensitive to which covariates are used; for example, Cowling (2000) obtained substantially different results just by using two different measurements of sea surface temperature. A related question is how much flexibility to allow in each covariate: is it reasonable to permit an estimated SST effect on sightability which first rises, then falls, and then rises again as SST increases? While there are some statistical tools available which help make these decisions, it is a big ask to rely on automatic model selection procedures to do the right thing when faced with a large range of covariates.

Finally, there is an important issue about whether each covariate affects sightability or abundance. Sometimes a covariate might affect both: for example, if depth were to affect (i) productivity and thus how likely a tuna is to be in the area, but also (ii) the amount of time a tuna spends below the surface. If a covariate is to appear in a model-based analysis, it is crucial to have an *a priori* judgment of whether it is a sightability covariate or an abundance covariate. Otherwise, it is impossible to know how to deal appropriately with the covariate when producing a standardised index. In the example of depth, for example, any change in SBT depth distribution from one year to the next would make it impossible to interpret unambiguously any changes in apparent abundance. If these issues are thought to be potentially serious, the resolution cannot come through analysis of survey data alone; it will be necessary to use results from other biological studies.

Methods and results of model-based analyses

There are four sections in this part of the report: one each for the analyses of BpP, PpS, and SpM, and one describing how the three are combined to produce an overall abundance index.

The single most important issue with model-based (or any) analysis of the SBT aerial survey, is how to estimate observer effects. Because observers flew in pairs between 1993 and 2000, essentially providing two comparative measures of the same quantities, a direct calibration of observers can be made from the pair-wise data without needing to account for other covariates. This gives *a priori* estimates of observer effects for BpP and SpM, independent of the main model-based analyses. Therefore, the BpP and SpM sections each contain several subsections: the first describes inter-observer studies based on observer pairs, while the remainder describe the main model-based analysis, including how the *a priori* estimated observer effects are incorporated.

Biomass per patch

Until 1998, the pilot and spotter both made individual assessments of the size of each patch that was detected. It is thus possible to estimate observer effects (by comparing pairs of assessments), as well as systematic effects of location, year, day of year, conditions, etc., on average patch size (by allowing for observer effects). In theory, it would be possible to do both analyses in one step. One way to describe the data, would be through a random-effects model that has a "patch effect" as well as an "observation effect". Unfortunately, with over 6000 patches in the data, current software simply cannot cope. It is necessary to run two simpler analyses, and combine the results. The two-stage approach actually sacrifices very little information; however, it does add some complexity when the two stages need to be combined.

To set up notation for describing the BpP analyses, it is helpful to consider what the conceptual form of an all-in-one random-effects model would be:

$$B_{it} = b_t \times \varepsilon_{it} = B\left(x_t;\theta\right) \times \eta_t \times \varepsilon_{it} \tag{2}$$

Here, B_{it} is observer i's estimate of the biomass of the t^{th} patch; b_t is the true biomass of that patch; ε_{it} is the observer's measurement error for that patch; x_t describes the time, place, and conditions of sighting; $B\left(x_t\right)$ describes the expected patch size at x_t as a function of unknown parameters θ; and η_t is the deviation of that patch from its expected value. In this multiplicative framework, η and ε are both random variables with mean value 1.

Direct calibration of observer effects

Estimates of "observer effect" can be made by direct pair-wise calibration of these assessments, so that we can say that observer X's biomass estimates are systematically 15% larger than observer Y's estimates of biomass. Provided that there is sufficient exchange of spotters and/or pilots across planes, it is in principle possible to inter-calibrate biomass estimates for all observers. A simple statistical model for this is

$$B_{it} = b_t \beta_i \varepsilon_{it} \tag{3}$$

where β_i is the observer effect. If observer j also assesses this patch, we have

$$\log\left(B_{it}/B_{jt}\right) = \log\beta_i - \log\beta_j + \eta_t \tag{4}$$

where $\eta_t = \log\varepsilon_{it} - \log\varepsilon_{jt}$ and all the η_t are independent. If one observer is chosen as the reference level with $\log\beta \equiv 0,$ then a linear model can be constructed with an appropriate design matrix of ±1's and 0's to explain the mean log-differences across all patches seen by a given pair of observers. Results are shown in **Table 1**.

Table 1: Observer effects on patch biomass, estimated by direct calibration. All effects are relative to A=100%.

Observer	Effect %	CV %
A	(100)	(0)
B	99	2.6
C	96	3.1
D	82	4.5
E	97	1.9

The only observer with substantially different estimates is D, whose estimates are on average almost 20% lower than the others. A 95% confidence interval on D's effect would fall well short of equality with the other observers. Overall, the model is only marginally significant compared to a null model of no observer effect, but this is presumably because 4 degrees-of-freedom are used while 3 of the spotters have almost identical performance. Note that the estimates have low CVs, even though the fit is based on only 8 combinations of observer and pilot. Direct calibrations such as this are effective because there are no confounding covariates.

Unfortunately, only these five observers could be inter-calibrated in this way, because the other five only joined the survey in 1999 and 2000, when there was just a single assessment of each patch's size.

It is reassuring that four out of five experienced observers appear to behave so similarly. Given that (pre-1999) each patch biomass was assessed twice, and the final "working biomass" of each patch is therefore an average of two corrections, the greatest correction required is only about 10%.

Preliminary model-based analysis

The observer effects estimated in **Table 1** can be used as prior information in a more general model of BpP that incorporates environmental covariates. A full development of this approach is given in the next section. However, as a preliminary step to illustrate the effect of environmental covariates, it is possible to simply use the direct calibration estimates as fixed offsets in a log linear regression. The results in this subsection are illustrative only, and are not used in constructing an overall abundance index.

The response variable is now the mean log biomass estimate for each patch $\frac{1}{2}\left(\log B_{it} + \log B_{jt}\right)$, i.e. the log of the geometric mean . Only one observation per patch is used, compared to two in the model proposed in equation (3). For 1999 and 2000, many patch size estimates come from an observer who cannot be directly calibrated with the other experienced observers. For illustrative purposes, I arbitrarily set the observer effect to 100% for that observer; in fact, subsequent analyses indicate that this is probably not appropriate.

For spatial structure, I divided the region into 5 longitude and 3 latitude bands roughly corresponding to depth strata (**Figure 1**).

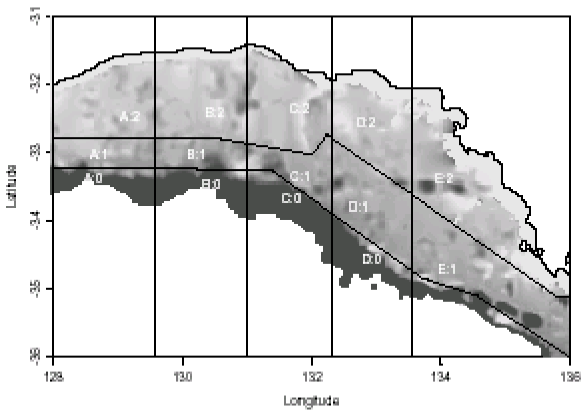


Figure 1: Area boundaries for stratification, with depth

Apart from spatial block, other variables considered for inclusion were: satellite sea surface temperature (SST), wind speed, wind direction, day-of-year, year. Based on stepwise AIC of a linear regression in log(BpP), the following explanatory covariates were chosen: year, longitude band, latitude band, and day-of-year (as a quadratic term). The full model and the reduced (AIC-selected) model show very similar estimates for the major effects (Figure 2), except perhaps for year in 1996-1997 where there is about a 10% difference. The fit improves significantly if the day-of-year effect is allowed to vary from year to year, but there has not been time to fully explore the implication of including such a term.

Confidence limits are fairly tight generally. The CV on year effect varies between about 5% and 8% depending on the year. The latitude (or depth-band effect) is quite weak, and much the strongest effect is the reduction in patch size in the easternmost longitude block. None of the terms shown could affect observer’s ability to estimate biomass, but latitude, longitude and day-of-year could all affect tuna behaviour as well as distribution.

There are a couple of points to bear in mind when interpreting **Figure 2**. First, a “year effect” does not correspond exactly to “mean log biomass in that year”. Rather, the year effect measures what was different about that year, after accounting for the other effects in the model. For example, if the spatial distribution of tuna is more westerly than average in a certain year, then it is likely that patches will be bigger anyway, because patches are generally bigger towards the west. The year effect measures only any changes above and beyond what would have been expected given the distributional shift. A fuller treatment of “year effect” is given in the following sub-section “Combining all the analyses”.

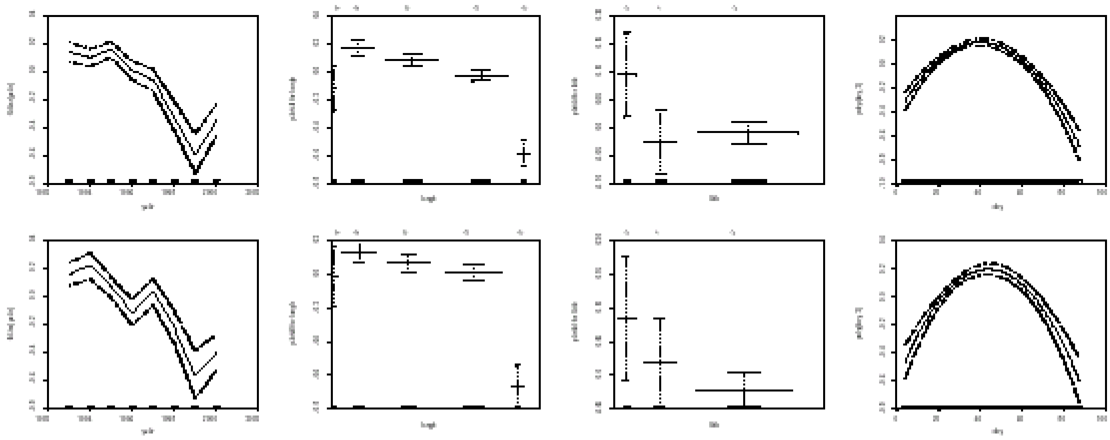


Figure 2: Estimated effects on biomass-per-catch of (by column): year, longitude, latitude, day-of-year, with 95% CI. Top row = AIC-selected model; bottom row=fuller model. Y-axis is log$_e$ scale; 0.1 units ~10% change.

Exactly this phenomenon occurred in 1998, when 40% of patch sightings were in longitude band B (shown in **Figure 1**); in other years, only 16% of patches on average were seen in that band. In fact, 1998 was unusual in other ways; the distribution of recorded patch sizes , although similar for smaller patches, has a much longer tail, with some huge patches being seen in that longitude band. This is apparent in **Figure 3**, which shows quantiles of patch biomass for 1998 plotted against the corresponding quantiles in other years. The log-biomass model underemphasizes the contribution of those patches to overall biomass-per-patch, and requires some refinement. This is provided by the indirect calibration GLM discussed in the next section.

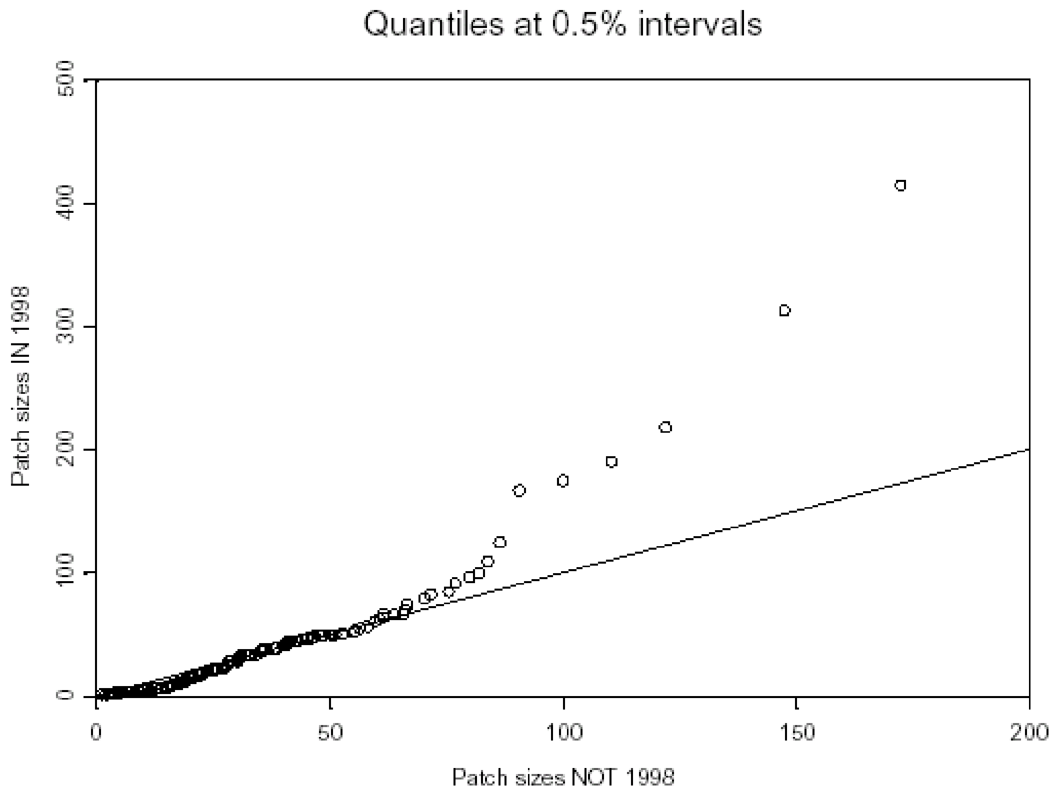


Figure 3: Patch sizes in 1998 vs other years.

These issues are not fundamental problems of modeling, but do highlight the need to be careful in building models, interpreting results and in producing "survey indices"; in particular, it is necessary to take account of real variations in spatial structure between years, rather than simply using "estimated year effects".

To investigate the limits to precision of this model, I bootstrapped the data based on the AIC-selected model and then re-fitted the model to each bootstrap realization but without re-estimating the coefficients for longitude, latitude, and day-of-year (i.e. keeping them fixed at their original estimates). This mimics the effects of having perfect information on "environmental" effects (and observer effects, since these are also kept fixed). The point of doing this is to show what might be attainable if a really long time series was available, or if other experiments could be used to fix environmental effects, rather than having to estimate them. The average improvement in year-effect CV is only about 10%, i.e. CVs improve from about 6% to about 5%. This is a tiny improvement, but the year-effect CVs are very small anyway.

Indirect calibration of observer effects

The model of the previous section needs to be modified to do three things: eliminate the bias caused by taking logs of biomass; become more flexible at capturing real changes in mean BpP that are local in space and time, e.g. those of 1998; and allow for the incorporation of observer effects from direct calibration. To do this, we can start with a modified model

$$\log \mathrm{E}\left[\overline{B}_t\right] = \frac{1}{2}\left(\log o_{i(t)} + \log o_{j(t)}\right) + \mathrm{lat}_t + \mathrm{long}_t * \mathrm{year}_t + \mathrm{DoY} + \mathrm{DoY\hat{}2} \qquad (6)$$

where $i(t)$ and $j(t)$ were the observers who assessed the t^{th} patch, and $o_{i(t)}$ represents the (multiplicative) effect of observer i. The o-values can either be fixed *a priori,* or estimated; the latter is "indirect calibration". Once the set of o_i that are to be fixed has been set at specific values, model 1 can be fitted by GLMs; a Gamma error model gives acceptable residual plot diagnostics. By fitting with many different sets of plausible o-values (as suggested by the direct calibration), and examining the best possible goodness of fit to the indirect calibration data given the current set of o-values, it is possible to formally synthesize the information in the direct and indirect calibrations. The technique is described in more detail in the next subsection; this subsection simply comments on the structure of model 1.

It is important to emphasize that "year effects" are not straightforward to calculate or interpret, especially when there are interaction terms involving year. As mentioned already, part of the reason is that the usual definition of a "year effect" filters out the effect of other variables which may be partly confounded with year, but which nevertheless have a genuine effect on patch biomass (rather than, say, an effect on observers' perception). When interaction terms appear, there is a further complication. The usual statistical definition of "year effect" when there is, say, a year*longitude interaction, is "the average across all longitudes in that year, of the corresponding year*longitude estimate". This is an unweighted average which does not take into account differences in the number of patches in each longitude band.

While there are sometimes good reasons for summarizing year effects as just described, it is certainly not appropriate to construct a biomass index based on such a calculation.

The most useful summary statistic is probably "estimated mean BpP in each year"; this is a weighted average reflecting that there are more patches in some strata than in others. The appropriate weights can be calculated based on the results from the PpS and SpM models, which together predict how many patches there really were in each stratum. This is the approach followed in the "Combining all the analyses" section.

Viewed in statistical isolation, model 1 is probably over-fitted (i.e. too many parameters); certainly, the model selected automatically by the AIC criterion does not include a year*longitude effect. Given the context, though, this over-fitting is less problematic than usual. For one thing, the incorporation of independent data from the direct calibration model helps to constrain all the parameters towards sensible values. Also, even though strata with very few observed patches will potentially have extreme and extremely imprecise estimates of BpP, the overall effect of these strata on annual indices will be small precisely because there were very few patches. It is the total biomass of all patches in the stratum that matters, so if patch density is low in one stratum, then the average size of those particular patches is not very important to overall abundance. Further, the original intention behind the aerial survey design, was to provide an unbiased time series based on empirical means within strata defined by latitude, longitude, and day-of-year block. Given this design, it is reasonable and perhaps desirable to use slightly-overfitted models that attempt as far as possible merely to correct observed stratum means for environmental/sighting effects, rather than trying to shrink of stratum means based on neighbouring values.

For the future, it would be worth trying a random-effect version of (6), with the year*longitude term (and possible some other interaction terms not tried here) modeled as a random effect. In essence, this would mean that estimates of mean BpP in strata where there are very few observations, would be "shrunk" back towards the mean (allowing for overall year and longitude effects). In principle, this would get round the problem of over-fitting. However, random effects would introduce extra complexity into the model synthesis discussed below. Further, although random-effect models can have markedly beneficial effects on estimation variance, they also induce some bias on a stratum-by-stratum basis unless the model is just right. It is not clear, for example, how a random-effect model would respond to the extraordinarily large patches seen in 1998 band B. Further investigation of the pros and cons of random-effect models for BpP would be useful.

Model compatibility and synthesis

It is important to check whether the direct and indirect calibrations are consistent. Inconsistency could potentially arise if, for example, observers on the same flight somehow influence each other's assessment of biomass; then the pair-wise comparisons would suggest less difference between observers than is really the case. Assuming that the models are shown to be consistent, it is also necessary to synthesize the two. The direct calibration on its own does not give sufficiently precise estimates of observer effects to simply fix these parameters before estimating year, space, and weather effects. There is also the issue of observer F, for whom direct calibration is impossible because of lack of overlap, but for whom indirect calibration is still feasible since B surveyed the same areas in the same years. It turns out that the same approximate Bayesian technique can be applied to both consistency-checking and model synthesis. The technical details are as follows.

A method for the approximate synthesis of two simple analyses

The goal of the method is to get a set of samples from the approximate posterior distribution $p_{\alpha\beta}(\alpha, \beta \mid d)$ of parameters α (in this case, observer effects that can be directly calibrated in a prior analysis) and β (in this case, all the environmental, spatial, and year effects), based on data d (in this case, only the indirect data). The direct calibration data is assumed to be already summarised in terms of a prior distribution π_α. The prior π_β on β is assumed flat for now. In order to avoid the technical difficulties associated with MCMC (e.g. convergence tests and choice of jump rules), and to avoid having to develop a combined model from scratch (which would preclude the use of standard model-fitting software), an importance-sampling approach can be used to get approximate samples of parameters from $p_{\alpha\beta}(\alpha, \beta \mid d)$.

Let $\Lambda(\alpha, \beta; d)$ be the log-likelihood of d. Suppose α is fixed at α^*, write $\Lambda^*(\beta) = \Lambda(d \mid \alpha^*, \beta)$ and let $\hat{\beta}^* = \hat{\beta}(\alpha^*)$ be the maximiser of $\Lambda^*(\beta)$. Now, the posterior distribution $p_{\alpha\beta}(.,. \mid d)$ can be approximated by taking a sample α^* from the marginal posterior $p_\alpha(. \mid d)$, and then a sample of β's from the conditional posterior $p_{\beta|\alpha}(. \mid \alpha^*, d)$. For the latter step, if $p_{\beta|\alpha}$ cannot be calculated exactly, samples β^{**} can be drawn from a known approximating distribution $\tilde{p}_{\beta|\alpha}$, and reweighted by $p_{\beta|\alpha}(\beta^{**}) / \tilde{p}_{\beta|\alpha}(\beta^{**})$. To sample from $p_\alpha(. \mid d)$, note that

$$p_\alpha(\alpha \mid d) = \frac{f(d \mid \alpha)\pi_\alpha(\alpha)}{f(d)}$$

We can sample α^* from π_α instead of p_α and re-weight the samples via an approximation to $f(d \mid \alpha^*)$. Again, if it is not possible to sample directly from π_α, samples can be taken from a convenient approximation and re-weighted. To approximate $f(d \mid \alpha^*)$, the adjusted profile likelihood of Cox (personal communication) can be used; if $I^*(\hat{\beta}^*)$ is the information matrix conditional on $\alpha = \alpha^*$, then for a flat prior π_β it can be shown that

$$\tilde{f}^* = f(d \mid \alpha^*) \approx e^{\Lambda^*(\hat{\beta}^*)} \times \left| I^*(\hat{\beta}^*) \right|^{-1/2}$$

Roughly speaking, this is a statistically consistent way of down-weighting those samples from the direct-data-driven prior π_α that provide a poor fit to the indirect data d.

Normally, sampling from a prior and then re-weighting in order to match a posterior, would not work very well; most of the weight would end up on a very few samples, because the posterior would be much more concentrated than the prior. In this case, though, the prior can reasonably be expected to carry much if not most of the information on α, and so the weights should be more evenly spread.

Results: compatibility

To check for consistency, 300 sets of possible observer effects were generated from the direct calibration using a weighted likelihood bootstrap (Newton & Raftery, 1994). For each set, the log-likelihoods of the direct calibration data and of the indirect calibration data (without F) were calculated, and plotted against each other in **Figure 4** The axis scores increase with increasing goodness-of-fit. Any y-value less than about 3 units below the maximum observed y-value, represents a bad fit for the indirect calibration data[3]; a similar interpretation applies to the x-values. Encouragingly, the graph shows that it is possible to get a relatively good fit to both datasets with the same set of observer effects, e.g. for the parameter values that correspond to the points near (-1,0). The feature to concentrate on, is the "missing triangle" in the top right-hand corner triangle, which shows the trade-off between getting a really good fit to one model but a worse fit to the other. In this graph, the "missing triangle" is fairly small (see footnote). If the datasets were inconsistent, the missing triangle would be bigger; or, worse, the diagonal boundary would be more L-shaped, with the only observer effects that fitted well for the direct data fitting poorly for the indirect data, and vice versa.

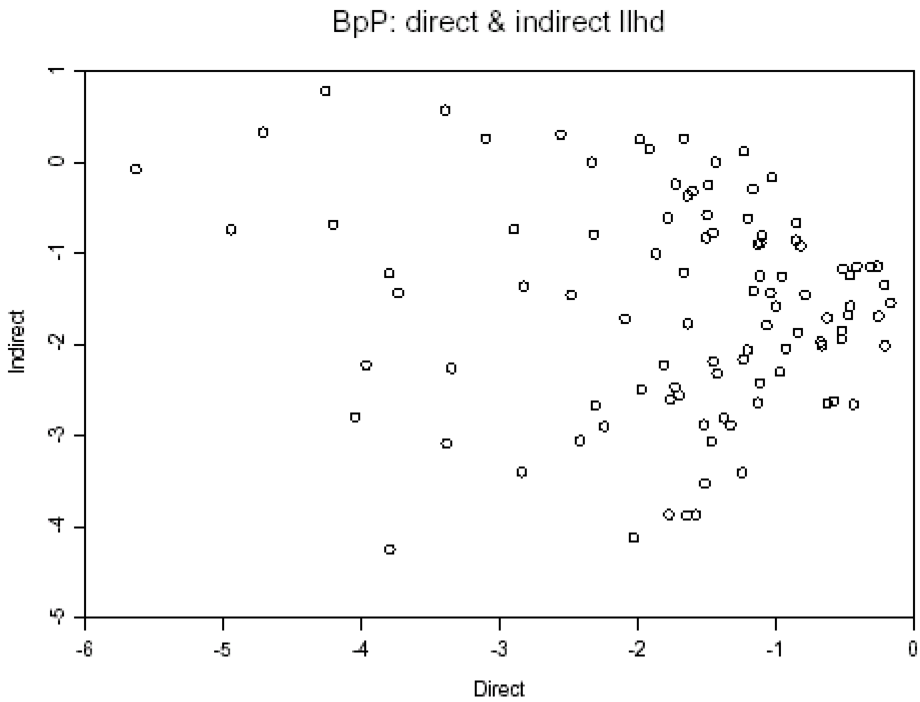


Figure 4: Log-likelihoods at different sets of parameter values, for direct and indirect BpP models.

The fact that the points are more spread out in the x-direction, reflects the higher precision obtainable from the direct calibration data, compared to the indirect calibration data. It is easier to find a set of observer effects that gives a good fit for the indirect calibration data, than it is for the direct calibration data; therefore the indirect data is less precise for calibration.

[3] Both models have four degrees-of-freedom for observer effects, because DH is always fixed at 0 for reference purposes. Wilk's χ^2 approximation to the likelihood ratio distribution with 4 DoF, puts the 80% confidence interval at almost 3 units of log-likelihood below the maximum. It is easy to get within 3 units of the maximum for one model while being close to the maximum for the other, so roughly speaking the models are at least "compatible at the 70% level".

Results: synthesis

For other environmental variables, mean estimated effects are close to those shown in Section 6.5. There is very little change in the estimated observer effects when the two models are synthesised; this is reassuring but not surprising, since the direct data are more informative (**Figure 4**). However, it does become possible to estimate an observer effect for the remaining experienced observer (F) who never flew with the other experienced observers, based on indirect comparisons with B who also flew in 1999 and 2000. The estimated effect for F is 20% higher than for most of the other observers (whereas the effect for D was 20% lower). The standard error on this estimate is 7%, higher than for the other observers (**Table 1**), but it is still interesting to note that the estimated range of observer effects is as high as 40%.

Summary of BpP results

On balance, model uncertainty and estimation uncertainty about BpP seems likely to make a small contribution to overall uncertainty of a survey index. Observer effects (for experienced observers; trainees were not used) can be moderately large, and can be estimated quite precisely *when pairs of observer fly together*. Calibration for F, who never flew with other experienced observers, remains imprecise. There are significant inter-year differences, after allowing for other variables. However, interpretation of "estimated year effects" is not straightforward; this point is revisited in the Combining all the analyses section, where a more meaningful BpP time series is developed.

It is interesting to recall the results of direct experiments in Cowling (2000, Appendix A), where biomasses of the same set of patches were estimated completely independently from two planes, as well as over varying sighting conditions. Detailed analyses of those data showed systematic differences between observers and between sighting conditions. However, different estimates of biomass for the same patch were clearly correlated, suggesting that meaningful estimates can be developed provided that calibration is possible. (Contrast this with the situation for fish size, where there is no relationship between different observers' estimates.) The models used in this section show how such calibration can be done.

Patches per sighting

Almost 50% of all sightings are of a single patch, and 90% consist of 7 or less, but 1.5% of sightings (15 in all) are "mega-sightings" with over 20 patches each. The two greatest mega-sightings had 76 and 53 patches. A crude test for year-to-year variations in the frequency of mega-sightings per year, is marginally significant ($p = 0.09$), with an indication of more mega-sightings in 1993–1996 than afterwards. More striking are the patterns with latitude and time of year: all but one of the mega-sightings are in the inshore band, and 45% occur in the first calendar fortnight compared to only 16% of all sightings.

In terms of biomass, 75% of total observed biomass comes from sightings with fewer than 15 patches (**Figure 5**). The cumulative curve for biomass lies above that for number of patches, showing that sightings with numerically *more* patches tend to have slightly *smaller* patches, but the effect is not large. An inverse-gaussian GLM, using as the covariate log(# patches) cut into four groups, gives only small variation in mean patch biomass ($\pm 12\%$) across the range of patches-per-sighting.

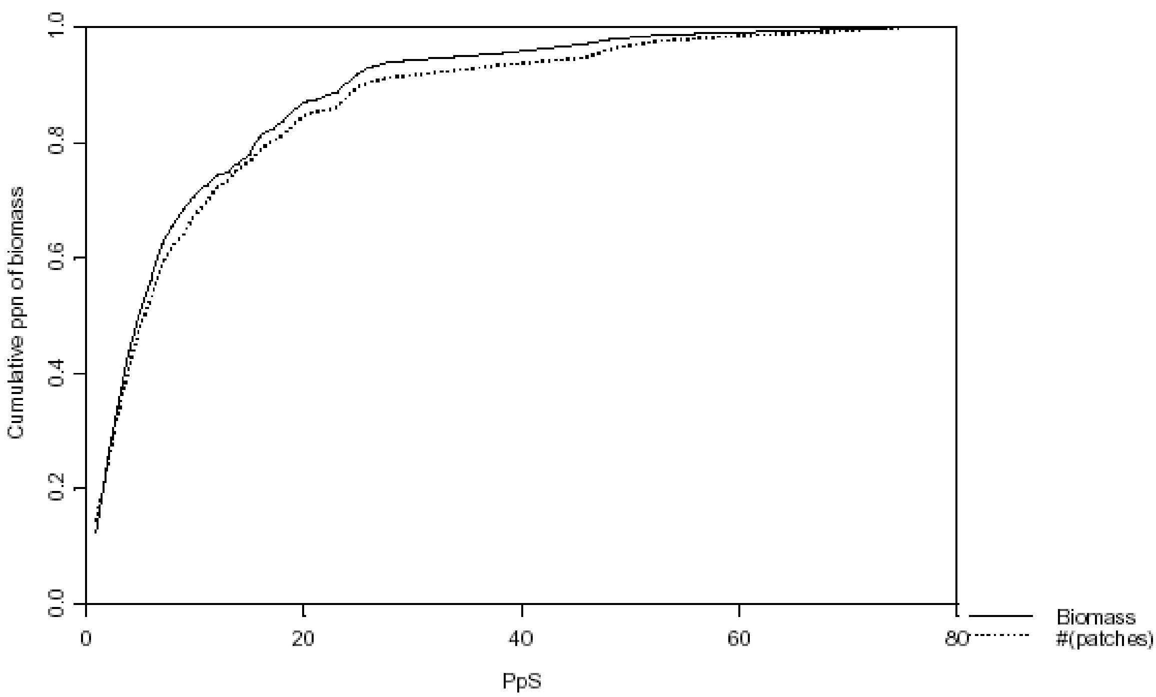


Figure 5: Cumulative proportions of biomass and number of patches, against patches-per-sighting.

Although sea surface temperature (SST) is expected to have some substantial effect *a priori* on sightability etc., about 30% of the sightings had no satellite SST estimate, mainly because of difficulties in interpolation when there is some cloud cover nearby. Cowling, 2002 concluded that aircraft-based measurements were not useful. Better interpolation schemes could ameliorate this difficulty in future; for the moment, though, the unacceptably high missing-data rate precluded the use of SST as a covariate for predicting PpS.

One difficulty in finding a good model, is that conventional model-choice diagnostics, such as AIC, may simply mislead if applied to such skewed data. Also, natural model extensions (such as random-effect models, to capture e.g. within-year variation in a parsimonious fashion) are suspect, because of the parametric assumptions involved. For the future, it would be worth developing models based on distributions that are more skewed than the exponential-family distributions available for GLMs. However, this would require fairly extensive work.

For the moment, it seems advisable to simply construct a richly-parameterised GLM that does a reasonable job of tracking the stratum means, while leaving enough observations within strata to allow meaningful estimation of nuisance covariates such as wind speed. The input to the overall abundance estimate, then consists basically of a nuisance-corrected stratum mean. Although no GLM performs well in terms of QQ plot diagnostics, an otherwise-reasonable pattern of residuals is obtained by fitting

$$(pcount - 0.8) \sim latb + longb * factor(year) + day * factor(year) + windspeed + \log(10 - depth) \quad (7)$$

using a Gamma GLM with a log link. (This is in fact the AIC-selected model, though the AIC choice is not to be trusted with such skewed data.) As shown in **Figure 6**, where each point is a fortnightly mean from a lat-long box, the PpS model does track the stratum means fairly well.

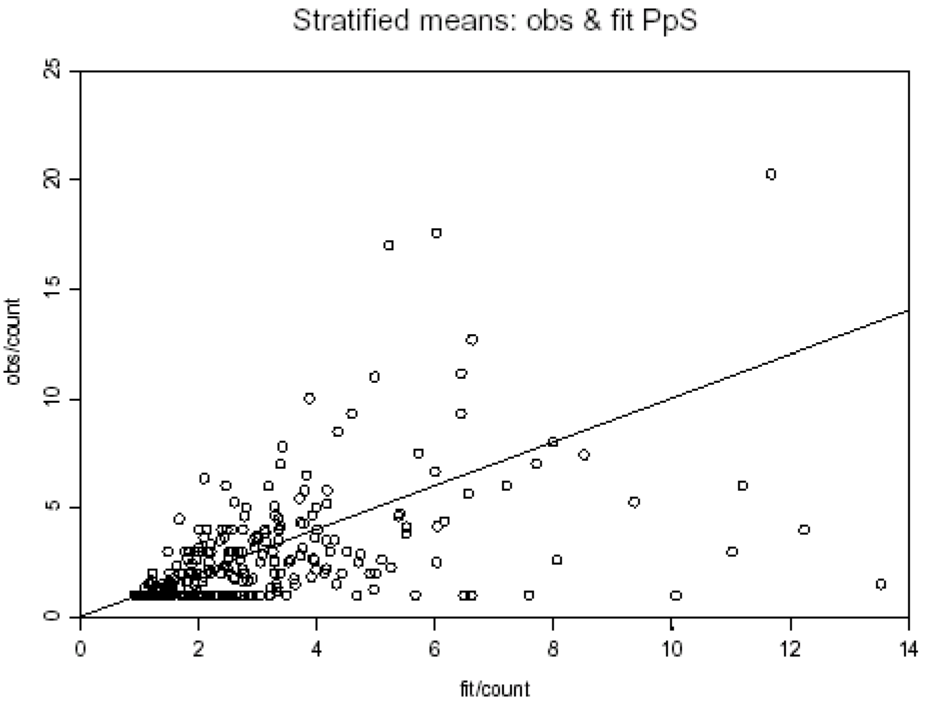


Figure 6:

Because of the interactions, the above model is rather hard to interpret. For descriptive purposes only, I also modeled $\log \text{PpS}$ by linear regression, using latitude, longitude, depth, day-of-year, SST, and wind speed. All else being equal, observed patches-per-sighting seems to increase with SST, decrease with wind speed, and decrease strongly with depth; all the really large sightings occur in the inshore region at relatively shallow depths. Estimated effects for the remaining terms are shown in **Figure 7**.

The most dramatic result from the regression, is the strong positive effect for 1996, with mean PpS being estimated at about 40% bigger than normal that year (SE about 10%). This is also apparent in the raw data (**Figure 8**). As with BpP, there is a caveat about interpreting “year effects” naively, because of possible interactions with distribution shifts. However, in this case the 1996 effect seems unambiguous.

Longitude effects are similar to those for BpP, with fewer patches per sighting in the far east and far west. However, the day-of-year effect shows the exact opposite to BpP: in the middle of the season, patches seem to be individually larger but less clustered, or at any rate with fewer of them to the sighting.

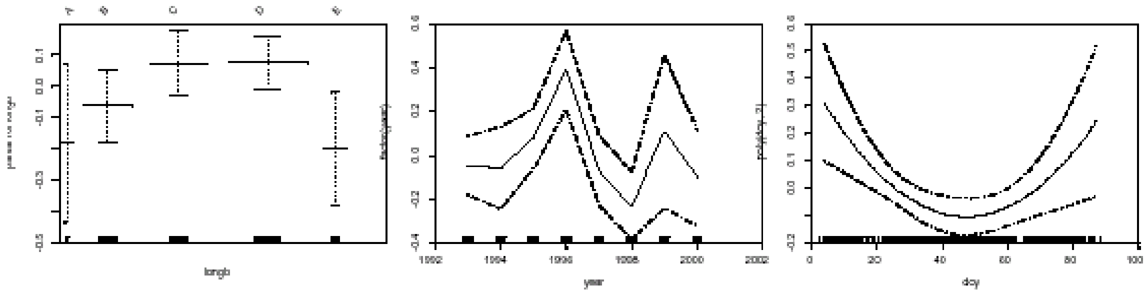


Figure 7: Estimated effects on log(#patches per sighting)

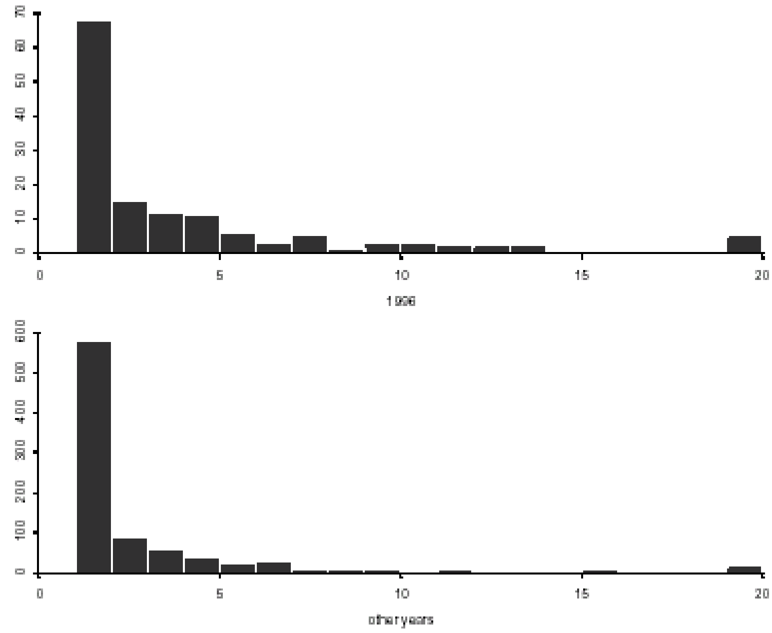


Figure 8: X-axis: number of patches in a sighting; Y-axis: number of sightings with that many patches.

Observer effects do not seem to be important for PpS. This is not surprising, since both observers are involved in counting patches once a sighting has been approached. Nevertheless, I tried using as a covariate the identity of the observer who first made each sighting. The point was to allow for the possibility that observers might vary in how they break up sets of patches into "sightings"; even if two observers are equally effective overall at seeing patches, one might have a lower effect for SpM and a higher effect for PpS. Sample sizes are too small to estimate fixed effects for all observers, but with a random-effects model, observer effects appear fairly small here (no more than about $\pm 5\%$).

The unmodeled skewness in the PpS data does not invalidate the predicted means, though it does suggest that more precise estimates could be made if a better underlying model could be found. However, the unmodeled skewness does imply that a bootstrap is the only way to get reliable CVs under this model. In fact, because there is no parametric likelihood available for this model (the GLM likelihood being inappropriate for such skewed data), I have used the bootstrap to generate an approximate posterior distribution of corrected stratum means from the PpS model, for subsequent incorporation in the overall abundance estimate.

Summary of PpS results

There is slightly more unexplained variability in the PpS model, than in the BpP model. In particular, standard errors on year effects are larger if year is used as a fixed effect, although CVs are still only around 9% on average. A random-effects model for year reduces the year effects by about 40% (so that the 1996 effect drops from 43% above normal to 25% above normal), but it is not clear what the effect on uncertainty is. The limits to precision of year effects is about 8%, based on the same approach as for BpP.

Experimental results in Section 6.1 discusses the consistency of patch counts from two planes. Overall patch counts in a high density area were similar, but inspection of position data showed that there were substantial discrepancies between the patches identified by the two planes. This does demonstrate that patch counting is an inexact science, but need not represent much of a problem for analysis of aerial survey data as a whole; significant discrepancies are only likely to occur in high density areas, but high density areas only constitute a small proportion of total sightings (and biomass) along the survey tracks.

Sightings per mile

<u>Direct calibration</u>

Because pilots and spotters operate in pairs, it is possible to cross-calibrate individual observers by comparing sighting rates *within* single flights, before having to take into account patterns of sighting rates across space, time, and weather. The rationale is that abundance and sighting conditions are roughly the same for both observers within a flight; since a flight involves both an inward and an outward leg, any side-specific differences, e.g. sun angle, are presumed to cancel out on average. Any within-flight differences between the numbers recorded by the two observers, are attributable to differences in observer sighting power (which is what we want to know), plus statistical noise (which we can adjust for).

The standard protocol for dual-observer line transect surveys, is for each observer i and j to remain unaware of what the other has seen (see e.g. Buckland *et al.*, 2001). After the survey has finished, each school seen first by i can be examined to check whether j also saw it eventually, and conversely. This permits a full mark-recapture analysis, in which the probability of both observers missing a school can be inferred by checking how many schools were missed by one observer, compared to how many were seen by both. However, in the cramped confines of the SBT aerial survey, this strict-independence protocol is impossible, and other methods of analysis must be developed.

As soon as one SBT observer makes a sighting, both become aware of it, and so there is no direct way to check whether the other observer would eventually have seen that sighting. This is important, because what matters for calibrating a time series is the actual "sighting power" (probability of not overlooking a potentially-visible school) of each observer, rather who sees a school first. If both observers are looking at the same stretch of sea, and i usually sees schools earlier than j, then i will record a much higher sighting rate even if j would eventually have seen almost all the schools that i saw first. There is an important distinction between instantaneous and overall sighting probabilities for observers — in the above case, i has a much higher instantaneous probability than j, but only a slightly higher overall probability. In fact, because observers have different fields of vision, it is possible to estimate overall as well as instantaneous probabilities, as follows.

The SBT survey protocol is for each observer to concentrate on their own side of the aircraft, although some sightings (about 35% overall) are "poached", i.e. made on the opposite side. (Note that "poaching" is not meant pejoratively here; the term is borrowed from other line transect work, where poaching can cause problems in analysis that do not apply here.) It is apparent that the extent of poaching varies between observers: in **Table 2**, for example, which compares sightings by two observers over a number of flights, the number of poached sightings is very similar, but the number of unpoached sightings differs significantly.

Table 2: Distribution of all sightings made by one pair of observers.

Nominal side of aircraft	Who saw it	Number of sightings
A	A	123
A	E	48
E	A	46
E	E	93

The simplest statistical model that allows for these effects is as follows:

$$\begin{aligned}
&\mathrm{P}\begin{bmatrix} i \text{ at last spots a school that's been potentially spottable} \\ \text{for } t \text{ seconds on } i\text{'s side, during the next } \delta t \ sec\text{onds} \end{bmatrix} = \alpha_i f(t)\delta t \\
&\mathrm{P}\begin{bmatrix} i \text{ at last spots a school that's been potentially spottable} \\ \text{for } t \text{ seconds on } j\text{'s side, during the next } \delta t \ sec\text{onds} \end{bmatrix} = \theta_i \alpha_i f(t)\delta t
\end{aligned} \tag{8}$$

where α_i is i's own-side efficiency, θ_i is i's poaching rate, and $f(t)$ is the function that determines "sightability" of a school in terms of the length of time t since it has been available. Note that these probabilities are only relevant to the *first* moment that an observer sees a school. To make the problem identifiable, we can choose the time scale so that $\int f(t)dt \equiv 1$ over the entire period of availability of a sighting. It can then be shown that, if a sighting is made on i's side, then the probability it was made by i is

$$\frac{\alpha_i}{\alpha_i + \theta_j \alpha_j}$$

Also, the total probability of a potential sighting being seen on i's side (by either i or j) is

$$1 - e^{-(\alpha_i + \theta_j \alpha_j)}$$

How do the data provide enough information to allow all these parameters to be estimated? A heuristic argument is as follows. Let n_{ij} be the number of sightings by j on i's side when the two are flying together, etc. Suppose that the pairs (ij) fly together. Then

$$E\left[\frac{n_{ii}}{n_{ii}+n_{ij}}\right]=\frac{\alpha_i}{\alpha_i+\theta_j\alpha_j}$$

etc. Equating observed and expected values, and rearranging, we obtain

$$\theta_j\frac{\alpha_j}{\alpha_i} \;=\; \frac{n_{ij}}{n_{ii}} \tag{9}$$

$$\theta_j\frac{\alpha_j}{\alpha_i} \;=\; \frac{n_{ij}}{n_{ii}} \tag{10}$$

$$\Rightarrow\ \theta_i\theta_j=\frac{n_{ij}n_{ji}}{n_{ii}n_{jj}} \tag{11}$$

i.e. the product of the θ 's is the product of the ratios of poached sightings. Now suppose a third observer k flies (separately) with i and with j. The same arguments give

$$\theta_j\frac{\alpha_j}{\alpha_k} \;=\; \frac{n_{kj}}{n_{jj}} \tag{12}$$

$$\Rightarrow\ \frac{\alpha_k}{\alpha_i}=\frac{n_{jj}n_{ij}}{n_{kj}n_{ii}}\ \text{ after division by (9)} \tag{13}$$

$$\theta_k\frac{\alpha_k}{\alpha_i} \;=\; \frac{n_{ik}}{n_{ii}} \tag{14}$$

$$\Rightarrow\ \theta_k=\frac{n_{ik}}{n_{ii}}\frac{n_{kj}n_{ii}}{n_{jj}n_{ij}}=\frac{n_{ik}n_{kj}}{n_{jj}n_{ij}}\ \text{ on substitution from (13)} \tag{15}$$

Once θ_k is determined, then it is easy to determine the other θ 's, and the ratios of pairs of α 's; note that at least three observers are needed. However, to set the absolute rather than relative values of the α 's, some different logic is required. If the α 's are all very large, then $1-e^{-(\alpha_i+\theta_j\alpha_j)}\approx 0$ and there is almost no chance of a sighting being missed. Thus the total numbers of sightings on the two sides of each flight should be similar, even if one α is bigger than the others. On the other hand, if the α 's are all very small, then $1-e^{-(\alpha_i+\theta_j\alpha_j)}\approx \alpha_i+\theta_j\alpha_j$, so that the side with the more powerful combination should record proportionately more sightings. By looking at relative numbers of sightings on the more powerful side vs. the less powerful side, the absolute values of the α 's can thus be determined. Note that when all α 's are small, there isn't enough information to set absolute values of α very accurately, so it is more useful to choose one observer as a reference and estimate only the relative values.

More formally, the α 's and θ 's can be estimated by maximizing a binomial likelihood based on the above probabilities. Imposing a constant θ causes a significant worsening of fit, so only variable θ results are shown here. The MLEs of α turn out to be small, so only values relative to A are reported below, but large values are also just about consistent with the data (i.e. slightly within the 95% confidence interval). In other words, the analysis suggests but does not prove that many potentially-visible tuna schools are not seen even under good conditions by experienced observers.

The variation between experienced observers is about 30%. The fit to the data is very good (**Table 3**), and shows no evidence of lack-of-fit according to a χ^2 test, at least for the experienced observers (top part of **Table 3**; $\chi^2 = 9.7$ on $32 - 10 - 8 = 14$ DoF; $p = 0.16$ where p close to 1 denotes bad fit). If the trainee observers are included, the test does become significant, with a nominal p value of 0.95, but the large number of small fitted values makes the p value dubious. At any rate, it is hard to pick out any gross or systematic failures in **Table 3**, so there seems no strong cause for concern about this model.

Table 3: Observed (LH) and fitted (RH) values for number of sightings, from direct calibration

	A:C		A:B		C:B		A:E		C:E		B:E		A:D		B:D	
Pside:P	4	4.5	76	71.8	29	25.8	123	129.4	14	17.0	31	31.5	47	44.0	39	39.1
Pside:S	4	2.7	22	16.8	7	6.4	48	42.8	5	5.9	8	12.1	12	11.3	11	11.7
Sside:P	1	1.5	22	23.0	12	16.5	46	41.8	14	10.8	11	8.5	11	14.2	11	10.6
Sside:S	4	4.3	53	61.3	24	23.4	93	95.9	14	13.2	29	27.0	36	36.5	38	37.5

	F:G		B:G		F:J		B:H		F:I		B:I		F:J		B:J	
Pside:P	3	6.2	9	5.5	7	5.3	6	5.6	5	6.8	7	4.3	10	6.7	18	16.1
Pside:S	3	1.0	0	1.0	2	2.5	5	3.0	2	1.9	1	1.4	2	1.6	4	4.4
Sside:P	2	0.8	0	1.5	0	0.6	0	1.5	1	0.8	0	1.2	0	0.8	0	4.4
Sside:S	1	1.0	0	1.0	2	2.5	2	3.0	6	4.5	2	3.2	0	2.9	11	8.1

Confidence intervals can be obtained either from binomial theory, or by bootstrapping at the level of individual flights. One *a priori* reason for using the bootstrap, would be to guard against possible flight-specific effects, which might make the binomial-theory intervals too narrow. But there is no evidence of the over-dispersion that flight-specific effects would likely produce, and the binomial-theory intervals in fact are generally somewhat wider than the bootstrap intervals (up to 50% wider), so the bootstrap does not seem to be adding any value. The intervals in Tables 4 and 5 are based on binomial theory.

Standard errors on the α 's and θ 's are fairly small, except for the observers who flew only in the last few years. Although F was an experienced observer, he never flew directly with the other experienced observers (B, C, E, A, D). The only way to calibrate F against the other experienced observers, is via the trainees who also flew with B. This is rather indirect and inevitably relies on rather little data, so F's standard errors are correspondingly wide.

Most of the experienced observers have estimated efficiencies close to A (chosen as a reference because he has the most data), but E's estimated efficiency is significantly lower. There are large and significant differences in the extent of poaching, even amongst the experienced observers. Note that, although the estimated own-side efficiencies α are very low for some of the trainees, two trainees have "poach factors" much higher than 1. Presumably, they were training by mostly watching the same side of the airplane as their experienced companion. It would not be surprising if there was some difference in behaviour from the experienced observer B in 1999 and 2000 if he was assisting trainees; B's estimated θ falls from 0.34 to 0.27 when these two years are included. Some further investigation might be warranted, but the amount of data is very limited.

To adjust for "observer effects" on sighting rates, what is needed is an estimate of the combined sighting power of observers i and j flying together (i.e. the probability that they will not miss an available sighting), at least up to some constant scaling factor. The combined sighting power is

$$\frac{1}{2}\left(1-e^{-\left(\alpha_i+\theta_j\alpha_j\right)}\right)+\frac{1}{2}\left(1-e^{-\left(\alpha_j+\theta_i\alpha_i\right)}\right) \quad (16)$$

since a potential sighting is equally likely to occur on either side. These can be calculated directly from point estimates and bootstrap distributions.

The full procedure for incorporating sighting power estimates into models of SpM, is discussed below. Less formally, it is also informative to indicate the overall impact of observer changes on sighting rates over the years of the survey, via the average of pairwise sighting powers (weighted by number of flights) as shown in **Figure 9a**. (Note that this is not showing the trend in sighting rates; it is simply showing how much of the "empirical trend" can be explained by changes in observers.) Each line represents one trajectory of sighting power according to the bootstrap (so all lines are equally plausible), with the mean fixed at 1. Note that the lines are almost parallel except for the last two, and that up to 1998 there is no more than a 10% variation between the "best" and "worst" years. In the last two years, uncertainty dominates, and the main reason for the shifts in earlier years is simply the effect of 1999 and 2000 on the overall mean. If these two years are removed, then the overall trajectory is seen to be very flat, with very tight confidence limits (**Figure 9b**).

Table 4: Estimated efficiencies relative to A

Obs	J	C	B	F	E	D	G	H	I	A
Eff	0.43	0.95	0.86	0.98	0.74	0.83	0.08	0.34	0.67	1
SE	0.17	0.20	0.11	0.43	0.09	0.14	0.08	0.21	0.35	0

Table 5: Estimated efficiencies on "poach" side relative to "own" side

Obs	J	C	B	F	E	D	G	H	I	A
Poach	0.55	0.64	0.27	0.12	0.45	0.31	3.00	1.75	0.38	0.32
SE	0.28	0.15	0.04	0.07	0.07	0.07	3.46	1.10	0.25	0.04

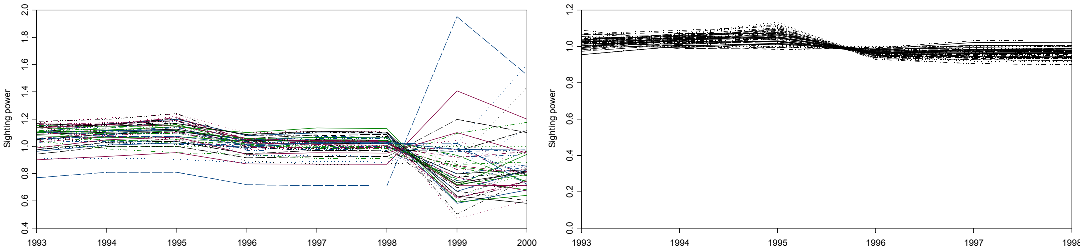


Figure 9: Average sighting power of flights each year, relative to long-term mean: (a) for 1993-2000, and (b) for 1993-1998.

Indirect calibration and a model for SpM

A natural way to model occasional events such as sightings, is through the probability of occurrence over a short time (or space) interval. For a flight path The mathematical formulation is

$$\mathrm{P}\left[\text{seeing something during time interval}\left[t, t+\delta t\right]\right] \tag{17}$$

$$\approx \left(\delta t\right) \times \left(\text{observer effect}\right) \times \left(\text{behavioural effect}\right) \times \left(\text{abundance effect}\right) \tag{18}$$

The effect terms on the right may depend on who was observing, on covariates such as wind speed which are known to affect sightability, on SST and other covariates that affect the proportion of time for which tuna are visible, and on covariates such as location and time of year that affect local density around the flight path. In a sense, model (17) is a continuous-time version of the presence/absence model used by Cowling, 2001a, who grouped data into quarter-lines for analysis. However, model (17) can work on a finer space/time resolution, does not require awkward decisions about exactly where to break transects, and is not sensitive to zeros; most of the time, nothing at all is being seen, and this is naturally accommodated in the model. Cowling & Laslett, 2000 suggest using a similar probabilistic formulation as part of their proposed modeling framework for sightings.

The strict implications of model 17 are that sightings follow a Poisson process, which implies a certain lack of patchiness in distribution; almost always, real sightings data show considerable clumping even after accounting for the covariates in model 17. Fortunately, this does not rule out the use of (17) as a basis for estimating the covariate effects, but it does imply that careful accounting is needed to get realistic estimates of variance. Full details can be found in Bravington, 1999b and Bravington, 1999a, and a similar model is described in Hedley *et al.*, 1999. In the present example, variance should be reasonably well accounted for by bootstrapping across flights, on the assumption that most observed clustering is either long-term (in which case it should be explicitly allowed for inside the abundance effect) or so transient that it would disperse before the next possible occasion a transect was flown.

Automatic covariate selection for (17) is difficult, because of clustered sightings (see next subsection). In this report, I have made no attempt to choose a “best” set, but have reported results for a reasonably full model incorporating within-year longitudinal movement, an overall year effect, wind speed, moon phase, sea swell, depth, and latitude band. Increasing wind speed seems to have a strong negative effect on sightability (or possibly on tuna surfacing behaviour), with probabilities halving at wind speed 4 relative to 0. Moon-phase suggests a 50% rise in sightability at full moon compared to new moon, but with wide confidence limits. Swell 2 seems to cause about a 40% drop in sighting rates. Low cloud cover (on a scale from 0 to 3) is very important, with sighting rates falling by 75% at level 3. Year effects — the real focus of interest — are shown in section 6.5.

The best way to report results for the space/time components of the abundance effect, is to show maps of estimated distribution at different times of the season (**Figure 10**). The maps show a gradual drift of tuna from east to west as the season progresses. If the shift is perhaps less than expected, this may be because the maps reflect an average annual pattern, which hides year-to-year differences.

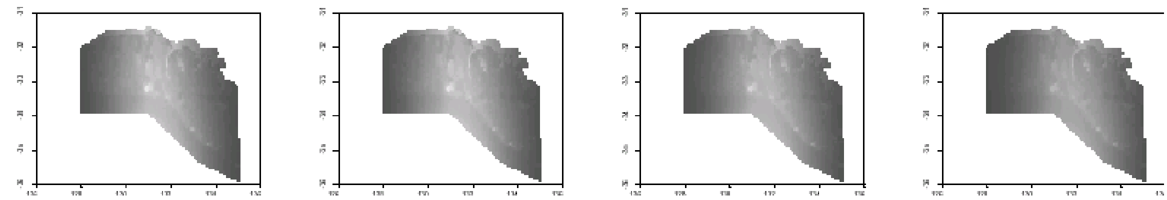


Figure 10: Normalised density of sightings, at 15, 30, 45 and 60 days into the year.

Potential gain in precision

As with the BpP and PpS models, I used a bootstrap to assess the potential improvement in CV for the SpM model if better information on the effects of environmental covariates and observers could be obtained. Results suggest that the CV of the estimated year effects on sighting rate could drop by 3-5 percentage points for an average year, out of a current CV of 22% averaged across years.

Overdispersion in the sighting rate model

It is usual for sighting rate models to reveal non-random levels of clustering among sightings, even after including all feasible explanatory covariates in the model. Clustering reflects transient clusters of prey, changes in sightability that are undescribed by any covariate in the model, persistent spatial features on too small a scale to be modeled, etc. Although clustering does not pose a serious problem for making point estimates, it does need to be allowed for assessing uncertainty.

The extent of residual clustering can be examined by comparing observed and expected numbers of sightings flight-by-flight. If there was no residual clustering, then a good fit should be obtained to observed sightings per flight by using a Poisson model with no parameters and an offset equal to the expected number of sightings for the flight. In fact, there is some over-dispersion. In terms of GLMs, a quasi-likelihood version of the small-α Poisson model leads to an estimated dispersion parameter ϕ of 8.52 (McCullagh & Nelder, 1989), with parameter variances about 8.52 times higher than if the same number of truly *independent* sightings had been seen. In other words, because sightings are clustered, the effective sample size (in terms of equivalent information about the underlying local density of clusters) of the aerial survey is only about $1/\sqrt{8.52} \approx 1/2.9 \approx 34\%$ of the nominal sample size.

Note that, when observer effects are fixed at the MLEs from the direct model, rather than being estimated just from indirect data, then the relative information content changes only slightly, to 33%. This is encouraging, as it suggests compatibility between the models; the fit to the indirect data is not worsened much by fixing the observer effects using other data. However, no formal compatibility checks have yet been carried out, and some substantial work would be required to devise a really good method.

Within a single analysis such as the indirect calibration of SpM, clustering can be allowed for by using a nonparametric bootstrap based on sampling units "big" enough to be statistically independent. Again, the flight is the natural unit here. Such a bootstrap will give a valid picture of uncertainty. When combining this model with the direct calibration, though, a little care is needed in applying the bootstrap, as discussed next.

Combining direct and indirect calibrations for SpM

A simple way to do this, is to allow the observer effects to be fitted parameters in the indirect model, but to place a penalty on the fitted observer effects according to how much they would worsen the fit to the direct model. This is akin to the approach used for BpP, but somewhat more approximate because the statistical models are harder to handle here. A statistically-reasonable approximate choice for the penalty, is the inverse of the covariance matrix from the direct model. However, for this to give statistically consistent results, the likelihood in the indirect model needs to be rescaled to take account of the over-dispersion; otherwise, there will be a distortion in the tradeoff between getting a good fit in one model versus the in other. A reasonable albeit approximate way to do this, is to divide the indirect model's log-likelihood by the sample size adjustment 2.9 above; this adjustment ensures that the analytical estimates of variances based on the rescaled log-likelihood, will roughly match the bootstrap variances.

Note that this simplification only works for the small-α (i.e. many schools missed) model; when α is substantial, then there is a non-linear relationship between the observer parameters α and θ in the direct model, and the combined efficiency for each pair of observers in the indirect model. The inverse-covariance penalty, which is quadratic in nature, does not deal well with strong non-linearity.

It is interesting to compare the direct, indirect, and combined estimates of overall efficiency $\alpha(1+\theta)$. **Table 6** shows MLEs for the indirect model (without any penalty based on the direct calibration results), direct, and combined models. Results are broadly comparable, and in most cases the combined MLE is between the direct and indirect estimates. Differences between direct and indirect MLEs do not seem unduly large given the CVs; the median CV for the combined MLEs is 19%. Concern might arise if there were differences between the "trainee effect" under the two models. There is no strong evidence of this: although three out of four estimated efficiencies are much higher under the direct model, precision is poor, and the fourth trainee (G) has substantially lower efficiency under the direct model.

Table 6: Comparison of estimated overall (both sides) sighting efficiency

Observer	F	C	B	D	E	G	H	I	J
Indirect	0.85	0.91	1.12	0.76	0.99	0.39	0.36	0.11	0.06
Direct	0.83	1.17	0.83	0.82	0.81	0.24	0.70	0.50	0.70
Combined	0.78	0.98	1.02	0.69	0.89	0.30	0.57	0.80	0.54

Detailed scrutiny of the table does reveal a few surprises. For example, observer J's combined MLE is much higher than under either individual MLE. There is at least a partial explanation, as follows. Almost all the information on J comes from direct calibration, which fixes J's efficiency relative to B and E (the only pilots he flew with). But much of the information on B comes from indirect calibration, which happens to raise B's combined MLE relative to his direct MLE. In turn, this raises J's combined MLE relative to his direct MLE. While this does not fully explain J's result, it does hint at the complexity of the correlation structure that hides behind the table. **Table 6**, which only shows MLEs, should not be over-interpreted. More sophisticated methods of combining the analyses (e.g. along the lines used for BpP) and of diagnosing discrepancies should be sought, but the task is not easy, partly because of the lack of a real likelihood for the indirect data.

The main messages from the SpM analyses are:

1. It is important to use data from direct as well as indirect calibrations;
2. Uncertainty is high for 1999 and 2000, because of the lack of overlap of observers with other years;
3. There are large differences between trainees and experienced observers, typically of the order of 50%;
4. There may be quite large differences even between experienced observers, of up to 30%.

Combining all the analyses

The idea behind the timing of the survey is that very few new fish should enter, and very few fish should leave, during the survey period. An index of abundance would be based on a completely synoptic survey during this period: take a "snapshot" of the whole GAB at some instant between those dates, correct the number of fish seen in each local region to account for local sighting conditions and who was observing there, and then add up across all local regions. The snapshot can be replaced by a prediction based on fitted models. In principle, this should give a consistent picture regardless of the exact date used, as long as the date is between the dates of immigration and emigration.

Recall that the local density near a position x_i at a time (or date) t is given by

$$density(x_i,t) = BpP(x_i,t) \times PpS\left(x_i,t\right) \times SpM\left(x_i,t\right) \tag{19}$$

The fitted models provide estimates of these three functions, in terms of parameters θ_{BpP}, θ_{PpS} and θ_{SpM}, collectively referred to as θ. For any particular value of θ, the predicted total abundance across the whole GAB at date t_1, is given by the sum of all the local densities:

$$abundance\left(t;\theta\right) = \int_x \left(BpP\left(x,t;\theta_{BpP}\right) \times PpS\left(x,t;\theta_{PpS}\right) \times SpM\left(x,t;\theta_{SpM}\right)\right)dx \tag{20}$$

The BpP, PpS and SpM models are all statistically independent. As described earlier, each model has been used to generate an approximate sample from the posterior distribution of its parameters, i.e. a set of values for θ that characterize the likely values and the extent of uncertainty for that model. Because the models are independent, the following approach can be used to combine the models:

1. Draw θ^*_{BpP}, θ^*_{SpM}, and θ^*_{PpS} randomly from their respective samples;
2. Predict BpP, PpS and SpM using the θ^*'s, at a grid of points across the GAB on specific days of the season and under standardised conditions, for each year between 1993 and 2000;
3. Multiply the three components at each grid point and date, to get predicted local densities;
4. Add up the predicted local densities at all the grid values and dates within each year, to form a time series over the 8 years;
5. Divide the time series by its mean (since we are only interested in a relative index).
6. Repeat steps 1-5 1000 times, and use the distribution of 1000 values of (say) 1996-index to construct a median and 90% confidence limits for I_{1996}.

The results (based on day-of-year 15, 30, 45, and 60, i.e. mid-January to start of March) are shown in **Figure 11**.

This figure is quite similar to those in Cowling, 2001a, p21 & p26; a dip in 1994 is followed by a rise around 1996, and a substantial fall thereafter to around 50% of the long-term average. The pattern of CVs is slightly different, though, although comparable in overall magnitude, with CVs smaller at the start and larger later relative to Cowling's results. CVs in the last two years are not directly comparable with Cowling's, because this analysis attempts to estimate observer effects for those years, and therefore has higher uncertainty since the trainee observers had fewer sightings on which to base estimates. In contrast, Cowling's model fixes the trainee effect at an assumed value in those two years; this leads to lower CVs but higher biases (as noted in Cowling, 2001a).

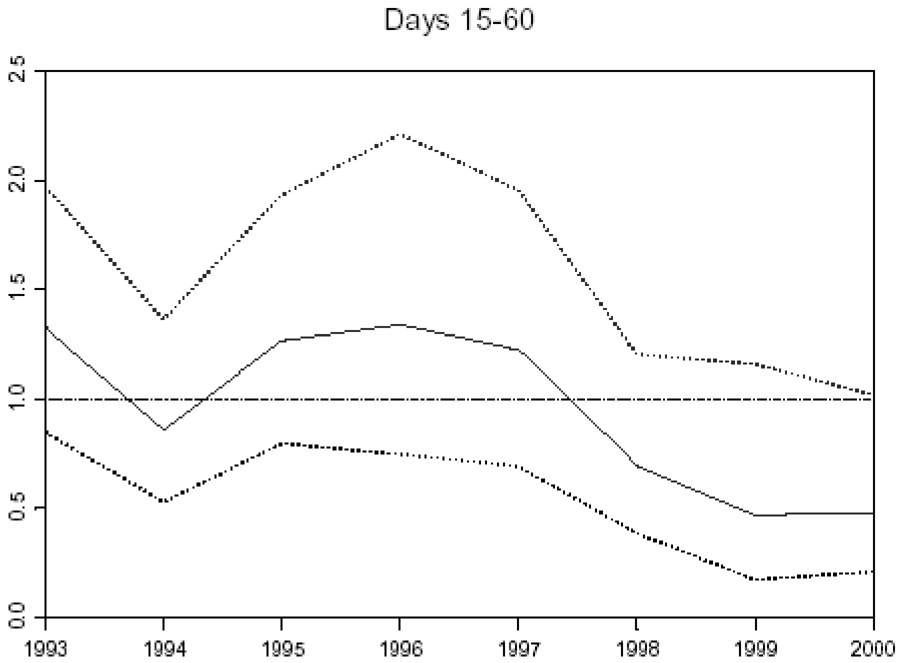


Figure 11: Abundance indices 1993-2000 with 90% confidence limits. Indices are standardised to a mean value of 1.

If the last two years are disregarded, CVs are about 2 percentage points lower (i.e. confidence limits are 7% tighter) in this analysis compared to Cowling's. Overall, the similarities between the results are much more striking than any differences.

Table 7: CVs on final estimate.

Year	1993	1994	1995	1996	1997	1998	1999	2000
Median combined estimate	1.33	0.86	1.27	1.34	1.22	0.7	0.47	0.48
Cowling's estimate	*1.48*	*1.31*	*1.10*	*1.51*	*0.84*	*0.73*	*0.44*	*0.59*
CV %	26	29	27	33	32	35	58	48
Cowling's CV %	*36*	*35*	*36*	*27*	*35*	*27*	*41*	*31*

As noted above, the basic assumption behind the survey is that most fish are in the GAB throughout the survey period. Therefore, in principle, the abundance series should look the same regardless of the date(s) when predictions are made. [An exception would be if within-season fishing mortality was very high, and variable from year to year.]. This can be checked by repeating the above exercise for different dates, as in **Figure 12**.

There are some differences between the four trends; the 1993 index declines through the season, while the 1996 and 2000 indices does the opposite. This suggests either some deficiencies in the underlying model, or some unexpected immigration/emigration of SBT during the survey season. It would be valuable to know if such movements are really plausible, because there would be major implications for survey design. If the differences in trends are a modeling artifact, it is not clear whether this should be a source of serious concern in a time-averaged model.

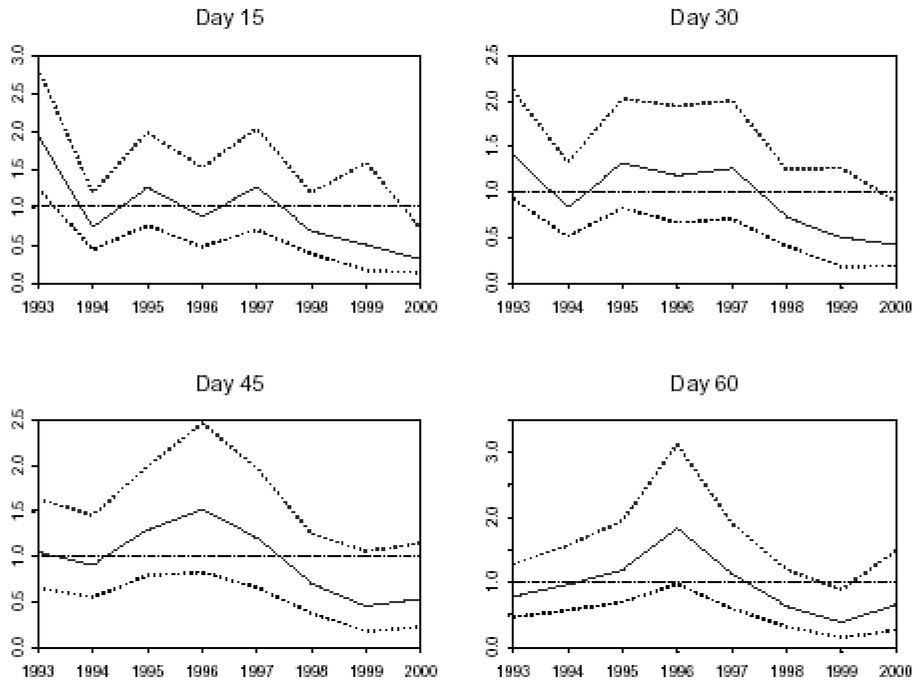


Figure 12: Time series of relative abundance, based on predictions at different moments in the season.

It is instructive to show the trends by abundance component (**Figure 13**). Comparing with **Figure 11**, it is clear that the overall abundance is not dominated by changes in any single component. Generally, there is less time variation in BpP than in PpS or SpM. In terms of precision, BpP is more precise than PpS is considerably more precise than SpM. Note that some care is needed in summarizing "average BpP", for example, because the results depend not just on the BpP model (which predicts average BpP *within* a particular stratum) but also on the PpS and SpM models (which predict *how many* patches there will be in that stratum).

Differences from previous analyses

As noted above, the overall abundance index series from this analysis is quite similar to that presented in Section 6.5. This is an encouraging sign of robustness, as there are some fairly big differences between the analyses. In particular:

1. The breakdown of "overall abundance" is different. Cowling uses "presence/absence within strip of fixed length" and "biomass conditional on presence". Instead, this analysis uses biomass-per-patch, patches-per-sighting, and sightings-per-mile. There are two reasons for the change. First, BpP etc. are much easier to interpret and relate to observers' experience. Second, only with BpP etc. is it possible to make systematic use of the observer-pair data, to improve calibration of observers.
2. This analysis uses extra data from pair-wise comparisons of observers, to aid in model calibration.
3. The assumptions of independence are different. Cowling assumes independence between quarter-transects. One difficulty with optimizing Cowling's analysis, is where to stop dividing the transects; her results suggest higher precision if shorter sub-intervals are used, but if the intervals are made too short, there will be significant statistical non-independence between consecutive sub-intervals, which

would invalidate the assumptions. Instead, the analysis in this section uses the "safer" assumption of independent flights for estimating precision, but with more detailed within-flight modeling.

4. The choice of covariates is different. For example, the SpM model here allows for within-year movements, via a longitude*day-of-year interaction. On the other hand, Cowling's analyses incorporate sea surface temperature (SST), avoided here because of the amount of missing data at the finer spatial scales of these models.
5. Model selection has been approached differently. Cowling uses stepwise selection based on significance tests to choose which covariates to include. This approach has been shown to perform poorly for predictive purposes, generally leading to under-fitting (see e.g. Burnham & Anderson, 1998). In these analysis, I started with AIC-based selection of covariates, which has generally better properties (*ibid.*). However, it is not clear that single-model AIC is optimal for the type of linked-model prediction required here. In order to minimize the risk of bias in the results— in other words, to keep the predictions close to the line-transect ideal of being a sum of corrected stratum means— I have used models that are deliberately over-fitted compared to AIC.

Overall, the similarity between results reflects the generally robust and sound design of the surveys to date, and lend weight to the viability of these data for quantitative purposes. It should be noted, though, that there are substantial differences between point estimates in two of the years (1994 and 1997); one way to think about this is that, even if the long-term trend is accurate, estimated CVs may not fully describe the short-term uncertainty. The analyses presented here make use of more data (via direct calibration), and offer perhaps a more fruitful basis for further work and use in assessment. However, the fact that there are some differences from Cowling's results, reinforces that the results are not completely "model-free", and underlines that the analyses here still need some further statistical investigation. Specific proposals on further analyses are given next.

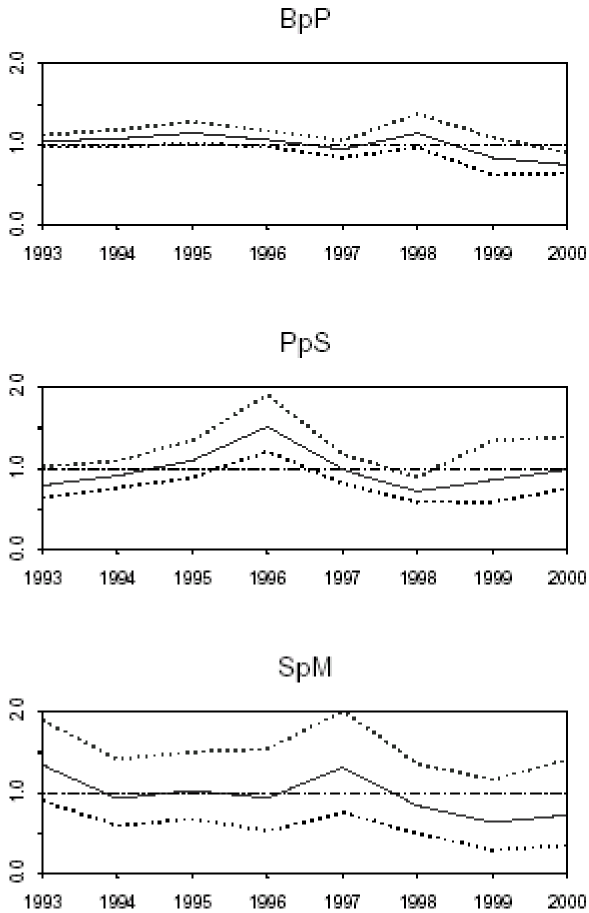


Figure 13:

Discussion

Future statistical issues

After a decade of data collection and dozens of reports, it might be tempting to ask why there still needs to be a section on "future statistical issues". In fact, it is not surprising for a dataset this complex, which behaved differently from initial expectations and required numerous techniques to be developed and explored. Indeed, it is unrealistic to expect that there will ever be a "final" method of analysis for such complex data. Statistical techniques keep improving, different types of data (e.g. on tuna surfacing behaviour) keep appearing, and the questions asked of the data also keep changing over time. However, none of this means that current analyses cannot or should not be used in relation to current management issues.

As a benchmark, the International "Decade" of Cetacean Research has been conducting annual line transect surveys of minke whales in the Southern Ocean for 24 years, comprising almost 3 complete circuits of the globe south of 60 ° S. At least since 1985, the survey protocols have been consistent and of high standard, under the oversight of the International Whaling Commission. But there is still no final agreement about appropriate methods of analysis, and in fact there are probably more serious questions today about appropriate methods for the IDCR data than about the SBT aerial survey. As the IDCR survey has progressed, the list of questions that need investigation has expanded; most recently, attention has focused about whether it is possible to explain apparent trends through behavioural/climatic shifts (Scientific Committee of the

International Whaling Commission, 2002). Addressing these questions requires analysis of the effects of covariates, more detailed investigation of the nature of individual sightings, and ancillary experiments (e.g. on dive times of minke whales, or on observer sighting abilities), so that more and different analyses and data have been required as times has passed. Yet, despite these difficulties, the IDCR data have been used many times for analyses relevant to management, and have been an invaluable resource in this regard.

In terms of prioritizing further work, though, the most pressing statistical issues are perhaps as follows:

1. Model selection and covariate choice

No systematic attempt has been made to optimize the choice of covariates in the models, e.g. by using AIC or attempting to minimize expected prediction error. Instead, in the interests of avoiding under-fitting, reasonably “full” models have been used, i.e. with a generous rather than parsimonious set of parameters. An exception is sea surface temperature (SST); in constructing the abundance index, SST proved awkward because of missing data, and was omitted; in future, efforts should be made to interpolate SST where data are missing. Otherwise, covariates (including the choice of time/area stratification) were included if there were reasonable *a priori* grounds to expect a substantial effect, either based on biological information or on observer reports and experiments. Even so, the range of models that could be examined was limited by time, and it remains possible that better results could be obtained e.g. by using finer spatial scales, or by allowing for different within-GAB migration patterns between years.

On the whole, though, the abundance derived here is more likely to rest on mildly over-fitted models, than on under-fitted models. The implication is that the abundance index is unlikely to suffer from serious bias (at least in purely statistical terms), but may be less precise than it could be if more systematic model selection was used. However, because the overall prediction is a composite taken from three models, it is not possible to merely "round up the usual statistical suspects". A significant amount of methodological development would be needed.

Because the final abundance indices entail heavy averaging across space and time, rather than predictions under conditions seldom seen, it seems *a priori* unlikely that results will be very sensitive to model selection. Nevertheless, model selection uncertainty is probably the most important statistical issue to tackle in further analysis. From a statistical point of view, analyses that are conditioned on one particular choice of model(s) and do not allow for model selection uncertainty, will always tend to underestimate the real uncertainty. From an end-user’s point of view, it is important not to leave results vulnerable to criticisms about the subjectivity of model choice.

Provided that an automatic model-selection algorithm such as AIC is available, techniques such as “bagging” (bootstrap aggregation; Breiman, 1996) can be used to incorporate model uncertainty in a fairly painless way. Practically, bagging is usually simple to apply and has shown impressive benefits in predictive power; conceptually, it sidesteps the criticism that results may be contingent on a particular model choice, by averaging across a range of models based on their goodness of fit. Adapting bagging to the SBT aerial survey poses a few challenges, however, not least in terms of computational demands.

2. Smooth vs. stratified models

A related issue is whether it is worthwhile moving to smooth models instead of stratified models. In reality, tuna density does not change sharply at the boundaries of statistical strata. Estimates of uncertainty, which are based on within-stratum variation, will interpret the true variation within the stratum as extra noise. In principle, therefore, models which avoid stratification and instead allow smooth transitions of BpP, PpS, and/or SpM in space and time, should be able to deliver more precise abundance estimates. As yet, though, experience in other survey contexts (e.g. minke whale abundance surveys in the Antarctic; Scientific Committee of the International Whaling Commission, 2002) suggests that further methodological refinement is required before such models can be confidently applied to abundance estimation. This work is being pursued elsewhere, and may become applicable to SBT within a couple of years.

3. Model synthesis & use of a priori information

For both BpP and SpM, it is necessary to synthesize the direct and indirect calibration models in order to get a properly-calibrated result. The methods of synthesis used in this report are fairly crude, and could be improved in one of two ways. The first is to write special-purpose estimation software for an integrated model that uses both direct and indirect data simultaneously. The problems are that it would be difficult to build in flexibility about choice of covariates, and also that tools for residual analysis, plotting, etc., would all need to be redeveloped. The synthesis of the SpM models in this report uses an approximation to this approach.

The second approach is to refine the approach used for the BpP models in this report, whereby simple models are fitted to each part of the data, and some kind of simulation is used to synthesize. The current method is quite computationally intensive, and is inefficient in that only a small proportion (20%) of the sets of possible parameter estimates actually receive significant weight.

As well as direct calibration data from paired observers, other kinds of information may become available for *a priori* incorporation in analysis. One example is biological information on the relationship between surfacing rates and space/time/weather covariates, from archival tag data. Reliable estimates from such data could in principle help to reduce overall variance, but further work would be needed to verify this (Cracknell, pers. comm..).

A hard but potentially important question, is whether the “law of conservation of fish” can be used in any practical way. If the population of juveniles in the GAB is closed within each year— i.e. no significant immigration or emigration within the survey period— then an ideal model would ensure that abundance indices constructed for (say) 1st January and 15th March in the same year have exactly the same value (neglecting fishing). In principle, imposing this constraint on models should reduce variance, but it is not at all obvious how to impose it. In any case, it is first of all important to be sure that the closed population assumption is biologically valid. If it is, then statistical attention could be given to improving the current model; if it is not, then the current basis for the survey estimates (a time-averaged estimate between 1st Jan and 1st March) needs to be reconsidered.

4. Incorporation of random-effect terms

This is an important variance reduction technique that could be applied for observer effects and perhaps for time/space interactions within years. In this section, random effect models were used in preliminary investigations of BpP and PpS data. However, there are two main complications that need to be resolved before random effect models can be used in an actual abundance index. The first is that the main SpM model does not have an exact likelihood, because of the clustered nature of sightings ("over-dispersion"). The second is how to link models that incorporate random effects, in the way that direct and indirect calibrations are linked in this report.

5. Improving the PpS model

Among the BpP, PpS, and SpM models, the PpS model is the least statistically satisfactory. The problem lies in the heavily-skewed distribution of PpS, even after allowing for covariates. Such data are not handled well by off-the-shelf statistical models (GLMs), and it is obviously inappropriate to simply "reject the outliers"; the observations are not mis-measurements. Although the GLMs used here do seem to perform reasonably well in tracking the stratum means, the results are sub-optimal in terms of robustness and efficiency. Further work is needed to develop bespoke models based on distributions that are more intrinsically skewed distributions.

Data, design and application

Comparison with other surveys

Despite the many concerns that have been raised over the years, the SBT GAB aerial survey seems to show promise for construction of a recruitment index. The CVs on annual estimates are not particularly high by the standards of fisheries data, and could be brought down further. With longer time series and/or better data from outside the survey itself on environmental and observer effects, the CV of each annual estimate could be reduced from over 35% to less than 30%. As a general guideline, fish abundance surveys typically report CVs in the range 20-50% (ICES CM 1991/D:40 p13); Pope (1983, in Gulland) alludes to 20% as "high precision".

What is more, fish survey CVs calculated from survey data alone (so-called "internal" CVs) tend to be considerably lower than the CVs calculated post-hoc based on the discrepancy between survey estimates and estimates based on other data. In many cases (e.g. ICES North Sea demersal fish surveys), the "internal" CV is not even routinely calculated because it is known that it will underestimate true uncertainty, although such a policy seems over-casual. Sometimes, an internal CV will be too low because of unacknowledged model uncertainty, e.g. if some covariates (such as environmental conditions) are not incorporated, but should have been. At any rate, discrepancies between different analyses (e.g. between the 1994 and 1997 estimates under Cowling's and under Bravington's model) are not unusual with survey data, but certainly do not preclude the use of a survey or a particular analysis in management. Ideally, an

Operating Model approach to management should be used, to safeguard against overconfident interpretation of data.

Compared to other line transect (as opposed to trawl or acoustic) surveys, the SBT line transect survey is not especially precise. CVs for marine mammal surveys are sometimes much lower, down to 10% for Northeast Atlantic Minke whale surveys since about 1990 (e.g. IWC). However, the 10% figure actually applies to a survey average calculated over 5 years, corresponding to an annual CV of 22%.

The problems of interpretation and index construction for the SBT GAB survey seem no worse than those encountered with CPUE standardization — something that is standard practice, albeit as a necessary evil, in many stock assessments — and much better in many ways because the experimental design cuts down the extent of confounding between covariates. The major remaining issue, though, is observer effects. For the 1993–1998 data, reasonably precise calibration is possible. However, because of limited overlapping, calibration of observers is difficult for the most recent years. If observers cannot be reliably calibrated, then there is no sense in continuing aerial surveys. Proper calibration and control of protocols would also improve the potential of the historical data for producing recruitment indices — something that is presumably still relevant to stock assessment, given the long lifespan of SBT. The issue of how to use indices in assessment is very important in considering how and whether to continue with aerial surveys in the GAB, and is revisited at the end of this discussion.

<u>Collection of other experimental data</u>

Some of the uncertainty in aerial survey indices arises because of the effects of uncontrollable environmental covariates such as SST. While it is possible to estimate these effects indirectly through the types of models in this report, it is definitely preferable if such effects can be fixed in advance. Technologies such as archival tagging offer the potential to do exactly this. Also, archival or acoustic tags might be useful in identifying the extent of any late immigration to, or early emigration from, the GAB during the survey season.

Of even greater importance is the role of covariates such as depth, which might in principle affect sightability as well as distribution. Faced with a dataset like the aerial survey, and noticing that depth is related to observed local abundance, it is possible to estimate a depth effect empirically. As long as the spatial distribution of SBT remains much the same, this estimated depth effect can be helpful in producing standardised indices. But if the distribution moves further offshore in one year, there is no way to tell whether there is really a different abundance or just a different surfacing rate. Basic biological information is required.

From the statistician's point of view, it would be helpful to get a biologist's input on how each component of sightability is likely to be affected by a known important covariate such as SST. Various possibilities can be conceived: higher temperatures may affect patch size because each patch is a transient made up of fish moving up from a deep-dwelling school, or may affect number of patches because fish decide en masse to make an excursion to shallower depths, or may affect sighting rate if entire large schools simply stay deep in cool conditions. Having an idea of not just the aggregate effect, but also the effect on particular components, make the analytical task much easier. Studies that concentrate on school visibility as well as individual visibility would be particularly informative.

There may also be scope for further experimental data related to observer's sightability, as reported in Cowling (2000). Opportunities might be sought out to integrate such studies with reports from commercial spotting data.

Role of commercial data

Commercial data offers two obvious possibilities for integration with future aerial surveys, the first being through calibration of patch size measurements. The second possibility is that estimation of the effects of environmental "nuisance" parameters on sightability may be possible, because commercial data will have huge numbers of sightings compared to transect data. However, differences in protocol between commercial and survey, and across different commercial operations, might make it difficult to apply the results with confidence to aerial survey data.

There are other more subtle ways in which commercial data may be useful. When choosing between several possible model-based analyses, it is important to have some notion of spatial scale, and the way in which spatial distribution changes over a season. Although it is possible in principle to estimate this from survey flights alone, it is really asking too much of a small number of sightings. Commercial data, with its dense coverage of certain areas, gives a much better picture of spatial scale and variation, could be a very useful tool in guiding appropriate model selection for survey data.

Historical protocol/ability changes

The value of a time series of estimates depends entirely on there being some consistency in measurement. In terms of the 3-stage decomposition of the data used here, it is difficult to imagine that PpS and SpM are problematic; although there may be changes in how patches have been grouped into sightings, the important statistic overall is the count of patches per mile, which should not be vulnerable to "observer drift". Changes in individual spotting ability over time are a possible concern, especially when relatively inexperienced spotters are used; more sophisticated direct calibration models could be employed to study this.

An individual's way of estimating individual patch biomasses, on the other hand, may conceivably change over time. The BpP model in **Figure 2** shows a fairly steady and strongly significant time trend in BpP, but without further insight it is impossible to say whether this is an artifact of observer drift (though this would have to apply to several observers together), or a real biological phenomenon. Since the inclusion or exclusion of this trend would have a substantial impact on any time series, further consideration should be given to whether this can be resolved, from commercial data or biological considerations. Since observers normally work for industry and must endeavour to provide patch biomass estimates where possible, one might be tempted to assume that there is some consistent basis in reality for their estimates.

One complication alluded to in Cowling (2001), is that there may also have been changes over time in the way certain covariates are measured, e.g. sea swell. Where possible, it would be desirable to replace such covariates with objectively-measured proxies (e.g. satellite data or weather predictions), even if the proxies are less strongly linked to tuna behaviour.

Importance of direct calibration and protocol control

The comparison of direct *vs* indirect calibrations for BpP and SpM, showed that direct calibration is more effective, both for obtaining more precise estimates of observer effects, and for avoiding disruption of e.g. year effects that are largely confounded with observer effects. The main source of uncertainty in an uncalibrated aerial survey would certainly be differences between observer's sighting rates, with second place probably going to differences between patch biomass estimates. It is fair to say that any future aerial survey which did not make plans for some kind of calibration (possibly in the future, to allow retrospective adjustments), would not be worth doing.

The reason that some kind of direct calibration has been possible in this report, is that the historical protocol has ensured some replication; two measurements of the same patch, and two sets of observations along the same flight path. The decision in 1999 and 2000 not to record two estimates for each patch, prevented calibration of the trainee spotters. Also, the limited crossovers between some subsets of experienced observers mean that some of the direct calibrations for sighting rate are quite uncertain.

Logistics aside, the best way to do calibration is to ensure that some flights carry two experienced observers, who do not collude or poach each others sightings. Some modifications in protocol might be considered, such as noting all observations of a "school" but not communicating the information to the other observer until the perpendicular waypoint is reached (so that the data records whether both observers really did see the school independently). Of course, not every flight needs to follow a calibration protocol; once the effects of observer X have been estimated to reasonable precision, it would be safe to take estimates from observer X alone for a few seasons.

Scope for changed survey design

The 2002 survey covered a restricted part of the historical survey area, between about 132.5°E and 134°E. Based on the models in this report, about 95% of SBT biomass in the GAB lies within this region. This suggests that a survey based on this limited set of transects might provide almost as effective an index of total abundance, as a GAB-wide survey. In fact, if the reduced spatial extent made more time available for surveying this higher-density region, the improved CV due to better information on sighting rate etc. might well be worth the slight uncertainty introduced about what animals may have been outside the range. If a model-based analysis is planned, it is even possible to consider a fairly radical change, such as running some transects east-to-west or along a depth contour. This might permit concentration of effort in areas of high abundance, where it is of greatest importance to reduce estimation uncertainty.

However, surveys that cover only part of the distributional range are vulnerable to changes in year-to-year distribution. The nightmare scenario for an analyst is to find a clump of high abundance on the edge of a surveyed region, because then doubts are raised as to how much biomass was cut off by the artificial boundary. It has not yet been possible to check what effect historical shifts in distribution would have had on a reduced-width survey indices compared with the full-width indices, although the calculations are fairly straightforward in principle.

Cowling (2001b) suggests a much more radical change in protocol, to a presence/absence survey that does not attempt to estimate biomass (or species) of patches encountered. She gives three main reasons: concerns about observer drift make it difficult to interpret indices which involve subjective estimates of biomass; the introduction of "biomass multipliers" into an annual index will inevitably increase the uncertainty; and time spent estimating patch biomass detracts from time available to search for patches, with patch density being a more important source of uncertainty than patch biomass.

Despite the attractions, this policy would carry some serious risks. Although CV might be reduced on paper, the reality would be that genuine uncertainty has simply been moved rather than removed: out of the *numerical* end of the problem, and into the *interpretational* end. It is not obvious how a P/A index would change as tuna recruitment changes; certainly, there is no reason to believe that the index will change proportionately if the changes in recruitment are large. In particular, if recruitment increases, a P/A index may "saturate", so that tuna will always be seen somewhere in each part-line within the core distributional area. Subsequent increases in recruitment would then not show up as increases in the index.

Even if P/A index did eventually prove to be proportional on average to recruitment over a wide range of recruitment, there could still be important year-to-year variability in the extent to which the tuna are spatially concentrated. The BpP analysis in this report shows that there may, at least, be quite major changes on the scale of individual patches, so that substantial changes in concentration over time are at least plausible. Getting the index "right on average" over the long term is not enough for stock assessment, where estimates of individual year-class effects are important. The key driver for any survey index that is destined for stock assessment is that the index should be as far as possible linearly proportional to abundance; variability is a secondary consideration.

For these reasons, if the aerial survey does continue, it seems imperative to keep to the basic protocol of estimating biomass as well as counting patches. Again, every effort should be made to directly calibrate estimates of patch biomass, both amongst observers but also between observers and the commercial fishery.

Role in assessment and management

Recruitment estimates are helpful to stock assessment and management in two main ways: by producing an annual index of cohort strength, and by allowing for detection of long-term trends in stock. The utility as far as the long-term index is concerned, can be studied by roughly calculating what level of change in recruitment over a 5-year block would be detectable with surveys of varying precision. SBT reputedly has comparatively stable recruitments compared to many fish species (the median across species being about 70%), so for the sake of argument assume that genuine CV of recruitment is 40%. Even with a perfect survey (no uncertainty), the mean recruitment over successive 5-year blocks would then fluctuate randomly with a CV of around 18%; this means that a real shift of 35% in long-term mean recruitment over 5-10 years would be on the margins of statistical significance. If the survey itself has a CV of 30%, then the 5-year mean will have a CV of around 22% (only slightly higher than with a perfect survey), so it would require about a 50% change in long-term mean recruitment over 5-10 years to achieve marginal significance. These guidelines need to be borne in mind when assessing whether year effects are "significant"; the power of any test to detect significant changes in recruitment is low unless the changes are quite drastic. The corollary is that statistical significance is often an inadequate guide as to whether managers should take action.

It is more complicated to consider how well cohort strength can be estimated from a given survey, because the SBT survey is an aggregate across three main age classes. A statistical approach can be developed, based on estimating the least variable sequence of cohort strengths that is consistent with the available indices and their standard errors. Assuming that the proportion π_x of x-year-olds entering the GAB is the same for all years (but not necessarily for all x), then the model is fully-estimable in principle, but it is unclear how precise the estimates will be in practice; results are also likely to be sensitive to the accuracy of the estimated CV. Some simulation work would be needed here. Additional information, such as archival tag information on π_x, would certainly be helpful to estimation of cohort effects.

A key issue for the utility of any GAB-based index, is how many fish simply don't go to the GAB. Hypothetical movements into the Indian Ocean from western Australia could take a significant proportion of the total population out of the range of any GAB survey. If this proportion varies substantially across years, then the relevance of a GAB-based index to management is questionable. Basic biological information on the likelihood of this scenario would be invaluable, though for fairly obvious reasons it would be hard to collect.

Finally, the utility of a survey-based recruitment index needs to be weighed up against the cost and the impact the index would have on management. This in turn depends on the quality of alternative data sources. It is certainly well beyond the scope of this report to discuss these issues, but it would be valuable to have discussion about e.g. the potential role of CPUE data and of juvenile tag returns, compared to any added value from an aerial survey index.

Summary

The SBT aerial survey dataset is very complicated. The analyses documented in this and earlier reports have evolved over many years, and have taken on many different forms. These changes, and the general complexity of the analysis, may give the impression that there can never be an outcome of quantitative value to management. That would be wrong. In fact, a fairly consistent picture of trend emerges from different analyses, and the precision of the indices (around 30% CV) is quite reasonable in global terms. It is inevitable that analyses will continue to evolve, and there may never be a final perfect answer for such complex data. Statistical techniques keep improving, different types of data (e.g. on tuna surfacing behaviour) keep appearing, and the questions asked of the data also keep changing over time. However, none of this means that current analyses cannot or should not be used in relation to current management issues.

In the case of the SBT data, the indications from the analyses here are that it can be used to deliver a reliable medium-term index of trends in GAB abundance. To get really good year-to-year estimates with reliable CVs, a little further methods development is still required, primarily over model selection. In saying this, though, it is important to keep a perspective on the precision of *other* data used in SBT assessment, particularly with respect to information about juvenile fish. The interpretation of SBT CPUE data, for example, is notoriously difficult, and model uncertainty affects the interpretation to a greater extent than with aerial survey data. The value of past aerial survey data, and the value of collecting such data in future, depends on the precise role it will be called to play in SBT management and assessment. This issue is very important, and requires considerable further input from those involved in stock assessment. For example, there is little precedent for using a mixed-age recruitment index in an assessment, so some development of potential assessment methods will be needed before it is possible to fully evaluate the long-term utility of an aerial survey index.

Further experimental work will be necessary, too, both to get reliable indices in future and to tidy up the 1999 and 2000 indices which at present are comparatively imprecise. This applies particularly to the intercalibration of observers. Further biological information on tuna visibility / surfacing behaviour, and on the extent of late arrivals and early departures from the GAB, would also help to improve precision. However, the fundamental questions of greatest importance are: what proportion of juveniles actually go to the GAB each year, and (if significantly below 100%) does this proportion vary with time? There is no current evidence to suggest that many juveniles go elsewhere, but equally the hypothesis cannot yet be disproved. If the answers did turn out to be "much less than 100%" and "yes", respectively, then the value for management of any survey in the GAB would be vastly reduced. It is essential to look at ways of collecting data that will answer this basic question.

References

Bravington, M.V. 1999a. *Coping with clustered sightings*. Scientific Committee IWC/51/RMP15. International Whaling Commission.

Bravington, M.V. 1999b. *Covariate models for continuous-time sightings data.* Scientific Committee IWC/51/RMP14. International Whaling Commission.

Breiman, L. 1996. Bagging predictors. *Machine learning*, **24**, 123-140.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Borchers, D.L., & Thomas, L. 2001. *Introduction to distance sampling*. Oxford University Press.

Burnham, K.P., & Anderson, D.R. 1998. *Model selection and inference: a practical information-theoretic approach*. Springer (New York).

Cowling, A.M. 2000. *Data analysis of the aerial surveys (1992-2000) for juvenile southern bluefin tuna in the Great Australian Bight*. Tech. Rept. Southern bluefin tuna recruitment monitoring program, Report of the 12th Workshop.

Cowling, A.M. 2001a. *Further data analysis of the aerial surveys (1993-2000) for juvenile SBT in the Great Australian Bight*. Tech. Rept. Southern bluefin tuna recruitment monitoring program. Report of the 13th Workshop.

Cowling, A.M. 2001b. *Issues in the design and implementation of an aerial survey of a pelagic species*. Tech. Rept. Southern bluefin tuna recruitment monitoring program. Report of the 13th Workshop.

Cowling, A.M., & Laslett, G. 2000. *Proposal for analysis of SBT line transect data using Gaussian random fields.* Tech. Rept. CSIRO Marine Research, Hobart.

Hedley, S.L., Buckland, S.T., & Borchers, .L. 1999. Spatial modeling from line transect data. *Journal of Cetacean Research and Management*, **1**(3), 255-264.

McCullagh, P., & Nelder, J.A. 1989. *Generalised Linear Models*. 2 edn. Chapman & Hall.

Newton, M., & Raftery, A. 1994. Approximate Bayesian inference with weighted likelihood boostrap. *Journal of the Royal Statistical Society.*

Research, CSIRO Marine. 2000. *Report of the aerial survey workshop, 14-19 February 2000, Port Lincoln.* Tech. Rept. MR-C 96-12. CSIRO Marine Research, Hobart.

Scientific Committee of the International Whaling Commission. 2002. *Annex G (In-depth assessment of whale stocks)*. International Whaling Commission.

7. Benefits

The Australian SBT industry will benefit from the research, as improved assessment of the SBT resource will provide a better basis for setting catch limits, and demonstrating the status of the stock.

The surface abundance indices from the aerial surveys were used in the stock assessments presented at the 4th CCSBT Scientific Meeting, August 1998. The report of that meeting includes the following in the Advice and Recommendations section:

> "The meeting also recognised that information on recent recruitment based on tagging studies, aerial surveys and possibly acoustic surveys was critical for providing timely advice on stock status and future management. Lack of future aerial survey information would seriously affect current and future assessments. The meeting strongly recommends that the Commission note the priority research needs, and in particular, urge each member to support continuation of the aerial survey."

The CCSBT continues to support the development of the index, and this is noted in the scientific reports for 2001.

8. Further Development

The SBT aerial survey dataset is very complicated. The analyses documented in this and earlier reports have evolved over many years, and have taken on many different forms. These changes, and the general complexity of the analysis, may give the impression that there can never be an outcome of quantitative value to management. That impression would be wrong. In fact, a fairly consistent picture of trend emerges from different analyses, and the precision of the indices (around 30% CV) is quite reasonable in global terms. It is inevitable that analyses will continue to evolve, and there may never be a final perfect answer for such complex data. Statistical techniques keep improving, different types of data (e.g. on tuna surfacing behaviour) keep appearing, and the questions asked of the data also keep changing over time. However, none of this means current analyses cannot or should not be used in relation to current management issues.

In the case of the SBT data, the indications from the analyses presented here are that it can be used to deliver a reliable medium-term index of trends in GAB abundance. To obtain good year-to-year estimates with reliable CVs, a little further methods development is still required, primarily with regard to model selection. In saying this, it is important to keep a perspective on the precision of other data used in SBT assessment, particularly with respect to information about juvenile fish. The interpretation of SBT CPUE data, for example, is notoriously difficult, and model uncertainty affects the interpretation to a greater extent than with aerial survey data. The value of past aerial survey data, and of collecting such data in future, depends on the precise role it will be called to play in SBT management and assessment. This issue is very important, and requires considerable further input from those involved in stock assessment. For example, there is little precedent for using a mixed-age recruitment index in an assessment, so some development of potential assessment methods will be needed before it is possible to fully evaluate the long-term utility of an aerial survey index.

Further experimental work will be necessary, both to get reliable indices in future and to tidy up the 1999 and 2000 indices that at present are comparatively imprecise. This applies particularly to the inter-calibration of observers. Further biological information on tuna visibility and surfacing behaviour and on the extent of late arrivals and early departures from the GAB, would also help to improve precision. However, the fundamental questions of greatest importance are: (i) what proportion of juveniles actually go to the GAB each year, and (if significantly below 100%) (ii) does this proportion vary with time? There is no current evidence to suggest that many SBT juveniles go elsewhere during the summer, but equally the hypothesis cannot yet be disproved. If the answers did turn out to be “much less than 100%” and “yes”, respectively, then the value for management of any survey in the GAB would be vastly reduced. It is essential to look at ways of collecting data that will answer this basic question.

Some individual-based modeling approaches are examining the likely variation in arrival and departure times from the GAB. As explained in this report, if different proportions of the juvenile stock are present in the GAB at different times during the summer, then the estimates of abundance in the aerial survey will be biased. Preliminary results of these analyses suggest that, if environmental variation influences the arrival and departure times, there is definitely the possibility of interannual variation in the summer residence time within the GAB. Archival tag data is also being analysed to determine when SBT arrive and depart the GAB, and the extent of interannual variability in the residence time.

Archival tagging of juvenile SBT throughout the winter range in the southern ocean is one way to address the question of the global location of juveniles throughout the summer. Current archival tagging has focused on tagging juvenile tuna that are already in the GAB during the summer, and while many return, fish that may not return are unlikely to be recaptured, and so determining if they spent summers in a different region is impossible. An archival tagging project, in concert with a large conventional tagging project under the auspices of CCSBT, has been proposed and would involve tagging and recovering juvenile SBT throughout a much greater range than the GAB. This study would address the question of what proportion of juvenile SBT use the GAB each summer, and may allow the GAB index to be related to the global population of juvenile SBT.

The cost and logistical constraints of the scientific aerial survey continue to be problematic. As a solution, a spatially reduced scientific survey was initiated in 2002, and covered the area where 95% of the SBT were detected in the full spatial survey of 1993-2000. Evaluation of this reduced survey and continuation of the time series of abundance are considered critical. Additional research is required to establish how to combine the two surveys. To offset the reduced spatial coverage, rigorous collection of data from commercial spotting operations (planes that assist the commercial domestic fishers to capture SBT in the GAB) was initiated in 2002 and will continue in 2003. Including this information in an index of abundance is a potential goal, although, as discussed in **Section 6.6,** this additional data may allow evaluation of differences between spotters and other similar validation processes that will further decrease the uncertainty in the scientific survey. The aerial survey and the analysis of the data will continue to evolve within the operational constraints to meet the objectives required to monitor the abundance of juvenile southern bluefin tuna.

9. Planned Outcomes

One of the planned outcomes of this project was to develop an index of abundance for juvenile southern bluefin tuna (SBT). This index was one of the main outputs of the project, and in fact several indices were developed and compared, each with differing units and assumptions. These alternative index constructions were necessary because we learned of problems with the data collection and interpretation through the project. The original goal was to develop an age-based index of abundance; it became apparent that the ability to reliably estimate the size of fish in the aerial survey was lacking. This ability was tested with experiments comparing two spotter planes surveying the same schools of fish. As a consequence, an age-aggregated biomass-based index was instead developed, again however, it became apparent in validation experiments that the estimation of school biomass was unreliable and imprecise. The next attempt was to develop a presence-absence index, based on the aerial detection of schools of SBT. This presence-absence index is a relative index of abundance and not an absolute index. It is valid as an index that can be used to monitor relative changes in abundance of juvenile SBT in the Great Australia Bight (GAB).

Throughout the project, the outputs were a compromise between the requirements for best-science and the practical realities of operating in the environment of the GAB. Weather conditions prevented the completion of surveys exactly as planned. A shortage of experience and trained spotters also compromised the project. Accounting for these changes in the aerial survey was a statistical challenge, and remains an area in which improvement is possible. For example, comparisons between spotters would allow correction of the possible differences, and as more data is gathered in the future, can be retrospectively done for the historical data. The domestic SBT industry has recognised the logistical problems associated with the survey, in particular the shortage of trained spotters. The industry has continued to support the development of an index, and recently supported the continuation of a spatially-reduced survey that solves many of the logistical problems.

The second major planned outcome that was achieved was the integration of environmental and archival data with the aerial survey data to improve the precision of the abundance index. One example was the inclusion of the surfacing rate under differing environmental conditions in one of the indices. The increase in knowledge about the topographical and environmental preferences of juvenile SBT resulting from analyses carried out during this project will continue to improve the precision of population estimates. These findings are already being included in subsequent projects to better understand the mechanisms that could lead to interannual variation in apparent abundance of SBT in the GAB.

Collectively, the outputs from this project do provide information on the status of the juvenile SBT population and in particular, should provide the first early warning signs for recruitment failure for SBT globally. The importance of the juvenile SBT abundance index has also been acknowledged by the international management body for SBT, the Commission for the Conservation of Southern Bluefin Tuna (CCSBT). Recent scientific meetings of the CCSBT have continued to support the development and emphasised the importance of the index.

Output from this project has also been recently used in a modelling project synthesizing environmental and archival tag data in an individual based model framework (Bestley and Hobday, 2002; Hobday and Bestley, 2002). The aerial survey data and the abundance index are used to condition and validate the model simulations. In particular, the individual-based model approach is aimed, in part, at evaluating the ability of the current survey design to detect changes in the SBT population. This fishery-independent dataset is critical for assessing such modelling simulations. The observations gathered during this project and the ancillary analyses of the companion studies will continue to be used prominently to understand the population dynamics of juvenile SBT within the GAB.

9.1. References

Bestley, S., Hobday, A. J. (2002). Integrated Analysis Project. Development of an individual-based spatially-explicit movement model for juvenile SBT (1) School size sub-model (RMWS/02/04 No. CSIRO Marine Research.

Hobday, A. J., Bestley, S. (2002). Integrated Analysis Project. Development of an individual-based spatially-explicit movement model for juvenile SBT (2) environment-based movement rules (RMWS/02/04 No. CSIRO Marine Research.

10. Conclusions

This project has been successful in developing a fishery-independent index of abundance for juvenile southern bluefin tuna (SBT) based on an aerial survey in the Great Australian Bight (GAB). A time series of eight years (1993-2000) of consistent data formed the basis of this index. Considerable and significant progress has been made in developing the statistical methodology underlying the indices of surface abundance derived from the aerial survey data. Environmental variables that may influence the detectibility of SBT at the surface or the presence in an area have been identified. Additional achievements have included developing an understanding of how environmental variation affects SBT surfacing behaviour and analysing SBT surfacing behaviour.

The success of the research in briefly summarised against the objectives for each of the original projects.

FRDC 1996/118

1. The aerial survey for SBT over the GAB was carried out each summer season from 1997 to 1999 and various surface abundance indices were estimated (**Section 6.1, 6.5, and 6.6**).
2. Statistical research was carried out such that
    - environmental variables that influence surface abundance were identified and incorporated into some of the estimates (**Section 6.2 and 6.5),**
    - the proportion of SBT at the surface under various environmental conditions was established and incorporated into some of the estimates (**Section 6.4 and 6.5**), and
    - the uncertainty in the index due to uncertainty in school size and fish size estimates was reduced through several approaches (**Section 6.5 and 6.6)**.
3. The usefulness of the indices of SBT abundance derived from the aerial survey was evaluated and compared with other fishery-independent surveys (**Section 6.6)**.

FRDC 1999/118

1. A range of statistical analyses of data from the archival tags and environmental archives were conducted to determine common responses in surfacing behaviour to environmental conditions through space and time (**Section 6.3)**
2. A range of statistical analyses of aerial survey data on surface distribution and surface abundance of juvenile SBT were carried out (**Section 6.2 and 6.4)** and environmental archives evaluated to develop a spatial model of abundance which allows for environmental variation through space and time (**Section 6.5**).
3. An integrated analysis of abundance of SBT in the GAB incorporating the surfacing behaviour, surfacing abundance and spatial distribution models was developed (**Section 6.5 and 6.6).**

In fulfilling the research objectives of the project, the amount of supporting data collected was far greater than originally planned. The importance, for example, of satellite-derived data such as sea surface temperature, modeled weather data from the Bureau of Meteorology, and validation data collected in multiple plane experiments all required new methodological development before they could be incorporated in the indices. These additional data allowed the development of more detailed understanding of the processes governing the appearance and detection of surface schools of SBT in the GAB and of the accuracy of the abundance estimates.

An original goal of the project was to produce an age-specific SBT biomass index. It is now apparent that the ability to differentiate SBT age classes in the aerial survey is limited, and so an aggregated index of juvenile abundance was developed in this report. Several methods of estimating surface abundance, incorporating surfacing rates of SBT and environmental conditions were developed. Integration of biological and environmental information reduced some of the uncertainty, and the CV associated with the final indices is sufficient for the index to be a useful measure of trends in abundance.

A useful index of SBT abundance based on this aerial survey data must be able to indicate trends in abundance. To allow the possibility of detecting trends, the CV around the estimates must not be too large, or else a trend will not be detectible against the uncertainty of the estimates. The final section of the results (**Section 6.6)** found that a CV of ~ 30% was possible and may be reduced below this level in the future. This level of precision is comparable to other fishery-independent surveys around the world that play an important role in the management of exploited stocks.

The utility of the SBT index with regard to the detection of long-term trends in juvenile SBT can be evaluated by roughly calculating what level of change in recruitment over a 5-year block would be detectable with surveys of varying precision (**Section 6.6)**. For the sake of argument assume that the genuine CV of SBT recruitment is 40%. Even with a perfect survey (CV = 0%), the mean recruitment over successive 5-year blocks would fluctuate randomly with a CV of around 18%. This means that a real shift of 35% in long-term mean recruitment over 5-10 years would be on the margins of statistical significance. If the survey has a CV of 30%, then the 5-year mean will have a CV of around 22% (only slightly higher than with a perfect survey), so it would require about a 50% change in long-term mean recruitment over 5-10 years to achieve marginal significance. These guidelines need to be considered when assessing whether year effects in an index are "significant"; the power of any test to detect significant changes in recruitment is low unless the changes are quite drastic. The corollary is that statistical significance is often an inadequate guide as to whether managers should take action.

The evolution in the methodological approach to the development of an abundance index has allowed for considerable flexibility in the final form of the index. Both a presence/absence index (**Section 6.5**) and a biomass-based estimate (**Section 6.6)** were developed and recommended. The patterns and trends in these two approaches were similar; they differed in the assumptions and the final units of the index. Overall, the same conclusion with regard to the trend in juvenile SBT biomass was reached. There has been a slight decline in the abundance of juvenile SBT over the period 1993-2000.

Continuation of the aerial survey is justified based on the results from this project. The aerial survey for juvenile SBT will continue to be an important monitoring tool, and in future the index derived from the survey data should be incorporated in the stock assessment of the global SBT resource.

11. Intellectual Property

No intellectual property is claimed.

12. Staff engaged on the project

Dr Ann Cowling
Dr. Alistair Hobday
Mr. John Gunn
Mr Colin Millar
Mr Jeremy O'Reilly
Dr. Mark Bravington

13. List of Appendices

Appendices are listed within each of the chapters covering the project results.