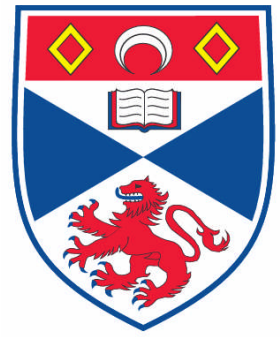
INVESTIGATIONS INTO THE USE OF QUANTIFIED BAYESIAN MAXIMUM ENTROPY METHODS TO GENERATE IMPROVED DISTRIBUTION MAPS AND BIOMASS ESTIMATES FROM FISHERIES ACOUSTIC SURVEY DATA

Ben Heywood

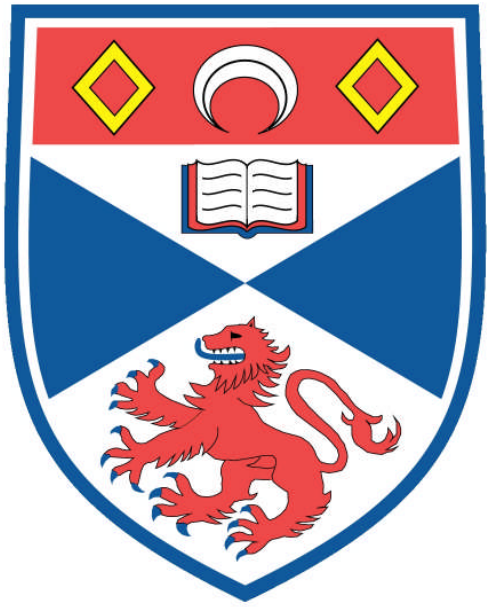
A Thesis Submitted for the Degree of MPhil
at the
University of St. Andrews



2008

Investigations Into The Use Of Quantified Bayesian Maximum Entropy Methods To Generate Improved Distribution Maps And Biomass Estimates from Fisheries Acoustic Survey Data

Ben Heywood



Thesis submitted for the degree of

MASTER OF PHILOSOPHY
in the School of Biology

University of St Andrews

September 2007

I, Benjamin Gordon Heywood, hereby certify that this thesis, which is approximately 21,400 words in length, has been written by me (except where explicitly stated otherwise), that it is the record of work carried out by me, and that it has not been submitted in any previous application for a higher degree.

Date 26/09/2007 Signature of Candidate.....

I was admitted as a research student in April 2004, as a candidate for the degree of MPhil in July 2007; the higher study for which this is a record was carried out in the University of St Andrews between 2004 and 2007

Date 26/09/2007 Signature of Candidate.....

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Master of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 26/09/2007 Signature of Supervisor.....

In submitting this thesis to the university, I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made available to any *bona fide* library or research worker.

Date 26/09/2007 Signature of Candidate.....

Contents

ABSTRACT ..... **IV**

CHAPTER 1: INTRODUCTION **1**

1.1 WHY DO WE NEED RECONSTRUCTIONS? 1
1.2 CURRENT STATISTICAL METHODOLOGIES 2
1.2.1 Simple interpolation *2*
1.2.2 Jolly and Hampton *2*
1.2.3 Kriging *3*
1.3 BAYESIAN AND FISHERIAN STATISTICS 3
1.3.1 What are Bayesian and Fisherian statistics? *3*
1.3.3 Real-world prior pdfs *6*
1.4 MAXIMUM ENTROPY IMAGE RECONSTRUCTION 7
1.4.1 The Principle of Maximum Entropy *7*
1.4.2 MemSys 5 ..... *7*
1.4.3 Why Image Reconstruction? *7*
1.4.4 MaxEnt Image Reconstruction versus Kriging *8*
1.5 RECENT WORK IN MAXIMUM ENTROPY IMAGE RECONSTRUCTION 9
1.6 DATA 10
1.6.1 The relevance of data collection methodologies *10*
1.6.2 Sources of data *10*

CHAPTER 2: METHODS **12**

2.1 MAXIMUM ENTROPY IMAGE RECONSTRUCTION AND MEMSYS 5 12
2.1.1 Choosing a prior *12*
2.1.2 The Likelihood function *12*
2.1.3 The Evidence *13*
2.1.4 Calculating the posterior probability *13*
2.2 SIMPLIFIED MEMSYS 5 ALGORITHM 14
2.3 THE INTRINSIC CORRELATION FUNCTION 15
2.4 MULTI-CHANNEL MAXENT 16
2.5 USER SPECIFIED INPUTS TO MEMSYS 5 16
2.6 ADDITIONAL METHODS 18
2.6.1 Changes made to the MemSys 5 software *18*
2.6.2 Other Software *18*

CHAPTER 3: EVALUATING THE ACCURACY OF MAXENT RECONSTRUCTIONS **20**

3.1 INTRODUCTION 20
3.2 BASIC VISUAL COMPARISON 21
3.3 A STATISTICAL COMPARISON METHOD 23
3.4 THE MAP COMPARISON KIT 26
3.4.1 Wavelet Analysis *27*
3.4.2 Image Quality Assessment *27*
3.4.3 Information weighted comparison *27*
3.4.4 Image Warping *27*
3.4.5 Bivariate Spatial Association *28*
3.4.6 Preliminary results *28*

CHAPTER 4: KRILL IN THE SCOTIA SEA **30**

4.1 INTRODUCTION 30
4.2 METHODS 33
4.2.1 What is Maximum Entropy Image Reconstruction? *33*
4.2.2 Parameter Estimation *35*
4.2.3 Calculating the Quantified MaxEnt Result *36*
4.2.4 Data Preparation *37*
4.3 RESULTS 39
4.4 DISCUSSION AND CONCLUSION 46

CHAPTER 5: BIOMASS AND DISTRIBUTION OF JELLYFISH AND FISH IN THE NAMIBIAN BENGUELA 50

5.1 WHY USE MAXENT? 50
5.2 JELLYFISH OVERTAKE FISH IN A HEAVILY FISHED ECOSYSTEM 50
5.3 SUPPLEMENTAL DATA 55

CHAPTER 6: DISCUSSION AND CONCLUSION 57

6.1 RECENT DEVELOPMENTS 57
6.2 SUGGESTED FUTURE WORK 58
6.3 CONCLUDING REMARKS 59
6.4 ACKNOWLEDGEMENTS 59

BIBLIOGRAPHY 60

Abstract

This thesis describes elements of the assessment and application of Bayesian Maximum Entropy (MaxEnt) image reconstruction techniques for the analysis of fisheries acoustic survey data. The objective is to investigate the utility of this approach in mapping density distributions and estimating biomass. The MaxEnt image reconstruction method derives originally from the field of astrophysics, and this thesis represents an attempt to apply the principles of MaxEnt to the field of ocean ecology. Essentially, what is required is to generate maps of the density distribution of pelagic species (species living in the water column) from extremely limited and sometimes skewed line-transect acoustic survey data. Techniques used presently are largely unsatisfactory for a variety of reasons, and are often inapplicable for data from surveys that do not follow a particular design strategy. This thesis investigates the usefulness of the MaxEnt technique in overcoming some of the difficulties of acoustic survey analysis. A study is made into the possibility of objectively testing whether these techniques offer improvements in accuracy over existing techniques, by attempting to reconstruct simulated data from a virtual survey. I find that plausible reconstructions are possible, and that statistical comparisons indicate these reconstructions are accurate. The technique is also applied quantitatively to real-world survey data, offering new insights into the abundance of Antarctic krill (*Euphausia superba*) in the Scotia Sea - raising abundance estimates from 109 million tonnes to 208 million tonnes - and into the relative abundance of fish and jellyfish in the Namibian Benguela, where it is shown that the biomass of jellyfish (12.2 million tonnes) now exceeds that of fish (3.6 million tonnes).

Chapter 1: Introduction

1.1 Why do we need reconstructions?

Biomass and distribution data for a wide variety of pelagic (i.e. living in the water column) species are collected by acoustic survey (e.g. Burczynski and Johnson 1986; Greenstreet et al. 1997; McGehee et al. 1998; Brierley et al. 1999); sound waves travel much further than light waves in sea water, and sonar enables a larger volume of the sea to be studied in a given time than trawl fishing (Maclennan and Simmonds 1992). Also, fishing vessels themselves use echosounders to locate schools of fish, and these data can sometimes be aggregated with the data from scientific surveys (Maclennan and Simmonds 1992). Survey data are usually recorded along the (approximately) straight-line, regular or semi-randomly-spaced transects (Jolly and Hampton 1990) traversed by the survey ship, and usually deliver data from only a fraction (0.5%-5%) of the survey area (echosounder beams are only a few degrees wide, to a depth in the order of 100m, and the transects are spaced a number of kilometres apart). In a perfect world, we would have either an infinity of survey vessels, or a stationary species to study, so that we could accrue all the data across the whole survey area and produce a fine-scale map of the density distribution of a species. In practice this is of course not the case, and hence methods are needed to extrapolate from the minimal along-transect data we have to reconstruct the overall distribution. Furthermore, since this extrapolation will be an approximation, it would be extremely useful to have a measure of our confidence in the reconstruction.

The problem is unarguably a difficult one, since the results are heavily under-constrained by the data. The difficulty is greatly increased due to the profoundly skewed (rare but biologically important extreme values from shoals) and exceedingly patchy (rare, dense, discrete swarms and schools) nature of pelagic species density distributions. Successful reconstructions, if they can be achieved, will be of obvious importance to fisheries management bodies (Hewitt et al. 2002), as well as greatly assisting understanding of the ecology of (and the man-made changes within) the pelagic realm.

The use here of a Bayesian Maximum Entropy (MaxEnt) technique has two objectives:

i) It is hoped that the MaxEnt method can be used to generate more accurate maps (see Christakos and Li 1998) of the density distribution of a target species than are currently possible with geostatistical methods (Maravelias et al. 1996; Rivoirard et al. 2000). Essentially, where a ship collects acoustic data along a straight-line transect, there is a need to 'fill in the gaps' between transects. Such maps are difficult to produce from the very limited data collected, but would be of obvious use to ecologists in showing the location and size of aggregations of animals. For example, a swarm of organisms observed in the same place (or in the same depth or temperature of water for example) over a number of surveys would have evident implications for our understanding of the behaviour of the creature.

ii) It is hoped that the MaxEnt method can be used to generate improved estimates (with, in particular, improved error-bars) for the overall biomass of the target species in the survey area, compared to the standard estimation method of Jolly and Hampton

(1990). Whilst a knowledge of total biomass is also important to ecologists, the most pressing need for this information is in fisheries management and quota determination. More accurate biomass estimates could generate more appropriate quota levels for the fishing community, enable faster and more definite identification of upward or downward population trends, and reduce the risk of over-fishing.

The project itself also has two aims:

i) to demonstrate the extent to which MaxEnt can (or can not) achieve these improvements over existing techniques;
ii) to apply the MaxEnt technique to data in the real world.

Accordingly, this thesis will proceed along the following lines. The remainder of this chapter will introduce the currently used techniques and the basic concepts of MaxEnt. Chapter 2 will provide the technical details of the application of the MaxEnt technique. Chapter 3 will discuss how accurate the MaxEnt reconstructions are, and Chapters 4 and 5 will show the use of the technique in real world situations. Finally Chapter 6 will look at recent developments, offer some concluding remarks, and discuss future possibilities.

1.2 Current statistical methodologies

1.2.1 Simple interpolation

The simplest method for generating a biomass estimate from our on-transect values is to take a simple mean of the data and apply it to the whole survey area. Because of the difficulties discussed above, and particularly because of the skewed data, this is far from satisfactory, since it makes no attempt to deal with the possibility that the transects may not pass through a representative sample of the survey area and essentially ignores any spatial structure in the data. Similarly, calculating a simple standard deviation from the data does not offer a good estimate of possible errors, again due to the highly skewed nature of the data.

Correspondingly, the simplest way to reconstruct the distribution is to interpolate between the known values in a linear manner. Again, we know that this is very unlikely to produce a distribution map that closely mirrors reality, due to the skewed data, and so we must examine other methods.

1.2.2 Jolly and Hampton

Jolly and Hampton (1990) describe a statistical method for generating biomass estimates from acoustic surveys (no attention is paid to generating distribution maps) based on following particular rules when designing the survey. This system has now become standard for the analysis of most acoustic surveys.

The method involves generating weighted means for a series of semi-randomly placed transects. The method is sound from a Fisherian statistical viewpoint (see ‘Bayesian and Fisherian statistics’ below), but can appear confusing. It is an example of the archetypal frequentist methodology of ‘plugging in the numbers’ – the method will generate a biomass estimate from given data, but not in a way that allows for

innovative ideas. The survey design must follow rigid guidelines, and there is no way to analyse or to incorporate data from non-standard surveys or from other external information.

In addition, this method offers no help in generating a distribution map from the survey data.

1.2.3 Kriging

Standard geostatistical analyses of line-transect data have generally failed, particularly where distributions are highly skewed and patchy (Maravelias et al. 1996; Murray 1996), which is normally the case for Antarctic krill and many other pelagic species. The most widely used modern geostatistical approach is *kriging* (after its inventor D. G. Krige), a method for reconstructing missing data originally developed in the mining industry. Kriging is the most direct alternative to the Maximum Entropy approach (for a full description of kriging methods, see Isaaks and Srivaslava 1989; Rivoirard et al. 2000). Essentially, kriging works by creating a matrix containing the variance (loosely, the uncertainty in the estimate of the value) at each point in the reconstruction area, and then using these variances as weighting functions when estimating the missing data values. The original data remain unchanged, although their associated variance does alter. The objective of the algorithm is to minimise the variances of the known data and hence estimate the 'best' value for each missing data point (RSINC 2000).

Kriging can be used to generate both distribution maps and biomass estimates. However, there are still some advantages to the MaxEnt approach. In order to discuss these, it will first be necessary to explain briefly the MaxEnt approach and in particular the Bayesian statistical ideas on which it is based.

1.3 Bayesian and Fisherian Statistics

1.3.1 What are Bayesian and Fisherian statistics?

There are two major contrasting approaches to statistics, known after their original proponents as Fisherian (Fisher 1925) and Bayesian (Bayes 1763; de Laplace 1812; Jeffreys 1939). Fisherian statistics are essentially frequentist and Bayesian statistics are probabilistic. Fisherian statistics have dominated in scientific research since their inception due to their *a priori* objectivity, as they deal with long-run frequencies that can be *measured*.

Frequentist statisticians can assign probabilities only to events, whether hypothetical or real, and not to arbitrary statements such as 'There was life on Mars a million years ago.' Frequentists cannot define a suitable sample space and experiment that could possibly relate to this question. Excluding such seemingly unscientific questions from Frequentist analysis has not prevented Fisherian statistics from achieving a great deal in the advancement of science, but in recent years (most particularly since the work of

Jaynes (see Jaynes 2003 for a summary)) the Bayesian approach has increasingly been investigated and applied.

Bayesianism is predicated on indisputable underlying principles – the axioms of probability theory, from which it is readily derived. It has been shown (Cox 1946) that these axioms follow directly from Boolean logic and ordinary algebra; and, further, that any other valid form of probabilistic inference is logically equivalent (i.e. there exists a one-to-one mapping) to these standard probability rules. Since everything is derived from a single solid argument, Bayesianism does not suffer from the apparently *ad hoc* "cook-book approach to data analysis" (Sivia 1996, p.1) that bedevils the teaching of Fisherian statistics. Bayesian statistics are not necessarily easier to use (in fact the opposite is quite generally the case, since one can usually select a ready-made frequentist approach from a 'tool-kit' of different methods) but Bayesianism does give a scientist a quantitative idea of *how much* faith they should have in their results, and it does give them the chance to more accurately fit their statistical methods to their situation, and to describe the probability of anything that can be described, whether a one-off event, historical conjecture, or whatever (for arguments in favour of frequentist statistics, see Efron 1986; Dennis 1996).

Bayesian statistics is, as mentioned, based upon probability theory, and in particular Bayes' theorem:

$$p(hypothesis \mid data, I) = [p(data \mid hypothesis, I) * p(hypothesis \mid I)] \; / \; p(data \mid I)$$

[the following terminology is often used for these probabilities:

$$posterior\ probability = (likelihood * prior\ probability) \; / \; evidence \quad]$$

where '*p*' denotes a probability, '|' means 'given', a comma means 'and', and *I* represents the entirety of all relevant background assumptions and information. It is formally correct to include *I* to remind oneself that probabilities are *always* conditional on our state of knowledge, but it is usually omitted from calculations to avoid algebraic clutter, and will generally be omitted in the rest of this thesis.

This formulation of Bayes' theorem can be very useful – what we normally want to know is the probability of our hypothesis in the light of our data (the posterior probability, $p(hypothesis \mid data)$) and the other three terms are generally much easier to assign, as we will see below. The prior probability represents our knowledge or ignorance about the hypothesis *before* we look at our data; this is modified via the likelihood function to give the posterior probability, which represents our adjusted beliefs in the light of our data. The evidence is simply a normalisation constant that can be used to discriminate between competing hypotheses – e.g., we may have a similar amount of faith in two incompatible hypotheses, but when they are applied to our data (via Bayes' theorem) it is likely that one hypothesis will generate a higher evidence value than the other, and this hypothesis will be preferred.

It is important to understand that in general we deal with a distribution of probability, or pdf (*probability density function*), rather than a point probability. Bayes' theorem for pdfs is identical to the point probability theorem shown above. The notation '*p'* will be used to denote a point probability, and 'Pr' will denote a probability density function.

The biggest challenge in a Bayesian analysis is to set the distribution of the prior pdf. This can be done with a little thought in simple cases: for example, a sensible prior pdf for the throw of a die that is assumed to be fair can be derived as follows:

p(of getting a 1) = p(2) = … = p(6) = 1/6
so the pdf Pr*(x)* (where x = 1 to 6) is [1/6, 1/6, 1/6, 1/6, 1/6, 1/6]

This covers all possible outcomes and is intuitively obvious, on the basis of symmetry and because (by convention) the total probability of all possible outcomes is 1.

1.3.2 Example of using Bayes' Theorem (adapted from Skilling 2003)

Suppose we observe a system X that could be in any of four states (X_1, X_2, X_3, X_4). These four states are equivalent to four competing hypotheses – one of them (and only one) is the actual situation. We would like to know, in the light of our data, the probability of each of these states, so we use Bayes' theorem (substituting X_j for *hypothesis*).

We would begin by assigning individual probabilities $p(X_j)$ to each of the four states. Assuming that we have no other relevant information (I, in the theorem) about which state is most likely, we must assign

$$\text{Prior pdf} = \Pr(X_j) = \text{All the } p(X_j) = [¼, ¼, ¼, ¼]$$

To observe X, we have equipment whose response can be described probabilistically as giving output (i.e. data) D_1, D_2 ... D_i when the real state is X_1, X_2 … X_j. This information would be obtained from a calibration exercise. For example,

$\Pr(D_i \mid X_j)$ = Likelihood =

7/10	1/10	4/10	3/10	
1/10	5/10	3/10	2/10	↓ D(i)
2/10	4/10	3/10	5/10	

X(j) →

This can be interpreted as follows: if state X_1 (for example) is occurring, we expect the equipment to respond with D_1 70% of the time, with D_2 10% and with D_3 20%. If we already had data, we would of course use only that row (D_1, D_2, or D_3), but here we're interested in the whole posterior distribution.

The analysis proceeds as follows. The first stage is to construct the joint probability of all the possible data values and 'real' values using the product rule *[Pr(a,b) = Pr(a|b)Pr(b)]* of probability theory:

Joint = (likelihood * prior) = $\Pr(D_i, X_j) = \Pr(D_i \mid X_j)\ \Pr(X_j)$ =

7/40	1/40	4/40	3/40
1/40	5/40	3/40	2/40
2/40	4/40	3/40	5/40

i

j

The columns $\Sigma_i \Pr(D_i, X_j)$ sum to the prior (each $X_i = ¼$). The rows $\Sigma_j \Pr(D_i, X_j)$ sum to the *evidence* $\Pr(D_i)$ [or *p(data)* in the notation for Bayes' theorem above]. This makes sense – the total probability for $\Pr(D_i$ **AND** $X_j)$ over all possible X_j should be just $\Pr(D_i)$. Then:

Evidence = $\Pr(D_i) = \Sigma_j \Pr(D_i, X_j) =$

15/40
11/40
14/40

We then use the evidence to *normalise* the joint probability to get the *posterior* distribution for X, which is what we've been after.

Posterior = *(likelihood * prior probability) / evidence = joint / evidence* =

$P(X_j \mid D_i) = (D_i, X_j) / P(D_i)$ =

7/15	1/15	4/15	3/15
1/11	5/11	3/11	2/11
2/14	4/14	3/14	5/14

i

j

So, for instance, if our measurement returns D_1, the probability that the real occurrence is X_1 is 7/15; of X_2 is 1/15; of X_3 is 4/15; of X_4 is 1/5.

Try to picture what's going on with a real-world example. Imagine, for example, that the measurement equipment described by the above likelihood function is a thermometer giving data $D1$ to $D3$ as 15.1°C, 15.2°C, and 15.3°C; and that states X_1 to X_4 are temperatures 15.20°C, 15.24°C, 15.26°C, and 15.29°C. From a given thermometer reading, we could then read off the probability of each possible temperature.

1.3.3 Real-world prior pdfs

Unlike in the simple assignment $\Pr(X_j) = [¼, ¼, ¼, ¼]$ above, in any more complex problem we find that other factors (the information I above) lead us to believe that a flat uniform prior pdf is not applicable – some possible results will be more likely than others. The simplest example of this would be a die that we had reason to believe may *not* be fair. In the real world, there may be a large number of apparently

justifiable but significantly different possible assignments for the prior pdf. Some non-arbitrary method of choosing between them is required.

1.4 Maximum Entropy Image Reconstruction

1.4.1 The Principle of Maximum Entropy

Assigning a prior in these more realistic situations can be very difficult. For theoretical reasons (Skilling 1988a; Sivia 1996 pp.113-120) when a variety of different pdfs are consistent with the constraints imposed by our knowledge *I*, we should choose the pdf which maximises the entropy *S*. This is the principle of Maximum Entropy (Cox 1961; Jaynes 1978): for different possible priors, maximise

$$S = \sum_{i=1}^{n} p_i \log_e [p_i]$$

where n is the number of competing hypotheses (infinity in continuous cases) and p_i is the single (i.e. not a pdf) prior probability *p(hypothesis)* for hypothesis *i*. (Note that possibilities 1 to n must be mutually exclusive and exhaustive of the probability space; i.e. all possibilities are covered exactly once and without overlap.) A derivation of the principle of Maximum Entropy can be found in Chapter 4, section 4.2.1.

1.4.2 MemSys 5

The application of Maximum Entropy methods has only become possible recently as computing power has increased, since they require a great many iterations to solve intractable equations numerically. The computer program *MemSys 5* has been developed by Steve Gull and John Skilling (see Gull and Skilling 1991) as a generic system to be used for a multiplicity of different problems. Various parameters can be set by the user to create a Maximum Entropy problem-solver for their particular data. Crucially, as well as producing a reconstruction using Bayes' theorem and the principle of Maximum Entropy, this software calculates the Bayesian *evidence* value, which enables objective comparison of different settings and hypotheses. The *MemSys 5* software is the main computing tool used in this project. Chapter 2 presents a more detailed description of MaxEnt and *MemSys 5*.

1.4.3 Why Image Reconstruction?

The problem of reconstructing a density distribution map over a given survey area is analogous to the reconstruction of an image. A set of point values on a two dimensional grid can naturally be viewed as a pixellated image. We can imagine the area of ocean involved in the survey as being divided into a grid of cells of a certain size, in just the same way as we see an image on a computer screen divided into pixels. A greyscale image can be defined by the *x* and *y* position, and the intensity, of each pixel; similarly, a distribution map of a survey area can be defined by the latitude and longitude of, and the density within, each cell.

It is common to display a map of density distribution as a colour-scale image (e.g. Hewitt et al. 2004a), where higher pixel intensity equates to higher density in that cell.

The reason this is so natural is that a cellular density distribution and a pixellated image are, at a basic level, mathematically identical.

The problem of filling in missing data in a survey area is, then, directly analogous to the problem of reconstructing a damaged photograph, or improving the resolution of astronomical images, and therefore reconstruction techniques from these disciplines may sensibly be applied to these biological data. Great success has already been achieved in image reconstruction using Maximum Entropy techniques in these other fields (e.g. Gull and Daniell 1978; Burch et al. 1983; Gull and Skilling 1984; Marshall et al. 2002).

In our case, an acoustic survey records values for those 'pixels' that the survey vessel passes over, and we then use the MaxEnt image reconstruction technique to generate a value for the density in every cell. Once this has been calculated, it is trivial to sum the biomass in each cell to arrive at an estimate of total survey area biomass.

1.4.4 MaxEnt Image Reconstruction versus Kriging

Although a fuller description of the MaxEnt method must wait until the next chapter, it should be possible with the brief outline given so far to discuss the possible advantages of MaxEnt. The biggest advantage is the ability to use the Bayesian evidence value to set parameters and to measure the relative accuracy of one reconstruction against another (this is obviously dependent on showing that the evidence value *does* reliably enable us to choose the most accurate reconstruction – see Chapter 3).

Both kriging and Maximum Entropy approaches provide some measure of the accuracy of the resulting reconstruction – in MaxEnt through the Bayesian *evidence* value, and in kriging through the output of the matrix of variances (in fact, if required, a similar matrix of errors can be derived for the MaxEnt reconstruction (Bontekoe et al. 1994)). The matrix of variances does give a good impression of which regions of the image are more believable, but the *single* evidence value from MaxEnt provides a simple criterion for choice between parameters, hypotheses, etc. Also, and importantly, a relative layman in the fishing industry or in politics can readily understand the concept of one reconstruction having better evidence than another. Another advantage of MaxEnt over kriging is its ability to deal with given data points very close to each other. In kriging, the algorithm requires the matrices it uses (internally, during calculation) to be invertible, but proximate data points can often result in singular matrices, rendering calculations impossible. It is necessary then to choose only one of the data points and discard the other. The MaxEnt techniques do not suffer from this problem.

One possible advantage of kriging over the MaxEnt approach is the constancy of the given data values. It seems intuitively reasonable not to change the values that were actually measured. As a rule, however, the measured values have an uncertainty associated with them, and it is easy to imagine that the overall probability of a reconstructed area can be increased by adjusting an outlying measurement. By fixing the input data exactly, we may sometimes force missing values in the vicinity of an improbable measurement to have unnecessarily large variances (of course, with schooling and swarming species, high isolated readings need not be outliers, and they

should not be overly dragged down by a low average reading. MaxEnt only adjusts such values via the *likelihood* function, which is dependent on the accuracy, σ, of the data. Any further changes to the original data values would make the reconstruction *inconsistent* with the data, and hence it would be rejected).

1.5 Recent work in Maximum Entropy Image Reconstruction

Biologists and ecologists have only begun relatively recently to apply MaxEnt techniques to their data analysis. This is not too surprising, as the landmark paper in astrophysics was only published in 1978 (Gull and Daniell) and the wider applications were not immediately obvious. Some work has now been undertaken, however, and it seems likely that exponentially more will be carried out in the future.

A 1998 paper (Vignaux et al.) successfully used MaxEnt techniques to produce distribution maps of the density of New Zealand hoki at a *finer* scale than the original unit of effort (in this case length of trawl, as this was not an acoustic survey). The authors achieved this by using the density information inferred from the points where different tows crossed each other. For example, two trawls which pass through the same high-density area will both have an above average catch, and vice versa. If enough trawls cross, as they do in the heavily-fished area off the west of New Zealand's South Island, then reasonably complicated inferences about points common to more than one trawl can be achieved. The authors conclude that the MaxEnt approach 'is a useful technique that allows fine-scale maps of fish density distribution to be generated from [commercial fishing] data.' (p.1226).

More recently, papers have been produced (Brierley et al. 2003b; Wafy et al. 2003) dealing specifically with the kind of reconstructions this project will study. The two papers present map reconstructions of krill distribution and density in two 80×100 km survey boxes near South Georgia. Inevitably, the plausible reconstructions detailed in these two papers cannot be compared to the actual distributions since these remain unknown, but they showed promise in a variety of ways. Firstly, it seems possible to generate a reconstruction in which the original transect positioning is not obvious. This is a clear prerequisite of a successful reconstruction, as we have no reason to expect on-transect values in the real world to differ significantly from the off-transect values. Secondly, using the reconstructions to calculate biomass and mean density produced plausible values. Thirdly, a visual inspection of the reconstruction showed that (biologically important) high density values were well preserved, and that qualitative features of the original data were reproduced in the reconstruction (for example, isolated high values, or regions of medium-high values). Fourthly, some known data points were left out of the analysis, and the reconstructions of these points were not significantly different from their true values (according to a Wilcoxon signed-rank test).

This project investigates whether there is a solid basis for the *prima facie* plausibility of the Wafy, Brierley et al. (2003) results. Those authors themselves suggest that further work should be done with heavily skewed, extremely patchy, realistic simulated data (i.e. simulated over the whole survey area, from which virtual surveys can be taken). Genuine comparisons can then be made between the reconstructions and the 'real' distribution from which the sample data were drawn..

1.6 Data

1.6.1 The relevance of data collection methodologies

This project focuses on pelagic data collected by acoustic surveys. There is a great deal of pre-processing associated with this data, notably in calibration and in the differentiation of species (Maclennan and Simmonds 1992; Hammond and Swartzman 2001; Korneliussen and Ona 2002). However, the methods employed to produce the data used in this project are almost irrelevant to the project itself, except insofar as they enable us to estimate statistical noise (as opposed to acoustic noise, which has been accounted for in the pre-processing) and errors in the data. The MaxEnt reconstructions deal only with numbers for x-position, y-position and value, and their correlations with each other – the numbers could represent krill density, herring density, pixel brightness; it makes no difference (except in that we expect the data to be skewed and patchy – smooth data wouldn't require this sort of reconstruction). Assuming that the species does not either avoid or follow the survey ship (see Brierley et al. 2003a), which would obviously make any reconstruction disastrously inaccurate, there are only two aspects of the data collecting *methods* that substantially affect this research:

a) The anticipated errors associated with the data, which constrain the accuracy of the reconstruction (and of course help to decide in advance what level of accuracy to *require* of the algorithm, to avoid wasting processing time);

b) The spatial resolution of the survey must be considered. Data are integrated at intervals along the cruise track to generate a biomass for that cell or pixel, and the length of along-track interval defines the resolution of the survey results. The resolution of the image reconstruction should obviously be based where possible on this information.

There are of course other factors that determine whether the data are *suitable* for this form of reconstruction. One is the mathematical distribution of the data, which must be capable of being transformed to an approximately Gaussian form. This is because the *Memsys 5* system assumes measurement noise to be Gaussian, and hence uses a Gaussian form for the likelihood function, as described in Chapter 2.

1.6.2 Sources of data

Simulated data (kindly provided by John Simmonds (Simmonds et al. 2002) - see Figure 1.1) is used to test the success of the MaxEnt approach, so that reconstructions generated from virtual surveys through the simulated data space can be compared with the known 'true' distribution. This mock data is based on a computer model of the movement of schools of herring in the North Sea, and is in fact a snapshot of that dynamic model.

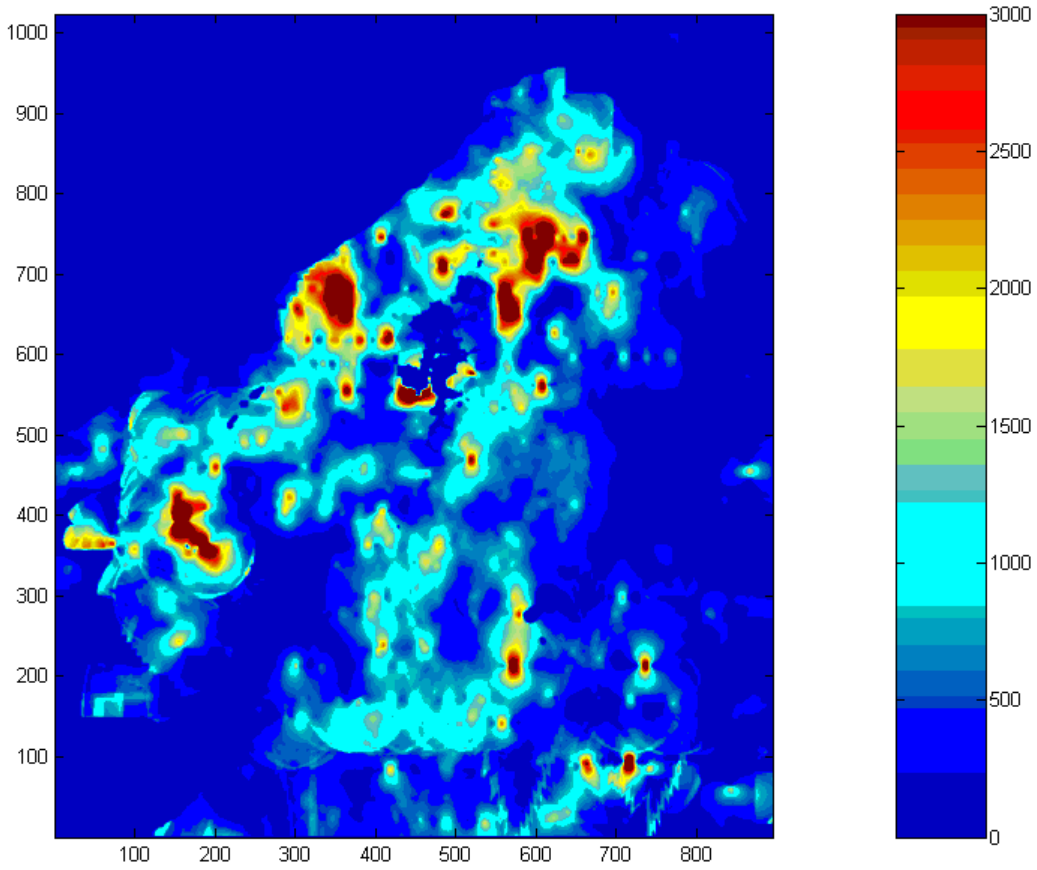


Figure 1.1: A Representation of simulated herring density distribution in the North Sea around the Shetland Islands
(Axes in pixels of the reconstruction [roughly 440m, origin at approx 58°N, 4°W]; Colour map: density of herring in arbitrary units; values over 3000 shown as 3000)

I have also had access to a number of real-world data sets. Two of these are analysed in Chapters 4 and 5. The first, kindly provided by David Demer, is the result of the 2000 CCAMLR (Commission for the Conservation of Antarctic Marine Living Resources) survey of the Scotia Sea (Trathan et al. 2001) and consists of point estimates of krill density. The second data set is the result of a survey, by the Pelagic Ecology Research Group, of the Benguela off the coast of Namibia (Lynam et al. 2004) and consists of estimates of the density of two jellyfish species and a number of fish species.

Chapter 2: Methods

2.1 Maximum Entropy Image Reconstruction and MemSys 5

This is a relatively in-depth description of the MaxEnt method. However, a fuller understanding can be gained from Gull and Skilling (1991) and Sivia (1996), and a more mathematical treatment can be found in Wu (1997).

2.1.1 Choosing a prior

Overall, we need the most probable image, i.e. the maximum of Pr(**h**|D), the posterior pdf of all possible distributions **h**, where each discrete h_i is the quantity in pixel i and D is the data.

To use Bayes' theorem to get Pr(**h**|D), we need first to assign the prior Pr(**h**). According to the Principle of Maximum Entropy (Jaynes 1978; Skilling 1988a; see also Sivia 1996 pp.113-120), the best set of proportions p_i (i = 1,2,…, L) on L *a priori* equivalent cells must be obtained by maximising the entropy

$$S(\mathbf{p}) = -\sum_{i=1}^{L} p_i \log p_i$$

The generalisation (to distributions **h** which need not sum to 1) is to find the most probable **h** by maximising

$$S(\mathbf{h}) = -\sum_{i=1}^{L} (h_i - m_i - h_i \log(h_i / m_i))$$

where **m** is the prior estimate or model, so that each m_i is the default value assigned to cell i (*n.b.* the justification for this formulation of the entropy can be found in Skilling (1988a)).

In this case, however, we need to do more than find the most probable **h**: we need the full prior pdf Pr(**h**). Now, if the most probable **h** is *always* found by maximising S, then Pr(**h**) must be a monotonically increasing function Φ of S

$$\Pr(\mathbf{h}) = \Phi(S(\mathbf{h}))$$

It can be shown (Skilling and Gull 1989) that Φ must be of exponential form

$$\Phi(S) \propto \exp(\alpha S)$$

where α is an unknown constant. Although as written here this is a proportionality, an equality can in fact be determined (Skilling and Gull 1989). However, the full equation is rather complicated and not of direct relevance; what matters is that we can formulate our prior, Pr(**h**), in terms only of **h** and one *unknown* parameter, α.

2.1.2 The Likelihood function

To use Bayes' theorem, we also need the likelihood. Noise in the data can reasonably be assumed to be normally distributed, and thus a Gaussian likelihood function will be adequate. Therefore, we take

$$\Pr(D|\mathbf{h}) = \frac{e^{-L(h)}}{Z_L}$$

where N is the number of measurements, and

$$Z_L = \int e^{-L(h)} d^N D$$

and

$$\begin{aligned} L(\mathbf{h}) &= \tfrac{1}{2} (D-\mathbf{F}(\mathbf{h}))^T [\sigma^{-2}] (D-\mathbf{F}(\mathbf{h})) \\ &= \tfrac{1}{2} \chi^2(\mathbf{h}) \end{aligned}$$

where $\mathbf{F}(\mathbf{h})$ are mock data generated from the current estimate of $\mathbf{h}$ (so that $L(\mathbf{h})$ is related to the mismatch between the real data and the mock data).

2.1.3 The Evidence

The Bayesian *evidence* Pr(D) can be determined from the prior and the likelihood:

$$\Pr(D) = \sum_h \Pr(\mathbf{h},D) = \sum_h \Pr(\mathbf{h}) \Pr(D|\mathbf{h})$$

2.1.4 Calculating the posterior probability

Once we have the prior, the likelihood and the evidence, we will be in a position to calculate the posterior probability using Bayes' theorem.

$$\Pr(\mathbf{h}|D) = \Pr(\mathbf{h}) \Pr(D|\mathbf{h}) / \Pr(D) = \Pr(\mathbf{h},D) / \Pr(D)$$

We only know the prior in terms of α, so that the joint probability $\Pr(\mathbf{h},D)$ should be written as

$$\Pr(\mathbf{h},D) = \Pr(\mathbf{h},\alpha,D) = \Pr(\alpha) \Pr(\mathbf{h}|\alpha) \Pr(D|\mathbf{h},\alpha)$$

$\Pr(\mathbf{h}|\alpha)$ is the entropic prior, and $\Pr(D|\mathbf{h},\alpha)$ [= $\Pr(D|\mathbf{h})$, since $\mathbf{h}$ itself induces the data] is the likelihood function; both of these are known. Therefore the only unknown in this expression of $\Pr(\mathbf{h},\alpha,D)$ is $\Pr(\alpha)$. According to Gull and Skilling (1991), with realistically large datasets "$\Pr(D|\alpha)$ is so sharply peaked that it overwhelms any plausible prior on α" so we select the 'best' value α_o [at the maximum of $\Pr(D|\alpha)$] and use this in our further calculation. Knowledge of the whole distribution $\Pr(\alpha)$ is then unnecessary.

To calculate the optimal reconstruction $\hat{\mathbf{h}}(\alpha_o)$ [*i.e.* the most probable **h,** given the most probable α] an iterative algorithm is used. Two more things are necessary to understand this algorithm as it is used by the software.

First, it is necessary to be able to find the best $\hat{\mathbf{h}}(\alpha)$, the best distribution $\mathbf{h}$ at a particular value of α. This is defined by the maximum of

$$Q(\mathbf{h}) = \alpha S(\mathbf{h}) - L(\mathbf{h})$$

(Gull and Skilling 1991) which is unfortunately intractable. A quadratic approximation to *Q* is used, which can be considered to be accurate within a trust region near **h**.

Secondly, we need some way to calculate when the correct value of α_o has been approached to sufficient accuracy. The correct stopping value can always be defined by an expression of the form

$$\Omega(\alpha) = 1 \quad \text{with} \quad d\Omega / d\alpha < 0$$

where $\Omega = G / (-2\alpha S)$, and *G* can be considered as the number of 'good' (i.e. properly-fitted) data (Gull and Skilling 1991, p.19).

2.2 Simplified MemSys 5 algorithm

Essentially, the algorithm proceeds as follows:

Set $\alpha = \infty$, at which point $\hat{\mathbf{h}}(\alpha) = \mathbf{m}$ (the default model)

Adjust α towards the stopping value α_o
(by an amount consistent with not shifting **h** *beyond the current trust region)*

Adjust **h** away from **m** toward $\hat{\mathbf{h}}(\alpha)$ for the current value of α
(*this is performed by the conjugate gradient method (Barrett et al. 1994; Wu 1997, p.179-80))*

Calculate and display various intermediate statistics – evidence [= $\log_e \Pr(D|\alpha)$], entropy etc., based on the current estimate of **h**

Repeat steps 2, 3 and 4 until a termination criterion is met.

[The termination criterion is based on the mismatch (calculated as the cross-entropy, see Rubinstein and Kroese 2004) between the 'correct' probability 'cloud' generated from $\hat{\mathbf{h}}$ and the 'cloud' generated from the current estimate of **h** (Maximum entropy clouds are explained in Gull and Skilling 1991, p.20). The result of some complicated mathematics is that the expected cross entropy *due to noise* equates to *G*/2, and therefore termination occurs when the cross entropy reaches *G*/2.]

This algorithm is crucial to the understanding of how the software works, and is easily misunderstood (Wu 1997, p.180 offers an incorrect summary, for example). The two iterations (on **h** and α) are not *nested* but *interlaced*. It is *not* the case that α is held constant while a value for $\hat{\mathbf{h}}(\alpha)$ is found – the two iterations are performed at the same time, so an adjustment to **h** (approaching $\hat{\mathbf{h}}(\alpha)$) is followed by an adjustment to α *before* $\hat{\mathbf{h}}(\alpha)$ is reached (Gull and Skilling (1991), Gull (pers. comm.)).

2.3 The Intrinsic Correlation Function

As described, the Maximum Entropy formalism does not account for any correlation between pixels. It is one of the axioms of Maximum Entropy that this should be the case (Skilling 1988a), but clearly in an image reconstruction problem some assumption of correlation is necessary; otherwise there would be no reason to change off-transect pixels from the default value. Fortunately, correlation (schooling and swarming behaviour) is a known feature of pelagic ecological systems, and so there is a sound biological basis for this assumption and a need for its inclusion.

In order for the entropic prior on **h** to be valid, **h** itself must be *a priori* uncorrelated. Therefore, the correlation must be encoded in a blurred version of a *hidden space* containing **h**. The blurring function used is called the intrinsic correlation function (*ICF*), and the blurred result exists in *visible space*. [The ICF currently used is a B-spline point-spread function, which approximates a Gaussian distribution but is considerably quicker to compute. The ICF takes a parameter for the *width* which determines the extent of the blurring, and the evidence value can be used to discriminate between different widths.]

In fact, three domains are required – hidden space, visible space, and *data space*. The blurring ICF transforms between hidden and visible space, and a function called OPUS (representing the response function of the instrument which generated the data) transforms between visible space and data space (see Figure 1.2). [Note that OPUS would represent any blurring due to the atmosphere above a telescope lens, or some other response function associated with gathering a whole image of data – it has nothing to do with the instrument that collected the acoustic data in our case, the errors in which are encoded in the likelihood function. The OPUS function used is identical to the ICF in our implementation of the *MemSys 5* software.] Both OPUS and ICF require reverse transforms, which can be considered as the *transpose* of the response matrix of these functions, and are called TROPUS and TRICF.

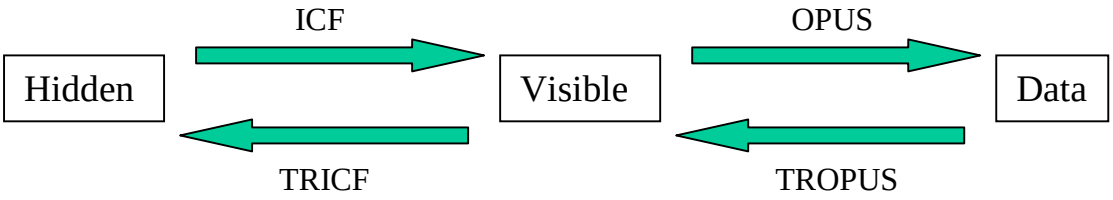


Figure 2.1

The Maximum Entropy prior remains valid because the correlation does not apply in the hidden space where **h** lies. The blurring *does* affect the posterior probability, however, since the mock data **F**(**h**) [used in our formulation of $L(\mathbf{h})$ above and which naturally have an effect on the shape of $Q = \alpha S(\mathbf{h}) - L(\mathbf{h})$, and therefore on $\hat{\mathbf{h}}$] are drawn from the data space. This is how the choice of ICF affects the choice of optimal solution.

2.4 Multi-Channel MaxEnt

The above should constitute a sufficient explanation of how the Maximum Entropy method is applied. There is, however, one major extension to the algorithm that should be mentioned.

The choice of ICF width – i.e. the range of the blurring – will cause the solution to prefer structure at a certain length-scale. In many problems, there are structures of many different length scales which are not noise, some of which may be unreasonably suppressed by the choice of ICF width (even though that width produced the highest overall evidence value).

Weir (1991) pioneered the use of multi-channel MaxEnt using the MemSys software. With this technique, a number of hidden reconstructions are created, each with a different blurring width, each of which will therefore emphasise structure at different length-scales. The final result is then a weighted combination of these images. The software currently produces, for *n* channels, reconstructions with ICF widths 0, 1, 2, 4, … ,2^{n-1}, with each weight a factor of four larger than the previous one; these are essentially *ad hoc* choices, and any combination of weights and blurring widths can be used. The discriminator between various possibilities, and indeed between the choice of single- or multi-channel operation, is again the improvement or otherwise in the evidence value.

2.5 User specified inputs to MemSys 5

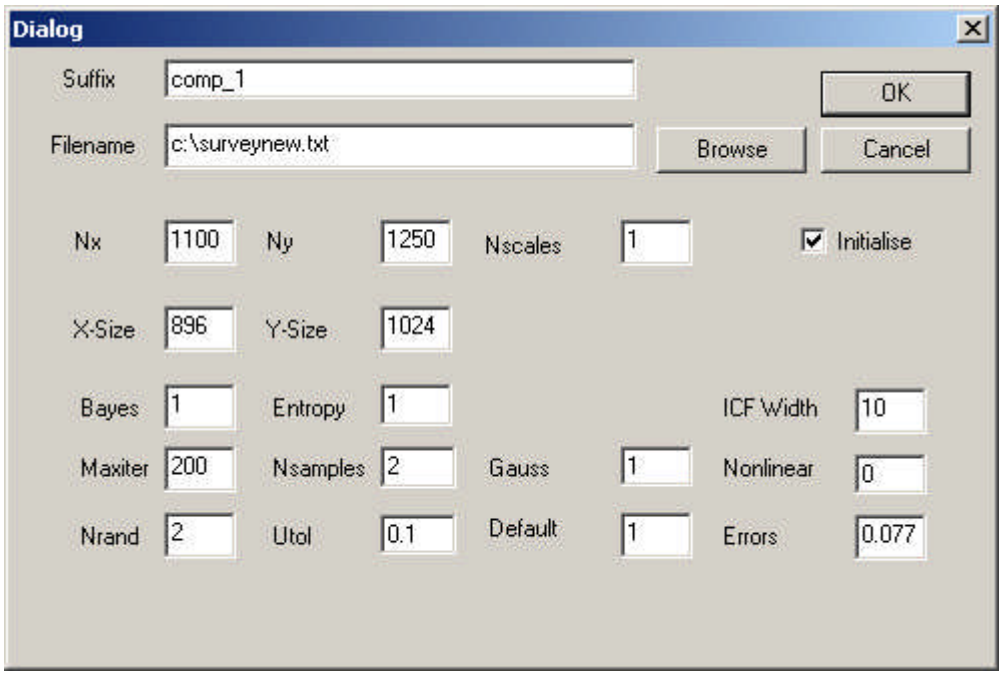


Figure 2.2
The *MemSys 5* input screen

Suffix: A text string to be appended to the filename under which the results are stored (e.g. *resultcomp_1.txt*), so that different reconstructions do not become confused.

Filename: The location of the input survey data (can be chosen with the *Browse* button).

X-size, **Y-size**: the dimensions of the survey area (in pixel units).

Nx, Ny: The dimensions of the reconstructed area. This is always larger than the survey area to avoid edge effects from the ICF. The output result is always the [X-size, Y-size] area in the centre of this larger reconstruction.

Nscales: The number of hidden reconstructions with different ICFs.

Initialise Button: Offers the option to continue a previous run or start a new one.

ICF Width: The parameter which determines the extent of the smoothing function. Note that this parameter is ignored if Nscales $\neq 1$, since each hidden reconstruction then has a preset ICF width.

Bayes: Offers a choice of stopping criterion, allowing for an extension to situations in which i) the noise is unknown and should be calculated from the data, ii) α is known or arbitrarily fixed, and iii) we wish to use *historic MaxEnt* (Gull and Skilling 1991), which stops when $\chi^2 = N$, where N is the number of data. Only Bayes = 1 (the standard stopping criterion described in Appendix A) and occasionally, for diagnostic purposes, Bayes = 2 (calculating the noise) were used in this project.

Entropy: Offers a choice of five slightly different entropy expressions, extending the formalism to deal with situations in which *i*) the distribution can be negative as well as positive, *ii*) the distribution has an upper limit (Fermi-Dirac distributions), *iii*) there is neither an upper nor a lower limit on values of h_i (whereupon MaxEnt reduces to least squares - Gull and Skilling 1991, p.34), and *iv*) there is a fixed total across the whole image. None of these are used in this project - the Entropy flag is always set to 1, which uses the standard entropy discussed in section 2.1.1.

Maxiter: The maximum number of iterations. The algorithm will stop at that point regardless of convergence.

Nsamples: Number of samples from the posterior cloud Pr(**h**|D) used in the mask function to integrate over a pixel.

Nrand: The number of random vectors used to calculate the evidence.

Gauss: 1 = Gaussian statistics, 2 = Poisson statistics. Poisson statistics are not applicable (Brierley et al. 2003b).

Nonlinear: Refers to the linearity or otherwise of the *mask* function (Gull and Skilling 1991, p.25), which integrates Pr(**h**|D) over each pixel to give a value for that pixel. We always use a linear function, and hence this flag is always set to 0.

Utol ($0 < \text{Utol} < 1$): The tolerance for the stopping criterion. Lower values make the iterated result closer to the true MaxEnt probability cloud, but this is computationally expensive.

Default: The uniform default value for all pixels in the absence of data. It is also possible to set up a non-uniform default map, if relevant prior information is known.

Errors: The uniform standard deviation expected in the data, which is overridden (and calculated internally) when Bayes = 2. Again, an errors map can be created instead if the error in each datum is known independently.

2.6 Additional Methods

2.6.1 Changes made to the MemSys 5 software

- Extending the code to cope with Nscales > 8 (see above), the previous limit. The current maximum is 10.
- Demonstrating that the code can deal with the enormous data spaces involved in this project (over 1,000,000 pixels * 10 hidden reconstructions = ~10 million pixels) and removing unnecessary code to increase execution speed. It had originally been envisaged that the survey space would need to be split into regions.
- Automating the code to run overnight batch jobs with a variety of different parameters.
- Adding code to enable the algorithm to be stopped partway through, whereupon certain parameters can be adjusted before continuing the iteration. This enables closer examination of the behaviour of the algorithm.
- Creating stand-alone versions of the updated software that can be run on Windows and Linux operating systems without the need for the source code or a separate C++ compiler.

2.6.2 Other Software

Various other pieces of software have been used to advance this research. In particular, Matlab has been used extensively to analyse and display the output from the *MemSys 5* algorithm. Various subroutines have been written in Matlab, performing such functions as displaying images in the correct colours, scales, and formats; collating and analysing statistical information (averages, standard deviations, maxima, the magnitude of changes to on-transect values, and much else); performing the Syrjala (1996) statistical test discussed in Chapter 3; rotating the survey in the x-y plane so that it fits into the smallest possible rectangle (thereby saving the algorithm from processing large amounts of empty space); collating the relevant information from hundreds of reconstructions together, making it easy to discern that with the highest evidence; generating a variety of virtual surveys through mock data; etc. Somewhere in the region of 50 such programs were written, ranging from just a few lines to many pages of code, and reproducing them here would extend this thesis unnecessarily. The results of these routines are, however, to be found throughout the thesis.

In order that the distances between pixels in the data, surveyed on the three dimensional surface of the earth, are properly represented in the two-dimensional space of the MaxEnt algorithm, it is necessary to convert from the latitude/longitude

format to an x-y format. This was achieved using the Proj.4 software (http://proj.maptools.org/). A Lambert Conformal Conic projection was employed.

VNC (virtual network computing) software (http://www.realvnc.com/) was used to control remotely a cluster of computers located in Cambridge. The use of these 20 processors enabled large numbers of reconstructions to be run simultaneously, and also left the office computer in St Andrews free to be used for analysis and other functions. The PuTTY (http://www.chiark.greenend.org.uk/~sgtatham/putty/) telnet software was also used to enable the remote connection used by the VNC client.

The version control software CVS (http://www.nongnu.org/cvs/) was used to control updates to the executable (.exe) version of the *MemSys 5* software as changes were made to the source code.

The changes to the source code itself, and the creation of executable versions, were made in Microsoft Visual Studio 6.0, using the C++ programming language.

Chapter 3: Evaluating the accuracy of MaxEnt reconstructions

3.1 Introduction

In order to justify the use of MaxEnt with real world data, it is necessary to be able to objectively test the usefulness of the technique. There are two difficulties with this:

i) We can never, as a matter of principle, know the exact at-sea distribution of a species that we are studying acoustically. The information simply is not available (but see section 6.2). Therefore, we cannot compare the results of our analyses with the real distribution and see how accurate our reconstructions are. For this reason, it is essential to use simulated data. With an appropriate simulated distribution (a virtual ocean) we can then run virtual surveys through this mock data, and try to recreate the whole data set from these few observed on-transect data points. Fortunately, such mock data has been made available (see Figure 1.1), enabling this problem to be surmounted.

ii) It is less than obvious on what criteria a reconstruction should be judged accurate. A balance must be struck between various useful aspects of a reconstruction – accurately calculating total biomass; accurately positioning large schools or swarms; accurately predicting the sizes or number of schools or swarms; etc. Some objective method of comparison must be arrived at, and this is far from trivial.

Unfortunately, a large amount of useful information on map comparisons (the Map Comparison Kit – see section 3.4) became easily available only towards the end of this project. Therefore, there are a limited number of useful results in this chapter since there has not been time to thoroughly explore the various comparison methods to the reconstructions.

This chapter, then, will discuss in detail the major method of comparison that was used during the project, and then briefly describe the (very probably much improved) methods that would have been used had they been easily available earlier, and which certainly represent the best way forward in the future.

As a general point about the comparison of reconstructions, it is important to note that we must objectively assess an objectively chosen reconstruction. It is clearly unreasonable to use our comparison method to choose the best reconstruction and then claim that MaxEnt produced an excellent reconstruction – this is a trivially circular piece of thinking. We must objectively choose a ‘best’ reconstruction first, and then compare this reconstruction to the original mock data. Specifically, it must be shown that the reconstruction with the highest value for the log of the Bayesian *evidence* value ($\log_e \Pr(D|\alpha)$ – see Methods) corresponds to the best reconstruction (where ‘best’ is defined in terms of the various properties of the reconstruction – biomass accuracy, distribution accuracy, etc.). If it turns out that some other reconstruction, not chosen by the evidence value, is spectacularly accurate, this would not be a success but a failure.

3.2 Basic visual comparison

Firstly, it does appear possible to generate plausible reconstructions from transect surveys through the simulated herring data. Generally, approximately 6 - 8 channels of hidden reconstructions are sufficient to 'fill in the gaps' between transects. Figure 3.1 shows a plot of the evidence from Nscales = 1 to 10, for reconstructions generated from two surveys through the simulated herring data. (The yellow graph represents a survey with 15 evenly spaced parallel transects, and the blue graph one with 45 evenly spaced parallel transects). The evidence value rises to a peak between 6 and 8 hidden channels, and then drops as further channels (and hence more smoothing) are introduced.

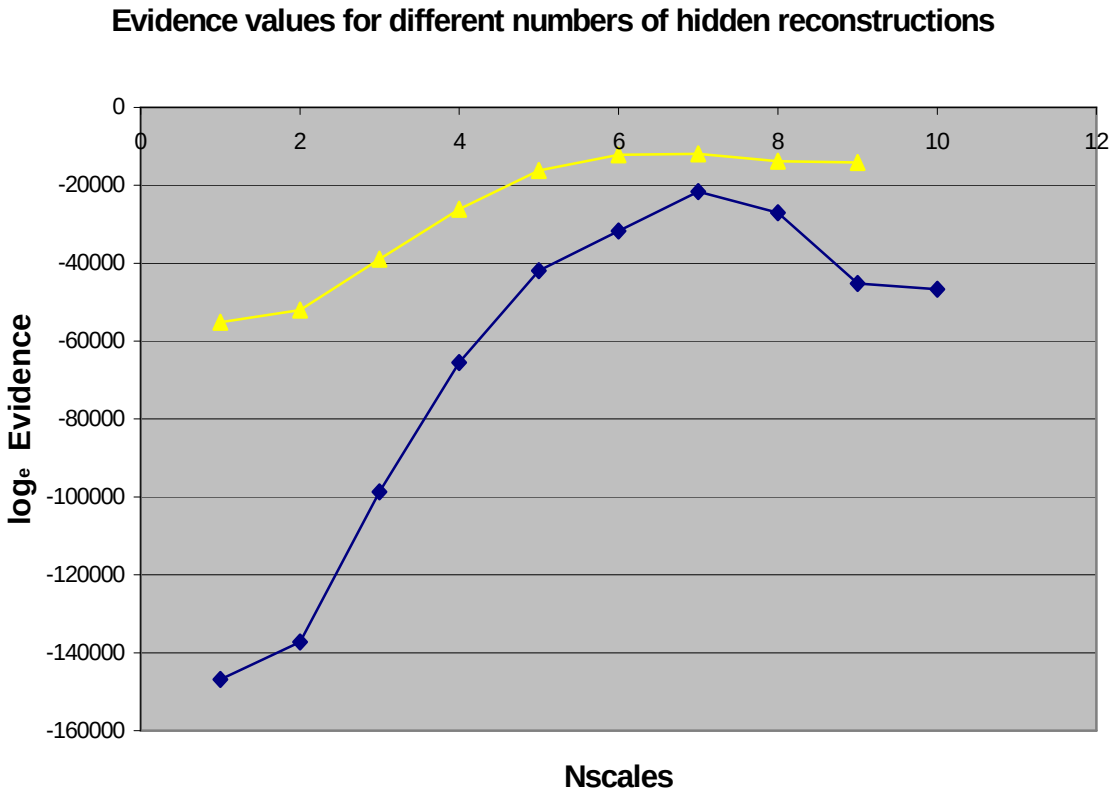


Figure 3.1
How the evidence value changes as more hidden reconstructions are incorporated.
(*Note how more transects (blue line) enable a sharper distinction of the 'best' reconstruction – with more data, there is less uncertainty in determining the best reconstruction.)*

The evidence falls away to the right of the maximum since the smoother images have a worse fit to the data, and falls away to the left as more (unnecessary) degrees of freedom are introduced. The maximum represents a balance between being able to fit the data and over-fitting them. (Brierley et al. 2003b did not show the evidence falling off to the right, as no more than seven hidden reconstructions were used in their analyses.)

Figure 3.2 shows reconstructions, based on the actual transects of a yearly North Sea survey, for 1 to 10 hidden channels (with all other parameters held constant). At first, the transects are individually discernible, clearly indicating that the smoothing has not been sufficiently applied – there is no reason to expect all the herring to be located on transects. It is reassuring that the point at which the gaps between data are visibly filled does indeed correspond to the highest evidence value, at Nscales=8 (for reconstructions from these transects – for the virtual surveys with 15 or 45 evenly spaced transects the highest evidence was for Nscales=7 (Figure 3.1)). This is consistently the case across all the many hundreds of reconstructions undertaken with these data, with a huge variety of parameter settings. Without fail, the highest evidence value matches the reconstruction that is *just* smooth enough to overcome the striped or boxy appearance due to the transects.

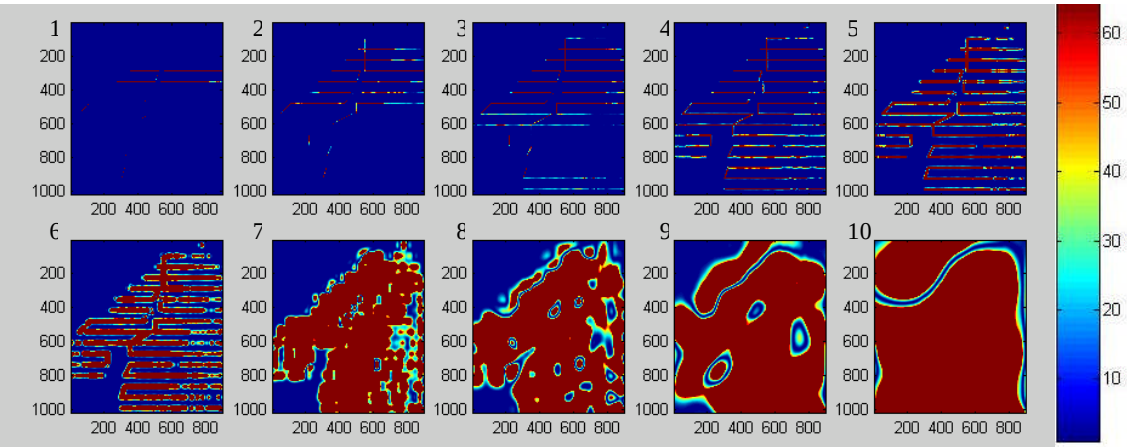


Figure 3.2
Sample reconstructions of the Figure 1.1 data for Nscales = 1 (top left) to Nscales = 10 (bottom right). The colour scale is in the arbitrary units of the simulated data. The highest evidence for these reconstructions was at Nscales = 8 (bottom middle), which corresponds to the visually best reconstruction.

Secondly, it appears visually that the reconstructions can correctly position the major aggregations of herring in the simulated data, and also that the use of regularly spaced transects is not hugely detrimental in comparison with the use of random sampling. Figure 3.3 is a surface plot of the actual simulated data at 1/32 resolution, Figure 3.4 is a reconstruction (again at 1/32) from approx. 11,000 randomly sampled data points, and Figure 3.5 is a reconstruction (at the same resolution) from the approx. 11,000 data points on the transects of the same survey used in Figure 3.2.

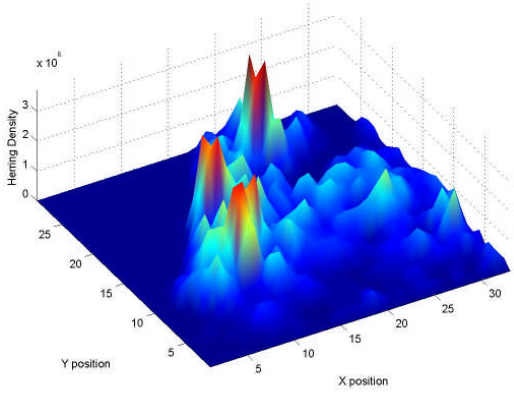


Figure 3.3
Simulated herring data (as in Figure 1.1) at 1/32 resolution.

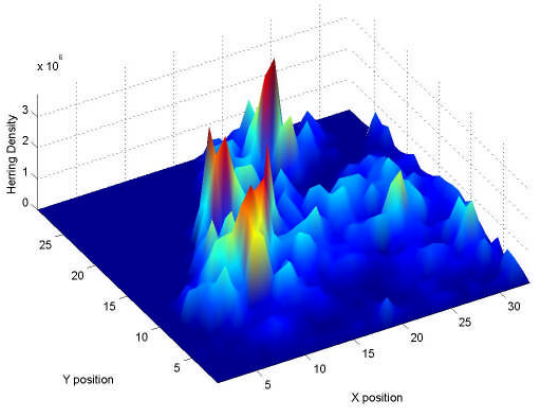


Figure 3.4
(from random sampling)

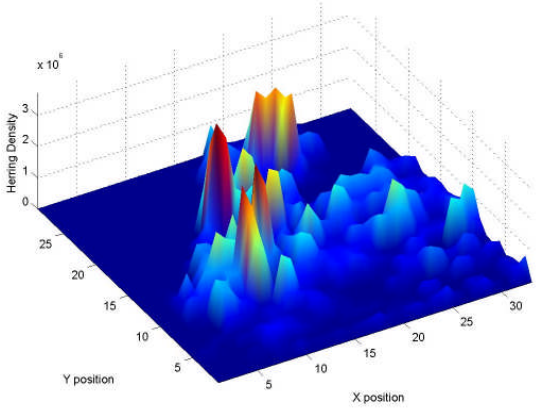


Figure 3.5
(from transect sampling)

We can see that at this resolution the visual comparison is excellent, and that the transected survey shows only minor drop in reconstruction accuracy compared to the random sampling. Other applications of MaxEnt (astronomical data, or photograph enhancement) often have the equivalent of randomly-sampled data, but random sampling is impossible using conventional research vessel-based approaches.

3.3 A statistical comparison method

Syrjala (1996) proposes a Fisherian statistical method for comparing spatial distributions. This is based on a bivariate generalisation of the nonparametric Cramér-von Mises test (Conover 1980) for a difference between two univariate probability distribution functions.

It is necessary to normalise the two populations to be compared (in this case, the original simulated herring data and our chosen reconstruction). Normalisation means that we are only comparing the distributions, and not the biomass estimates. In an

ideal world, we would find a method to simultaneously compare both the distribution and the magnitude. However, given that it is so simple to compare the magnitude of the biomass – it's just a number – it is worthwhile employing a method that purports to resolve the much harder problem of objectively comparing the distributions. The Syrjala method will not work without normalised data. We proceed as follows:

i) Choose a corner of the rectangular image as the origin. Let (x_k , y_k) denote the coordinates of the *k*th pixel (k=1, ... ,K); let $d_i(x_k , y_k)$ denote the density at the *k*th sampling location of the *i*th image (where *i*=1 is the simulated data, *i*=2 the reconstruction).

ii) Normalize the data. We divide each pixel density by the total of the density values in every pixel of that particular image, so that

$$\gamma_i(x_k, y_k) = \frac{d_i(x_k, y_k)}{\sum_{k=1}^{K} d_i(x_k, y_k)}$$

iii) Calculate the *cumulative distribution function* for each pixel. This is defined as the sum of all normalized density distributions $\gamma_i(x_k ,y_k)$ whose location (x, y) is such that $x \le x_k$ and $y \le y_k$. Therefore the cumulative distribution function for the *i*th population at the *k*th pixel is

$$\Gamma_i(x_k, y_k) = \sum_{\forall x \le x_k, \forall y \le y_k} \gamma_i(x, y)$$

iv) By analogy with the Cramér-von Mises test, a useful test statistic is the square of the difference between the two cumulative distribution functions, summed over all pixels, i.e.

$$\Psi = \sum_{k=1}^{K} [\Gamma_1(x_k, y_k) - \Gamma_2(x_k, y_k)]^2$$

v) Unlike the traditional univariate Cramér-von Mises test, the statistic Ψ is not invariant with respect to our choice of origin. Following Zimmerman (1993), we simply perform the calculation four times with each of the four corners as the designated origin, and then take the average of these four values to be our final value for Ψ.

vi) We now have a statistic related to the similarity in distribution of the two maps we are comparing. In order to assign a level of significance to this statistic, we randomly swap some values from map 1 (the simulated data) with values from the same pixel location in map 2 (the reconstruction), and then calculate the test statistic for the two new maps thus created. We repeat this many times, and then rank the test statistics that are generated. (It is not possible to calculate the test statistics for every possible pairwise permutation of the data sets – that would involve calculating Ψ some 2^k times, which in our case (k=1024x896) would be a truly enormous number. Following Syrjala, we perform the test on 999 random pairwise permutations of the two maps, which, including the test on the real maps, gives 1,000 values for Ψ). We then need to look at the position in the ranking of the actual test statistic from the real maps – if the actual test statistic is one of the highest, this indicates that the difference between

the distributions of the two maps is much greater than would be expected randomly. On the other hand, if many of the randomised permutations generate higher test statistics than the actual test statistic, this indicates that the difference in distribution between the simulated data and the reconstruction is not significant.

From a Fisherian point of view, we would say that we have a null hypothesis stating that the normalised distributions of the simulated data and the reconstruction are the same. The P value is the proportion of the 1000 test statistic values that are greater than the actual test statistic. A very low P value (less than 0.05) indicates that the null hypothesis is invalid (i.e. there IS a difference between the distributions).

The Syrjala comparison method was applied extensively to test for a difference in density distribution between the simulated herring data and the many reconstructions generated.

The results are encouraging, but not conclusive. We consider, for a variety of reconstruction parameters, from a variety of numbers of transects, whether the evidence and the Syrjala test agree on which number of hidden channels (1 to 10) creates the best reconstruction.

The reconstruction chosen by the highest evidence value always achieves a higher P-value as compared to the reconstructions with low values for evidence. Typically, reconstructions with only 1-4 hidden channels, which are visually very different in distribution from the simulated data, do indeed have P-values of less than 0.05, indicating that the Syrjala method has detected a difference in the distributions. Also, reconstructions with between 6 and 8 hidden channels consistently show no difference in distribution to the simulated data, as determined by the Syrjala test.

However, it is not always the case that the highest evidence achieves the highest P-value, which would be a nice indicator of successful reconstruction. The P value of the ‘best’ reconstruction, as determined by the evidence, is always within 2% of the highest P value for any reconstruction in that set of 10, but it is not always actually the highest. This may be because we consider only 1,000 random permutations when calculating the P value, when theoretically we should consider every permutation.

Nevertheless, what we do find is that where the evidence value struggles to choose between competing reconstructions (the flattening of the line in figure 3.1), the Syrjala test is equally unable to choose between the different reconstructions. It may be that a statistical test which takes into account both distribution and biomass estimate is necessary to truly test whether the evidence value is a useful objective selector of the best reconstruction, but such a test was not available through much of the duration of the work..

In terms of a comparison between our chosen ‘best’ MaxEnt reconstruction and a kriged reconstruction, the Syrjala technique unfortunately seems too blunt an instrument. The technique is not sensitive enough to define either the MaxEnt or kriging result as ‘better’ – it merely shows that both results are not significantly different from the original simulated data. Different statistical tests will be needed.

3.4 The Map Comparison Kit

It now seems that such statistical tests are indeed available, though their very novelty means that they are relatively untested. It would obviously have been very worthwhile to attempt to apply these methods in this project, but unfortunately, time does not permit. They offer a fruitful future avenue for research.

The *map comparison kit* (http://www.riks.nl/mck/ - see Figure 3.6) is a piece of free software developed by the Research Institute for Knowledge Systems in the Netherlands. It is the result of an in-depth literature scan across seemingly unrelated disciplines for the very latest in map comparison techniques, and offers a user friendly interface to enable complicated comparisons to be undertaken. Within the last few months, a significant extension to the software has enabled the comparison of continuous-valued maps like the ones in this project – previously, the software dealt only with category maps (e.g. land-use maps) with a finite number of distinct categories. Some of the more promising new techniques are described briefly below, and some very early results are given.

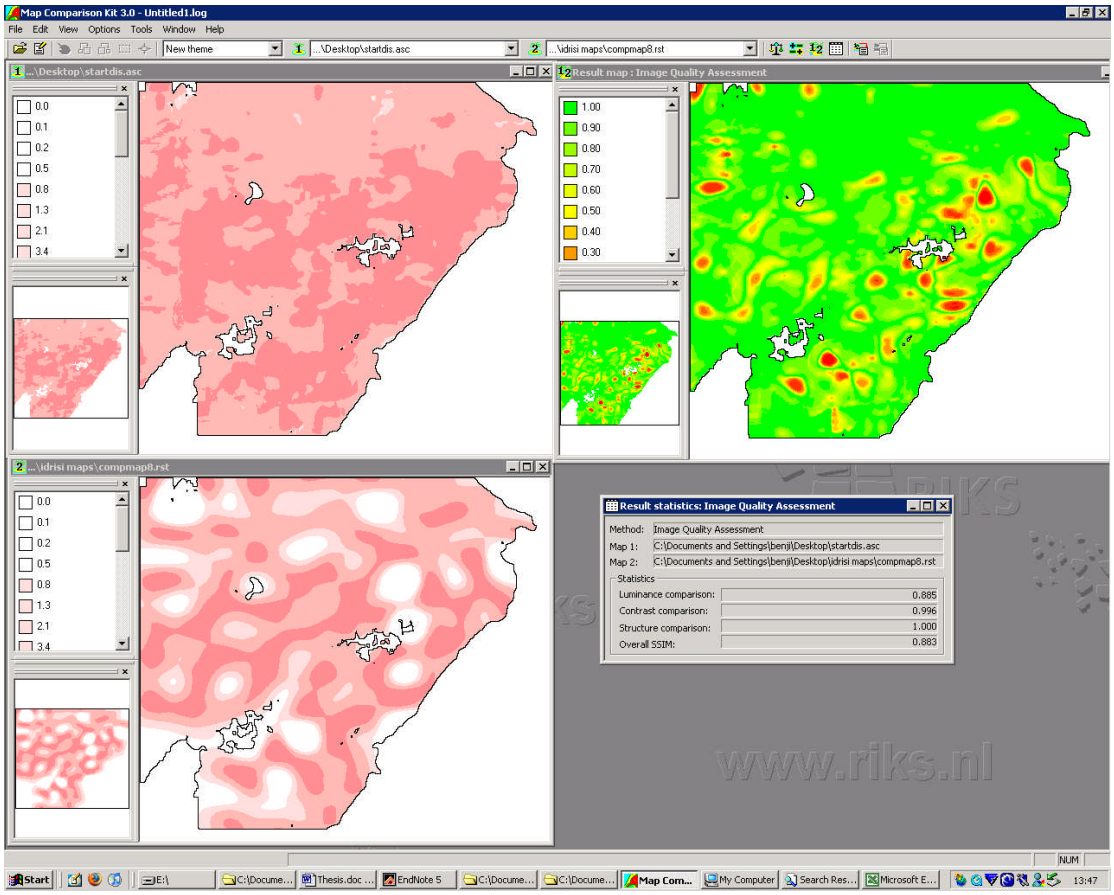


Figure 3.6: A screen shot of the Map Comparison Kit software. The map in the top left is of the simulated data from figure 1.1 (rotated due to the way the software reads the data in, but this does not affect the comparison); bottom left is a MaxEnt reconstruction; top right is the comparison map generated by the Image Quality Assessment method, showing regions of disagreement in darker colours; bottom right is a selection of relevant statistics.

3.4.1 Wavelet Analysis

Working in the field of weather forecasting, Briggs and Levine (1997) suggest decomposing each map, using a discrete wavelet transform, into a series of maps at different scales. The simpler maps can then be compared more easily, as they contain far less information, and something as simple as a root-mean-square comparison should be sufficient. The combined results of these comparisons at different scales can then be aggregated, and an overall comparison arrived at. This method is interesting, but has two major flaws with regard to its application to our data. Firstly, it is relatively arbitrary which pixels fall under a peak in the wavelet transform – moving the whole map one pixel to the left could give very different results. Secondly, the *MemSys 5* software created the reconstruction using different scales, and so it would be somewhat circular to deconstruct the map once again before comparison.

3.4.2 Image Quality Assessment

Working in the field of image processing, Wang et al. (2004) created the concept of image quality assessment. The method is highly mathematical, but essentially the idea is that a 'moving window' is used. In effect, we study only the part of the map visible through an imaginary window of, say, 20 pixels by 20 pixels, and then repeat this with the window moving one pixel at a time. Again, the smaller image under consideration enables much simpler mathematical analysis of the relation between the two maps.

3.4.3 Information weighted comparison

Again from the field of image processing, Tompa et al. (2000) consider another method. The basic idea is that variations that occur within value ranges that are common in the map are weighted less than variations that occur within uncommon ranges, i.e. it is the outliers that are most taken into account, and small variations around the common values are ignored. This seem to be a useful idea in the case of our pelagic reconstructions since, in general, the rare and extreme values are the most interesting as they contribute enormously to biomass and are extremely important from a biological perspective.

3.4.4 Image Warping

The method of Reilly et al. (2004), from the field of biometry, is extremely unusual. The idea here is that instead of directly comparing the two maps, one map is distorted until it closely matches the other map, and the amount of distortion required is a measure of the similarity. For example, if a school of herring, in our data, is predicted in slightly the wrong place or is the wrong size, the reconstruction can be adjusted to better match the original, and some measure of how much adjustment is required can be calculated. Unfortunately, this method struggles when there is a large discrepancy between the two maps, and may in fact be intractable in many cases. On the other hand, where the maps are relatively similar, for example in the situation above in which the Syrjala algorithm could not distinguish between reconstructions, perhaps this method may be of use.

3.4.5 Bivariate Spatial Association

Lee (2001) offers a statistical measure of spatial association based on a integration of two accepted statistical measures – an aspatial measure of bivariate association, and a univariate measure of spatial association. The mathematics is reasonably complicated, but in essence the method offers a single comparison statistic (called 'L'), based on both the magnitude and the distribution within the maps.

3.4.6 Preliminary results

As mentioned, this software has only recently become available, and so the full range of possibilities has not been explored. Each comparison method has a range of parameters that can be set, and each offers a resulting statistic that must be understood differently, so clearly there is great scope for further work. Figures 3.7 and 3.8 show how various statistics generated using the Map Comparison Kit vary in a broadly similar way to the evidence – both the evidence and the various statistics reach a peak for reconstructions with 7-9 hidden channels.

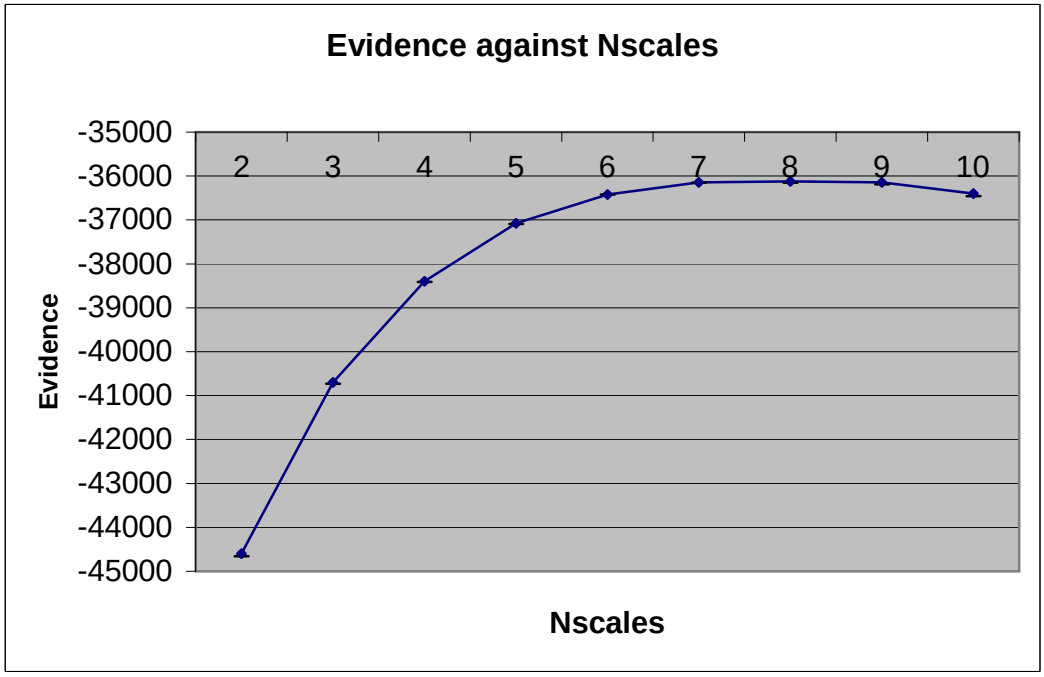
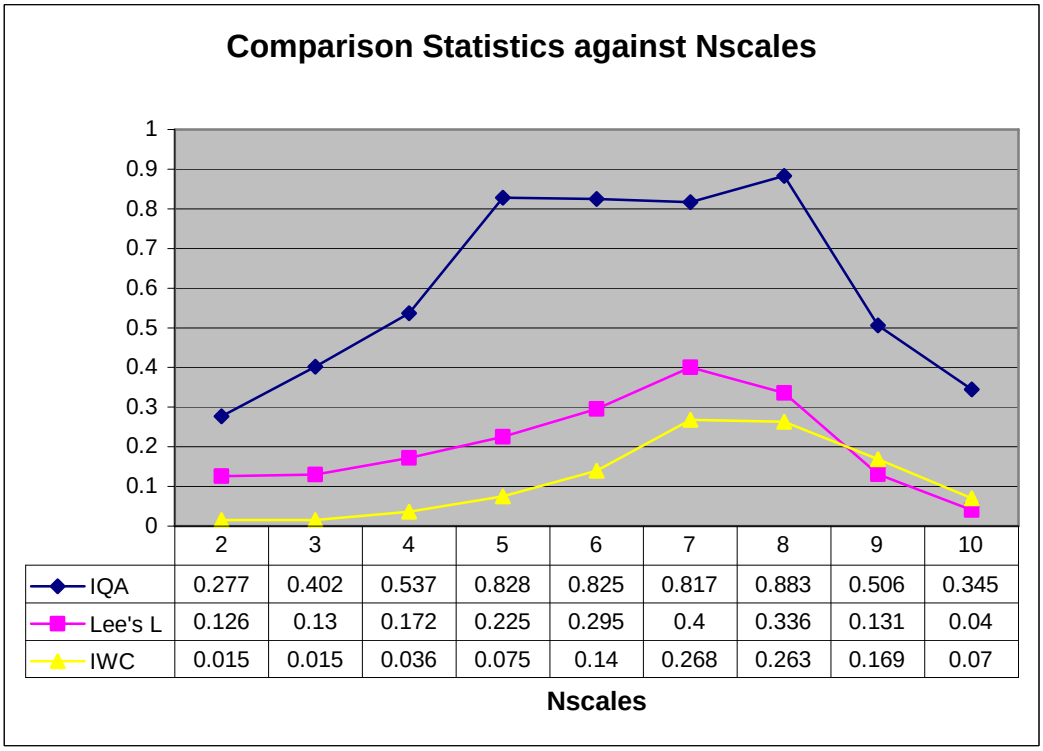


Figure 3.7: The Bayesian evidence value for ten MaxEnt reconstructions with 1-10 hidden channels, with all other parameters held constant.



	2	3	4	5	6	7	8	9	10
IQA	0.277	0.402	0.537	0.828	0.825	0.817	0.883	0.506	0.345
Lee's L	0.126	0.13	0.172	0.225	0.295	0.4	0.336	0.131	0.04
IWC	0.015	0.015	0.036	0.075	0.14	0.268	0.263	0.169	0.07

Figure 3.8: The resulting statistics from the comparison of the ten reconstructions of Figure 3.7 with the original simulated data, by three different comparison methods – Image Quality assessment (IQA), Bivariate Spatial Association (Lee's L *statistic), and the Information Weighted Comparison (IWC)*

It appears that, at a very broad level, these methods offer some support for the use of the evidence value in choosing the 'best' reconstruction, but a great deal more work must be done. A change in the number of hidden channels can produce a very different reconstruction, and so we would expect the evidence to be easily able to distinguish the 'best' one. It would clearly be necessary to look at the evidence and the comparison statistics for reconstructions that are more similar (for example with different weighting factors – see figure 4.3 below) to see exactly how accurately the evidence value can discriminate between competing reconstructions. Also, it will very much be worth looking beyond the 'headline' statistics of the various comparison methods, and looking at how successfully different regions are reconstructed, how the results vary with changes to the comparison parameters, and so on. The bottom line, however, is that the 'best' maps are identified.

Chapter 4: Krill in the Scotia Sea

This chapter is a reprint of a paper published in CCAMLR Science Vol. 13 (2006) pages 97-116, by Heywood, Gull, and Brierley. The manuscript has been through the peer-review process and been published in the journal; therefore it is presented here in this format, even though it may repeat some small parts of the previous method chapters of this thesis. In terms of intellectual property, this paper is entirely my own work, with supervisory and editorial assistance from Dr Andrew Brierley. Dr Gull is listed as an author, but had no additional specific input beyond providing the *MemSys 5* software and instruction in how to use it.

A QUANTIFIED BAYESIAN MAXIMUM ENTROPY ESTIMATE OF ANTARCTIC KRILL ABUNDANCE ACROSS THE SCOTIA SEA AND IN SMALL-SCALE MANAGEMENT UNITS FROM THE 2000 CCAMLR ACOUSTIC SURVEY

B.G. Heywood, S.F. Gull, and A.S. Brierley

4.1 Introduction

The CCAMLR 2000 survey (Trathan et al. 2001; Hewitt et al. 2004a) of the Scotia Sea employed acoustic techniques to measure the density distribution of Antarctic krill. Due to inevitable pressures of time and expense, only a very small fraction (0.56%) of the total survey area was acoustically sampled directly, and hence some method of estimating total abundance from these limited data is necessary. Hewitt et al. (2002) used the Jolly and Hampton (1990) statistical method and calculated a total biomass across the survey area of 44.3 million tonnes. The Jolly and Hampton method involves generating weighted means for a number of semi-randomly placed transects, and the CCAMLR 2000 survey was designed in accordance with such transect placing. Demer and Conti (2005) used the same method, but their updated krill target-strength model led to a biomass estimate of 109.4 million tonnes. We present here a biomass estimate, based on the Demer and Conti (2005) target-strength model, derived using a Bayesian Maximum Entropy (MaxEnt) technique rather than the standard Jolly and Hampton (1990) method. MaxEnt maps of the density distribution are presented as an alternative to the standard kriged maps (a good description of the kriging approach can be found in Rivoirard et al. (2000)) of density distribution, as given by Hewitt et al. (2004a).

In addition to the survey total, we also present biomass estimates derived by the MaxEnt method for each of the krill small-scale management units ('SSMUs' - Hewitt et al. 2004b - see Table 4.1 & Figure 4.1). These SSMUs are ecologically crucial areas around South Georgia, the South Orkney Islands and the South Shetland Islands within which land-based predators forage, and are likely to be particularly important from an ecosystem management perspective. Since the densities estimated by the Jolly and Hampton (1990) method are per stratum of the survey, previous estimates of biomass within certain SSMUs have been based on the density determined for nearby strata (Hewitt et al. 2004b), despite the fact that only parts of these strata fall within

the SSMUs. The density distribution maps generated by MaxEnt enable biomass values for the SSMUs to be inferred on a more appropriate spatial scale.

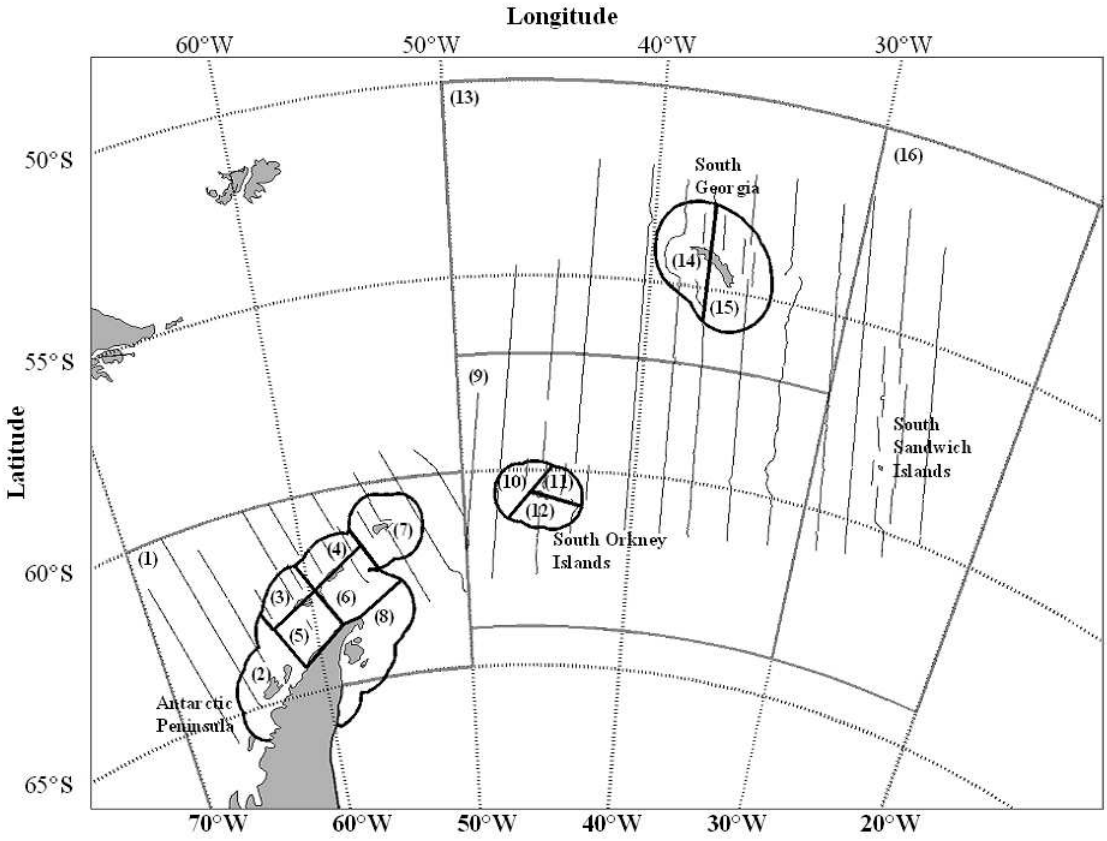


Figure 4.1: Small-scale management units (CCAMLR Sub-Areas 48.1 - 48.4 as solid grey lines; smaller SSMUs in black) in CCAMLR Area 48, numbered following Hewitt et al. (2004b), and the CCAMLR 2000 survey transects.

Table 4.1: List of the SSMU designations.

SSMU	Number	Full Title
APPA	**1**	Antarctic Peninsula Pelagic Area (Sub Area 48.1)
APW	**2**	Antarctic Peninsula West
APDPW	**3**	Drake Passage West
APDPE	**4**	Drake Passage East
APBSW	**5**	Bransfield Strait West
APBSE	**6**	Bransfield Strait East
APEI	**7**	Elephant Island
APE	**8**	Antarctic Peninsula East
SOPA	**9**	South Orkney Pelagic Area (Sub Area 48.2)
SOW	**10**	South Orkney West
SONE	**11**	South Orkney North East
SOSE	**12**	South Orkney South East
SGPA	**13**	South Georgia Pelagic Area (Sub Area 48.3)
SGW	**14**	South Georgia West
SGE	**15**	South Georgia East
	16	Sub Area 48.4

The MaxEnt image reconstruction method has been widely and successfully used to generate complete images from sparse point data in many disciplines, from astrophysics (e.g. Weir and Djorgovski 1991) to medicine (e.g. Charter and Gull 1991). The inference of krill density values in grid squares across the CCAMLR 2000 survey area can be treated as an exercise in image reconstruction since the data consist of point estimates of krill density derived from echo integration at a given latitude and longitude, and hence our data space is two-dimensional. A set of point values on a two-dimensional grid can naturally be viewed as a pixellated image or picture. The problem of filling in missing data is then directly analogous to the problem of reconstructing a damaged photograph, or improving the resolution of astronomical images, and therefore reconstruction techniques from these disciplines may sensibly be applied to these biological data. The CCAMLR 2000 data can be considered as an image in which krill densities are plotted on a rectangular grid, with correspondingly brighter dots for higher density values. We then endeavour to reconstruct the missing parts of this image grid, creating a map from which total biomass can be estimated. The intensity of each pixel in our reconstructed image corresponds to an estimate of krill density (g m^{-2}) in that pixel, where each pixel represents a 1 nautical mile (nm) × 1 nm cell of the survey area. MaxEnt image reconstruction has previously been used to generate maps and biomass estimates from smaller scale surveys of krill around South Georgia (Brierley et al. 2003b; Wafy et al. 2003).

Line-transect data contain potentially valuable information on spatial distribution, which is ignored by conventional (e.g. Jolly and Hampton 1990) statistical techniques. We believe that the MaxEnt method offers advantages over the Jolly and Hampton (1990) method because it makes explicit use of this spatial information (Brierley et al. 2003b). The other commonly used techniques for biomass estimation are geostatistical (Rivoirard et al. 2000), but Maravelias et al. (1996) showed that such methods are unsatisfactory when the distribution of biomass is heavily skewed, which is very much the case with the krill density data from the CCAMLR 2000 survey. Densities of almost 24,000 g m^{-2} were reported for individual 1 nm^2 regions, but only 0.8% of measurements exceeded 1,000 g m^{-2}, and two-thirds were under 10 g m^{-2} (see Figure 4.2).

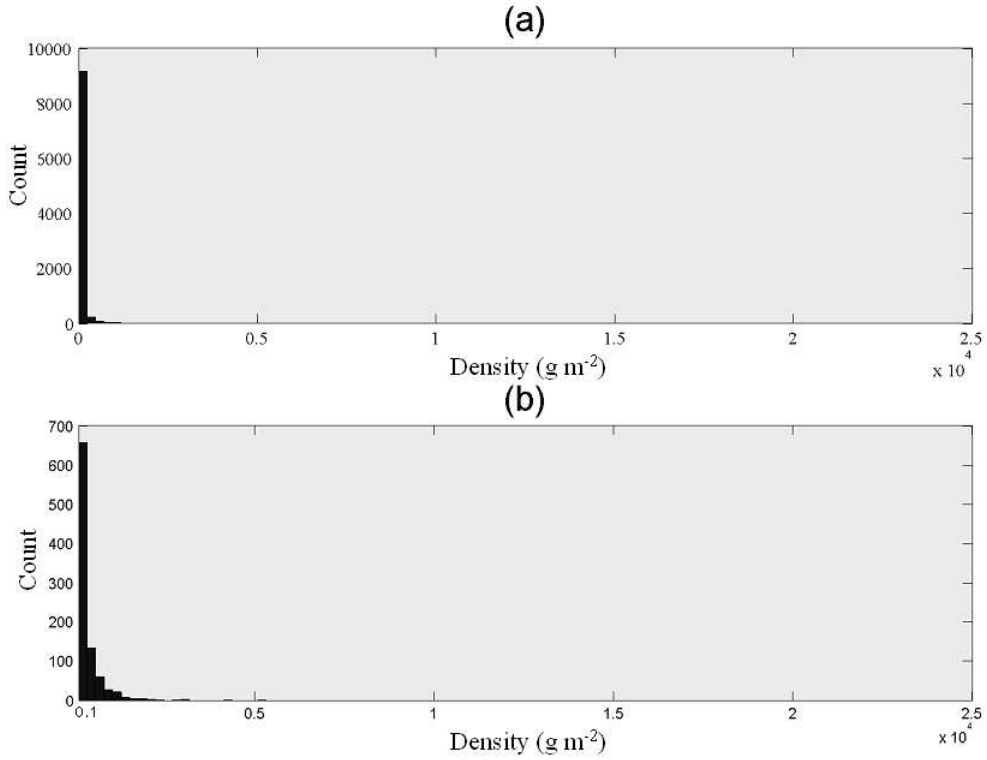


Figure 4.2: Histograms of density (g m^{-2}) from the CCAMLR 2000 survey, showing the extreme skewedness of the data. Histogram (a) shows all data; histogram (b) shows only those density values over 100 g m^{-2}, demonstrating that the skewedness is inherent throughout the distribution and not just created by a large number of zeros.

4.2 Methods

4.2.1 What is Maximum Entropy Image Reconstruction?

Image reconstruction, in this case, is the inference of missing values in a grid framework. MaxEnt describes the statistical, probabilistic framework under which this is achieved. Johnson and Shore (1980; 1983) and Tikochinsky et al. (1984) offer clear theoretical and mathematical justifications for the use of MaxEnt, and the method has been placed robustly in a rigorous Bayesian framework (Skilling 1988a; Skilling and Gull 1989; Skilling and Sibisi 1990; Gull and Skilling 1991; Skilling 1991; see Sivia 1996 for an introduction to Bayesian data analysis; and Clark 2005 on why ecologists are becoming Bayesians). We present here just one relatively non-mathematical argument for the use of MaxEnt (drawn from Skilling 1992), and concentrate on its application to biomass estimation and distribution mapping from fishery acoustic data.

We make use of Bayes' Theorem (Bayes 1763; Cox 1946; Jaynes 2003):

$$\Pr(\boldsymbol{h}|\text{data}) \propto \Pr(\boldsymbol{h}) \Pr(\text{data}|\boldsymbol{h})$$

where '|' means 'given', and $\boldsymbol{h}$ is the set of all possible images [i.e. krill distributions] h, each consisting of the intensity [i.e. density] values in m pixels [i.e.1 nm × 1 nm cells of the survey area], h_1, h_2 ... h_m. The posterior inference $\Pr(\boldsymbol{h}|\text{data})$ measures how

closely trial images *h* are in accord with the survey data, given any prior information, Pr(***h***). The other factor, Pr(data|***h***), is the Bayesian Likelihood. Since noise in the data can tolerably be described by Gaussian statistics (Gull and Skilling 1991), a Gaussian likelihood function is adequate (Gull and Skilling 1991).

To use the above formula, we need to assign a sensible prior distribution of images, Pr(***h***). This specifies our original ideas, without the data, about the plausibility of various images ***h***. We choose a prior according to the Principle of Maximum Entropy (Jaynes 1978). One relatively non-technical explanation of why the concept of entropy should be fundamental to the process is as follows:

Suppose an enormous number *N* of individual krill are thrown, one by one and at random, into an empty Scotia Sea, which has been conveniently split into 1 nm × 1 nm cells. This imaginary experiment is repeated many times. We wish to quantify our preferences for different possible distributions of krill, without reference to (i.e. prior to) any data. Fortunately, not all distributions are equal – some are more probable than others. For example, a total of seven is the most likely outcome of the roll of two dice, even though the value on each die is completely random, because there are more ways to make seven from two dice than any other number. Similarly, there are more ways of distributing the krill evenly across of the Scotia Sea (i.e. an equal number in each 1 nm × 1 nm cell) than there are of distributing them in any particular uneven pattern. In fact, the prior probability of any particular pattern being produced is proportional to the number of permutations Ω of *N* krill that result in that particular pattern, which is given by

$$\Omega(n) = \frac{N!}{n_1!n_2!...n_m!}$$

where $n_1...n_m$ are the number of krill in cells *1, 2, ... m*.

It is mathematically much more convenient to work with the logarithm of Ω, called the *entropy, S*.

$$S = \log\Omega \approx \sum n_i \log n_i$$

Our prior expectation, Pr(***h***), incorporating our expectation of randomness and unpredictability (which expresses our lack of prior knowledge) can - after some mathematical work (see Skilling 1992) which is beyond the scope of this paper - be written in terms of the entropy *S* as

$$\Pr(\boldsymbol{h}) \propto \exp(S)$$

Thus we have an entropy-based prior distribution which expresses a preference for smoothness (since an equal distribution of krill across all cells has the highest number of permutations). A more rigorous mathematical approach (Skilling 1988a; Skilling 1988b) confirms and justifies this result, and properly calculates the scaling constants to remove the proportionality.

Using our chosen prior and likelihood, we are now free to calculate the most probable posterior image by means of Bayes' theorem. This most probable image is the traditional MaxEnt result. However, since our prior is a probability *distribution* of images, so is our posterior; we can in fact not only calculate the maximum of the posterior distribution (the most probable image), but can also sample, from the posterior probability space, a selection of the images *near* that maximum. This then enables calculation of the standard deviation of each pixel in the reconstruction. This is the quantified MaxEnt result.

4.2.2 Parameter Estimation

Each image reconstruction requires the choice of a small number of parameters, for example the width of the blurring function, described below, which is used to capture spatial autocorrelation in the data. Crucially, we can calculate the normalising constant of Bayes' theorem, a value which is often known as the *evidence*, *p*(data), for each set of parameters:

$$p(\text{data}) = \sum_{h} \Pr(\boldsymbol{h},\text{data})$$

where $\boldsymbol{h}$ now contains *only* those images consistent with our chosen parameters. *p*(data) is used to indicate that this is a single probability value (rather than a distribution) calculated from a particular reconstruction attempt with particular input parameters. The evidence value is a unitless probability between zero and one, and is usually presented as its logarithm.

The evidence value is used to discriminate between prior images, much as the likelihood discriminates between posterior images (Gull and Skilling 1991). Note that the prior distribution Pr($\boldsymbol{h}$) is a *distribution of prior images*, and not an individual prior image. This distribution Pr($\boldsymbol{h}$) tells us which images are more likely in advance - before the data - and the comparison of a number of values for evidence tells us which selection of prior images, defined as those consistent with the chosen parameters, produced the most likely MaxEnt result, *after* considering the data.

By running a number of reconstructions with different parameters (i.e. different *assumed* prior information; in effect a different set of prior images), and choosing the one with the highest evidence value, we can progressively approach the optimal values of any unknown parameters.

The use of the evidence value to objectively decide between possible parameter values is a major strength of MaxEnt. The chosen reconstruction must be that with the highest evidence, regardless of the prejudices of the researcher. This applies to any alterable parameter of the reconstruction, not only those estimated in this paper but also more fundamental elements such as the shape of the blurring function discussed below.

The analysis was undertaken using the software *MemSys 5* (Gull and Skilling 1991).

4.2.3 Calculating the Quantified MaxEnt Result

The data gathered by an acoustic survey of a defined area are usually, like almost all data, incomplete. This is certainly the case for the CCAMLR 2000 survey. There is in principle no mathematical transform that can be applied to the data that will result in the actual krill densities in every 'pixel' since there is not enough information in the data – we cannot transform from data to image. Therefore, it is necessary to approach the problem from the other direction – we generate a trial image (in this case a possible krill density distribution across the whole survey area) and transform this into mock data (the set of on-transect values implied by this trial density distribution). After comparing this with the actual data, another trial image is generated, which has been updated so that the next mock data set will be a better fit.

The first trial image is simply the uniform image. This is the most likely distribution in the absence of data. Since the final image evolves from this smooth starting point, any structure in the reconstructed image must be introduced by the data itself, and cannot be an artefact of our first trial image. This smooth starting image is then iteratively updated by comparison with the data (i.e. the on-transect density values from Demer and Conti (2005)), and becomes progressively less smooth. The iterations stop when the fit of the mock data (from the latest trial image) to the actual data is optimal, where optimal is defined in terms of the balance between the entropy (which decreases as we move away from the completely smooth image) and the likelihood (which increases as the fit to the data becomes more exact). This balance between entropy (upon which the prior distribution is based) and likelihood is directly analogous to the formulation of Bayes' theorem above. Stopping the iterative process too soon will mean that some genuine data are not fitted, and stopping too late will mean that noise in the data begin to be fitted, resulting in unwarranted structure in the reconstruction. The MaxEnt stopping criterion is chosen on solid mathematical and probabilistic grounds (Gull and Skilling 1991).

The transform applied to generate the mock data from a trial image depends heavily on the particular application of the MaxEnt method. In astrophysics, for example, the data from an instrument may be a Fourier transform of the real-world image, and therefore such a transform would need to be applied to each trial image in order to approach the correct result. Similarly, complex transforms are sometimes necessary with biological data. Lizamore (1995) used commercial trawl data to reconstruct density distributions for New Zealand hoki. A transform between data space (trawl length and position, weight of catch) and image space (density in each pixel) was required.

In the application to fishery acoustic surveys, there is no need for such complication, since both the data and the image reside in the same mathematical space and share the same units. Thus, the only transform applied between image and data is a blurring function, which characterises our expectation that, on biological grounds, some local smoothness (spatial autocorrelation) should exist in the final image (Weber et al. 1986).

The blurring function used here is simply an approximation to a Gaussian point spread function. It is necessary to choose a width for this blur, and it has been found that in practice a particular smoothing width tends to emphasise structure of a similar

width in the reconstruction. Since it is reasonable to expect local structure (in this case, krill swarms and/or clusters of swarms) to have varying sizes, this is unwelcome. The solution proposed by Weir and Djorgovski (1991), subsequently incorporated into *MemSys 5*, is to concurrently produce a number of 'hidden' reconstructions, each with a different blurring width. These separate but concurrent reconstruction 'channels' are then convolved to produce a single image. Generally, precedent has suggested that each separate hidden reconstruction should have a blurring width twice as wide, and a weighting for the convolution four times smaller than (i.e. one quarter as much as), the previous hidden channel (which we term a *scaling factor* for the blur equal to *two*, and a *weighting factor* for the convolution equal to *four*). Thus with four hidden reconstructions, the blurring widths would be 1, 2, 4 and 8 units, and the 2^{nd}, 3^{rd}, and 4^{th} channels would be 4, 16, and 64 times less important to the convolved reconstruction than the first channel.

Previous papers (Brierley et al. 2003b; Wafy et al. 2003) used these essentially *ad hoc* values (for the scaling and weighting of hidden channels) for their reconstructions of krill density around South Georgia. However, the evidence value, as described above, can be used to objectively select not just the *number* of hidden channels (as Brierley and Wafy used it) but also to objectively select appropriate scaling and weighting. Recent changes we have implemented to the software interface used to perform MaxEnt mean that these values can be chosen at run-time. This has enabled much deeper exploration of the effect of these values than was available to the authors of these previous papers.

The standard deviation for each pixel value can be calculated under MaxEnt. A sampling of the posterior distribution generates a number of images, all very nearly as probable as the best estimate, but not necessarily similar in shape or total intensity. Stable, well-predicted pixels will be very similar in almost every reconstruction, whereas those about which greater uncertainty exists will fluctuate. For each pixel, there is therefore a population of values (one from each sample image) from which standard deviation (in g m^{-2}) can very simply be calculated. From these individual pixel standard deviations, the standard deviation of a given region or of the whole survey is simply the sum of the standard deviations of the pixels within it.

4.2.4 Data Preparation

The on-transect data used here are exactly as used by Demer and Conti (2005) to calculate a 109.4 Mt krill biomass estimate. In order to process the data as an image (a rectangular image, in fact, for ease of computation) it was necessary first to convert the positional stamps for this data (Demer and Conti 2005) from latitude-longitude format to Cartesian (x,y) format so that a consistent spatial scale existed across the grid. A Lambert Conformal Conic projection was used. The coordinates defining the survey bounds and the SSMUs, identical to those used by Hewitt et al. (2002; 2004a; 2004b), were transformed by the same projection, so that the biomass inferred for any area could easily be found by summing the biomass of all pixels in that area. [Pixel biomass is $1{,}852^2$ m^2 × density (g m^{-2}), divided by 10^{12} to convert from grams to million-tonnes.] For those SSMUs that extend beyond the bounds of the survey, this method cannot apply; instead the mean biomass of the pixels within these areas was multiplied by the total area of the SSMU, using the same SSMU area data as Hewitt et al. (2004b). Both these methods were applied to the SSMUs around South Georgia to

check that the simple summing of pixel biomass did not induce any systematic bias. The estimates generated by the two methods differed by no more than 1%.

In addition, since the density data were extremely heavily skewed (see Figure 4.2), much better evidence values were obtained by normalising the distribution of the data somewhat before input to the algorithm. This is due to the normal shape of both the Likelihood function and the blurring function, which naturally act more evenly on data with a normal-shaped distribution. Therefore, the second root of the density ($\sqrt{\sqrt{data}}$) was input to the algorithm, and the result was squared twice before output. It should be noted that proponents of the kriging technique have also used data transforms to reduce the skew of the input data (Rivoirard et al. 2000); both kriging and MaxEnt are similarly challenged by skewed data and have developed similar solutions to the problem. Further experiments with the shape of the blurring function, which have not yet been undertaken, may reduce the need for such measures under MaxEnt.

4.3 Results

Approximately two hundred reconstructions were undertaken in total, and some statistics relating to a selection of these are presented in Table 4.2. The two parameters that had most effect on the reconstructed result were the number of hidden channels and the weighting factor (the rate of change of convolution weighting from one hidden channel to the next). The scaling factor (the rate of change of the blurring width) was found to have almost no effect on the result. One hundred of the above-mentioned reconstructions were therefore undertaken with all integer values of weighting factor 1 through 10 for all numbers of hidden channels 1 through 10 (each with an arbitrary scaling factor of two), although they are not all shown here, in order to check for possible multiple maxima of evidence. Figures 4.3 and 4.4 show some sample reconstructions created in order to assign values to these parameters. In fact, the distribution of evidence values for the reconstructions was well behaved and had a single peak, corresponding to eight hidden channels and a weighting factor of seven. This is the most probable MaxEnt reconstruction and is shown in Figure 4.5. For comparison, a previously published kriged estimate of density distribution from the CCAMLR 2000 survey is reproduced in Figure 4.6.

Table 4.2: A selection of statistics relating to the reconstructions shown in Figures 4.3 and 4.4. The highlighted rows show the reconstruction chosen, based on the highest evidence.

Channels	Weighting	Evidence	StdDev	Mean Value	Mean Error	Max Value	Max Error	CV(%)	Biomass (Mt)
8	***5***	-12595.1	101.6	17.4	5.5	5657.1	191.3	26.8	42.81
8	***6***	-12533.6	367.0	48.7	5.4	12261.6	207.7	12.1	94.61
8	***7***	-12508	994.7	99.5	4.9	43606.5	209.2	4.9	207.98
8	***8***	-12508.8	1365.9	150.1	4.8	83340.9	182.0	3.0	341.59
8	***9***	-12512.3	2074.7	225.8	4.9	130628.3	161.8	2.0	500.50
8	***10***	-12516.1	4608.1	418.5	5.0	188896.3	147.4	1.5	699.19
5	7	-13123.4	250.0	10.2	1.3	925329.7	68.5	6.3	88.44
6	7	-12751.3	505.4	23.5	1.8	627868.5	114.7	3.4	218.79
7	7	-12561.4	20604.3	464.1	2.6	3637959.7	81.9	0.8	1290.40
8	7	-12508	994.7	99.5	4.9	43606.5	209.2	4.9	207.98
9	7	-12546.9	59.2	8.7	13.3	1011.8	104.5	157.1	6.12

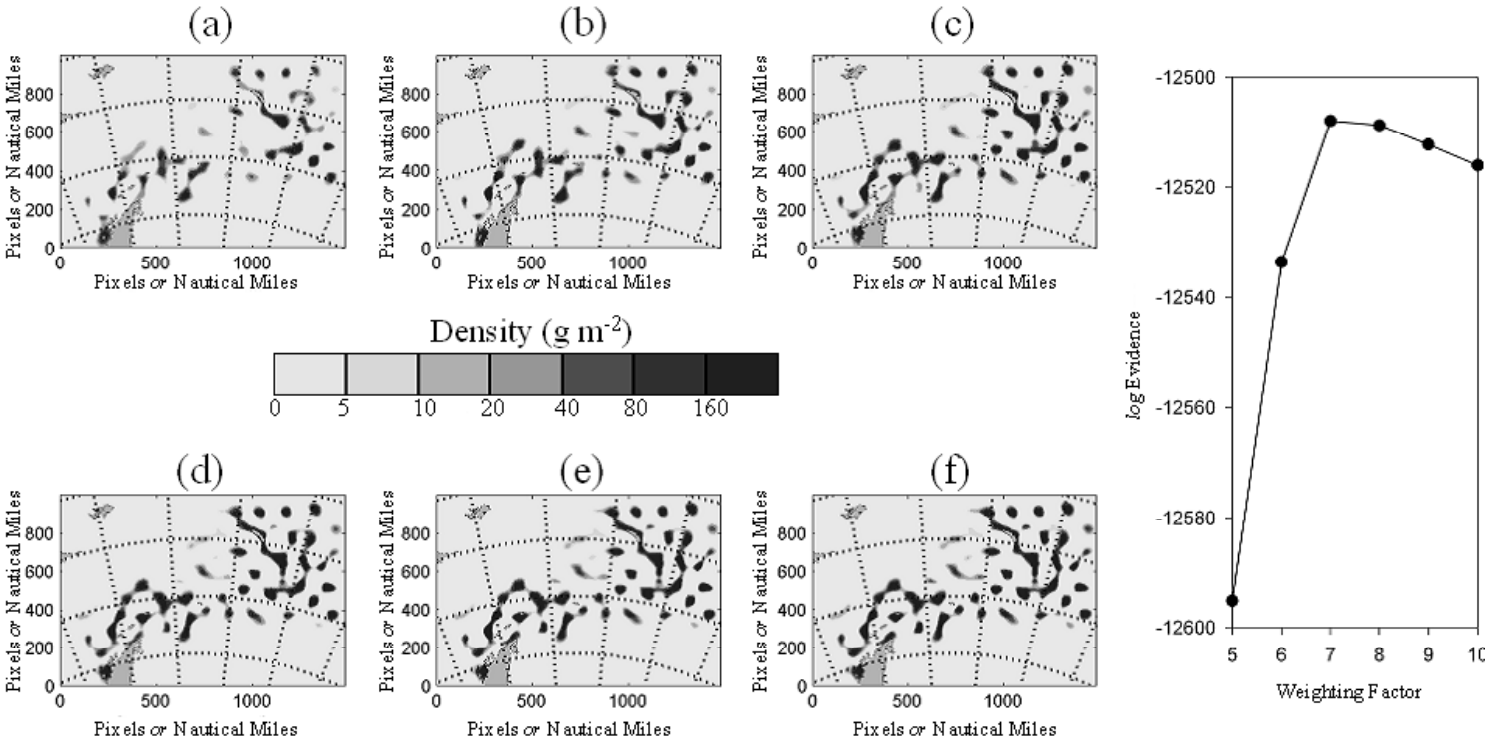


Figure 4.3: Reconstructions with values of a)5, b)6, c)7, d)8, e)9 and f)10 for the weighting factor. The associated graph shows the highest evidence to be associated with (c).

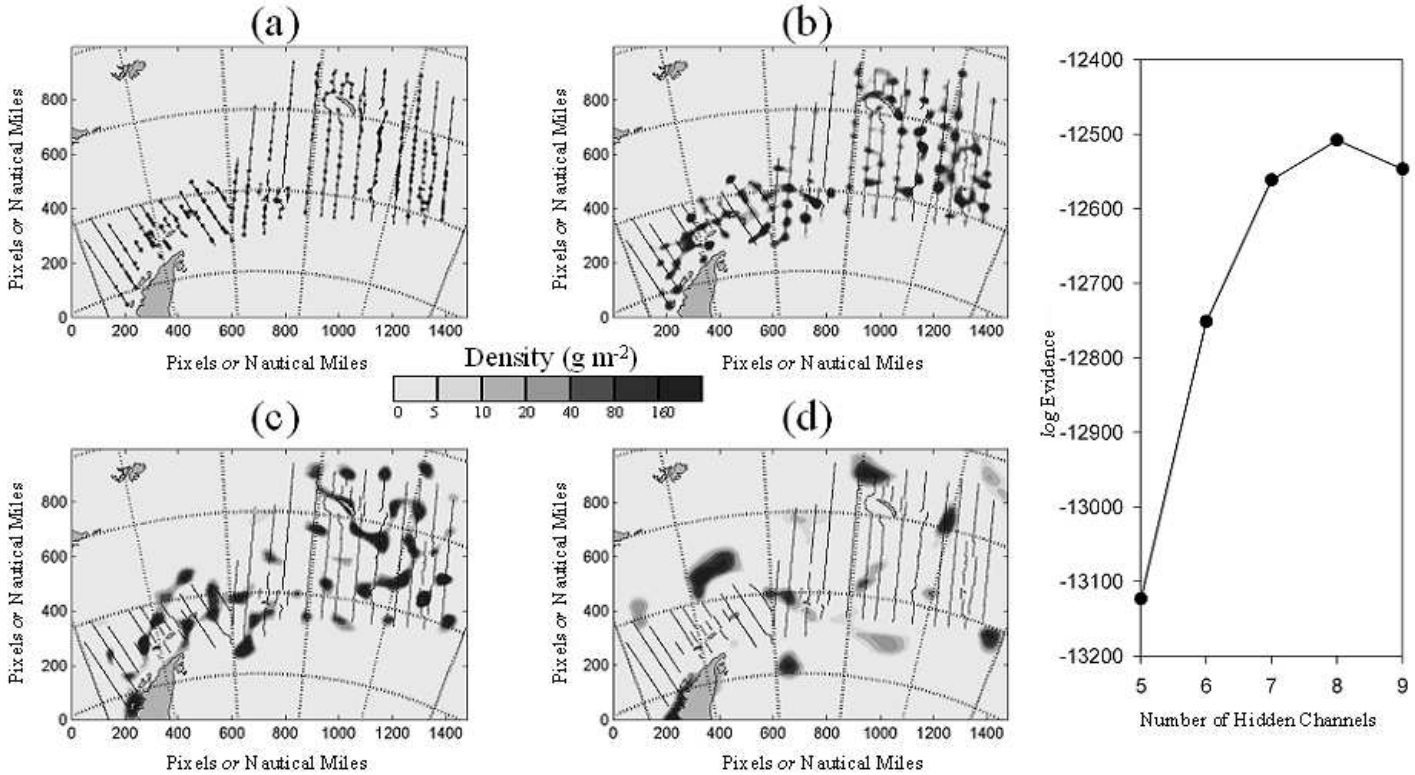


Figure 4.4: Reconstructions with a) 6, b) 7, c) 8, and d) 9 hidden channels. The associated graph shows the highest evidence to be associated with (c).

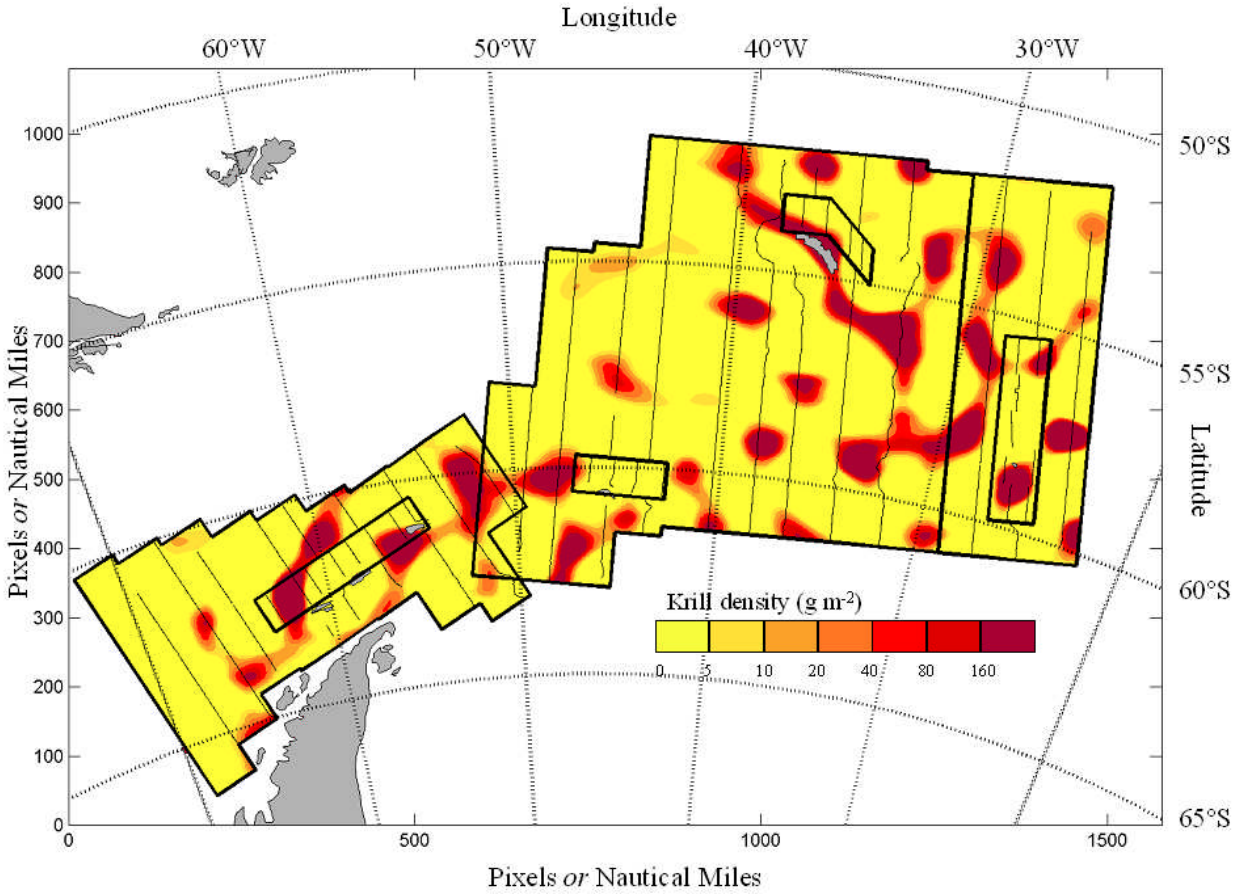


Figure 4.5: The MaxEnt result, with the bounds (thick lines) and transects (thin lines) of the CCAMLR 2000 survey. The colour scale is chosen to facilitate comparison with Figure 4.6

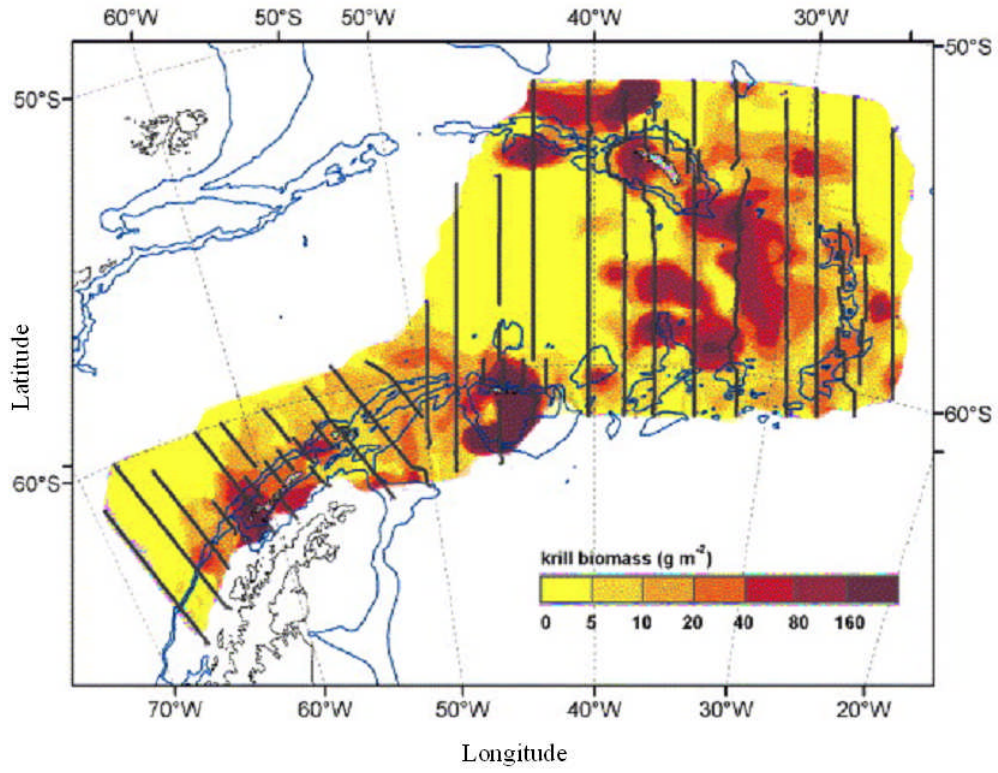


Figure 4.6: An estimate of krill density distribution from the CCAMLR 2000 survey, shown in approximately the same colour scale as the MaxEnt reconstructions, reprinted with permission from Hewitt et al. (2004a).

The total krill biomass estimate from this result is 207.98 million tonnes, with a standard deviation of 10.08 million tonnes. Biomass estimates from this result are given for each of the sixteen small-scale management units (Figure 4.1) in Table 4.3. The standard deviations calculated for each individual pixel value of the image are shown in Figure 4.7. For some pixels the predicted standard deviation may be larger than the predicted biomass, thereby apparently suggesting the possibility of a negative biomass in that pixel; which is obviously impossible. This is a common problem with noisy data, and may in the case of MaxEnt be related also to a local failure of the Gaussian approximation to Pr(***h***|data) that is used in the calculation of standard deviation (Gull and Skilling 1991).

It is prudent to note, when looking at the biomass and standard deviation values, that issues of calibration, target strength, krill orientation, species identification, sea-bottom detection etc., whilst of great importance to the final biomass and error estimates, lie beyond the scope of this paper – we merely wish, given previously calculated on-transect density values, to reconstruct the most probable off-transect distribution. For a discussion and calculation of the errors involved in the gathering and processing of the CCAMLR 2000 data, see Demer (2004)

Table 4.3: Biomass estimates for the SSMUs in CCAMLR Area 48.

		Mean Density		Max Density			
Survey		60.4		23,832.7			
Entire Reconstruction		101.1		43,569.5			
SSMU	***AREA***	***Mean Density (MaxEnt) g m^{-2}***	***Mean Density (Hewitt) g m^{-2}***	***Max Density (MaxEnt) g m^{-2}***	***Biomass (MaxEnt) Mt***	***Biomass (Hewitt) Mt***	***CV (Maxent) (%)***
APPA	***1***	134.9	11.2	28,558.9	65.192	5.414	5.1
APW	***2***	20.5	37.7	237.6	0.753	1.384	70.9
APDPW	***3***	2,270.5	37.7	28,558.9	35.874	0.596	2.4
APDPE	***4***	24.2	37.7	1,990.3	0.397	0.618	16.2
APBSW	***5***	7.6	37.7	89.6	0.167	0.829	336.5
APBSE	***6***	48.1	37.7	2,306.2	1.381	1.082	2.7
APEI	***7***	302.4	37.7	8,195.3	10.946	1.365	0.8
APE	***8***	0.0	37.7	0.3	0.003	2.322	655.5
SOPA	***9***	38.1	24.5	4,886.0	30.799	19.816	16.5
SOW	***10***	61.4	150.4	947.2	0.988	2.421	17.4
SONE	***11***	0.5	150.4	6.4	0.005	1.624	8,751.1
SOSE	***12***	15.7	150.4	136.4	0.243	2.331	274.5
SGPA	***13***	61.7	24.5	5,385.9	57.194	22.721	3.5
SGW	***14***	40.7	39.3	612.5	1.742	1.682	1.2
SGE	***15***	64.4	39.3	2,998.6	3.554	2.169	1.6
Sub Area 48.4	***16***	235.3	n/a	43,569.5	198.163	n/a	2.2

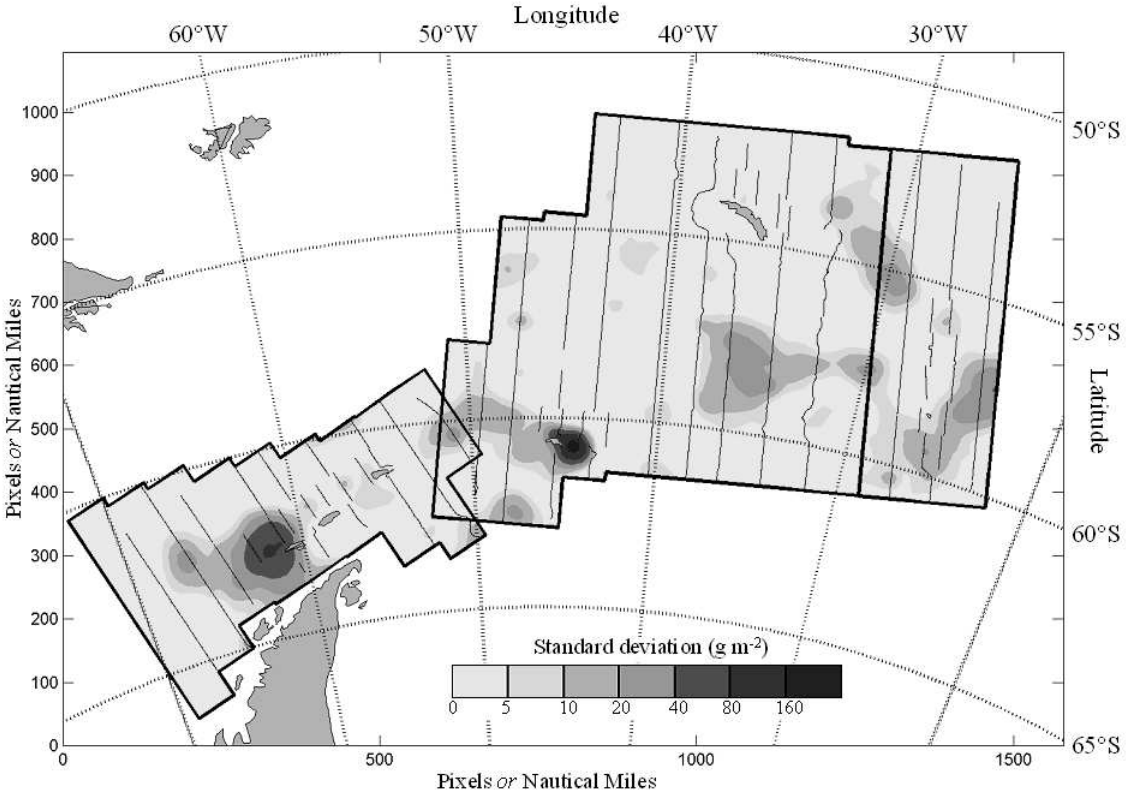


Figure 4.7: The calculated standard deviation for the pixels of the chosen MaxEnt reconstruction, with the survey transects overlaid.

Even with this exceptionally large data-space (1578×1094 = 1,726,332 pixels, of which we have data for 9,586 pixels or just 0.56%) results were obtained in 30-40 iterations of the algorithm, taking in total around 40 minutes on a 2.8 GHz Pentium 4 PC running Windows 2000. This is a sufficiently short time to allow the processing of the number of reconstructions needed to identify the image with the best evidence. Very thorough attention was paid to creating a wide range of possible reconstructions in order to ensure that the chosen result was indeed justified. However, our experience with this data has supported the expectation that there is a single maximum for evidence (Gull and Skilling 1991). This being the case, there is no need to continue to increase or decrease parameters such as weighting factor after a maximum has been passed, and very many fewer reconstructions are required to choose a result. Therefore, for confident reconstructions of, for example, images from any subsequent survey of CCAMLR Area 48 with a similar extent of data coverage, ten to fifteen reconstructions would be sufficient.

Direct testing of the quality of the reconstruction is not straightforward, since the 'truth' of the off-transect distribution of krill density is not and cannot be known. Concurrent with this work on the CCAMLR 2000 data, however, is an investigation of the success of the MaxEnt technique in recreating a simulated data set (based on the distribution of herring in the North Sea - see Simmonds et al. 2002). Although the area of this reconstruction from simulated data is very much smaller than the reconstruction area of the CCAMLR data (50,625 nm^2 as opposed to 1,726,332 nm^2), the data are integrated on a much finer scale (440 m as opposed to 1 nm), and hence

the reconstruction of 1024×896 pixels is similar in magnitude to the 1578×1094 pixel reconstruction of the CCAMLR 2000 data. Values are known for every pixel in this simulated data set, and hence distributions reconstructed from virtual surveys along imaginary transects can be compared with the 'true' distribution. Many different virtual surveys can be generated, covering a higher or lower percentage of the total data space, with different transect spacing and orientation, and with different levels of noise added to the data, and so on. The preliminary results of this work so far show that the MaxEnt technique is capable of robust and accurate reconstructions of images from acoustic survey data with skewed distributions. Specifically, a statistical hypothesis test (Syrjala 1996) failed to find significant difference spatially between the original simulated data and our chosen reconstruction, generated from a virtual survey which provided values for approximately 0.56% of the pixels, the same percentage as is available for the CCAMLR 2000 reconstructions.

4.4 Discussion and Conclusion

The resulting 'best' reconstruction (Figure 4.5) exhibits a number of the qualities we might wish to see, and indeed which we might ordinarily use to select the best reconstruction. It is reassuring that, using only the Bayesian evidence value, we have been led to choose this reconstruction. Firstly, the reconstruction does not show structure parallel to the original survey transects. Naturally, a failure to properly fill the gaps between transects would lead to an unsupported bias toward higher values on-transect than off-transect. Such transect-related structure was evident in krill reconstructions prepared in previous papers (Brierley et al. 2003b; Wafy et al. 2003), and suggests that a further increase in the number of hidden channels would have improved their results. It is a minimum requirement of a plausible reconstruction that the gaps be filled; equally, any further smoothing beyond the point at which transect-related structure disappears would unnecessarily reduce the information content of the image. It is informative that as we look at reconstructions with 6-9 hidden channels (Figure 4.3), the first image in which transect-related structure is not evident is at eight hidden channels, which corresponds to the highest evidence value. Furthermore, the excessive smoothing with nine hidden channels does indeed result in a fall in the evidence value.

Secondly, the result demonstrates that MaxEnt can assign density maxima off-transect (Figure 4.5, for example around 40°W, 56°S). Certainly there is no reason to believe that the survey transects happened to pass through all the regions of highest krill density, so this behaviour is very welcome. Inevitably, the position and size of such maxima is only probabilistically determined from very limited data. However, since the use of the MaxEnt prior ensures that any structure in the reconstruction *must* be based on the data (Gull and Daniell 1978; Gull and Skilling 1991), some level of support for these off-transect maxima *must* exist within the data set. This is a clear example of the spatial information which would be ignored by and hence lost to the conventional Jolly and Hampton (1990) approach.

Thirdly, the resulting total biomass estimate for the whole survey area of 207.98 ± 10.08 Mt is plausible, compared with the Demer and Conti (2005) estimate of 109.4 ± 11.38 Mt generated from the same on-transect data values by the Jolly and Hampton (1990) method. The MaxEnt estimate is substantially larger, but not implausibly so, given the tiny number of data involved and the fact that higher off-transect densities

are possible. There is considerable doubt about what lies between transects, and it is reasonable to expect different statistical analysis methods to produce different biomass estimates. One measure of support for this higher biomass estimate is that it accords much better with estimates that predator populations require a krill biomass of between 150-300 million tonnes to sustain them (Priddle et al. 1998; Smetacek and Nicol 2005).

It is worth remembering that the CCAMLR 2000 survey was designed specifically with the Jolly and Hampton (1990) approach in mind. Survey transects were pseudo-randomly placed within chosen strata, in accordance with the Jolly and Hampton formalism. Conversely, from a MaxEnt point of view, it would be more suitable to have the transects evenly spaced. Those regions where the transect spacing was as high as 175 km will inevitably present greater challenges to reconstruction than areas where the separation was as low as 75 km (or even lower, where extra survey effort was concentrated, for example to the North of South Georgia). We would suggest that future surveys be designed on a regular grid.

Summary statistics of krill biomass are shown (Table 4.3) for each of the SSMUs shown in Figure 4.1 and listed in Table 4.1. However, substantial parts of regions 1, 2, 8, 9, 13 and 16 lie outside the survey area. In these cases, the mean density, calculated over a small part of the SSMU, has to be applied to the whole SSMU to generate biomass estimates. It is likely that these mean densities are not representative of the whole SSMU, and hence the biomass estimates may not be reliable. For example, in Antarctic Peninsula East (SSMU 8) the biomass estimate of 0.003 Mt reflects the fact that very few krill were found in the tiny part of SSMU 8 that was in the survey area (in fact, although the survey bounds do encroach into SSMU 8, no actual transects do – see Figure 4.1). In the absence of more data, little can be done to better calculate biomass for such regions.

Hewitt et al. (2004b) took the mean for *all* SSMUs 2-8 (and 10-12, 14-15) and applied this to each of the individual areas (see Table 4.3). This is perhaps the best overall option available, particularly since Hewitt et al. required solid figures to continue their analysis of possible catch limits, but may not provide the best answer for those SSMUs (e.g. 3, 4) where more data are available. We have chosen to report all biomass estimates as calculated, whilst accepting that some values are subject to severe uncertainty. What Table 4.3 indicates most clearly is the level of ignorance about these density and biomass values. Bearing in mind that we used data based on the Demer and Conti (2005) target-strength model, which led to a biomass estimate about 2.5 times larger than previous estimates (Hewitt et al. 2002), it is fair to assume that the SSMU biomass and mean density estimates from Hewitt et al. (2004b) are underestimates also. However, the MaxEnt biomass estimates are sometimes appreciably lower than those reported in Hewitt et al. (2004b) calculated using the Jolly and Hampton (1990) method, even in regions (e.g. 10-12) with reasonable survey coverage. It seems that any putative ‘best estimate’ of biomass in SSMUs is subject to enormous uncertainty, and estimates from the CCAMLR 2000 data may not lead to good decisions about catch limits. The application of the ‘precautionary principle’ would suggest that, in this state of relative ignorance, catch limits for at least some of the SSMUs should be set at extremely low levels.

In addition to generating biomass estimates, the second stated ambition of this research is to generate accurate, useful maps of krill distribution. Figures 4.5 and 4.6 show the MaxEnt reconstruction and an estimated map taken from Hewitt et al. (2004a). There are clear differences between these maps. The question of which map better represents the actual krill distribution is hard to resolve in the absence of more data. Hewitt et al. (2004a) offer only a very short paragraph to explain how their map was created, since that paper was much more concerned with estimating krill biomass than with mapping dispersion. It is of course not impossible that the map in Hewitt et al. (2004a) is more correct than the MaxEnt map, but our analysis provides reason to believe the MaxEnt solution to be more probable. We believe the MaxEnt solution to be the most accurate map yet published from the CCAMLR 2000 data.

I submit that any map from which biological or stock-management inferences are to be drawn should be very carefully considered, since situations of sparse data allow so many different possible maps to be consistent with that data. We further suggest that the MaxEnt formalism, with its preference for smoothness and its sound probabilistic basis, is a useful framework for refining our best estimates.

The errors calculated for the reconstruction, shown in Figure 4.7, do not exhibit large amounts of unwanted structure. High values of standard deviation are generally found only where there are high density values, and thus represent small percentage errors. The exception is to the west of the South Orkney Islands. In terms of the MaxEnt calculations, this means that there are large variations in the density estimates in this area between the chosen solution and almost-as-probable candidate solutions. This may be due to a lack of information in the local data; at this time we are not able fully to explain this error peak. It is possible that approximations in the algorithm, used to overcome intractability in some of the calculations, are less than sufficiently accurate for data in this region.

There are undoubtedly improvements still to be made to our reconstruction algorithm. Specifically, the Gaussian blurring function is used simply for ease of computation and has no particular basis in biology. One of the main aims of our research in the coming months is to redress this situation. Since the MaxEnt formalism treats the blurring function as entirely separate from the MaxEnt prior distribution, we are free to adjust it without fear of compromising the validity of the technique. In fact, different blurring functions will produce different evidence values, and selection between alternative functions becomes just another exercise in parameter estimation. As a first step, future investigations will use different blurring widths in different directions, allowing better reconstruction of data exhibiting significant anisotropy, the effects of which have not been closely considered in this paper.

Another possible route to better estimates could be found by using more information that is external to the data set in question. Any information derived from a survey cannot be used to define a prior for analysis of that survey, since such an idea is obviously circular; however, information from other sources, other surveys or experiments, may legitimately form part of the prior information for an analysis. There is a relatively straightforward way to include such information in the formalism. Currently, a uniform first trial image is used (see ‘Calculating the Quantified MaxEnt Result’ above), since we claim no prior knowledge of the distribution of the species. However, there may be relevant knowledge available – for

example, a known relationship between water depth and species density. In such circumstances, we may be justified in basing our first ‘guess’ on some function of the bathymetry of the survey area. In a very simple case, for example where species density could be thought approximately linearly related to water depth, our starting trial image would simply consist of the water depth in each cell of the survey area scaled by an appropriate constant. Note that the MaxEnt prior distribution was a formula expressing our relative belief in different possible prior images – whether we start from a uniform or non-uniform trial image, that formula still expresses our expectations about the relative probability of changes to that trial image made after comparison to the data.

In the case of krill, there is evidence (Trathan et al. 2003) that krill density is normally significantly higher in on-shelf than in off-shelf regions. Using this information, we can choose a first trial image, based on the bathymetry of the area, with very low density values in off-shelf areas. This starting image would then ensure a bias towards higher on-shelf densities that may not be deducible from the CCAMLR 2000 data directly, but which we are entitled to predict based on other available information. This can reasonably be expected to further refine our estimates of krill biomass and density distribution, and is a logical next step for our research. The ability to make use of such external information represents one of the strengths of the MaxEnt formalism.

One of the aims of this paper has been to consider the possible advantages of MaxEnt as an alternative to kriging and to Jolly and Hampton (1990) analysis. In reference to kriging, we believe that the use of the Evidence value offers a chance to objectively compare competing MaxEnt reconstructions in a way that is not always available when comparing, for example, reconstructions from two different types of kriging. Additionally, the possibility of including external information such as the bathymetry may represent a far bigger advantage over traditional geostatistics.

With respect to the Jolly and Hampton (1990) method, the obvious advantage of MaxEnt is in making use of the spatial information contained in the data. This advantage may not be pronounced in this study, due to the enormous transect spacing of the CCAMLR 2000 survey, but can be expected to be more crucial for smaller scale surveys such as those studied by Brierley et al.(2003b).

In conclusion, it is our belief that the MaxEnt procedure shows significant promise as a reconstruction technique, and also as an alternative to the Jolly and Hampton (1990) method of calculating total biomass. The CCAMLR 2000 data is a particularly strong challenge for the algorithm, but it so far appears that plausible reconstructions are possible. Our attempts to generate biomass results demonstrate the shortage of data in certain SSMUs. This should emphasise that, regardless of apparent errors calculated by this method or by the Jolly and Hampton (1990) method, enormous uncertainty exists in these biomass estimates, and this must be taken into account when setting allowable catch limits.

Chapter 5: Biomass and Distribution of Jellyfish and Fish in the Namibian Benguela

This chapter comprises the text of a paper published in the journal Current Biology Vol. 16, No. 13, by Lynam, Gibbons, Axelsen, Sparks, Coetzee, Heywood and Brierley. My input in this paper was in generating the maps and biomass estimates from the survey data, which would have been difficult and statistically questionable using frequentist statistics due to the non-standard survey design. I include the whole paper to give context to the work, along with a short introduction discussing the particular reasons why MaxEnt was employed.

5.1 Why use MaxEnt?

The survey data used in this study are not compatible with the frequentist Jolly and Hampton analysis, forming as they do a zigzag pattern (see figure 5.1 C-F below) rather than a series of parallel, randomly-spaced transects. Because the transects are nearer together as we approach the corners of the zigzag survey, assumptions about the statistical independence of the data do not hold. In particular, it is an underlying assumption of the Jolly and Hampton analysis that transects are independent. As long as the transect spacing is much larger than the autocorrelation range of the data, this assumption can be said to hold. However, when transects become very close together, as at the corners of a zigzag survey, this condition is breached. The accepted way to make the data (approximately) compatible with the Jolly and Hampton analysis would be to calculate the autocorrelation range of the data, and truncate the transects by at least this distance at either end, ignoring data collected near the corners. In the case of the Benguela survey, a significant fraction of the survey data would be ignored. This is clearly unsatisfactory, both because of the waste of useful data, and also because the data near the corners can in fact contain the *most* spatial information about the scale of swarms - the corners of the zigzag survey have a much greater effective survey effort than the middle of each transect.

Therefore, since biomass estimates were so vital to the work, we decided to use the MaxEnt technique, which does not require such starting assumptions. On a technical note, all the reconstructed maps in the article (chosen because they generated the highest evidence value) were produced at Nscales=8, although the optimum weighting factor varied between 3 and 5 for the different species.

5.2 Jellyfish overtake fish in a heavily fished ecosystem

Christopher P. Lynam, Mark J. Gibbons, Bjørn E. Axelsen, Conrad A. J. Sparks, Janet Coetzee, Benjamin G. Heywood and Andrew S. Brierley.

Over the past half century fishing has led globally to a reduction in the mean trophic level of commercially landed species, with a significant decline from large predatory fish toward plankton-eating pelagic species and low trophic-level invertebrates (Pauly et al. 1998). An implied endpoint of this ‘fishing down marine food webs’ is a

proliferation of previously suppressed gelatinous plankton (jellyfish) (Pauly et al. 2002) thriving on the food no longer consumed by fish. We report here that, in the heavily exploited northern Benguela off Namibia, a transition towards this endpoint has occurred, and jellyfish biomass (12.2 million tonnes (MT)) now exceeds the biomass of once-abundant fish (3.6 MT). This is a profound ecosystem change, with possible consequences from carbon cycling to fish stock recovery.

The northern Benguela is a highly productive eastern-boundary ecosystem fertilised by upwelling, nutrient-rich waters. Historically the region supported large stocks of fish, including sardines (*Sardinops sagax*) and anchovies (*Engraulis encrasicolis*), but heavy fishing pressure has reduced stocks, and total landings have fallen from around 1.7 MT in the late 1970s to just 0.1 MT now (Figure 5.1A). Prior to this period of heavy exploitation, large jellyfish (Scyphozoa and Hydrozoa) were not prominent in the Benguela ecosystem: reports of extensive plankton sampling in the 1950s and 1960s do not mention large jellyfish, although numerous small gelatinous species (e.g. ctenophores) were observed (for example (Hart and Currie 1960)).

Following early collapses of pelagic fish stocks (in the 1960s), reports of the large and conspicuous jellyfish *Chrysaora hysoscella* (mean umbrella diameter ~27 cm (Brierley et al. 2001)) and *Aequorea forskalea* (mean umbrella diameter ~13 cm (Brierley et al. 2001)) became increasingly common (Venter 1988). Since the 1990s, reports of these jellyfish have been ever-increasing, particularly because of the nuisance they now cause to fishing (bursting trawl nets, spoiling catches), power generation (blocking power station coolant intakes) and diamond mining (blocking alluvial sediment suction). Despite their present prevalence (the term ‘jellyfish explosion’ has been used (Heymans et al. 2004)), the ascendance of jellyfish has not been quantified, and ecosystem studies have had largely to ignore jellyfish because of a lack of quantitative data on biomass and distribution (Heymans et al. 2004).

We have conducted a series of research cruises to study *C. hysoscella* and *A. forskalea* (previously *A. aequorea*) in the northern Benguela (for example (Brierley et al. 2001) and (Sparks et al. 2001)). The most recent, in August 2003, was a survey to map distribution and estimate biomass. We used multi-frequency scientific echosounders and trawl nets to sample jellyfish and fish (see Supplemental data below) along the entire Namibian shelf — between the Angolan and South African borders and the 25 m and 350 m depth contours, an area of 33,710 square nautical miles (Figure 5.1B–F). We estimate that the biomass of jellyfish was 12.2 million tonnes (99% by mass *A. forskalea*, mean jellyfish density 361 T n.mi^{-2}, standard error 22 T n.mi^{-2}), and that the total biomass of fish was 3.6 MT (Cape horse mackerel 1.1 MT, mean 33 T n.mi^{-2}, SE 1.5 T n.mi^{-2}; Cape hake 1.7 MT, mean 50 T n.mi^{-2}, SE 2.3 T n.mi^{-2}; clupeids 0.8 MT, mean 23 T n.mi^{-2}, SE 1.0 T n.mi^{-2}). Our fish biomass estimates are consistent with independent fish stock assessments conducted by others for fishery-management purposes (see Supplemental data) and our maps of jellyfish distribution are consistent with recent qualitative surveys (Sparks et al. 2001).

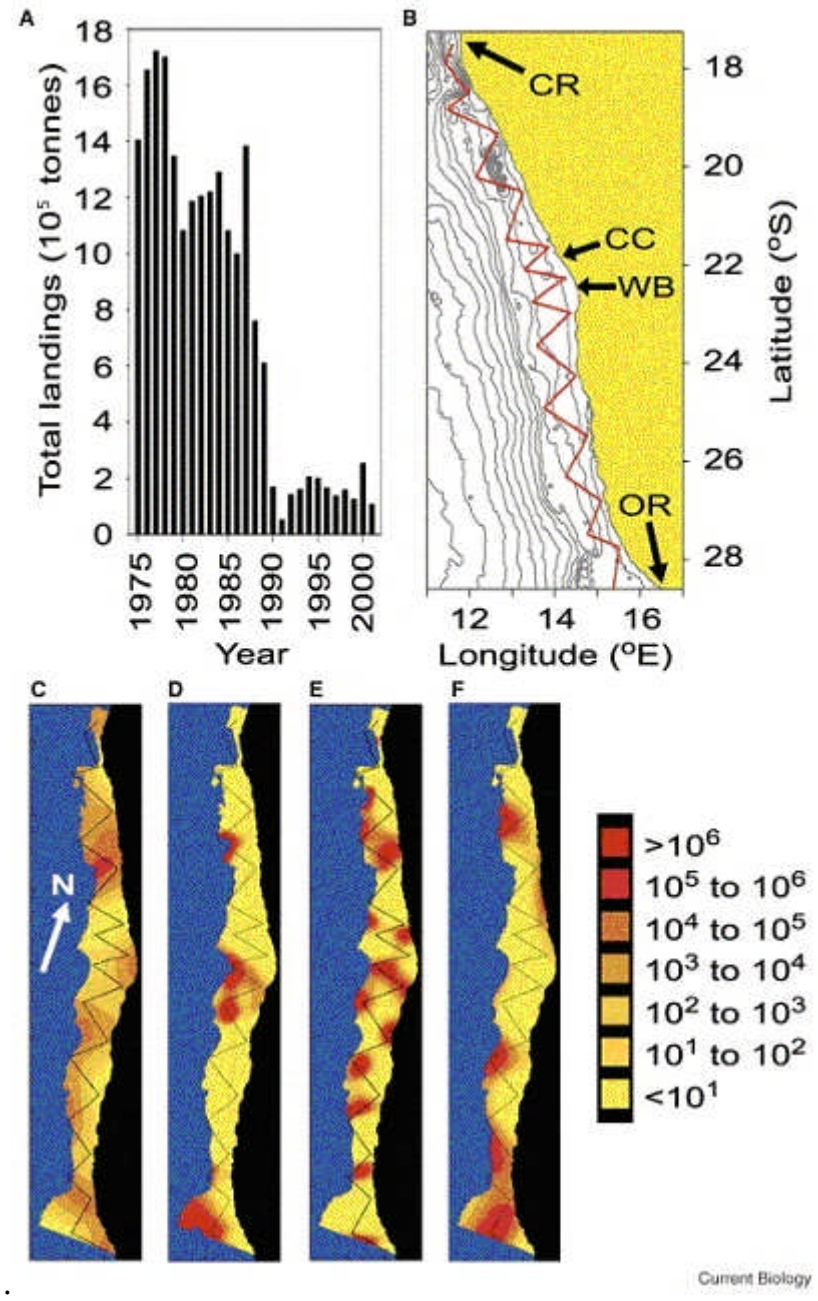


Figure 5.1.
Fish and jellyfish in the Namibian Benguela. (A) Time series of total fish landings from the northern Benguela (from FISHSTAT+ www.fao.org/fi/statist/FISOFT/FISHPLUS.asp. Data are Capture Production in the South East Atlantic Major Area 47, Western Coastal Subarea, Divisions 1.3 Cunene, 1.4 Cape Cross, and 1.5 Orange River, covering 15°S to 30°S and from the coast to 10°E). FISHSTAT+ data extend from 1975 only, so do not cover the large sardine crash in the 1960s (Boyer & Hampton 2001). (B) Bathymetric map (grey contour lines at 100, 200, 300, 500, 700, and 1000 to 4500 in 500 m increments) and the cruise track (solid red line) followed southward from the Angola–Namibia border to the Namibia–South Africa border. CR, Cunene River; WB, Walvis Bay; CC, Cape Cross; and OR, Orange River. (C–F) maps of distribution of jellyfish and fish from 17°15'S 11°28'E to 28°45'S 15°50'E: (C) Chrysaora hysoscella; (D) Aequorea forskalea; (E) Cape horse mackerel/Cape hake; and (F) clupeids (sardine, anchovy and round herring combined). Colour scale is density, tonnes per nautical mile^2.

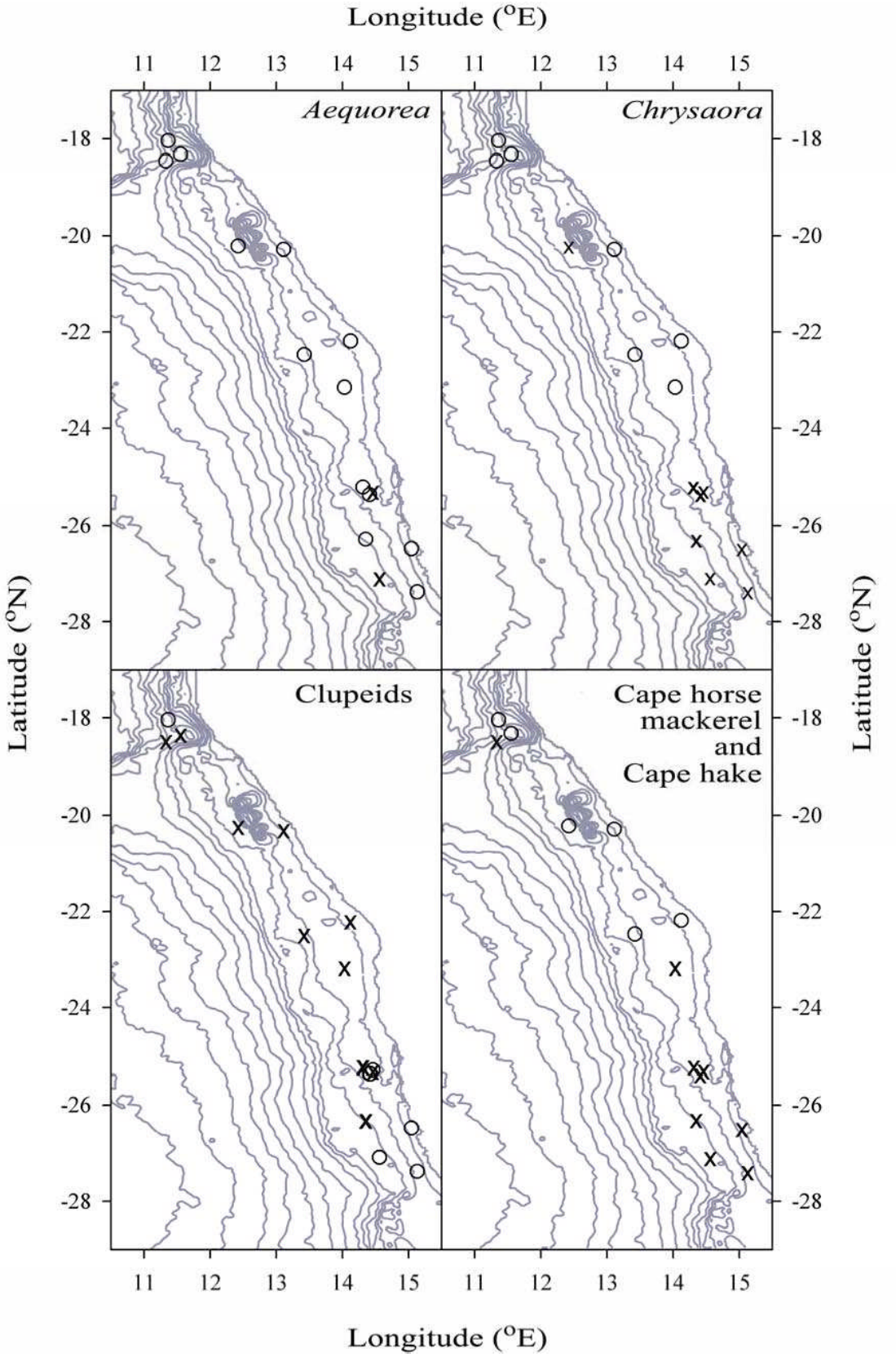


Figure 5.2

Presence (circles) and absence (crosses) by location in trawled samples of jellyfish (Chrysaora hysoscella and Aequorea forskalea) and fish (Cape horse mackerel/Cape hake and clupeids: anchovy, sardine and round herring). Two or three samples were taken using the pelagic trawl at each location. Bathymetry is as in Figure 5.1.

Jellyfish biomass has increased substantially in several locations worldwide, perhaps as a consequence of fishing (Mills 2001). Ecosystem shifts from dominance by fish to dominance by jellyfish may be irreversible as jellyfish are predatory upon fish eggs and larvae, and strong competitors for fish food (Lynam et al. 2005) and (Sommer et al. 2002): jellyfish may thus impede the recovery of fish stocks even after a cessation of fishing. Jellyfish proliferation may also be climatically driven (Mills 2001), either directly or in response to the impact of environmental perturbations on abundance and distribution of fish: an El Niño event in the Benguela in 1963 contributed to the sharp decline in sardine biomass (Boyer and Hampton 2001), and may have presaged the early establishment of jellyfish. In the north Atlantic there is a significant link between climate, as encapsulated by the North Atlantic Oscillation index, and jellyfish abundance. During the last boreal winter (2005/06) the North Atlantic Oscillation was in a pronounced negative phase (see http://www.cgd.ucar.edu/cas/jhurrell/indices.data.html#naopcdjfm) and, if previous patterns persist (Purcell 2005), this could result in outbreaks of jellyfish in coastal waters on both sides of the ocean this summer (2006). Such outbreaks have consequences for fisheries (Lynam et al. 2005) and are trophic dead-ends (jellyfish have few predators) with consequences for carbon cycling. Jellyfish play potentially major controlling roles in marine ecosystems and, in this era of apparent jellyfish ascendancy, marine ecosystem managers and modellers cannot afford to ignore them.

5.3 Supplemental Data

[The following supplementary section was published online, as was Figure 5.2.]

JELLYFISH OVERTAKE FISH IN A HEAVILY FISHED ECOSYSTEM
Christopher P. Lynam, Mark J. Gibbons, Bjørn E. Axelsen, Conrad A. J. Sparks, Janet Coetzee, Benjamin G. Heywood, and Andrew S. Brierley

Supplemental Methods

Survey methods - We surveyed the Namibian continental shelf (28°38'S to 17°15'S, Figure 5.1B) between August 20 and 31 2003 from the Research Vessel *Dr Fridtjof Nansen*. We sampled continuously with calibrated echosounders (Simrad EK500s operating at 18, 38, 120 and 200 kHz) along an approximately 1800 nautical mile zig-zag track (Figure 5.1B) from c. 10 m depth to the seabed. We fished periodically with pelagic and demersal trawls (codend mesh 22 mm) to identify acoustic targets and obtain jellyfish and fish for size-frequency analysis. Species caught most frequently were the jellyfish *A. forskalea* (25% of total biomass caught, present in 72% of trawls) and *C. hysoscella* (43% biomass, 42% trawls), and the fish Cape horse mackerel (*Trachurus trachurus capensis*) (14% biomass, 17% trawls), Cape hake (*Merluccius capensis*) (10% biomass, 19% trawls), and clupeid species (sardine *Sardinops sagax*, anchovy *Engraulis encrasicolis*, round herring *Etremus whiteheadi*, each 1.5 % biomass). Total and mean trawl catch compositions differ from total and mean acoustic estimates of community biomass but, for several reasons, this is not unexpected. Indeed, part of our motivation for studying jellyfish acoustically over the past several years (Brierley et al. 2001; Brierley et al. 2004; Brierley et al. 2005) has been to develop a method that can sample jellyfish effectively, overcoming some of the well-recognised limitations of net sampling for this group. Reasons for differences apparent here between the net-based and acoustic measures include: 1) whereas the acoustic survey track was essentially random, trawls were not targeted randomly but were aimed at acoustic features that required confirmation of identity; 2) trawls were, of necessity, deployed through the surface layer that was unsampled acoustically – the jellyfish acoustic biomass estimate (particularly the *C. hysoscella* component) may actually be biased low because on occasion jellyfish were aggregated in the very near surface; 3) *Chrysaora* and *Aequorea* have different size distributions and structural rigidities, so are not necessarily both sampled with the same net efficiency.

Biomass estimation and distribution mapping - Echo energy allocated to each species/group was integrated at nautical mile intervals along track and scaled by TS to produce density estimates per species/group per nautical mile. A Bayesian Maximum Entropy technique (Brierley et al. 2003b) that takes account of spatial autocorrelation was used to map species distribution (Figure 5.1C-F) and infer biomass (standard errors were determined by sampling from the posterior distribution). Bayesian analyses were conducted using MemSys5 Quantified Maximum Entropy software.

Supplemental Discussion

Our fish biomass estimates are consistent with independent stock estimates (Boyer et al. 2001; Boyer and Hampton 2001; Cury et al. 2005), and our Bayesian-inferred maps of fish and jellyfish distributions derived from acoustic data are consistent with species distributions evident from trawl data (Figure 5.2). We are confident that our

acoustic data provide a robust view of the distributions and abundances of jellyfish and fish. Our data suggest that in the northern Benguela jellyfish biomass now exceeds that of the once-abundant fish. It has been argued that the increases in *Chrysaora hysoscella* off Namibia may not be recent (Mills 2001), but the lack of any sightings or collections of this species during the time of the Great Expeditions (Hart and Currie 1960) seems remarkable if the species was abundant then (given that many far smaller and less conspicuous gelatinous species were collected). Our interpretation of the present survey data, in conjunction with the known timescale of the demise of commercial fish stocks (Figure 5.1A) and subsequent increasing reports of jellyfish, is that jellyfish have proliferated in the northern Benguela following release from competition with fish. Whether fishing was the sole direct cause of fish stock decline, and thus, as predicted (Pauly et al. 2002), fishing was the ultimate cause of jellyfish ascendancy, we cannot say: jellyfish increase may be a proximate consequence of fish stock decline brought about for other reasons (perhaps including temperature and oxygen fluctuations (Boyer and Hampton 2001)). Irrespective of the ultimate cause of their ascendance, however, the fact remains that jellyfish biomass on shelf in the Namibian Benguela now appears to be greater than the biomass of once-abundant commercially important fish.

Jellyfish are voracious predators on fish eggs and larvae and consequently it may be difficult for fish populations to become re-established even following reductions in fishing effort: the Benguela may have switched to an alternate stable state where jellyfish dominate [cf. (Sommer et al. 2002)]. Jellyfish have few natural predators, and gelatinous plankton are thus essentially watercolumn trophic dead ends. When they die, jellyfish sink, serving to sequest carbon and transport nutrients to the seabed. In the Southern Ocean, environmentally-mediated switches between krill (crustacean) and salp (gelatinous) dominance of the plankton have the potential to alter carbon flow radically (Loeb et al. 1997), with severe consequences for krill-dependent higher predators such as penguins and seals. In the Arabian sea, by contrast, jellyfish falls may actually be important food sources for the deep sea (Billett et al. 2006). More research is required in order to understand the full impact of the shift to jellyfish dominance in the northern Benguela, but consequences for higher fish predators including Cape fur seals and gannets could be severe [cf.(Boyer and Hampton 2001)]. Although jellyfish are fished in some areas of the world (Kingsford et al. 2000), jellyfish proliferation is not generally considered to be a desired management outcome. However, if developments in the Benguela are a portent of more general consequences of over exploitation of traditional fishery resources, jellyfish dominance may become a familiar phenomenon in years to come. This possibility should be viewed with concern not just for Africa, where per capita fish consumption has declined since the mid-1980s (Delgado et al. 2003), but globally given the inexorable demand for protein by the rising human population.

Chapter 6: Discussion and Conclusion

6.1 Recent Developments

Maximum Entropy methods continue to be used in an increasing number of diverse fields (e.g. de Nazelle and Serre 2006; Law et al. 2006; Lorenz-Fonfria and Kandori 2007). MaxEnt work in the specific area of pelagic ecology remains limited, although research is in progress on describing stock abundance of prawn and scallop in Queensland, by Dr. Norman Good at the Southern Fisheries Centre, Deception Bay, Queensland, Australia. As yet, his research is incomplete and unpublished. His intention is to use *MemSys 5* to analyse abundance distribution using daily catch data and hourly position data of a large fleet of fishing vessels.

In fisheries acoustics in general (as opposed to just MaxEnt applications) Makris et al. (2006) offer a very exciting alternative to traditional acoustic surveys. Their method enables the instantaneous imaging of large areas (thousands of square kilometres) of the ocean. The technique achieves results so fast that an image of the area can be taken every 50 seconds or so, and thus a real-time study of behaviour is possible in a way that has not been seen before. The method uses the continental shelf as an acoustic waveguide, enabling the acoustic pulses to travel enormous distances, using sound some three orders of magnitude less intense than an ordinary acoustic survey.

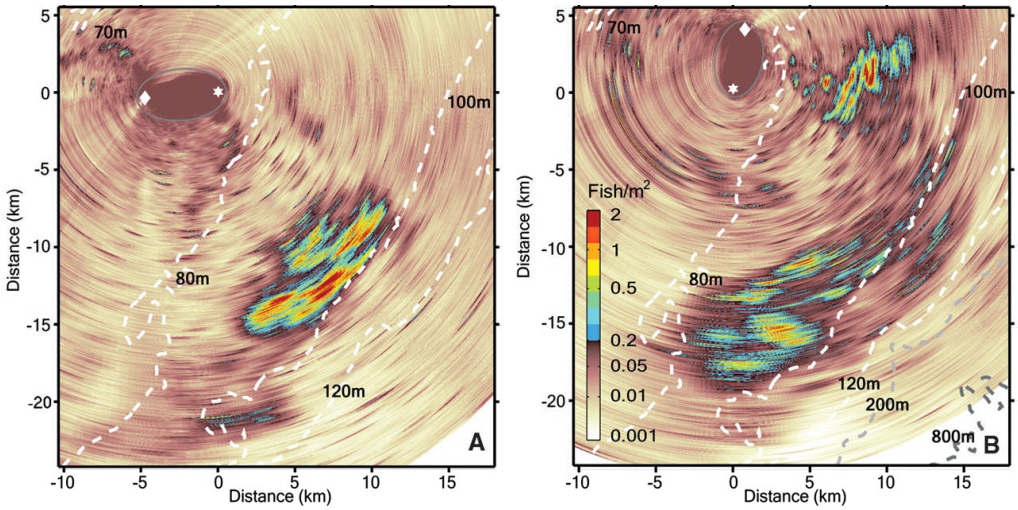


Figure 6.1

Images of the distribution of fish shoals off the coast of Long Island, New York, USA, on consecutive mornings in May 2003, reprinted from Makris et al. (2006). The star and the diamond (top left of each image) represent the moored acoustic transmitter and the receiver vessel respectively. Data cannot be collected within the grey ellipse surrounding the receiver and transmitter; this is a blind spot.

This new method offers great possibilities. The proportion of the area that is directly surveyed is many, many times greater than with a traditional, transect-based acoustic survey. In terms of creating maps of density distribution, this technique is obviously vastly superior to traditional survey techniques – not only does it sample the whole area, but it does so in real time, removing problems such as species movement during the survey period, and even allowing the creation of an animation of the movement of a species over a period of a few hours (Makris et al. 2006, supplemental online data).

Figure 6.1 shows the large differences in density distribution that occur in a 24-hour period, and highlights a problem for traditional surveys that may last a number of weeks. Fortunately, many species studied by acoustic surveys exist far more abundantly on-shelf than off-shelf, and so the fact that the technique is limited to on-shelf areas (because the sea-bed acts as a waveguide) will be a relatively minor problem in many cases, and the technique may be adopted widely.

In the short term, the new technique does not offer a simple way to generate biomass estimates. This is because the target-strength and species differentiation models used in standard acoustic survey analysis are designed for looking at the pelagic realm from above. The new technique involves the sound waves travelling horizontally through the ocean, and entirely new models must be created (a two-degree change in the assumed swimming angle of Antarctic krill can result in a near-doubling of the biomass estimate (Demer and Conti 2005), so clearly it is not safe to use existing target strength models in this new context).

It can be expected that techniques for reconstructing density distributions from traditional acoustic surveys will continue to be useful for at least the following two reasons:

i) future surveys of off-shelf areas, and
ii) analysis of existing survey data, which will remain vital in determining ongoing trends in populations.

It is likely to be many years before this new technique becomes usable as a replacement for traditional survey methods, and research into improved analysis of traditional survey data remains of great value.

6.2 Suggested Future Work

Naturally, the most pressing need for further work is in the analysis of the accuracy of the MaxEnt technique. As outlined in Chapter 3, there are a number of methods that it would be worthwhile applying.

One interesting idea is to use the work of Makris et al. (2006) to provide the kind of fine-scale data required to test the MaxEnt reconstructions thoroughly. Running virtual surveys through this genuine data would offer a very direct test of the MaxEnt reconstruction method. There could be no question of whether the data was similar enough to the real world to provide a robust test since, unlike the herring data used in this thesis, it would *be* real world data.

The herring data were produced using a simulation, which is itself a statistical process. There is therefore an obvious circularity in using any statistical method to reconstruct it, and in particular, since geostatistics were actually used in this simulation, a geostatistical method such as kriging could not be fairly compared to the MaxEnt method. This would not be a problem with data from the Makris et al. (2006) surveys. Nick Makris has been kind enough to discuss the possibility of this avenue of research, but unfortunately the work has not been done in time for inclusion in this thesis.

Other possible future work could involve adjusting the Internal Correlation Function (ICF) used by the MaxEnt method. In theory, any smoothing algorithm could be used, and the evidence value would offer an objective way to choose the one that performs best. The ICF currently used (a simple B-spline approximation to the Gaussian (Gull and Skilling, 1991)) exists because it was sufficiently accurate for the various uses to which the *MemSys 5* software has been put in the past. It may be that other functions are more suited to biological data, and this could be investigated.

6.3 Concluding Remarks

This project has advanced the work of Brierley et al. (2003b) and Wafy et al. (2003) considerably, and has further developed the use of MaxEnt for fisheries acoustic data. It has been shown clearly that plausible reconstructions are achievable, and also that the evidence value is useful in objectively choosing the reconstruction that also seems best by other important criteria. Unarguably, more work is required in the testing of the MaxEnt reconstructions, with regard to the accuracy both with respect to the original simulated data, and with respect to the performance of other reconstruction methods.

6.4 Acknowledgements

Thanks firstly to my supervisor, Dr Andrew Brierley, for help and cajoling throughout.

Thanks to the Envrionmental Mathematics and Statistics program, supported by the National Environment Research Council and the Engineering and Physical Sciences research Council, and to the School of Biology at the University of St Andrews, for the funding which made this work possible,

Thanks to Maged Wafy for teaching me the basics of MemSys 5 at the beginning of this project.

Thanks to Steve Gull for help with the MemSys 5 software, and for processing time on his computer cluster.

Thanks to David Demer for the original on-transect krill data, Eric Appleyard for the SSMU coordinates, Cairistiona Anderson and Claire Waluda for providing the coordinates of the survey bounds, all used in Chapter 4.

Thanks also to John Simmonds for the North Sea herring mock data used to test the veracity of MaxEnt reconstructions (Chapters 3 and 4), and for the opportunity to join a two-week research cruise in the North Sea to gain an insight into acoustic data collection.

Bibliography

Barange, M., Hampton, I., et al. (1994). "Determination of Composition and Vertical Structure of Fish Communities Using in-situ Measurements of Acoustic Target Strength." Canadian Journal of Fisheries and Aquatic Sciences **51**(1): 99-109.

Barange, M., Hampton, I., et al. (1996). "Empirical determination of in situ target strengths of three loosely aggregated pelagic fish species." ICES Journal of Marine Science **53**(2): 225-232.

Barrett, R., Berry, M., et al. (1994). http://mathworld.wolfram.com/ConjugateGradientMethod.html.

Bayes, T. (1763). "An essay towards solving a problem in the doctrine of chances." Philosophical Transactions of the Royal Society **53**: 370-418.

Billett, D. S. M., Bett, B. J., et al. (2006). "Mass deposition of jellyfish in the deep Arabian Sea." Limnology and Oceanography **51**(5): 2077-2083.

Bontekoe, T. R., Koper, E., et al. (1994). "Pyramid Maximum-Entropy Images of IRAS Survey Data." Astronomy and Astrophysics **284**(3): 1037-1053.

Boyer, D. C., Boyer, H. J., et al. (2001). "Changes in abundance of the Northern Benguela sardine stock during the decade 1990-2000, with comments on the relative importance of fishing and the environment." South African Journal of Marine Science **23**: 67-84.

Boyer, D. C. and Hampton, I. (2001). "An overview of the living marine resources of Namibia." South African Journal of Marine Science **23**: 5-35.

Brierley, A. S., Axelsen, B. E., et al. (2004). "Single-target echo detections of jellyfish." ICES Journal of Marine Science **61**(3): 383-393.

Brierley, A. S., Axelsen, B. E., et al. (2001). "Acoustic observations of jellyfish in the Namibian Benguela." Marine Ecology-Progress Series **210**: 55-66.

Brierley, A. S., Boyer, D. C., et al. (2005). "Towards the acoustic estimation of jellyfish abundance." Marine Ecology-Progress Series **295**: 105-111.

Brierley, A. S., Fernandes, P. G., et al. (2003a). "An investigation of avoidance by Antarctic krill of RRS James Clark Ross using the Autosub-2 autonomous underwater vehicle." Fisheries Research **60**: 569-576.

Brierley, A. S., Gull, S. F., et al. (2003b). "A Bayesian maximum entropy reconstruction of stock distribution and inference of stock density from line-transect acoustic- survey data." ICES Journal of Marine Science **60**(3): 446-452.

Brierley, A. S., Ward, P., et al. (1998). "Acoustic discrimination of Southern Ocean zooplankton." Deep-Sea Research Part II-Topical Studies in Oceanography **45**(7): 1155-+.

Brierley, A. S., Watkins, J. L., et al. (1999). "Acoustic estimates of krill density at South Georgia, 1981 to 1998." CCAMLR Science **6**: 47-57.

Briggs, W. M. and Levine, R. A. (1997). "Wavelets and field forecast verification." Monthly Weather Review **125**(6): 1329-1341.

Burch, S. F., Gull, S. F., et al. (1983). "Image-Restoration by a Powerful Maximum-Entropy Method." Computer Vision Graphics and Image Processing **23**(2): 113-128.

Burczynski, J. J. and Johnson, R. L. (1986). "Application of Dual-Beam Acoustic Survey Techniques to Limnetic Populations of Juvenile Sockeye-Salmon

(Oncorhynchus-Nerka)." Canadian Journal of Fisheries and Aquatic Sciences **43**(9): 1776-1788.
Charter, M. K. and Gull, S. F. (1991). "Maximum-Entropy and Drug Absorption." Journal of Pharmacokinetics and Biopharmaceutics **19**(5): 497-520.
Christakos, G. and Li, X. Y. (1998). "Bayesian maximum entropy analysis and mapping: A farewell to kriging estimators?" Mathematical Geology **30**(4): 435-462.
Clark, J. S. (2005). "Why environmental scientists are becoming Bayesians." Ecology Letters **8**(1): 2-14.
Conover, W. J. (1980). Practical nonparametric statistics. New York, John Wiley & Sons.
Cox, R. T. (1946). "Probability, frequency and reasonable expectation." American Journal of Physics **14**: 1-13.
Cox, R. T. (1961). The Algebra of Probable Inference. Baltimore, Johns Hopkins University Press.
Cury, P. M., Shannon, L. J., et al. (2005). "Trophodynamic indicators for an ecosystem approach to fisheries." ICES Journal of Marine Science **62**(3): 430-442.
de Laplace, P. S. (1812). Théorie analytique des probabilités. Paris, Courcier Imprimeur.
de Nazelle, A. and Serre, M. L. (2006). "Ozone exposure assessment in North Carolina using Bayesian maximum entropy data integration of space time observations and air quality model prediction." Epidemiology **17**(6): S189-S189.
Delgado, C. L., Wada, N., et al. (2003). Outlook for fish to 2020 - Meeting global demand. Penang, International Food Policy Research Institute.
Demer, D. A. (2004). "An estimate of error for the CCAMLR 2000 survey estimate of krill biomass." Deep-Sea Research II **51**: 1237-1251.
Demer, D. A. and Conti, S. G. (2005). "New target strength model indicates more krill in the Southern Ocean." ICES Journal of Marine Science **62**: 25-32.
Dennis, B. (1996). "Discussion: Should ecologists become Bayesians?" Ecological Applications **6**(4): 1095-1103.
Efron, B. (1986). "Why Isn't Everyone a Bayesian?" American Statistician **40**(1): 1-5.
Fisher, R. A. (1925). Statistical methods for research workers. London, Oliver and Boyd.
Greenstreet, S. P. R., Tuck, I. D., et al. (1997). "An assessment of the acoustic survey technique, RoxAnn, as a means of mapping seabed habitat." ICES Journal of Marine Science **54**(5): 939-959.
Gull, S. F. and Daniell, G. F. (1978). "Image reconstruction from incomplete and noisy data." Nature **272**: 686-690.
Gull, S. F. and Skilling, J. (1984). "Maximum-Entropy Method in Image-Processing." IEE Proceedings - F Radar and Signal Processing **131**(6): 646-659.
Gull, S. F. and Skilling, J. (1991). Quantified Maximum Entropy MemSys 5: Users' Manual.
Hammond, T. R. and Swartzman, G. L. (2001). "A general procedure for estimating the composition of fish school clusters using standard acoustic survey data." ICES Journal of Marine Science **58**: 1115-1132.
Hart, T. J. and Currie, R. I. (1960). "The Benguela Current." Discovery Reports **31**: 123-298.

Hewitt, R. P., Watkins, J. L., et al. (2004a). "Biomass of Antarctic Krill in the Scotia Sea in January/February 2000 and its use in revising an estimate of precautionary yield." Deep-Sea Research II(51): 1215-1236.
Hewitt, R. P., Watkins, J. L., et al. (2002). "Setting a precautionary catch limit for Antarctic krill." Oceanography **15**(3): 26-33.
Hewitt, R. P., Watters, G., et al. (2004b). "Options for allocating the precautionary catch limit of krill among small-scale management units in the Scotia Sea." CCAMLR Science **11**: 81-97.
Heymans, J. J., Shannon, L. J., et al. (2004). "Changes in the northern Benguela ecosystem over three decades: 1970s, 1980s, and 1990s." Ecological Modelling **172**(2-4): 175-195.
Horne, J. K. (2000). "Acoustic approaches to remote species identification: a review." Fisheries Oceanography **9**(4): 356-371.
Isaaks, E. H. and Srivaslava, R. M. (1989). Applied Geostatistics. Oxford, Oxford University Press.
Jaynes, E. T. (1978). Where do we stand on Maximum Entropy? E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics. R. Rosenkrantz. Dordrecht, Reidel**:** 211-314.
Jaynes, E. T. (2003). Probability theory: the logic of science. Cambridge, Cambridge University Press.
Jeffreys, H. (1939). Theory of probability. Oxford, Clarendon Press.
Johnson, R. W. and Shore, J. E. (1980). "Axiomatic Derivation of the Principle of Maximum-Entropy and the Principle of Minimum Cross-Entropy." IEE Transactions on Information Theory **26**: 26.
Johnson, R. W. and Shore, J. E. (1983). "Axiomatic Derivation of the Principle of Maximum-Entropy and the Principle of Minimum Cross-Entropy - Comments and Correction." IEE Transactions on Information Theory **29**(6): 942-943.
Jolly, G. M. and Hampton, I. (1990). "A Stratified Random Transect Design for Acoustic Surveys of Fish Stocks." Canadian Journal of Fisheries and Aquatic Sciences **47**(7): 1282-1291.
Kingsford, M. J., Pitt, K. A., et al. (2000). Management of jellyfish fisheries, with special reference to the Order Rhizostomeae. Oceanography and Marine Biology, Vol 38. **38:** 85-156.
Korneliussen, R. J. and Ona, E. (2002). "An operational system for processing and visualizing multi-frequency acoustic data." ICES Journal of Marine Science **59**: 293-313.
Law, D. C. G., Bernstein, K. T., et al. (2006). "Modeling a syphilis outbreak through space and time using the Bayesian maximum entropy approach." Annals of Epidemiology **16**(11): 797-804.
Lee, S. I. (2001). "Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I." Journal of Geographical Systems **3**(4): 369-385.
Lizamore, S. C. "Topics in maximum entropy applications" MSc thesis, Victoria University
Loeb, V., Siegel, V., et al. (1997). "Effects of sea-ice extent and krill or salp dominance on the Antarctic food web." Nature **387**(6636): 897-900.
Lorenz-Fonfria, V. A. and Kandori, H. (2007). "Bayesian maximum entropy (two-dimensional) lifetime distribution reconstruction from time-resolved spectroscopic data." Applied Spectroscopy **61**(4): 428-443.
Lynam, C. P. "Ecological and acoustic investigations of jellyfish (Scyphozoa and Hydrozoa)" PhD Thesis, University of St Andrews

Lynam, C. P., Brierley, A. S., et al. (2004). Pinging down the food web: Multi-frequency acoustic discrimination of jellyfish and fish, ICES Report CM2004/R:06, ICES.
Lynam, C. P., Heath, M. R., et al. (2005). "Evidence for impacts by jellyfish on North Sea herring recruitment." Marine Ecology-Progress Series **298**: 157-167.
Maclennan, D. N. and Simmonds, E. J. (1992). Fisheries acoustics. London, Chapman and Hall.
Makris, N. C., Ratilal, P., et al. (2006). "Fish population and behavior revealed by instantaneous continental shelf-scale imaging." Science **311**(5761): 660-663.
Maravelias, C. D., Reid, D. G., et al. (1996). "Spatial analysis and mapping of acoustic survey data in the presence of high local variability: Geostatistical application to North Sea herring (*Clupea harengus*)." Canadian Journal of Fisheries and Aquatic Sciences **53**(7): 1497-1505.
Marshall, P. J., Hobson, M. P., et al. (2002). "Maximum-entropy weak lens reconstruction: improved methods and application to data." Monthly Notices of the Royal Astronomical Society **335**(4): 1037-1048.
McGehee, D. E., O'Driscoll, R. L., et al. (1998). "Effects of orientation on acoustic scattering from Antarctic krill at 120 kHz." Deep-Sea Research Part II-Topical Studies in Oceanography **45**(7): 1273-1294.
Mills, C. E. (2001). "Jellyfish blooms: are populations increasing globally in response to changing ocean conditions?" Hydrobiologia **451**(1-3): 55-68.
Murray, A. W. A. (1996). "Comparison of geostatistical and random sample survey analyses of Antarctic krill acoustic data." ICES Journal of Marine Science **53**(2): 415-421.
Pauly, D., Christensen, V., et al. (1998). "Fishing down marine food webs." Science **279**(5352): 860-863.
Pauly, D., Christensen, V., et al. (2002). "Towards sustainability in world fisheries." Nature **418**(6898): 689-695.
Priddle, J., Boyd, I. L., et al. (1998). "Estimates of Southern Ocean primary production - constraints from predator carbon demand and nutrient drawdown." Journal of Marine Systems **17**(1-4): 275-288.
Purcell, J. E. (2005). "Climate effects on formation of jellyfish and ctenophore blooms: a review." Journal of the Marine Biological Association of the United Kingdom **85**(3): 461-476.
Reilly, C., Price, P., et al. (2004). "Using image and curve registration for measuring the goodness of fit of spatial and temporal predictions." Biometrics **60**(4): 954-964.
Rivoirard, J., Simmonds, E. J., et al. (2000). Geostatistics for estimating fish abundance, Blackwell.
RSINC (2000). Kriging Description, Research Systems Inc. http://www.rsinc.com/services/techtip.asp?ttid=2192.
Rubinstein, R. and Kroese, D. (2004). The Cross-Entropy Method. Berlin, Springer-Verlag.
Simmonds, E. J., Fernandes, P. G., et al. (2002). School based model of the spatial distribution and dynamics of an acoustically surveyed herring distribution in the North Sea.
Sivia, D. S. (1996). Data analysis: A Bayesian tutorial. Oxford, Clarendon Press.
Skilling, J. (1988a). The axioms of maximum entropy. Maximum entropy and Bayesian methods on science and engineering. G. J. Erickson and C. R. Smith. Dordrecht, Kluwer. **1**.

Skilling, J. (1988b). Classic Maximum Entropy. Maximum Entropy and Bayesian Methods. J. Skilling. Cambridge, Kluwer Academic Press**:** 45-52.
Skilling, J. (1991). "Bayesian Reasoning." Nature **353**(6346): 707-708.
Skilling, J. (1992). "Quantified Maximum-Entropy." American Laboratory **24**(15): J32-M32.
Skilling, J. (2003). Inference - a plea for simplicity.
Skilling, J. and Gull, S. F. (1989). Bayesian Maximum Entropy. Proceedings of the AMS-IMS-SIAM Conference on Spatial Statistics and Imaging, Bowdoin College, Maine, 1988.
Skilling, J. and Sibisi, S. (1990). "Fundamentals of Maxent in Data-Analysis." Institute of Physics Conference Series(107): 1-21.
Smetacek, V. and Nicol, S. (2005). "Polar ocean ecosystems in a changing world." Nature **437**(7057): 362-368.
Sommer, U., Stibor, H., et al. (2002). "Pelagic food web configurations at different levels of nutrient richness and their implications for the ratio fish production: primary production." Hydrobiologia **484**(1-3): 11-20.
Sparks, C., Buecher, E., et al. (2001). "Observations on the distribution and relative abundance of the scyphomedusan *Chrysaora hysoscella* (Linne, 1766) and the hydrozoan *Aequorea aequorea* (Forskal, 1775) in the northern Benguela ecosystem." Hydrobiologia **451**(1-3): 275-286.
Syrjala, S. E. (1996). "A Statistical Test For A Difference Between The Spatial Distributions Of Two Populations." Ecology **77**(1): 75-80.
Tikochinsky, Y., Tishby, N. Z., et al. (1984). "Consistent Inference of Probabilities for Reproducible Experiments." Physical Review Letters **52**(16): 1357-1360.
Tompa, D., Morton, J., et al. (2000). Perceptually based image comparison. Paper presented at the 2000 International Conference on Image Processing, Vancouver, Canada.
Trathan, P. N., Brierley, A. S., et al. (2003). "Oceanographic variability and changes in Antarctic krill (Euphausia superba) abundance at South Georgia." Fisheries Oceanography **12**(6): 569-583.
Trathan, P. N., Watkins, J. L., et al. (2001). "The CCAMLR-2000 Krill Synoptic Survey: A description of the rationale and design." CCAMLR Science **8**: 1-23.
Venter, G. (1988). Occurrence of jellyfish off Southwest Africa/Namibia. Long-term data series relating to Southern Africa's renewable natural resources. I. Macdonald and R. Crawford, S. African Nat. Sci. Progs.
Vignaux, M., Vignaux, G. A., et al. (1998). "Fine-scale mapping of fish distribution from commercial catch and effort data using maximum entropy tomography." Canadian Journal of Fisheries and Aquatic Sciences **55**(5): 1220-1227.
Wafy, M. H., Brierley, A. S., et al. (2003). "Maximum entropy reconstructions of krill distribution and estimates of krill density from acoustic surveys at South Georgia, 1996-2000." CCAMLR Science **10**: 91-100.
Wang, Z., Bovik, A. C., et al. (2004). "Image quality assessment: From error visibility to structural similarity." IEE Transactions on Image Processing **13**(4): 600-612.
Weber, L. H., Elsayed, S. Z., et al. (1986). "The Variance Spectra of Phytoplankton, Krill and Water Temperature in the Antarctic Ocean South of Africa." Deep-Sea Research Part a-Oceanographic Research Papers **33**(10): 1327-1343.
Weir, N. and Djorgovski, S. (1991). "High-Resolution Imaging of the Double QSO 2345+007." Astronomical Journal **101**(1): 66-70.
Wu, N. (1997). The Maximum Entropy Method. Berlin, Springer-Verlag.

Zimmerman, D. L. (1993). "A bivariate Cramer-von Mises type of test for spatial randomness." Applied Statistics **42**: 43-54.