UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

Department of Mathematics

Modelling and Predicting Decompression Sickness: An Investigation

by

Jotham P. K. Gaudoin

Thesis for the degree of Doctor of Philosophy

July 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES
Department of Mathematics

Doctor of Philosophy

MODELLING AND PREDICTING DECOMPRESSION SICKNESS: AN INVESTIGATION

by Jotham P. K. Gaudoin

In this thesis, we shall consider the mathematical modelling of Decompression Sickness (DCS), more commonly known as 'the bends', and, in particular, we shall consider the probability of its occurrence on escaping from a damaged submarine.

We shall begin by outlining the history of DCS modelling, before choosing one particular model-type - that originally considered by Thalmann et al. (1997) - upon which to focus our attention. This model combines tissues in the body sharing similar characteristics, in particular the rate at which nitrogen is absorbed into or eliminated from the tissues in question, terming such combinations 'compartments'. We shall derive some previously unknown analytical results for the single compartment model, which we shall then use to assist us in using Markov Chain Monte Carlo (MCMC) methods to find estimates for the model's parameters using data provided by QinetiQ. These data concerned various tests on a range of subjects, who were exposed to various decompression conditions from a range of depths and at a range of breathing pressures. Next, we shall consider the multiple compartment model, making use of Reversible Jump MCMC to determine the 'best' number of compartments to use.

We shall then move on to a slightly different problem, concerning a second dataset from QinetiQ that consists of subjective measurements on an ordinal scale of the number of bubbles passing the subjects' hearts (known as the Kisman-Masurel bubble score), for a different set of subjects. This dataset contains quite a number of gaps, and we shall seek to impute these before making use of our imputed datasets to identify logistic regression models that provide an alternative DCS probability.

Finally, we shall combine these two approaches using a model averaging technique to improve upon previously generated predictions, thereby offering additional practical advice to submariners and those rescuing them following an incident.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Jotham P. K. Gaudoin , declare that the thesis entitled *Modelling and Predicting Decompression Sickness: An Investigation* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission.

Signed:.....

Date:.....

Acknowledgements

Firstly, I would like to thank the EPSRC and QinetiQ, without whose financial support this research could not have been undertaken.

Next, I would like to thank my supervisors: Professor Jon Forster, Dr. Alan Kimber and Dr. Robin Mitra at the University of Southampton, together with Geoff Loveman at QinetiQ, for their constant support, patience and encouragement and their suggestions towards this thesis. Without them, I would not have been able to complete this work. I would also like to thank my advisor, Professor Dave Woods, for his input and pastoral guidance.

Third, I would like to thank my many pedagogues over the years from Señor Maurice Thompson and Mr. Mark Harley at school to the members of staff in the Department of Russian at St. Andrews, Dr. Claire Whitehead, Professor Ian Press and the late Professor Stefan Pugh, for encouraging me to pursue postgraduate study, though I suspect this is not the subject they had in mind at the time, being neither translation studies nor Slavic philology! Thanks are also very much due to Dr. Rob Penfold who was my Open University tutor for a third of my OU degree. Not only did he provide outstanding tuition throughout my degree and write me a great many references, he also helped me to learn how to teach via correspondence as my mentor in my Associate Lecturer position with the OU. For all of these things I am especially grateful.

Next come my many friends for their hours of board-gaming and putting up with listening to me talk at them for many hours about things they likely had no interest in, particularly my long-suffering flatmates Tom and Helen. Outside the department, I have enjoyed many hours spent chatting to Chloe, Miriam, Laura and Cody over Skype and to Kez, Dawn and Raya on the phone throughout my time doing a PhD about matters both work-related and otherwise. Within the department, I must thank my many

friends who, throughout my time at Southampton, have given me a mixture of useful insights, interesting coding problems, entertaining conversations, cake and a great deal of support.

Fifth, I would like to thank my family for their support both moral and financial over the years. My mother for her tireless proof-reading and both practical and moral support, my grandparents for their financial help, my brother Japheth and uncle Dominic for providing a dose of reality, and my uncle Lucian and auntie Sheila for their frequent logistical support.

Last, and by no my means least, I would like to thank my wonderful wife, Heidi, my beautiful daughter, Beatrice, my twins, Astrid and Freyja, and any future children for giving me a reason to press on with my thesis and enriching my life every day. This thesis is for you.

Chapter 1

Introduction

1.1 Motivation

Decompression Sickness, henceforth referred to as DCS and more commonly known as 'the bends', is an ailment that is believed (see, for example, Brubakk and Neuman (2002), Chapter 9) to be caused by nitrogen bubbles forming in the bloodstream and tissues of the body. The precise mechanism of this process is a highly complex one and we shall not attempt to describe the aetiology of the condition in detail. In divers, the bubbles occur when moving from a region of higher breathed air pressure to one of lower breathed air pressure - i.e. from deep water to the surface - in too short a space of time. Symptoms occur soon after the dive has finished and, because the nitrogen bubbles can form in different parts of the body, vary from individual to individual. The diver may complain of headache or vertigo, abnormal tiredness or fatigue. They may have a rash, pain in one or more joints, tingling in the arms or legs, muscular weakness or paralysis. Less commonly, breathing difficulties, shock, unconsciousness or death may occur. Immediate treatment is 100 per cent oxygen on site and during transportation, followed by a period of time in a hyperbaric chamber that recompresses the diver and forces the nitrogen bubbles back

into solution, followed by a gradual 'ascent' in the chamber. DCS affects not only divers and submariners, who are the main focus of this thesis, but also caisson workers - a caisson being a watertight retaining structure used, for example, to work on the foundations of bridge piers, tunnels, concrete dams or for the repair of ships. Caissons are constructed such that the water can be pumped out, keeping the working environment dry. Here, the difference between breathing the compressed atmosphere within the caisson and rapidly re-entering normal (uncompressed) atmospheric conditions causes DCS. Moreover, airmen and astronauts moving from areas of standard sea-level pressure to regions of lower pressure in the upper atmosphere and beyond can also suffer from the same deleterious effects.

To date, all models for the decompression process itself have been deterministic and the parameters of such models have been either empirically estimated or, more recently, estimated using maximum likelihood methods. Models for DCS have then built on these deterministic models by attempting to evaluate the probability of contracting DCS using these deterministic models as a basis for survival analysis methods. However, the uncertainty in the models themselves has not been accounted for. In particular, in modelling the decompression process, the lack of a stochastic element is clearly unrealistic as, given the same environmental, atmospheric and biological conditions, a given subject may on one occasion experience DCS while not experiencing it on another. Further, each model is very much standalone from any other so that any information not included in the model under consideration is lost when making predictions. This is an undesirable state of affairs as it is quite possible that information gained about DCS through one model may not be incorporated in another and we are therefore not making use of all possible information from our data. To alleviate this, we seek to use Bayesian methodology to improve the inference available for these models and also to obtain measures of uncertainty regarding their estimation and usage. Our goal is to provide improved inference and measures of uncertainty so as to be able to make more informed predictions

about DCS for a given dive, as well as trying to use information from more than one model at a time so that we might further improve such predictions. Thus, we seek to make use of both decompression theory and Bayesian theory to provide more accurate predictions and measures of uncertainty that can be used when making predictions for particular dives.

In this chapter, we introduce the basic concepts of Bayesian theory that we will use throughout the thesis while in the following chapter we will consider the various physical and physiological elements of decompression theory that we draw upon. Rather than referencing each section separately, we note here instead that we have drawn on amalgamation of Lee (2004) in Section 1.2, O'Hagan and Forster (2004) in Section 1.3, 1.4 and 1.5, Robert and Casella (2004) in 1.4 and Ando (2010) in Section 1.6 within this introduction.

1.2 Bayesian Theory

It is important to note first that, under the Bayesian view of probability, we consider the observed data as being fixed, while any parameters in a given model are unknown variables with their own probability distributions. This is to be contrasted with the frequentist perspective on probability where the data are seen to be drawn from probability distributions conditional on any covariates, with any parameters being fixed and simply unknown (so that they do not possess a distribution but have a single 'true' underlying value). To illustrate Bayes' Theorem, suppose that $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ is a vector of observations from some process with a sampling density $f(\boldsymbol{y}|\boldsymbol{\theta})$. Here, $\boldsymbol{\theta}$ is a vector of model parameters that are both unknown and possesses a distribution, the sample space of which we designate $\boldsymbol{\Theta}$. Equivalently, $f(\boldsymbol{y}|\boldsymbol{\theta})$ is often called the likelihood for $\boldsymbol{\theta}$ and this is written $l(\boldsymbol{\theta}|\boldsymbol{y})$. We must also place a prior distribution $f(\boldsymbol{\theta})$ on $\boldsymbol{\theta}$. Here we may incorporate

any information on $\boldsymbol{\theta}$ that we have from, for example, expert opinions, judgements or practical experience if we have any. In particular, some of the parameters in the models we shall consider have physical interpretations that allow us to put sensible limits on their possible values. Alternatively, we may use a diffuse prior that allows a wide range of values for $\boldsymbol{\theta}$. This can lead to some difficulties, however, as we shall discuss later.

By using our likelihood $f(\boldsymbol{y}|\boldsymbol{\theta})$ together with our prior $f(\boldsymbol{\theta})$ we may write the joint probability density function (pdf) of $\boldsymbol{y}$ and $\boldsymbol{\theta}$ as

$$f(\boldsymbol{y}, \boldsymbol{\theta}) = f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}). \tag{1.1}$$

Now, we ideally want an expression for $f(\boldsymbol{\theta}|\boldsymbol{y})$ as we have observed the data $\boldsymbol{y}$ and, given this information, we wish to find the *posterior* distribution $f(\boldsymbol{\theta}|\boldsymbol{y})$ using the likelihood $f(\boldsymbol{y}|\boldsymbol{\theta})$. The key to doing this is Bayes' theorem, which states that

$$f(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}, \boldsymbol{\theta})}{f(\boldsymbol{y})} = \frac{f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta})}{f(\boldsymbol{y})},$$

where the last equality is obtained using (1.1). Further, the quantity $f(\boldsymbol{y})$ is termed the marginal likelihood and is obtained as

$$f(\boldsymbol{y}) = \int f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta})\, d\boldsymbol{\theta}.$$

1.3 Making Use of Posterior Distributions

As mentioned in the previous section, we wish to consider the *posterior* distribution of the parameters $\boldsymbol{\theta}$ given the data $\boldsymbol{y}$, which we label $f(\boldsymbol{\theta}|\boldsymbol{y})$.

Unless the likelihood and prior are drawn from particular *conjugate* distributions (so that their distribution is of some known, tractable form), it is highly unlikely that we will be able to obtain a closed form for this posterior distribution. We must therefore make use of simulation methods to simulate draws from the desired posterior distribution and this requires a large amount of computation. Using these draws, we may then approximate to a reasonable degree of accuracy suitable summaries of the posterior distribution, such as its mean and median. In the case of the mean, for example, supposing we have N samples $\theta_1, \theta_2, \ldots, \theta_N$ from our posterior, then we may approximate the mean of our posterior by

$$E[\boldsymbol{\theta}|y] \approx \frac{1}{N} \sum_{i=1}^{N} \theta_i.$$

We can also use estimation techniques (such as kernel density estimation) to interpolate our draws from $f(\boldsymbol{\theta}|\boldsymbol{y})$ in some way into a smooth estimate of the posterior density $f(\boldsymbol{\theta}|\boldsymbol{y})$. Using this density estimate, we can subsequently find the approximate mode(s) of $f(\boldsymbol{\theta}|\boldsymbol{y})$ which may then be used as point estimate(s) of $\boldsymbol{\theta}$ if required. Also, we might calculate a $100(1-\alpha)\%$ *credible region* for $\boldsymbol{\theta}$. This is a region in which $\boldsymbol{\theta}$ lies with probability $100(1-\alpha)\%$. Such an interval, which we might label A, can be obtained by finding a suitable set A such that

$$P(\boldsymbol{\theta} \in A|\boldsymbol{y}) = \int_A f(\boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{\theta} = 1 - \alpha.$$

Setting $\alpha = 0.05$ would, for example, yield a 95% credible interval.

Alternatively, we could calculate a *Highest Posterior Density* (HPD) region for $\boldsymbol{\theta}$. A $100(1-\alpha)\%$ HPD region is one with posterior probability $100(1-\alpha)\%$ such that the minimum density of any point within the region is at least as large as the density of any point outside the region. Note that HPD

regions may not be continuous single intervals but are unique providing that the posterior is not uniform on any given neighbourhood within the support of the posterior (see Zellner (1971) for further details).

1.4 Simulation Methods

As noted in the previous section, we are likely to require draws from our posterior distributions to obtain suitable summaries as we are not likely to have available a closed form of these posteriors, particularly given the forms of models that we shall be using. We shall therefore require methods to simulate from our posterior distributions in some way. Such simulation can be difficult, particularly if these distributions are of high dimension. One oft-used technique that we shall be employing extensively is that of Markov Chain Monte Carlo (MCMC). MCMC methods allow us to create a sequence of *dependent* samples from the desired posterior distribution by creating a Markov chain whose steps form a dependent sample from our posterior $f(\boldsymbol{\theta}|\boldsymbol{y})$.

1.4.1 The Metropolis-Hastings Method

Originally proposed by Metropolis et al. (1953) and subsequently expanded upon by Hastings (1970), the Metropolis-Hastings method allows us to draw dependent samples from our posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{y})$ as we desire, providing that we can compute some function proportional to the desired density. That is, we need not calculate the normalizing constant of the posterior, which is given by

$$f(\boldsymbol{y}) = \int f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta})\, d\boldsymbol{\theta},$$

but rather need only be able to compute $f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})$, which is much easier to obtain given that it is simply the likelihood of $\boldsymbol{\theta}$ given the data $\boldsymbol{y}$ multiplied by our desired prior distribution.

To outline the algorithm, suppose that we have a Markov chain for $\boldsymbol{\theta}$ which, at the ith state, we label $\boldsymbol{\theta}^{(i)}$. We require a proposal distribution $q(.,.)$ which, ideally, is similar in shape to the desired target (i.e. posterior) distribution. The level of similarity between the proposal and target distributions is a key factor in how quickly the algorithm converges. Given a proposal distribution q, we generate a proposed value for the chain from $q(.,.)$ and label this $\boldsymbol{\theta}^*$. With probability α, where

$$\alpha = \min\left\{\frac{f(\boldsymbol{y}|\boldsymbol{\theta}^*)f(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i)})}{f(\boldsymbol{y}|\boldsymbol{\theta}^{(i)})f(\boldsymbol{\theta}^{(i)})q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^*)}, 1\right\} \tag{1.2}$$

we accept the proposal and set the new value of the chain (i.e. its value in the $i+1$th state) to $\boldsymbol{\theta}^*$ while with probability $1-\alpha$ we reject the proposal and set its value in the $i+1$th state to be equal to the value in its ith state. Note that in computing (1.2) we do not need to evaluate the normalizing constant of our posterior distribution as this has been cancelled by taking a quotient - a most helpful state of affairs when calculating as this constant is difficult to compute. Note that if the proposal is chosen to be a *symmetric* distribution then our expression for α may be simplified to

$$\alpha = \min\left\{\frac{f(\boldsymbol{y}|\boldsymbol{\theta}^*)f(\boldsymbol{\theta}^*)}{f(\boldsymbol{y}|\boldsymbol{\theta}^{(i)})f(\boldsymbol{\theta}^{(i)})}, 1\right\}$$

as with a symmetric proposal distribution we will have that $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i)}) = q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^*)$ so that the ratio $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i)})/q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^*)$ is 1 for all $\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^*$.

1.4.2 Convergence

In order to be drawing samples from the desired posterior distributions, we require that the Markov chains in question have converged to their respective stationary distributions. It is not necessarily easy to determine when this occurs and indeed Cowles and Carlin (1996) note that 'many of the [available] diagnostics produce results that are difficult to interpret, and [are] potentially misleading even in (...) idealized settings'. Thus, to determine whether convergence to a stationary distribution has been achieved is difficult. Informally, we often examine the *trace plot* of the chain - that is, a time series plot of the various states of the chain. This allows us to identify the *burn in* of the chain, or the number of iterations that should be discarded as the chain has not yet settled to its stationary distribution. This will usually depend on the starting point of the chain. If we start far from regions of high posterior probability for our posterior distribution then we are liable to have a longer burn in period. Such trace plots will also help us to gauge whether convergence to the desired stationary distribution has indeed been achieved and several such trace plots are given in this work to illustrate (non-)convergence to desired stationary distributions.

Once we have obtained a chain that converges successfully to the desired stationary distribution for our parameters $\boldsymbol{\theta}$, we may then make use of the samples as outlined in the Section 1.3. Before making use of these samples, however, it may be wise to *thin* the chain. Here, we take only every nth iteration, where n is to be determined by looking at trace plots containing only every nth iteration of our Markov chain, and seeing whether the chain appears to mix well (that is, whether we have autocorrelation in the chain or not) or by looking at autocorrelation functions for the chain. We expect that thinning to every nth iteration will be appropriate if we have an autocorrelation close to 0 at lag n. For some of our chains, we shall see that there is a high degree of autocorrelation between subsequent

states of the chain and we must take, for example, every 50th iteration so as to obtain approximately independent samples from the desired posterior distributions. We would hope that at lag 1 for our final thinned chain there is a low autocorrelation so that we have approximately independent samples from our posterior.

1.5 Kernel Density Estimation

In order to estimate the shape of the posterior density given an MCMC sample, we need some way in which to smooth our points to provide a smooth curve given our sampler datapoints. To this end, we employ Kernel Density Estimation - for full details see Wand and Jones (1995). We provide an outline below.

Suppose that $x_1, x_2, \ldots, x_n$ are n datapoints drawn from a given distribution, but one for which no closed form is available. Then we may estimate this distribution by a function f_h

$$f_h(x_1, x_2, \ldots, x_n) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right),$$

where K is a (non-negative) function with mean 0 that integrates to 1 with respect to x over the real line and $h > 0$ is a parameter termed the 'bandwidth'. In fact, Wand and Jones note that the biggest factor in controlling the approximation is not the kernel K, but rather the bandwidth h. This parameter controls the relative smoothness of the estimate. For a large value of h, there is *oversmoothing*, whereby it is impossible to see features in the desired distribution clearly as they have been 'levelled out'. On the other hand, small values of h may produce *undersmoothing*, resulting in a very 'bumpy' and awkward posterior approximation that is overly affected by individual data values and tries to fit these too closely.

We select a Gaussian kernel. To select a bandwidth, we note that Silverman (1986) suggests the use of

$$h = 0.9 \times \min\left(\sqrt{\text{Var}(\boldsymbol{x})}, \frac{\text{IQR}(\boldsymbol{x})}{1.34}\right) \times n^{-1/5},$$

where $\boldsymbol{x}$ is our vector of data observations, $\text{Var}(\boldsymbol{x})$ the variance of the observations, $\text{IQR}(\boldsymbol{x})$ the interquartile range of the observations and n the number of observations. Silverman demonstrates that this proves to be optimal in the sense of producing the least possible mean square error in the approximation. For that reason, we adopt this as our standard bandwidth function given for our posterior kernel density estimates.

1.6 Model Comparison and Model Averaging

Suppose that, rather than a single model, we have several models - a total of M models, say - to deal with. Then, we find that we must often choose between the M models or, better still, somehow use the information from all M models to obtain the inference that is in some sense 'best possible'. To do this, we first assign a prior probability to each model. These prior probabilities are often taken to be equal for each model, unless we have some reason to believe a priori that this is not the case. Then, using the standard laws of conditional and joint probabilities, and designating the parameters in model m as $\boldsymbol{\theta}_m$, we may obtain the expression

$$f(\boldsymbol{y}, \boldsymbol{\theta}_m, m) = f(\boldsymbol{y}|\boldsymbol{\theta}_m, m) f(\boldsymbol{\theta}_m) f(m).$$

The posterior distribution of the mth model given our data $\boldsymbol{y}$ is then, by application of Bayes' Theorem,

$$f(\boldsymbol{\theta_m}, m|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta}_m, m)f(\boldsymbol{\theta}_m|m)f(m)}{f(\boldsymbol{y})}$$

and further decomposition gives

$$\begin{aligned} f(\boldsymbol{\theta_m}, m|\boldsymbol{y}) &= f(\boldsymbol{\theta_m}|m, \boldsymbol{y})f(m|\boldsymbol{y}) \\ &= \frac{f(\boldsymbol{y}|\boldsymbol{\theta}_m, m)f(\boldsymbol{\theta}_m, m)}{f(\boldsymbol{y}|m)} \times \frac{f(\boldsymbol{y}|m)f(m)}{f(\boldsymbol{y})}, \end{aligned}$$

where $f(\boldsymbol{y}|m)$ is known as the marginal likelihood under model m.

Note that using Bayes' Theorem we also have that

$$f(m|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|m)f(m)}{f(\boldsymbol{y})} = \frac{f(\boldsymbol{y}|m)f(m)}{\sum_{m=1}^{M} f(\boldsymbol{y}|m)f(m)},$$

which is the posterior probability of the mth model given the data and will be useful when we wish to calculate the probability of the mth model being (in some sense) the 'best' model given our data.

Given the above, it is also possible to average over all the models (a procedure known as model averaging) by using as weights their posterior probabilities given any data. This avoids the need to select an optimal model and may allow the inclusion of information that would otherwise be lost and thereby improve any predictions we may wish to make. Thus, if we have some parameter β, say, then its posterior distribution given the data under this model averaging procedure would be

$$f(\beta|\boldsymbol{y}) = \sum_{m=1}^{M} f(\beta|m, \boldsymbol{y})f(m|\boldsymbol{y}).$$

1.7 AIC, BIC and DIC

In the course of this thesis, we shall require some way to select a model which is in some way 'best' from among a group of models. In order to do this, we shall at various points make use of what are termed *Information Criteria*. These information criteria effectively compare likelihoods between different models while also introducing some form of penalty for decreasing parsimony of model and are designed so that we should select the model with the lowest value of the given criterion as the 'best' model. We shall make use of three standard information criteria: AIC (Akaike's Information Criterion), BIC (the Bayesian Information Criterion) and DIC (the Deviance Information Criterion). We shall briefly outline each of these here.

AIC was originally proposed by Akaike (1974) and is defined as

$$AIC = 2k - 2L,$$

where k is the number of model parameters and L is the log-likelihood of the model. We can see that both likelihood and parsimony of model will affect the value of the criterion - the former providing a positive (and thus undesirable) contribution with increasing k, and the latter a negative (and thus desirable) contribution with increasing likelihood. However, it was found by various researchers (see, for example, McQuarrie and Tsai (1998) and Burnham and Anderson (2002)) that AIC has a tendency to overfit models. That is, it has a tendency to select overly complex models, rather than to select the true model (in those cases where the true model is known). In order to counteract this, Hurvich and Tsai (1989) suggested a small sample size modification to AIC - denoted by AIC_c and termed the

corrected AIC - which is defined as

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1},$$

where k and AIC are as defined above while n is the number of datapoints in the given dataset. This corrected AIC, as noted by Burnham and Anderson (2002), is much less likely to overfit, particularly in those cases where n is small and/or k is large, and tends to the standard AIC as the number of available datapoints increases (for k fixed). It is this corrected AIC that we use throughout.

Next, we consider BIC which was proposed in Schwarz (1978). With notation as used above for AIC, we have

$$BIC = -2L + k(\ln n - \ln(2\pi)).$$

Again we see that increased model dimension is penalised while improved likelihood is encouraged. It is noted by Burnham and Anderson (2004) that BIC is less prone to overfitting than AIC but does often produce similar results to AIC_c.

We shall also have use for the Deviance Information Criterion or DIC. Noting that the likelihood is a function of the parameters $\boldsymbol{\theta}$ of a model given data $\boldsymbol{y}$, we have

$$DIC = p_D + \overline{D} = D(\overline{\boldsymbol{\theta}}) + 2p_D,$$

where D is the deviance of the model defined as $D = -2L(\boldsymbol{\theta}|\boldsymbol{y})$, $\overline{D}$ is the expectation of the deviance with respect to $\boldsymbol{\theta}$, $D(\overline{\boldsymbol{\theta}})$ is the deviance evaluated at the posterior mean of the parameters and p_D is a measure of the effective number of parameters in the model given by $p_D = \widehat{\text{var}}(\boldsymbol{\theta})/2$ (as suggested in Gelman et al. (2003)). Here, $\widehat{\text{var}}(\boldsymbol{\theta})$ is an estimate of the variance of the posterior distribution of $\boldsymbol{\theta}$.

1.8 Generalised Linear Models and Logistic Regression

Since we will be dealing with the prediction of probabilities for events occurring or not occurring in this thesis, we shall have cause to make use of the Generalised Linear Model framework first introduced by Nelder and Wedderburn (1972). In this framework, there is a dependence of the data of some function of the linear predictor, the latter being defined as $\boldsymbol{X}^T\boldsymbol{\beta}$ for some covariate vector $\boldsymbol{X}$ and some coefficients $\boldsymbol{\beta}$. That is, if $\boldsymbol{Y}$ represents our response distribution, and $\boldsymbol{\mu}$ its mean then we have

$$E[\boldsymbol{Y}] = \boldsymbol{\mu} = g^{-1}(\boldsymbol{X}^T\boldsymbol{\beta}),$$

where g is known as the *link* function, as it links the linear predictor with our response vector. In particular, if we take our link function to be the logit function given by

$$g(x) = \log\left(\frac{x}{1-x}\right)$$

then if our response vector lies between 0 and 1 being, say, a probability, then g maps this probability one-one onto the real-line. Using such a link is termed logistic regression and enables us to model the probability of an event occurring given covariates. We shall have several uses for this procedure in this thesis, as we shall be dealing with binary events and thus the modelling probabilities shall occur frequently.

Note that when using logistic regression, it is difficult to check for goodness of fit as there are few distinct residual values given by the procedure, since we are estimating not for binomial proportions but individual Bernoulli $0-1$ observations. Further, the deviance (defined as being twice the difference between the likelihood of the model in question and the likelihood under a fully saturated model) is not helpful to us in the binary case as it may

be shown that the data do not appear at all in the expression for this model's deviance, meaning that the estimated probabilities of success or failure cannot be compared with the data easily.

To ameliorate this problem, Hosmer and Lemeshow (1980) suggest grouping the data by predicted success probability (into, for example, deciles) and then using a *binomial* regression (for which the deviance may be analysed as having approximately a χ^2_{n-p} distribution for n datapoints and p model parameters) to analyse the goodness of fit of the model for the grouped data. We shall make use of this technique later in the thesis, where we shall have cause to check the goodness of fit of some logistic regression models.

1.9 Survival Analysis

In this thesis, we are concerned with the occurrence (or otherwise) of a particular event - DCS. To this end, we shall require some useful results on modelling data of this type. We draw from Davison (2008).

Suppose, then, that T be a random variable indicating the time at which DCS occurs. We define the survivor function, S, as being

$$S(t) = P(T \geq t) = 1 - F(t),$$

where F is the cdf of T.

We also define a general hazard function, h say, as being the risk of DCS occurring at some time t and is the instantaneous failure (i.e. DCS incidence) rate at time t, conditional on DCS taking place at or after t. If f is

the pdf of T then for a small time interval $[t, t+\delta t)$, we have

$$h(t) = \lim_{\delta t \to 0} \frac{P(t \leq T < t+\delta t | T \geq t)}{\delta t} = \lim_{\delta t \to 0} \frac{F(t+\delta t) - F(t)}{S(t)\delta t} = \frac{f(t)}{S(t)},$$

where the conditioning in the second equality is evaluated using the definition of the survivor function itself. Note that h can be *any* non-negative function but if $\lim_{t\to\infty} S(t)$ is strictly positive then this indicates that there will be long term survivors who never experience the event in question.

Next, there is a further useful form of the hazard function to be noted. Since we have $S(t) = 1 - F(t)$, we may also write

$$\begin{aligned} h(t) &= \frac{f(t)}{S(t)} \\ &= \frac{f(t)}{1 - F(t)} \\ &= -\frac{d}{dt} \log(1 - F(t)) \\ &= -\frac{d}{dt} \log S(t), \end{aligned}$$

so that we may write the hazard function in terms of the derivative of the logged survivor function. Alternatively, integrating and rearranging, we have

$$\log S(t) = -\int_0^t h(u)\, du$$

or, more usefully,

$$S(t) = \exp\left(-\int_0^t h(u)\, du\right).$$

In particular, for a *binary* event where we do not know at what time the event occurs, but only whether or not it does in fact occur, we find that

the probability of event occurrence is given by

$$S = \exp\left(-\int_0^\infty h(u)\,du\right)$$

and we have cause to use this result in particular in what follows in Chapters 3 and 4.

1.10 Missing Data

1.10.1 Ordinal Data Imputation

In Chapter 5, we shall have cause to deal with a dataset comprised of ordinal variables with missing data and we shall need to make imputations to form a completed dataset. Van buuren et al. (2006) note that there are two primary means to achieve this: joint modelling, whereby a distribution (often a multivariate normal distribution) is specified for a latent variable that is supposed to underlie all observations. Its parameters are then estimated either by a maximum likelihood procedure or a Bayesian method. The other method involves specifying a full conditional distribution for each missing covariate for each datapoint (conditional on the observed covariates). The parameters of these conditional distributions are then estimated using a Gibbs Sampler.

In our case, we shall wish to make predictions in a time-critical situation, and the latter approach would involve running our Gibbs Sampler for each new observation found. For that reason, we shall instead use the joint modelling approach whereby we shall consider the idea of an underlying latent variable and use a model involving the multivariate normal distribution. Details on this approach will follow in Section 5.3.

1.10.2 Multiple Imputation

We note that there is inaccuracy in any single imputation and, in order to take account of this imputation inaccuracy, Rubin (1987) suggested that we should carry out several imputation runs and aggregate over these in a particular fashion. This led him to the process of *Multiple Imputation* whereby we impute the missing values multiple times to generate multiple imputed datasets, analyse these separately and then combine our results according to a set of rules which are now popularly known as Rubin's rules, which we outline in the next subsection. By creating multiple imputed data sets, we may gauge the variability both within and between imputed datasets for the parameters of interest in our regression models.

Now, while Rubin suggested that three to five imputations were likely to be sufficient to have the desired effect of providing more stable parameter estimates than a single imputation, more recent work (see, for example, Graham et al. (2007)) has shown that the number of imputations required is a function of the amount of data that are missing. In particular, they recommend that 40 imputations is a reasonable number of imputations to use when there is no more than 70% missing data, which is the case for our dataset. We thus accept this recommendation and make use of 40 imputations to reach our conclusions in this section.

1.10.3 Rubin's Rules

As promised above, we now outline Rubin's rules as given in Rubin (1987) for combining the results of analyses on several different imputed datasets to obtain overall parameter estimates and appropriate standard errors.

Suppose we perform m imputations, yielding m datasets, and m results for whatever analysis it is we wish to perform. Suppose that we are interested in making inferences about some parameter Q, say. Then, we treat each imputed data set as if it were the original fully observed data and obtain the usual point and variance estimates for the parameter in each imputed data set, which we denote by $\hat{Q}_j$ and U_j respectively, in data set j. Then if $\hat{Q}_j$ estimates the desired result of interest (such as a regression coefficient) for the jth dataset, and U_j is its standard error, then Rubin's rules given that the overall average may be obtained as expected as

$$\overline{Q} = \frac{1}{m} \sum_{j=1}^{m} \hat{Q}_j.$$

Further, the mean within imputation variance is given by

$$\overline{U} = \frac{1}{m} \sum_{j=1}^{m} U_j$$

and the between imputation variance is

$$B = \frac{1}{m-1} \sum_{j=1}^{m} \left(\hat{Q}_j - \overline{Q} \right)^2.$$

This gives a total variance of

$$T = \overline{U} + \left(1 + \frac{1}{m} \right) B,$$

where the additional term $\left(1 + \frac{1}{m}\right) B$ represents the additional variability due to the presence of missing values.

Our estimate for Q is then distributed as a t_q-distribution on

$$q = (m-1)\left(1 + \frac{m\overline{U}}{(m+1)B}\right)^2$$

degrees of freedom, so that we may test the hypothesis that $Q = 0$ by comparing this t_q distribution with the observed t-value $t = \overline{Q}/\sqrt{T}$.

1.11 The Problem

In this thesis, as initially mentioned in Section 1.1, we are interested in models for the occurrence of DCS, as well as its prediction. This is particularly important in military and commercial applications. For example, submariners, in the case of emergency, must quickly choose whether to attempt to escape from their submarine or wait for help to arrive. Where possible, it is preferable to wait for rescue to arrive. However, if the situation in the submarine is deteriorating and escape is possible with a low DCS risk, then the submariners must weigh the choices of remaining on the submarine that is in poor condition against that of contracting DCS. Submariners must be able to make a rapid, informed decision and must be given easy, clear guidance on the matter that can be used in an operational situation. Only recently has any indication of probability of DCS been given as operational information - prior to this there was a simple depth and pressure at which submariners were told that escape was either safe or unsafe. Clearly, however, this is not the case as the process is a random one with many factors needing to be accounted for. Not everyone even with identical height, weight and so on will contract DCS even in identical diving conditions. Indeed, the same individual may not even contract DCS for a second time under such identical dive conditions. Thus, we must provide

the best possible prediction of the probability of this occurring, so that the submariners may make informed decisions as to whether or not to surface.

We now require the idea of a *dive profile*. A dive profile gives the depth of a diver at any given time during a dive. One important point about submarine escape dives is that they have a rather particular dive profile. In the case of such dives, the diver will likely have no (or very little) time for stopping while ascending during the dive, so the dive profile will indicate a (near-)linear ascent. This means that our results will generally be inapplicable to more general type of dives, as these would require a rather different ascent strategy and thus would require different data from which to work. Further, we do not aim to produce safe decompression schedules (which is the aim of most decompression modelling). Rather, we seek to give a clear indication of how dangerous a particular ascent may or may not be with respect to the likelihood of contracting DCS while undergoing this given ascent. This will then further the ability of military or commercial submariners to make informed decisions during emergencies. These models could easily be integrated into onboard computers (such as those currently used to decide when to make safe decompression stops) for use in emergencies. The work in finding such parameters, as set out here, requires considerable computational effort. However, once the parameters in the models have been obtained, finding the probability of DCS for a given dive is a quick procedure that does not require great computational power.

1.12 The Data

There are two datasets available to us. The first is provided by QinetiQ and gives information on various tests performed at QinetiQ between 1990 and 2005. The subjects were subjected to decompression from various depths, and at various breathing-air pressures. In this dataset, the data

provided include covariates as follows: the date of the dive, the identity of the subject in question (some subjects were used more than once), the saturation depth used (i.e. the partial pressure of nitrogen with which the subjects' bodies were saturated), the depth from which subjects were exposed to escape conditions (where this was the type of pressure exposure undertaken) and the subject's body mass. Also included was the binary response variable indicating whether the subject was observed to have any signs of DCS. It should be noted that subjects were left to rest for at least a month between dives, and it is thought that there are no carryover effects between experiments after such a long period between successive trials on the same subject. For this dataset, no detailed pressure data are given - that is, we do not know the precise profiles over the course of the dives. We call this dataset, which consists of 4396 observations and has no missing covariate or response observations, the first QinetiQ dataset. Of these, 2411 are standard, non-saturation or *bounce* dives, 995 are sub-saturation dives and 990 are saturation dives (these terms will be explained in Chapter 2).

The second dataset is also provided by QinetiQ and gives information on various tests performed at QinetiQ. These subjects were also exposed to decompression from various depths, and at various breathing-air pressures. In this dataset, the data provided include covariates as follows: the identity of the subject in question (some subjects were used more than once), the saturation depth used (i.e. the pressure of nitrogen with which the subjects' bodies were saturated), the depth from which subjects were subjected to escape conditions (where this was the type of dive undertaken), and a vector of length 25 giving the Kisman-Masurel (KM) bubble scores (effectively a subjective measure of the number of bubbles heard passing the heart) at each of the following numbers of minutes following the start of the ascent from the dive: 2, 5, 15, 30, 45, 60, 75, 90, 105, 120, 150, 180, 210, 240, 270, 300, 330, 360, 390, 420, 450, 480, 540, 600 and 1440. More information on these bubble scores is given in Section 4.1. Also included in the dataset is the binary response variable indicating whether the subject was observed

to have any signs of DCS, together with the time of onset of DCS after the start of the ascent for those dives where DCS occurred. It should be noted that subjects were again left to rest for at least a month between dives (Loveman, Personal Communication, 2010). There are a total of 986 dives in this dataset, which we call the second QinetiQ dataset.

1.13 Computational Resources

All computations were performed on a machine with an Intel Core i7 920 processor (stock clock 2.66 GHz, overclocked to 3.7 GHz) with 6 GB of DDR3 Memory clocked at 1480 MHz. All code for the models in Chapters 3, 4 and 6 was written in C. Extensive use was also made of the R language, particularly in Chapters 3 and 5, and the outputs from all MCMC runs were analysed using the R coda package. The analytic solutions derived in Chapter 3 were obtained with the help of the MAPLE Computer Algebra System.

1.14 The Thesis: An Outline

We now proceed to outline the structure of the thesis. In Chapter 1, we have provided a general introduction to Bayesian theory and to the problem at hand.

In Chapter 2 we shall give a brief account of the biological background to the problem and the decompression models available to date.

In Chapter 3 we shall focus on one class of model in particular and will derive some previously unknown analytic results for this class of model. We

shall then use these new analytic results to enable us to employ Bayesian methods in a problem that has previously only been tackled in a frequentist manner. Furthermore, we shall attempt to ascertain whether there is a considerable increase in the usefulness of the data obtained when a survival time is provided, rather than just a binary output variable. This knowledge would be useful to experimenters as, if there is no value to be gained by collecting these, then some experimental effort can be saved. Here, we will use the first QinetiQ dataset.

In Chapter 4 we shall expand upon the model class examined in Chapter 3 and again make use of Bayesian methods, which have not previously been used in the area of this problem, to find parameter estimates for models of this type. In addition, we shall use a reversible jump sampler to help us to estimate what might be a reasonable number of components for this model class and to quantify the uncertainty of our knowledge of which model is 'best'. Again, we shall use the first QinetiQ dataset here.

In Chapter 5 we shall move on to a different but related problem but this time using the second, previously unanalysed, QinetiQ dataset. Here, we shall use a Gibbs sampling method to impute missing response values (bubble scores) into the dataset. We shall then use the imputed bubble scores as predictors for DCS using a logistic regression model.

In Chapter 6 we shall make predictions from our model obtained in Chapter 4. Firstly, we shall use model averaging to account for our uncertainty in both our parameter estimates and also which model is most appropriate from our given model class. Secondly, we shall combine the two types of model from Chapters 4 and 5 so as to form a better prediction of DCS than would be possible using either model alone. We shall provide demonstrably improved DCS predictions compared with the models currently in use.

Finally, in Chapter 7 we shall outline some possible directions for future work.

Chapter 2

More on the Problem, its Biological Background and Previous Work

2.1 The Problem Continued

Having provided some mathematical background in Chapter 1, we now give a more detailed explanation of the problem here.

First, note that the standard air pressure at sea-level is 101325 Pascals. Suppose now that at time $t = 0$ a submarine is damaged in some way and becomes stranded at a depth of $E > 0$ metres. We term this the escape depth. We also define the saturation depth, S say, as being the internal air pressure within the submarine. If damage has caused the submarine to become flooded to some extent then this will cause it to have an internal air pressure of $S > 101325$ Pascals, otherwise the submarine is pressurised to standard pressure and we have $S = 101325$ Pascals.

The submariners now have a number of choices available to them depending

on their tactical situation and the chance of DCS occurring for an ascent at this level of breathing-air pressure and current depth. If the submarine is in a safe state then rescue is the preferred option as then neither DCS risk nor issues on surfacing - survival on the surface being very difficult due to the likelihood of drowning or hypothermia - apply. If the submarine is in a somewhat safe but deteriorating state then, then, if DCS is extremely likely to occur, then the submariners will also likely elect to remain aboard the submarine and try to await rescue. Conversely, if it is highly unlikely then they may well choose to escape.

The submariners' decision becomes considerably more difficult if the probability of DCS occurring is not close to 0 or 1. In this case, they must weigh up the possible plans based on their situation and the probability of DCS and make a decision based on this information. Thus, we see that a good assessment of DCS probability for a range of E and S values will be useful in helping to decide on what course of action the submariners are advised to take. Producing such an assessment is the first problem we shall tackle in this thesis. In the following sections in this Chapter, we outline the models that have already been used to model decompression and we shall go into more detail of one particular model that might be used to help us produce the desired probabilistic assessment in Chapters 3 and 4.

The second problem is of a related but slightly different nature. Suppose that the submariners *have already escaped* from the submarine. Depending on the size of the submarine, there may be quite a number of submariners for any rescuers to treat for injuries and, importantly, some of them may need to be treated for DCS.

Most importantly, however, DCS may take some time to become established in a possible patient. Further, there are only a limited number of treatment spaces available as DCS can only be treated by recompression in a hyperbaric chamber, followed by a more controlled ascent with stops

made in an attempt to prevent bubbles forming in the body . Treatment in these chambers is not only expensive but is also restricted by available space in the chamber. It is very useful, therefore, to be able to *predict* who is likely to be susceptible to DCS based on whatever data are readily available post-surfacing. We shall concentrate on this problem in Chapter 5 of the thesis.

2.2 Some Biological Background

In this chapter, we outline the models that have been used to date in decompression modelling, together with the methods that were used for their analysis. First, we observe that DCS is an ailment that may affect divers either upon ascent from a dive, or when changing breathing gases during a dive. It is thought to be caused by bubbles forming in the tissues and subsequently moving into the bloodstream (Brubakk and Neuman, Chapter 9). The severity of the condition varies widely, depending upon how extreme the dive is with respect to depth and ascent rate. Symptoms may range from a mere tingling sensation, often called the 'niggles', to a loss of blood flow to the brain and subsequent death. In order to avert such consequences, divers usually make stops at various points in their ascent to allow any excess of gas to be exhaled without forming bubbles, although it is not universally agreed whether the bubbles are a symptom of DCS or its actual cause.

Modellers such as Boycott, Damant and Haldane (1908) and Thalmann et al. (1997), then, seek to consider the process of absorption (or uptake) and elimination (or off-gassing) of nitrogen in the blood that is respectively caused by descending and ascending while diving. On diving, the higher external pressure outside the body, due to the weight of the water (which differs between fresh and salt water), forces additional nitrogen into the

breathing mix and the lungs take in this additional nitrogen. This leads to a subsequent build-up of dissolved nitrogen within the tissues. This process is not dangerous in and of itself while the nitrogen remains dissolved in the tissues. The nitrogen tension in the tissues will continue to build until it equals the partial pressure of nitrogen in the lungs. At this point, the tissues are said to be *saturated* with nitrogen. Further, since it is the difference between the partial pressure of nitrogen in the lungs and that in the tissues that results in a net movement of nitrogen into the tissues, then as this difference is reduced, the net rate of absorption of nitrogen into the tissues falls following exponential kinetics. If the diver ascends too quickly then the excess nitrogen will leave the tissues by diffusion as the tissues will now have a higher nitrogen partial pressure than the the hydrostatic pressure, which will have been reduced due to ascent, but will not be able to do so in a slow and regulated manner and simply be absorbed into them. Instead, there will be the forced formation of bubbles in the bloodstream and/or tissues of the body themselves. These bubbles can, for example, form in the joints (causing pain), or, more dangerously, in the spine or brain.

Something that we should bear in mind at all times is that of course a small amount of excess nitrogen in the tissues due to the force of water - which we term the nitrogen *supersaturation* - may apparently be tolerated without any danger at all, after all, one does not contract the bends from going swimming in a swimming pool! Thus, there is likely to be a difference between the amount of nitrogen that is actually present in the tissues and the amount of that which is dangerous. We refer to the level of nitrogen in the tissues as the *nitrogen partial pressure* in the tissues, while we call the excess amount above the safe limit the *(nitrogen) gas burden*. Note that we refer to the nitrogen *partial* pressure in the previous sentence as there are also other gases (such as oxygen and carbon dioxide) present in the tissues. We refer to the amount of oxygen present in the tissues as the *oxygen partial pressure* while we refer to the sum total of all gases present as the *total tissue gas pressure*.

There are then two major types of decompression model: firstly, those that focus on the *amount* of (excess) nitrogen present in the body's tissues, such as Boycott, Damant and Haldane (1908) and Thalmann et al. (1997) as previously mentioned, and secondly those that focus on the *actual formation of (idealised) nitrogen bubbles* within the tissues, such as Yount and Hoffmann (1986) and Wienke (1990). Both of these approaches have their merits and difficulties. In this thesis, we focus our attention on the models that deal with the excess nitrogen in the tissues as these are faster to compute and, once we have seen how these methods respond to our Bayesian analysis, our parameter estimation methods may then be applied to the bubble-based models as well (these having similar numbers of parameters). In all cases, the models used to date have all been *deterministic*. That is, for any given input, the models will always produce the same output with regard to the amount of nitrogen in the tissues at the time, or the number and/or size of bubbles in question. This would seem to be somewhat unreasonable as the same subject, even under identical conditions, has often been found not to have identical results at each trial. Thus, it seems that in the future there may well be room to introduce a stochastic element into the physical models themselves. For example, the level of nitrogen in each compartment could be modelled in part with the aid of a Brownian Motion with drift whose level is dependent on the depth at a given time in the dive, resulting in Stochastic Differential Equations (SDEs) rather than the ODEs in the current models. In the case of the early, Haldane type models, the models were fully deterministic in their overall outcome. The postulation here relied on the number and depth of stops during ascent (or lack thereof) together with ones rate of ascent. Thus, one either did or did not contract DCS. In the later models, while the level of nitrogen or bubbles in the tissues was deterministic, the actual outcome of the model (i.e. whether DCS was or was not contracted) is probabilistic (i.e. we obtain a probability for DCS occurring, given the dive parameters).

The key to all the models in question is the consideration of different bodily

'compartments' in which nitrogen is caused to behave differently by the attributes of those compartments. By 'compartment' we mean a grouping of tissues that behave similarly with respect to their uptake and elimination of nitrogen. As the gas is taken up at and expelled at different rates in dissimilar tissues, the compartments have different levels of tolerance for change in ambient pressure. It should be noted that even though the various tissues in the body itself do have a wide range of properties, such as poorly perfused adipose tissues being slow to become saturated with nitrogen and highly blood perfused brain-tissue being becoming quickly saturated, there exists no one-one correspondence between these compartments and *actual* tissues in the body. Each of these compartments is assumed to only receive nitrogen by *the active passing of nitrogen from the blood into the tissue via the capillaries*. It is assumed that the *diffusion rate* (i.e. rate of movement of nitrogen from a region of a higher concentration of nitrogen into a region with a lower concentration by thermal energy) is very high relative to the perfusion rate, due to the very short diffusion distances involved, though Goldman (2007) relaxed this requirement. We do not consider this as there is a substantially increased computational requirement associated with this method, as it requires the solution of a set of several coupled differential equations. Further, the easier-to-handle standard differential equations that we discuss in Chapter 3 become considerably more difficult partial differential equations if we include diffusion, due to such diffusion-based motion occurring in both time and space. We shall consider the specifics for the single compartment model in Chapter 3, particularly Section 3.3, and the multiple compartment approach of Chapter 4 will then simply apply 'in series' as it were. In Chapter 4, we shall see there that, when considering the risk of DCS due to several compartments, we simply find the total hazard across all compartments by summing.

2.3 A Brief History of Decompression Modelling

Attempts to model the effects of decompression on the body began in 1908, when Boycott, Damant and Haldane (1908) observed that many caisson tunnel workers were suffering from, at that time, an unidentified sickness whose symptoms involved aching and pains in the joints, back and limbs. This illness was termed the bends, a label which was derived from the bodily contortions its sufferers undergo when atmospheric pressure is abruptly changed from high to low. In their model, which we term the Haldanean model, this group supposed that the body was comprised of five separate 'tissues' - which we now, in modern decompression terminology, term *compartments*. This classification came about because, as noted before, there is not a one-one correspondence with body tissues. Previous experiments by exponents such as Heller, Mager and von Schrötter (1900), which were unsupported by any modelling, had concluded that "perfectly uniform decompression at the rate of 20 minutes an atmosphere would always be safe." Boycott et al., however, claimed that 'it is evidently a mistake to assume that a given rate of uniform decompression, such as 20 minutes per atmosphere, is either necessary for safety in all cases, or would be actually safe except from some limit of pressure. From a pressure below this limit the rate will be unnecessarily slow, and from above it dangerously fast.' The Haldanean model in Boycott et al. (1908), then, uses the idea of nitrogen half-times of 5, 10, 20, 40 and 75 minutes in a process of staged decompression. However, beyond the idea of half-times, they did not make use of any further direct mathematical modelling and all tissues were assumed to have an equal tolerance for nitrogen. Nevertheless, this was a significant step forward from the work of Heller et. al.

Essentially, to use the model of Boycott et al., suppose one starts at a tissue

partial pressure of u Pa. If descending from a depth of d_1 metres to a depth of d_2 metres (so that $d_2 > d_1$), then it requires 20 minutes to make the 20-minute compartment half-saturated with nitrogen, giving a new, increased tissue partial pressure of $u+(d_2-d_1)/2$ after this 20 minutes. Alternatively, if $d_2 < d_1$ so that ascent is occurring then it takes 20 minutes to lose half of the difference between the nitrogen partial pressures at the old and new depths. Therefore, after 20 minutes one again has a new tissue partial pressure of $u + (d_2 - d_1)/2$ but this time $d_2 - d_1$ will be negative, so there is a decrease in partial pressure in the tissue. Now, at each discrete stage of pressure adjustment, we recalculate the current internal tissue partial pressure for all tissues based on the time spent at the new depth. They then hypothesised that providing the nitrogen partial pressure in *all* the tissues was never more than double that in the external environment then DCS would not occur.

Two substantial improvements to this model were made by Workman (1965). Firstly, Workman realised that there were not enough compartments with sufficiently long half-times to represent what happens during the latter portion of any given ascent and that DCS was occurring more than would be expected from the Haldanean model in deep saturation dives, where the slower compartments would take longer to release all of their stored nitrogen. To this end, Workman added additional compartments with half-times of 120, 160, 200 and 240 minutes. Secondly, Workman observed that DCS was occurring less than expected under the Haldanean model for short, shallow dives. He therefore proposed that the ratio of 2 : 1 of internal compartment nitrogen partial pressure to external pressure was not appropriate, but that instead the allowable ratio $M_j : 1$, say, for $j = 1, \ldots, 9$ of internal compartment nitrogen partial pressure to external pressure should be allowed to vary linearly with depth and by compartment, so that the faster compartments had higher allowable ratios and the slower compartments had smaller allowable ratios. Workman showed that this improved the performance of the model and termed these adjusted ratios 'Maximum

Values' or M-values (though they are in fact linear functions, rather than single values), and allocated one to each compartment.

A further notable improvement came in 1984 when Bühlmann, building on Workman's findings, updated the number of compartments to 16 and considered the effects of diving at altitude (notably in the lakes in the Swiss Alps). He collected these new conclusions together with previous results and published them as a book, Bühlmann (1984).

Following this, modelling continued in two different directions. In the first instance, some modellers began to work with more than simple half-times and extended the compartment-by-compartment modelling to allow for a richer compartmental model, in which uptake and release of nitrogen did not necessarily follow the same archetype. This approach was started by Thalmann in Thalmann (1984), and we shall explore it in more detail in Chapter 3. It is well-suited for our purposes as, while being a more realistic model than the previous models, it is amenable to the identification of exact solutions and relies only on *ordinary* differential equations to reach its conclusions. In the second instance, other modellers such as Yount and Hoffman (1986) took a more physiologically oriented approach to the problem. Having used Doppler ultrasound to detect venous gas emboli (bubbles) in the blood flowing through the hearts of subjects following pressure exposures, they suspected that it was in fact these bubbles that caused DCS (though of course it is the bubbles that are *not* detected that cause the actual damage as they are trapped). To that end, they used *partial* differential equations to explicitly model bubble radius development over the time course of each dive on a compartment-by-compartment basis. However, this approach did not allow gas transfer between bubbles and supposed that gas transfer into or out of a bubble was possible only during those times when some cut-off pressure was reached. Wienke then extended this with his Reduced Gradient Bubble Model (usually termed RGBM) in Wienke (1990). This model does allow for transfer of gas between bubbles

and also allows for a more nuanced consideration of when nitrogen transfer into or out of each bubble is possible, by allowing this possibility to vary based on the current bubble surface tension. We do not pursue this avenue of modelling further, as the use of partial differential equations that do not admit analytic solutions in these models makes them unsuitable for our purposes, given that the currently used iterative approach to their solution would demand far too much computational time compared to the possibilities offered by a Thalmann-type model.

Chapter 3

Exploratory Data Analysis and the Single Compartment Supersaturation Model

3.1 Exploratory Data Analysis

It is instructive to consider the basics of the data with which we shall be working over the course of this thesis. As the response variable is binary, scatterplots are not especially useful to achieve this. However, we can use some boxplots to show a few key features of the data. Note that the 0.1 responses are indicative of a 'partial DCS' case where the the subject suffered very minor symptoms but these resolved themselves before a treatment decision could be made.

We see from Figure 3.1 that DCS is considerably more common in dives with longer bottom times (these being saturation dives where the tissues in the divers' bodies have been 'filled' with nitrogen that has been forced out of gaseous form into the tissues). This is reassuring and is most definitely as expected as, anecdotally, deeper, longer dives have been associated with

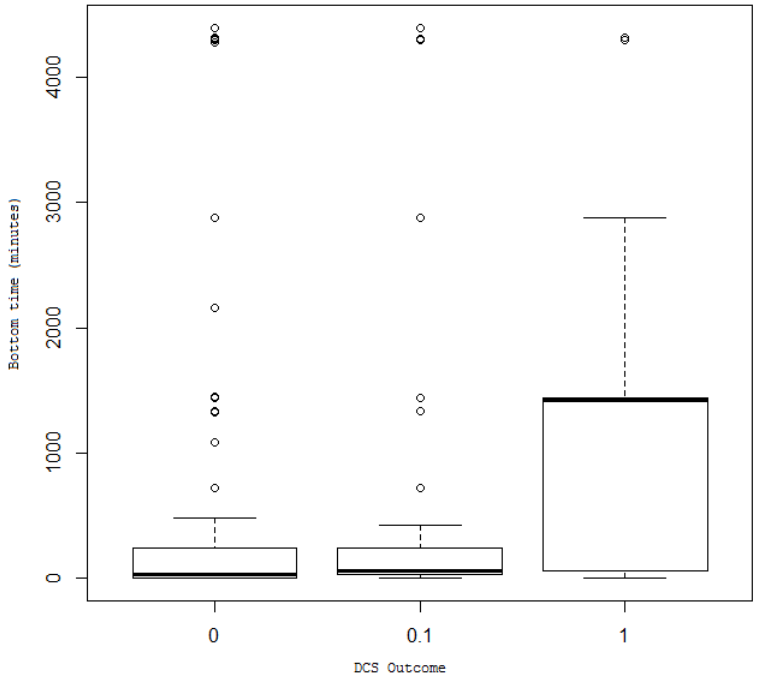

Figure 3.1: Bottom time (minutes) against DCS outcome

a higher proportion of DCS. On the other hand, Figure 3.2 would seem to indicate that DCS is less common at higher maximum dive depths. In fact, this is somewhat misleading, as most of the saturation dives with long bottom times were conducted with lower maximum dive depths, and the bottom-time is more readily associated with positive DCS outcomes than is maximum dive depth (where the maximum depth may only be sustained for a very short time).

In what follows in this chapter, it may not be immediately evident why we need to consider the possible models for our different data types (i.e. saturation, sub-saturation and bounce dives) separately. However, consider that the different dive-types are unlikely to be homogeneous in nature.

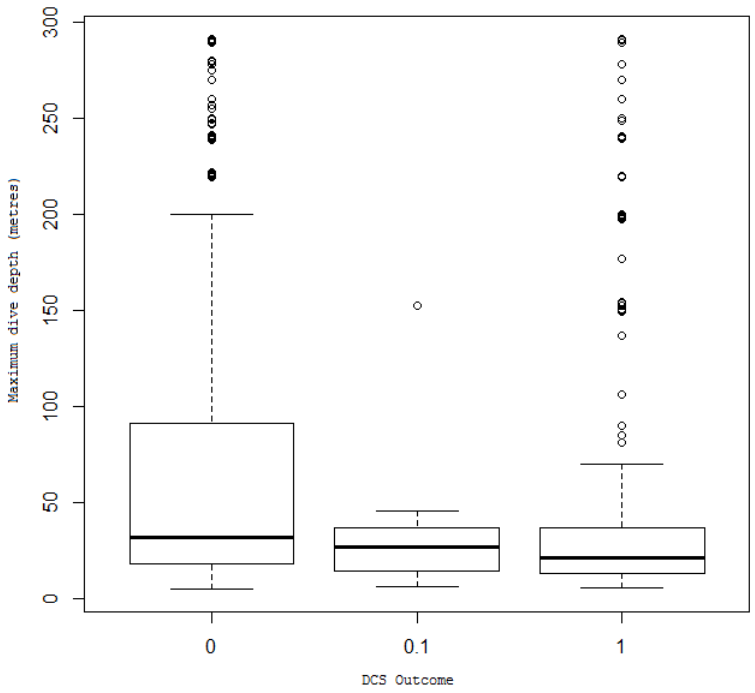

Figure 3.2: Maximum dive depth (metres) against DCS outcome

Indeed, we would be most surprised if they were! If this *were* the case then a single model would give optimum parameter estimates for any particular dive profile as in our model there is no allowance for an 'interaction' type effect. Furthermore, on accounting for maximum dive depth, we would expect similar proportions of DCS across the various groups of dive-types. But consider Figure 3.3 in which each row relates to a particular dive-type (being saturation, sub-saturation and bounce respectively), where we have binned the data into four different subsets depending on whether the maximum dive depth falls below the first quartile of maximum dive depths (16.5 m) for our total dataset (first column), between the first (16.5 m) and second (30.5 m) quartiles (second column), between the second (30.5 m) and third (50.9 m) quartiles (third column) or above the third quartile

(fourth column) in which we show the proportion of that type of dive for which DCS occurred.

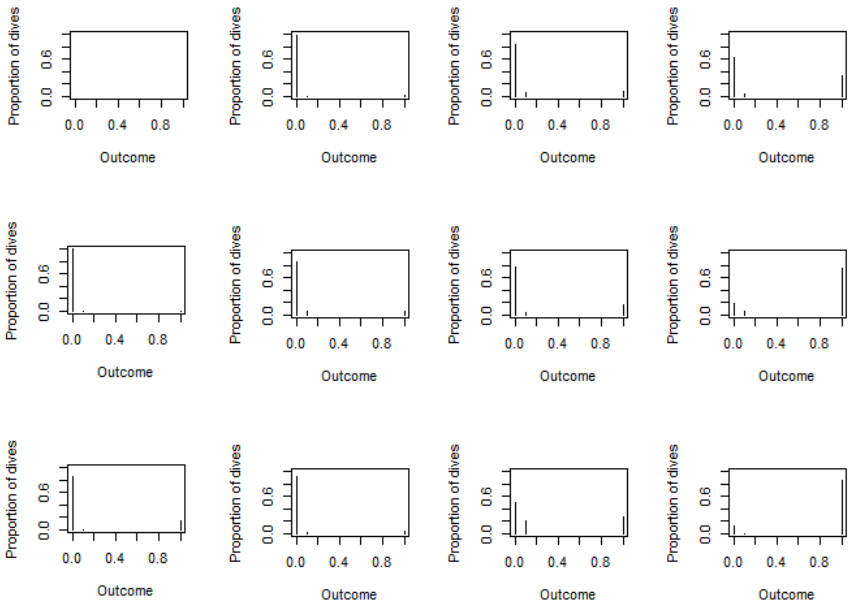

Figure 3.3: DCS incidence as a proportion of total incidence in each bin by dive-type and depth. Rows signify dive-types bounce, sub-saturation and saturation respectively. Columns signify maximum dive depth: maximum dive depth below first quartile of maximum dive depths, between first and median, between median and upper quartile and above upper quartile respectively.

From this diagram (Figure 3.3), it is plain to see that the distribution of decompression sickness for each of the different dive-types is quite different. There is a considerably higher proportion of DCS cases within the saturation dives than in the sub-saturation or bounce cases. Further, DCS incidence clearly increases with maximum dive depth. It follows that a single model is unlikely to be suitable for all three dive-types as the dives as a whole certainly do not form an iid sample from a single population and there is systematic variation between them. We need to account for this somehow when making our predictions. However, our data will often not include specific bottom times, so that (a) estimates of these must be

made and (b) they will not be accounted for as a covariate, when in fact they clearly should be as there is variation across the different dive-types. To this end, we allow for three sets of parameter estimates for each model - one representing the parameters when considering saturation dives, one for sub-saturation dives and one for bounce dives. This will allow us to take into account the possibility of different parameters being needed for different types of dive without requiring any further covariates in our model.

3.2 Baseline Models for Comparison

In what follows in this subsection, we bin the 'marginal' DCS cases with the no-DCS cases as they are incompatible with the logistic model framework which only allows for binary successes and failures and are considered by experts (Howle et al., 2009) to be near to non-events.

There are two baseline models that we might use to compare our models against. The first is the basic null model whereby we simply model the probability of DCS as a constant - the mean incidence of DCS within the dataset. This assumes that the covariates have no effect on the incidence or otherwise of DCS. We should expect that this model will provide a very poor fit, given our observations in Section 3.1, but it is still useful to use as a comparator. This model gives the estimated probability of DCS occurring on any dive as being 0.1344 and has an AIC_c of approximately 3473.

The other, less naïve but still basic, model we can use as a comparator is that given by a logistic regression of DCS outcome using maximum dive depth and saturation depth as covariates. This model is less simplistic than the null model (as is to be expected given that it takes into account what we fully expect to be useful covariates) but, as we shall see later, is still not as good as the models that we shall consider subsequently. We do not

include an interaction term here as this is the current 'default' model used by QinetiQ and they do not include an interaction term. We do include one in the models below that are separated by dive-type as these are new models. The linear predictor for this particular model, then, is given by

$$-2.3148886 + 0.1379169x_1 - 0.0034904x_2,$$

where x_1 is the dive saturation pressure (in Pascals) and x_2 is the maximum dive depth (also in Pascals). These estimates are provided with their standard errors in Table 3.1.

We may obtain an estimate of $P(DCS)$ by applying the inverse logit function

$$\text{invlogit}(x) = \frac{1}{1+e^{-x}}$$

to this linear predictor. This model has an AIC_c of approximately 2923 - a considerable improvement on the null model, but still not as good as the later models that we shall use.

	Estimate	Std. Error	z-value	p-value
Intercept	-2.3149	0.0751	-30.176	$< 2 \times 10^{-16}$
Saturation Depth	0.1379	0.0077	18.210	$< 2 \times 10^{-16}$
Escape Depth	-0.0035	0.0008	-4.845	1.27×10^{-6}

Table 3.1: Estimates for regression model using saturation pressure and maximum dive depth as covariates

Suppose we now split the data into the three dive-types and estimate a full factorial regression model incorporating dive-type, saturation depth and escape depth. This model has an overall AIC_c of 2444 - a further improvement on our previous models. The parameter estimates for this full model are included in Table 3.2 (in this model, note that the saturation dive-type is taken to be the reference level).

	Estimate	Std. Error	z-value	p-value
Intercept	-1.97176	0.2252	-8.605	$< 2 \times 10^{-16}$
Saturation Depth	0.13934	0.0199	7.229	4.85×10^{-13}
Escape Depth	-0.01365	0.0037	-3.682	0.000232
dive-type=Bounce	-4.39094	0.26624	-16.492	$< 2 \times 10^{-16}$
dive-type=Sub.Saturation	-2.02869	0.66224	-3.063	0.001096
Saturation Depth:Escape Depth	0.00117	0.00042	2.655	0.007929
Saturation Depth:(Bounce)	1.82256	0.0875	20.8292	$< 2 \times 10^{-16}$
Saturation Depth:(Sub.Saturation)	0.216382	0.032066	6.7480	7.49×10^{-12}
Escape Depth:(Bounce)	0.020827	0.00392	5.313	5.39×10^{-8}
Escape Depth:(Sub.Saturation)	0.093402	0.02465	3.789	7.56×10^{-5}
Saturation Depth:Escape Depth: dive-type=Bounce	0.065101	0.00131	49.695	$< 2 \times 10^{-16}$
Saturation Depth:Escape Depth: dive-type=Sub.Saturation	-0.006929	0.002651	2.614	0.004474

Table 3.2: Estimates for regression model for saturation dives using saturation, maximum dive depth and dive-type as covariates with all interactions

3.3 The Basic Supersaturation Model

The compartmental model for DCS is based on work by Thalmann (1984) and was subsequently further investigated in Thalmann et al. (1997). This model splits the body into a number, n say, of compartments that share similar characteristics such as blood flow rate. It should be noted, however, that such compartments do not correspond to any particular parts of the body. We consider the case here when $n = 1$ so that there is a single compartment to be considered. Thalmann then considers the inert (i.e. non-metabolised) gas (normally nitrogen or helium) in the breathing mix, which is the gas most likely to form bubbles, as it is not metabolised by the body. He hypothesises that inert gas is taken into the tissue at an exponential rate and is expelled at either a linear or an exponential rate (depending on a threshold parameter), with all transport occurring by

perfusion, rather than by diffusion. He then derives an expression for the amount by which a diver at a given depth is overpressurised relative to the current external pressure and terms this the supersaturation ratio. Note that we always take the inert gas to be nitrogen as this is the inert gas used in the tests whose results are provided in the QinetiQ datasets (because the air breathed in the submarine and during escape is air, which is comprised of oxygen and nitrogen, rather than helium).

In what follows, the necessary notation for which is summarised in Table 3.3, P_a represents the ambient pressure outside the body, P_b the inert gas burden (i.e. the gas partial pressure that would exist if any extant gas bubbles were forced back into the tissue, combined with actual gas partial pressure within the tissue), T some threshold of supersaturation that the tissue can tolerate without any risk of DCS and P_m the constant pressure of metabolic gases such as oxygen and carbon dioxide that are always present in the body (Tikuisis and Gerth (2002)). Here, P_a is known and given in our dataset at multiple timepoints, as is P_m, which takes a very similar constant value for each individual. He then finds that a suitable supersaturation ratio, R, is given by

$$R = \frac{P_b - P_a - T + P_m}{P_a}.$$

Here, the numerator represents the difference in pressure in the tissue and the ambient environment less the threshold gas pressure that the tissue can tolerate (we do not, for example, experience DCS after swimming in a pool). The whole expression then gives the relative over (or under-)pressure of the internal tissues compared with the ambient external pressure. This is an important ratio as DCS is much more likely to occur when this ratio is large, since bubbles are likely to be created in the more highly pressurised internal tissues.

Identifier	Type	Definition
$\dot{Q}$	Parameter (but subsumed by τ)	The rate of blood flow to the compartment per unit of compartment volume (min^{-1})
α_t	Parameter (but subsumed by τ)	Tissue-nitrogen solubility coefficient (ml/Pa)
T	Parameter	Amount of nitrogen that compartment can safely tolerate (Pa)
τ	Parameter	Rate of exchange of nitrogen into and out of compartment (min)
g	Parameter	A dimensionless scaling parameter for the hazard function h
P_L	Parameter	Kinetics switching parameter (Pa)
x_1	Covariate	The saturation depth (Pa)
x_2	Covariate	The pressure at the escape depth (Pa)
x_3	Covariate	The bottom time (minutes)
P_a	Derived Covariate (Derived from x_2)	The arterial nitrogen partial pressure (Pa)
P_b	Function	The compartment nitrogen burden (Pa)
P_g	Function (a function of P_b)	The compartment nitrogen tension (Pa)
m	Constant	The number of datapoints
P_m	Constant	The total partial pressures of all metabolic gases (Pa)
α_b	Constant	Blood-nitrogen solubility coefficient (ml/Pa)
t	Variable	The time after decompression begins (minutes)
R	Function	The supersaturation ratio
h	Function	The derived hazard function
L	Function	The derived likelihood
S	Function	The derived survivor function

Table 3.3: Functions, parameters, covariates, variables and constants used in Chapter 3, together with their definitions

Now, in Thalmann et al. (1997), it is found that P_b is the solution of the differential equation

$$\frac{dP_b}{dt} = \frac{\alpha_b}{\alpha_t}\dot{Q}\left(P_i - P_g(P_b)\right), \tag{3.1}$$

subject to the initial condition $P_b(0) = P_S$, where P_S is the saturation pressure for a particular set of parameters, which will be described shortly, and where ascent begins at time $t = 0$. That is, initially, the compartment's Nitrogen partial pressure will be equal to that at the relevant saturation pressure.

Now, in (3.1), P_g is the inert tissue gas tension (i.e. the actual inert gas partial pressure present in the tissue), given by

$$P_g(P_b) = \begin{cases} P_b & \text{for } P_b < P_a + P_L - P_m \\ P_a + P_L - P_m & \text{for } P_b \geq P_a + P_L - P_m, \end{cases} \tag{3.2}$$

where P_L is the excess inert gas partial pressure at which linear expulsion kinetics begins to apply. That is, once the threshold pressure P_L is reached, slow, linear offgassing will occur. Once sufficient offgassing has taken place, the kinetics will become exponential as in the gas-uptake phase. The appropriate P_g must be substituted into (3.1) before we attempt to solve (3.1) itself. Furthermore, P_i is the arterial inert gas pressure, $\dot{Q}$ is the rate of blood flow to the tissue per unit of tissue volume, α_b is the blood inert gas solubility coefficient (this is a constant known fairly accurately for each inert gas used in diving) and α_t is the tissue inert gas solubility coefficient. As noted in Tikuisis and Gerth (2002), α_t varies for different tissues and, given that the tissues in the model do not correspond to physiological tissues, may be difficult to determine.

Before we continue, observe that by considering the covariate x_2, which is the pressure (in Pa) at the escape depth for the dive, and supposing a linear

rate of rise of the subject of 3 ms^{-1} and consequent decrease of 30,000 Pa per second in the ambient pressure as suggested by Waters et al. (2007), we may write that the arterial (external) nitrogen partial pressure at time t is given by

$$P_a(t|x_2) = \begin{cases} x_2 - 30000t & \text{for } 0 \leq t < x_2/30000 \\ 0 & t \geq x_2/30000, \end{cases} \tag{3.3}$$

with the case for $t \geq x_2/30000$ corresponding to the subject having surfaced, and this is approximately equal to the nitrogen partial pressure in the lungs (though, in fact, it is marginally less due to the increased level of water vapour in the lungs).

Parameter estimation is then achieved by making use of the data to determine suitable parameter values for the model in question. Thalmann et al. (1997) do this by rewriting (3.1) as

$$\frac{dP_b}{dt} = \frac{1}{\tau}\left(P_i - P_g\right), \tag{3.4}$$

substituting for P_g using (3.2) as appropriate, where τ is defined by

$$\frac{1}{\tau} = \frac{\alpha_b}{\alpha_t}\dot{Q}$$

and then solving (3.4) subject to the initial condition $P_b(0) = x_1$.

In order to proceed, we shall need some analytic results not derived in the original papers by Thalmann and Thalmann et al. There, it was possible to solve the differential equations for tissue gas burden at time t using iterative (step-based) methods and continue to compute in small increments as was suggested by Thalmann et al. Unfortunately, we now need so many likelihood evaluations that that approach - while previously merely slow (Loveman, Personal Communication, 2009) - is now untenable. Thalmann

et al. note that the required equations for switching to and from linear kinetics 'can only be solved for t if there is no depth change or the pressure of O_2 [not merely its rate of change] is constant'. However, for the types of dives in which we are interested, we *can* provide such solutions by considering what happens at each stage of our dive. This will be essential to gain sufficient likelihood evaluations and reduces the time required for a single likelihood evaluation to approximately one second.

To begin with, one problem with the saturation pressure (as opposed to the saturation *depth*) is that it in fact depends on the parameter τ as well as on covariates. Thus, we must solve (3.4) using the initial condition

$$P_b(0) = P_S(\tau|S_D, D, t_s, O_2, t_b) \tag{3.5}$$

where P_S is the initial saturation pressure. We shall find an expression for this shortly. However, as this initial condition does not depend on t itself, it is, given the dive parameters and the parameter τ, a constant. In all cases below, pressures are *absolute* rather than gauge - that is, we include the pressure imposed by the atmosphere.

To solve (3.4) subject to the initial condition (3.5), we observe that there are two cases to consider, corresponding to which part of the piecewise function $P_g(P_b)$ given in (3.2) is to be applied (i.e. whether there is a transition to Linear Kinetics). In the case where $P_b < P_a + P_L - P_m$, so that $P_g = P_b$, we find that the desired solution is (as a function of time) given by

$$P_g(t) = P_o e^{-t/\tau} + (P_{a_o} - P_H)(1 - O_2)(1 - e^{-t/\tau}) + (1 - O_2)U(t + \tau(e^{-t/\tau} - 1)),$$

where P_H is the (constant) partial pressure of water at $37°C$.

This gives the nitrogen level at time t, given that it was P_0 at time 0, where P_{a_o} is the ambient pressure at time 0 and that $U = U(t_a, S_D) = S_D/t_a$,

which depends upon t_a the time to ascend and S_D the saturation depth, is the rate of rise of the subject. Where data on t_a and S_D are missing, we use the (physically observed) estimate suggested by Waters et al. (2007) of 30,000 Pa/sec.

Alternatively, in the case where $P_b \geq P_a + P_L - P_m$, we have that

$$P_g(t) = P_o - \frac{1}{\tau}\left(t(O_2(P_{a_o} - P_H) + P_H + P_L - P_m) + \frac{O_2 U}{2}t^2\right) \tag{3.6}$$

Note that (3.6) may only ever apply during *ascent* as we only want the slower linear kinetics to be possible during the offgassing ascent. Using this observation, we note that the nitrogen pressure in the compartment immediately after submerging may be found to be given by

$$\begin{aligned} P_{S_1}(\tau|S_D, t_D, t_s, t_a, O_2) = (1 - O_2)e^{-t_s/\tau} + (A - P_H)(1 - O_2)(1 - e^{-t_s/\tau}) \\ + UO_2(t_s + \tau(e^{-t_s/\tau} - 1)) \end{aligned}$$

where A is the (constant) surface pressure, O_2 is the fraction of inspired oxygen, t_s is the time spent descending to the saturation depth, S_D is the maximum dive depth and D is the time taken to reach the maximum dive depth. Note that we assume that the time to ascend and descend to a given depth are the same, so that we may use U here as well. Barring external forces, this is highly likely to be reasonable to within a good order of accuracy.

Further, given that a time of t_b is spent at the maximum depth, we find that the initial saturation pressure, P_S is given by

$$\begin{aligned} P_S(\tau|S_D, t_D, t_s, O_2, t_b, t_a) = P_{S_1}(\tau|S_D, D, t_s, O_2)e^{-t_b/\tau} \\ + (S_D - P_H)(1 - O_2)(1 - e^{-t_b/\tau}) \end{aligned} \tag{3.7}$$

and we use the initial condition (3.5) to solve (3.4) with P_S as given in (3.7).

First, consider the case where no change to linear kinetics applies and where the origin time is after saturation has occurred (to whatever degree is applicable). In this case, we obtain the solution

$$N(t) = \begin{cases} P_S e^{-t/\tau} + (E - P_H)(1 - O_2)(1 - e^{-t/\tau}) & t \leq E/U, \\ \quad +(1 - O_2)U(t + \tau(e^{-t/\tau} - 1)), & \\ (P_S e^{-E/(U\tau)} + (E - P_H)(1 - O_2)(1 - e^{-E/(U\tau)}) & t > E/U, \\ \quad +(1 - O_2)U(t + \tau(e^{-E/(U\tau)}) - 1))e^{-(t-E/U)/\tau} & \\ +(A - P_H)(1 - O_2)(1 - e^{-t/\tau}), & \end{cases}$$

where we must split into the two cases where $t < E/U$ and $t \geq E/U$ as at time $t = E/U$ the diver surfaces and there is no longer any change in the ambient pressure.

Now, in order to determine the required solution for the whole dive in terms of the covariates, we need to find the nitrogen level in the compartment as a function of t, taking into account the possibility of a switch to linear kinetics. We need to find when such a change may occur. Observe first that the equation governing the linear kinetics is quadratic in t provided that the surface has not yet been reached, for if the surface has been reached there is no longer any change in the ambient pressure so that $U = 0$ and hence the coefficient of the square term is 0. It follows that, in fact, if a change to linear kinetics is to occur then it must occur *before* $t = E/U$. To see this, suppose that this is not the case. As a change to linear kinetics would occur when $P_b = P_a + P_L - P_m$, we would first require a change into linear kinetics when this level was first reached. But then, if we suppose that our linear kinetics function is described by N_2, we would have $P_a(t) + P_L - P_m = N_2(t)$. But as this is linear in t there could only be a transition into linear

kinetics and not away from it. But this is impossible as we *must* finish with exponential kinetics once the nitrogen level in our compartment has dropped below the threshold level. Therefore, it must indeed be the case that a change occurs before $t = E/U$.

Next, we have two possibilities to consider. There may be a switch to linear kinetics before $t = E/U$, followed by a switch back to linear kinetics before $t = E/U$. Alternatively, there may be a switch to linear kinetics before $t = E/U$ followed by a switch back to exponential kinetics after $t = E/U$. We label the function for this first case N_a and for the second N_b. Further, we designate the time at which we switch to linear kinetics as t_1 and the time at which we switch away from this kinetics as t_2. The equation that must be satisfied by t_1 in both cases is given by

$$P_S e^{-t_L/\tau} + (E - P_H)(1 - O_2)(1 - e^{-t_L/\tau}) \\ - (1 - O_2)U(t_L + \tau(e^{-t_L/\tau}) - 1) = E - Ut_L + P_L - P_m. \tag{3.8}$$

Note that the maximum amount of nitrogen in the compartment occurs at time $t = t^*$ where

$$t^* = \tau \ln\left(\frac{O_2 P_H + U\tau + E - O_2 U\tau - EO_2 - P_H - S}{(1 - O_2)U\tau}\right)$$

In solving (3.8), we shall have need of a special function - namely the Lambert W-function (originally used by Corless et al. (1996)). This function is defined to be the (many-valued) function satisfying the equation $z = W(z)\exp(W(z))$. For real arguments z, it has one real-valued branch for $z > 0$ and two real-valued branches for $-1/e \leq z < 0$. We require the principal branch as this branch is continuous at $z = 0$.

The solution to (3.8) is then given by

$$t_L = \frac{1}{O_2 U}\Bigg(W\Bigg(\frac{1}{O_2 U\tau}\Bigg((-O_2U\tau - EO_2 + O_2P_H + U\tau + E - P_H - S) \\ \times \exp\left(\frac{-O_2U\tau - EO_2 + O_2P_H + U\tau + P_m - P_H - P_L}{O_2U\tau}\right)\Bigg)\Bigg)O_2U\tau \\ + O_2U\tau + EO_2 - O_2P_H - U\tau - P_m + P_H + P_L\Bigg).$$

To ensure that our condition that linear kinetics is only followed during ascent and never during descent, we must set $t_1 = \max(t_L, t^*)$.

While this may not seem to be an immediately useful solution, the Lambert W-function may be evaluated efficiently using, for example, the routines provided in the open source GNU Scientific Library and thus - while not elementary - our solution is in fact useful.

Now, suppose we switch back to exponential kinetics at time $t_2 < E/U$. In this case, we obtain the following solution for the overall nitrogen level $N(t)$ in the compartment at time t

$$N(t) = \begin{cases} P_S e^{-t/\tau} + (E - P_H)(1 - O_2)(1 - e^{-t/\tau}) & t \le t_1, \\ \quad +(1 - O_2)U(t + \tau(e^{-t/\tau} - 1)) & \\ P_1 - ((O_2(E - Ut_1 - P_H) + P_H + P_L - P_m) & t_1 < t \le t_2 \\ \quad \times(t - t_1) + 0.5O_2U(t - t_1)^2)/\tau, & \\ P_2 \exp(-(t - t_2)/\tau) + (E - Ut_2 - P_H)(1 - O_2) & t_2 < t \le E/U \\ \quad \times(1 - \exp(-(t - t_2)/\tau)) + (1 - O_2)((t - t_2) & \\ \quad +\tau(\exp(-(t - t_2)/\tau) - 1)), & \\ P_3 \exp(-(t - E/U)/\tau) + (P_o - P_H)(1 - O_2) & t > E/U, \\ \quad \times(1 - \exp(-(t - E/U)/\tau)), & \end{cases}$$

where P_o is the pressure at the surface and $P_1 = N(t_1), P_2 = N(t_2), P_3 = N(E/U)$. Here, t_2 is the larger solution of the quadratic

$$E - Ut_2 + P_L - P_m = P_1 - \frac{1}{\tau}(O_2(E - U_t - P_H) + P_H + P_L - P_m)(t_2 - t_1) + \frac{O_2 U}{2}(t_2 - t_1)^2.$$

In the case where we switch back to exponential kinetics at time $t_2 > E/U$ we instead obtain the following solution

$$N(t) = \begin{cases} P_S e^{-t/\tau} + (E - P_H)(1 - O_2)(1 - e^{-t/\tau}) & t \le t_1, \\ \quad +(1 - O_2)U(t + \tau(e^{-t/\tau} - 1)) & \\ P_1 - ((O_2(E - Ut_1 - P_H) + P_H + P_L - P_m) & t_1 < t \le E/U \\ \quad \times(t - t_1) + 0.5O_2U(t - t_1)^2)/\tau, & \\ P_2 - (O_2(P_o - P_H) + P_H + P_L - P_m)(t - E/U), & E/U < t \le t_2 \\ P_3 \exp(-(t - t_2)/\tau) + (P_o - P_H)(1 - O_2) & t > t_2 \\ \quad \times(1 - \exp(-(t - t_2)/\tau)), & \end{cases}$$

where P_o is again the pressure at the surface while $P_1 = N(t_1), P_2 = N(E/U), P_3 = N(t_2)$. Here, t_2 is the solution of

$$P_L - P_m = P_2 - \frac{1}{\tau}\left(O_2(P_o - P_H) + P_H + P_L - P_m\right)\left(t_2 - \frac{E}{U}\right)$$

which is given by

$$t_2 = \frac{EO_2P_H - EO_2P_o - P_2U\tau + P_LU\tau - P_mU\tau - EP_H - EP_L + EP_m}{U(O_2P_H - O_2P_o - P_H - P_L + P_m)}$$

Note that in both cases we have ensured continuity of our solutions at each of the crossover points between the piecewise continuous components of the functions N.

Having done this, a hazard for the compartment is now given by

$$h(T,\tau,g,P_L,t|x_1,x_2,x_3) = g\times \max\left\{\frac{P_b(\tau,P_L,t|x_1,x_2,x_3) - P_a(t|x_2,x_3) - T + P_m}{P_a(t|x_2,x_3)}, 0\right\} \tag{3.9}$$

so that the risk in the compartment is proportional to the relative supersaturation of that compartment above the tolerable threshold (with the constant of proportionality, g, being a parameter that is to be estimated and is known as the 'gain').

Note that T is reasonably physically interpreted as the amount of nitrogen that a compartment can tolerate without any risk of DCS being incurred and that τ could be interpreted as the exponential time constant for gas exchange in and out of the tissue. However, g and P_L do not possess such interpretations and are parameters that are merely useful for modelling purposes, rather than being physically interpretable parameters.

3.4 Using the Single Compartment Model

3.4.1 The Likelihood Function

By using the standard survival analysis in Section 3.4.1, Weathersby, Homer and Flynn (1984) show that the probability of not contracting DCS is given by

$$S(T,\tau,g,P_L|x_1,x_2,x_3) = \exp\left(-\int_0^\infty h(T,\tau,g,P_L|x_1,x_2,x_3,t)\,dt\right), \tag{3.10}$$

which we term the survival probability.

If there are m dives in the dataset then, by treating each dive j for $j = 1, 2, ..., m$ as a Bernoulli trial with success probability S_j and taking the product of the pdfs of all m dives, we arrive at the likelihood

$$l(T, \tau, g, P_L | x_1, x_2) = \prod_{j=1}^{m} S_j^{1-\delta_j} (1 - S_j)^{\delta_j}, \tag{3.11}$$

where $\delta_j = 1$ for the jth dive if DCS occurs and $\delta_j = 0$ if DCS does not occur. Note that in creating this likelihood, we assume that all dive observations are independent, which would seem to be a reasonable assumption given that most of the dives are undertaken by separate subjects, while where a subject is reused, there is at least a month's gap between their trials to avoid carryover effects (Loveman, Personal Communication, 2010).

3.4.2 Estimating integrals

The integral in the survivor function (3.10) (and thus the likelihood) is costly to compute and may not be evaluated analytically, whence we require a numerical approximation for it. We compute it by first transforming to the interval $[-1, 1]$ by using the transformation $t = (1-u)/(1+u)$. To see how this gives the desired result, observe that for this transformation we have

$$\frac{dt}{du} = \frac{-2}{(1+u)^2}$$

so that the integral becomes

$$\begin{aligned} \int_0^\infty h(t)\, dt &= \int_1^{-1} h\left(\frac{1-u}{1+u}\right) \times \frac{-2}{(1+u)^2}\, du \\ &= \int_{-1}^{1} \frac{2}{(1+u)^2} h\left(\frac{1-u}{1+u}\right)\, du. \end{aligned}$$

We then employ Gaussian Quadrature using Legendre polynomials to evaluate this integral. This method ensures that the integral approximation

$$\int_{-1}^{1} p_k(t)\,dt = \sum_{i=1}^{k} w_i p_k(x_i),$$

where the weights w_i are given constants and the x_i, known as nodes, are given by the roots of the kth Legendre polynomial, is exact for polynomials p_k of degree at most $2k-1$. This approximation generally provides good approximations when p_k is replaced by some other function f, say. For details see, for instance, Powell (1981).

This evaluation was found to be more accurate than using quadrature on the original interval, for which the approximation

$$\int_{0}^{\infty} e^{-t} p_k(t)\,dt = \sum_{i=1}^{k} q_i p_k(x_i)$$

holds with weights q_i, where the x_i are the roots of the kth degree Laguerre polynomials on the original interval $[0, \infty]$. To obtain the integral of a function f, we must write

$$\int_{0}^{\infty} f(t)\,dt = \int_{0}^{\infty} e^{-t} e^{t} f(t)\,dt = \sum_{i=1}^{k} q_i f(x_i) e^{x_i}$$

for this approximation to be applicable. However, as the exponential e^t is large for large t, this approximation subsequently produced unstable numerical evaluations. Moreover, attempting to use only a low degree quadrature so that only small nodes were used also gave poor accuracy compared with the previously mentioned Legendre quadrature.

It should be noted that Thalmann et al. (1997) proposed an alternative hazard function of the form

$$\int_0^\infty h(T, \tau, g, P_L, t|x_1, x_2)\, dt$$

to represent the fact that risk may be delayed and accumulate over time, rather than decreasing monotonically as the original function h does. Clearly, this will result in a double integral in the survivor function, which would be still more costly to compute. Indeed, for a full dataset, Howle, Weber and Vann (2007) note that full optimization requires a month of computing time with such a risk function. As they did not find any benefit in this non-monotonic risk function over that which was originally proposed, we shall not consider it further at this time, given the considerable increase in computational time that would then be required.

3.4.3 Bayesian inference for the single compartment model

We seek to implement a Bayesian approach to parameter estimation through Markov Chain Monte Carlo (MCMC). That is, we consider the parameter values to be random and seek to generate a dependent random sample from the posterior distribution of the parameters, given the data that we have observed. To achieve this, we place a prior distribution on the logarithm of each of the parameters T, τ, g and P_L. That is, we seek to move in the space $\log T$, $\log \tau$, $\log g$, $\log P_L$. In this way, we may move in an unrestricted parameter space (that of the logarithms of the desired positive parameters) as, after running MCMC over the logged space followed by using exponentiation, any of the parameters is guaranteed to be positive. In each case, a highly diffuse prior was used so that few assumptions were made about the possible parameter values. Note that we place independent priors on the

parameters in the original scale, rather than on the log scale. The reason for this is exemplified in Gelman (2006), where the author notes that this leads to less bias than placing priors on the parameters on the log scale. We also observe that while a Jeffreys prior would be theoretically pleasing in this context, obtaining it from this model would be possible only via numerical computation and is not available analytically. While this might be reasonable in the one compartment case with four parameters, in our later work in Chapter 4, it would become unreasonably cumbersome. We therefore take our priors as being $T \sim U(0, 10^6)$, $\tau \sim U(0, 10^4)$, $g \sim U(0, 1)$, $P_L \sim U(0, 10^6)$.

We next seek to construct a Markov Chain Monte Carlo sampling scheme to enable us to draw from the desired posterior distributions. In constructing this sampling scheme, we have thus far restricted ourselves to a standard Metropolis-Hastings transition scheme. We first choose a starting point. We then suppose that the transition probability of moving from one point in the parameter space $\boldsymbol{\theta} = (\log T, \log \tau, \log g, \log P_L)$ to another proposal point, $\boldsymbol{\theta}^* = (\log T^*, \log \tau^*, \log g^*, \log P_L^*)$, say, is given by a Gaussian random walk proposal centred at the current point. We have also used a small variance as, with a large variance, the T parameter becomes rather harder to estimate. Thus, under this proposal, the samples will tend to be rather close together and highly correlated. We choose to modify each parameter in turn, rather than moving all parameters simultaneously. Thus, we first update $\log T$ followed by $\log \tau$, then $\log g$ and finally $\log P_L$. As such a proposal is symmetric, we have that the probability of accepting such a move at the $(t+1)$th iteration is given by (see, for instance, Robert and Casella (2004))

$$\alpha = \min \left\{ \frac{l(\log \tau^* | \log T^{(t+1)}, \log g^{(t)}, \log P_L^{(t)}) \tau^*}{l(\log \tau^{(t)} | \log T^{(t+1)}, \log g^{(t)}, \log P_L^{(t)}) \tau^{(t)}}, 1 \right\},$$

for τ, say, where the factor $\tau^*/\tau^{(t)}$ accounts for the fact that we are making proposals on the log scale and subsequently exponentiating these for insertion into our routine to obtain the log-likelihood, L - the requirement for this factor being noted by Browne (2006). Hence, we propose a move using the desired Gaussian distribution and then accept this move with probability α. Note that α is generally better computed as

$$\alpha = \min\left\{\frac{\tau^*}{\tau^{(t)}}\exp(L(\log\tau^*|\log T^{(t+1)}, \log g^{(t)}, \log P_L^{(t)}) - L(\log\tau^{(t)}|\log T^{(t+1)}, \log g^{(t)}, \log P_L^{(t)})), 1\right\},$$

where $l = \log L$ is the log-likelihood noted above so as to avoid underflow, as L is generally very small. By generating samples in this way, we move around the sample-space in a dependent manner and build up a picture of the posterior distributions of the parameters and their behaviour.

3.5 Results

3.5.1 The First QinetiQ Dataset

In the case of the first QinetiQ dataset (as described in Section 1.7), we first find the maximum likelihood estimates of the four parameters for our model and these are given in Table 3.4.

Having obtained these, we can now give the results obtained by using Bayesian computation. In this case, we selected a normal proposal distribution for the logged parameters and tuned the proposal variances for each parameter so that the acceptance rates for the parameters were between around 0.1 and 0.4 per run, as is seen to be ideal by many (see, for

	Saturation	Sub-Saturation	Bounce
T	11041.71 (2667.33)	13935.04 (2981.27)	12521.17 (2879.74)
τ	985.14 (151.04)	718.61 (137.65)	498.11 (117.83)
g	0.004147 (0.00076)	0.002001 (0.00068)	0.002079 (0.00076)
P_L	10114.89 (1079.03)	13159.18 (1216.09)	14601.12 (1246.18)

Table 3.4: Maximum likelihood estimates (standard errors) for the single compartment model for the first QinetiQ dataset

example, Robert and Casella 2004, Chapter 7). Initially, we used a run of 10,000 iterations but this was barely enough for the Markov chains in question to converge to their stationary distributions. Instead, after initial exploration, a run of 200,000 iterations was made, with the first 10,000 iterations being discarded as burn-in. This gave the trace plots and kernel density estimates in Figure 3.4.

In order to provide these results, we need to consider how best to summarise our posterior distributions. Clearly, their standard errors will need to be given. However, we must then consider whether to summarise using the mean, median or mode as our primary point estimator of the requisite parameters as these will be required for later computations. Given the appearance of the posteriors below, it would seem that the mean is an unsuitable measure as the posterior distributions of the parameters are skew and so the mean will be a poor summary of the distribution. We must therefore choose between the posterior median and posterior mode. While the posterior median brings the advantage that is invariant to (legitimate, monotonic) transformations of the data, the posterior mode brings the advantage that, firstly, it is perhaps more easily interpretable than the posterior median. In addition, under a uniform prior, the posterior mode will be directly comparable to the MLE. For the reason of ease of comparison, then, we select the posterior mode as our summary measure of choice but note that the posterior median would do equally well.

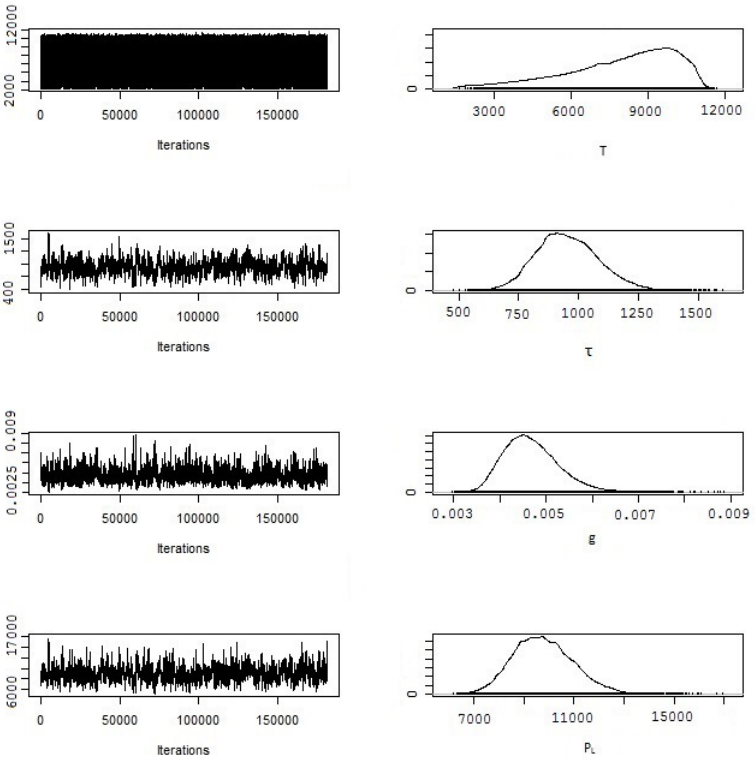

Figure 3.4: Trace plots and kernel density estimates for the single compartment model for the first QinetiQ dataset

In this case, we obtain the approximate estimates in Table 3.5 for the posterior modes of the parameters.

	Saturation	Sub-Saturation	Bounce
T	10872.59 (2631.18)	13732.08 (2902.56)	12733.61 (2885.01)
τ	939.55 (154.23)	714.83 (142.05)	513.36 (122.65)
g	0.004249 (0.00082)	0.002341 (0.00073)	0.002117 (0.00074)
P_L	10138.02 (1080.22)	13657.18 (1227.76)	14533.69 (1265.20)

Table 3.5: Posterior modes (standard errors) for the single compartment model for the first QinetiQ dataset using unthinned chains

We see here that there is a great deal of dependence between each of the

states of the chain and the next state. That is, there is a very high level of autocorrelation in the chain. To reduce this and obtain approximately independent samples, we take every 50th iteration - that is, we consider the effect of *thinning* the chains. This gives the trace plots and kernel density estimates in Figure 3.5.

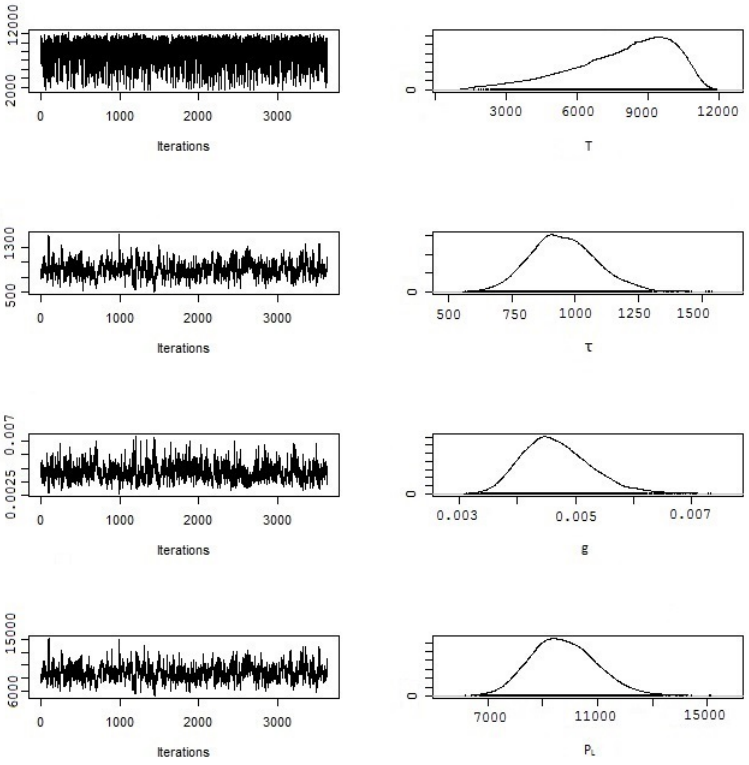

Figure 3.5: Thinned trace plots and kernel density estimates for single compartment model for the first QinetiQ dataset

We now obtain the estimates in Table 3.6 for the means and standard deviations of the parameters.

We now see that there is considerably better visible mixing within all the chains and that the kernel density estimates look rather smoother. However, in what follows we shall nevertheless *avoid* thinning our chains. The

	Saturation	Sub-Saturation	Bounce
T	10865.51 (2622.32)	13735.01 (2891.37)	12712.61 (2880.83)
τ	940.12 (152.25)	711.48 (139.87)	513.174 (122.58)
g	0.004241 (0.00070)	0.00231487 (0.00072)	0.00204714 (0.00072)
P_L	10187.98 (1065.60)	13645.64 (1212.72)	14556.42 (1250.83)

Table 3.6: Posterior modes (standard errors) for the single compartment model for the first QinetiQ dataset using thinned chains

reason for this is that we only reduce our effective sample size by removing available samples - even if these samples are dependent. This suggestion is supported by Link and Eaton (2012) who note that the only desirable reason to thin is when storage of all iterates causes a computer memory storage issue - something which will likely only be an issue in very high-dimensional problems with large numbers of iterates. In this case, we may store only every kth iteration in order to thin by a factor of k, thus reducing by a factor of k the amount of storage space required. In our case, we shall not have an issue with this, even when considering higher-dimensional models later on and we therefore choose to keep all of our samples at each stage and avoid thinning.

Here, the AIC_c for the model for the bounce data is 1223, for the sub-saturation data is 659 and for the saturation data is 949 (all to the nearest integer). Now, while these figures are of course not directly comparable (as the dataset for each is different), we can compare them with the results from Section 3.3, where we found initial logistic regression models for each sub-dataset. Indeed, we see that the logistic regression models perform rather better than each of the individual single compartment models here. Our single compartment model is clearly, then, not sufficiently good for our purposes as we would have a better model fit using the simple logistic regression model outlined earlier. We shall see in Chapter 6 that our more detailed models there perform rather better than the logistic regression models and far better than the single compartment model given in this

subsection. However, these models have been useful as an introduction to the type of model to be constructed and to our estimation procedure.

3.6 Some Validation

In order to validate the above results to some extent, we consider the saturation dives and split these into a training dataset (80% of the total dataset) and a test dataset (20% of the total dataset). We partition the data in this ratio at random and then refit the one compartment model to the saturation data using only the training dataset and then use the results to make predictions for the test dataset.

It is helpful at this stage to consider a pseudo-R^2, suggested in Tjur (2009). This is a simple, but intuitive, summary measure found by calculating the average predicted probability for each of the three possible outcomes (DCS, no DCS, again counting marginal DCS as a no DCS outcome) and taking the absolute difference between them. Essentially, this is rather sensible as we would expect events with a negative outcome (i.e. no DCS) to have predicted probabilities close to 0 and events with a positive outcome (i.e. DCS does occur) to have predicted probabilities close to 1. Note that this summary measure can only take values between 0 and 1 (inclusive), and we term this pseudo-R^2 R^2_T.

We would hope to obtain similar parameter estimates as we did using the whole dataset when using this restricted training dataset, as well as a similar R^2_T value for the predictions on the test dataset.

Fitting the model to the training dataset yields the parameter estimates in Table 3.7 and Figure 3.6. We see that, as hoped, these parameter estimates are quite similar to those obtained in Table 3.5 for the saturation dives

(although oddly, the standard errors are rather less, with the exception of g - this could be a peculiarity of the training dataset selected). In addition, the R_T^2 value for the test dataset using these parameter values is 0.2449. The value obtained for the whole dataset is 0.2631, and this similarity is again reassuring (despite the fact that the value is quite low, indicating that the model is not a good one).

	Saturation
T	10872.59 (1431.18)
τ	939.55 (54.23)
g	0.004049 (0.00082)
P_L	10138.02 (580.22)

Table 3.7: Posterior modes (standard errors) for the single compartment model for the simulated training dataset

3.7 Are survival times helpful?

Collecting detailed pressure data during a dive requires additional effort, man-hours and expense compared with the time required during a standard experiment, where the only covariates measured are body mass, saturation pressure and saturation depth. It is therefore of interest to examine whether anything is added by obtaining time-to-event data, rather than simple binary data (i.e. did DCS occur or did the subject remain healthy). In this section, we therefore consider this question by making use of the First QinetiQ dataset (as outlined in Section 1.7) and comparing our results with and without use of time-to-event data, attempting to see if our model is improved with the additional information provided by the time-to-event data.

Firstly, note that interval censored time-to-event data are available for approximately 95% of data in the first QinetiQ dataset. By interval censored

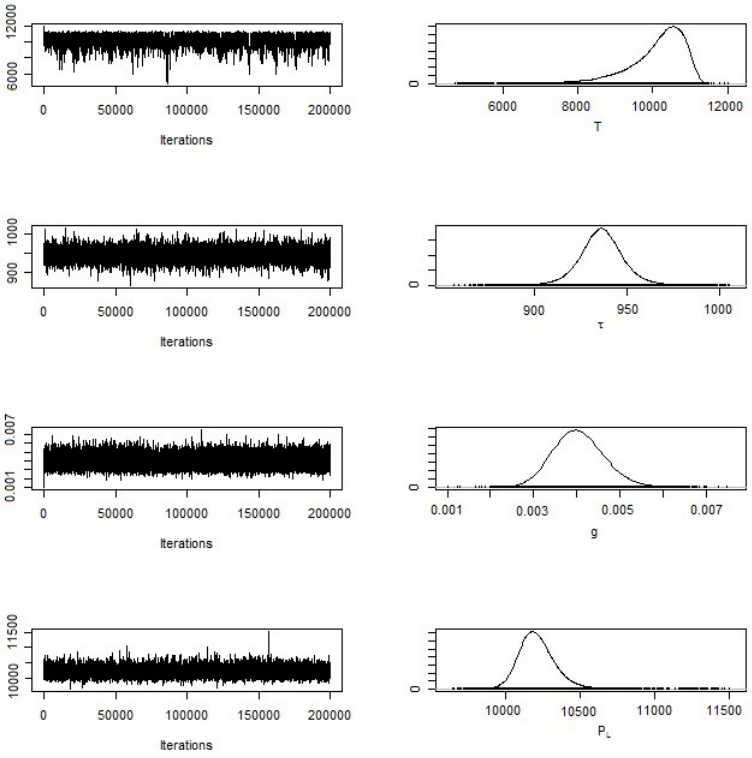

Figure 3.6: Trace plots and kernel density estimates for the single compartment model for the training dataset

data, we mean that if DCS occurs then we do not observe exactly when it occurs, but rather that it has not occurred by a time, t_1, say, but has occurred by some time t_2 so that the event occurs in the interval (t_1, t_2). The subjects for which these data are missing were not systematically biased in any way (in some cases, for example, experimenters were called away and unable to observe the DCS time or sometimes made less frequent observations as the trial progressed (Loveman, Personal Communication (2009)).

Our survivor function (3.10) remains the same in this case for dives where no DCS occurs, but for those where it does occur, it becomes

$$S(T,\tau,g,P_L|x_1,x_2,x_3,t_1,t_2) = \left(\exp\left(-\int_0^{t_1} h(T,\tau,g,P_L,t|x_1,x_2,x_3)\,dt\right)\right) \times \left(1-\exp\left(-\int_{t_1}^{t_2} h(T,\tau,g,P_L,t|x_1,x_2,x_3)\,dt\right)\right),$$

with our likelihood remaining the same with this new survivor function being used in cases where DCS occurs.

With this survivor function in place, we obtain the results in Table 3.8.

	Saturation	Sub-Saturation	Bounce
T	9483.41 (2190.04)	11921.72 (2365.56)	10437.94 (2208.54)
τ	937.45 (158.66)	710.905 (141.20)	463.43 (131.74)
g	0.00150768 (0.00088)	0.00183061 (0.00075)	0.0013584 (0.00070)
P_L	13475.66 (1191.01)	14078.72 (1198.07)	14872.16 (1236.40)

Table 3.8: Posterior modes (standard errors) for the single compartment model for the first QinetiQ dataset using interval censored survival times

We see that our parameter estimates are fairly similar to those obtained without using the survival data. Of course, we cannot directly compare the AIC_c values that we obtained previously to those obtained here, as the dataset here has been slightly modified to remove those few points for which we are missing the time-to-event data. Therefore, we recalculate the parameter estimates the AIC_c values obtained when we omit these data from our calculations. This gives us table 3.12.

This would seem to be paradoxical. By adding more information, we have actually *increased* our AIC_c value. However, to see why this might happen, consider the new form of the survivor function. It may well be that the

	Saturation	Sub-saturation	Bounce
AIC_c for binary data	949	659	1223
AIC_c for time-to-event data	1546	1241	1867

Table 3.9: AIC_c values (to the nearest integer) for single compartment models for the first QinetiQ dataset with and without using interval censored survival times

parameter values appropriate in $(0, t_1)$ are not appropriate in (t_1, t_2). Suppose, then, that we allow the gain parameter g to vary separately in $(0, t_1)$ and (t_1, t_2) so that we have parameters g_l in the first interval and g_u in the second interval, say. We retain the same priors as before for the parameters that we had previously and take priors $g_l \sim U(0, 1)$ and $g_l \sim U(0, 5000)$ for the new parameters. Indeed, while we might anticipate that a $U(0, 1)$ prior would have been suitable for g_u, in fact we find that this does not work and that g_u must be able to be much larger for this method to be successful. In this case, we obtain the parameter estimates in Table 3.10 and the AIC_c values in Table 3.11.

	Saturation	Sub-Saturation	Bounce
T	9483.41 (2190.04)	11921.72 (2365.56)	14392.17 (2482.67)
τ	937.45 (158.66)	710.905 (141.20)	522.05 (133.54)
g_l	0.00384992 (0.00099)	0.00121931 (0.00081)	0.00129032 (0.00069)
g_u	1444.13 (112.43)	1204.97 (104.20)	1335.51 (105.44)
P_L	10654.00 (1043.09)	13123.15 (1145.44)	12903.20 (1198.28)

Table 3.10: Posterior modes (standard errors) for the single compartment model with extra gain parameter for the first QinetiQ dataset using interval censored survival times

These results are much more in line with what we might expect from introducing new information and we have improved our AIC_c slightly, in part due to the fact that the interval $(0, t_1)$ is generally rather longer than (t_1, t_2), so that the hazard function requires a different scaling parameter. In addition, it is very likely that the hazard will behave rather differently in

	Saturation	Sub-saturation	Bounce
AIC_c for binary data	949	659	1223
AIC_c for time-to-event data	841	611	1184

Table 3.11: AIC_c values (to the nearest integer) for single compartment models with extra gain parameter for the first QinetiQ dataset with and without using interval censored survival times

the two intervals. Thus, in order to make full use of these data, we would have to double the dimensionality of our parameter space. For the work that follows in Chapter 4, this would be impractical as we already have tens of parameters in the largest models that we consider and computation time is already long. We therefore do not pursue this avenue of investigation further, but note that it could be expanded on in future work.

Chapter 4

The Multiple Compartment Model

4.1 Introduction

Having previously considered the single compartment model, we now move on to consider the multiple compartment model. Rather than having only a single set of four parameters to be varied within the hazard function, this model allows for multiple sets of four parameters to be involved in a mixture to create the hazard function in a manner similar to the mixing of normal distributions to provide a model for some situations where a single normal distribution would not be appropriate. The reasoning for this in our case is as follows: in the single compartment model, it is assumed that all of the tissues in the body behave in the same way. That is, the inert gas nitrogen (we are avoiding consideration of helium dives as this is not applicable to the submarine escape scenario) is taken up in the body and released at the same rate, regardless of which tissues in the body we might be considering.

However, this is clearly not the case - brain tissue, for instance, is highly perfused (see, for instance, Vann ed. (1989)) and will thus both absorb and release gas quickly. On the other hand, adipose tissues are less well

perfused and therefore very slow to be both permeated by our inert gas and then subsequently release any such gas. That is to say, the body is clearly not homogeneous in this respect and it is thus much more appropriate to consider a possible mixture of compartments here which, while not directly being representative of tissues within the body, will enable the modelling of this inhomogeneity within our framework. Note that, while this framework is not new and has been used before by, for example, Yount (1997) in his Variable Permeability model and Wienke (2001) in his Reduced Gradient Bubble Model, together with their own models for the hazard function, as well as by Thalmann et al. (1997), several important questions have gone unanswered in each of these. In particular, given our Bayesian framework, what uncertainty is there in our parameters? More importantly, how many compartments do we need in our model and how should we select the number of compartments to be used? In Section 4.2, we shall make an initial consideration of the second of these points - just how many compartments are useful in our model? Remember, of course, that the more parameters we must estimate the more computational time will be needed and, furthermore, only so many parameters can be estimated from our data given the sparsity of DCS events within the dataset, so this is an important question.

In Section 4.5, we will lay the groundwork for considering how many compartments (and thus parameters) are needed by outlining a reversible jump algorithm (the idea of which was initially proposed by Green in his seminal paper (Green 1995)). Here, we use MCMC with moves based on splitting and merging somewhat similar to that in Richardson and Green (1997) where a mixture of normal distributions with unknown parameters is considered. In this thesis, we are considering an unknown mixture of *compartments* used to form the hazard function for our model with an already known simple likelihood (once given this hazard function) which is very similar to that in Chapter 3. The reasons for this will be explained later but are mainly due to properties of the key nitrogen exchange parameter which effectively 'defines' each compartment to some extent.

We shall also later consider Bayesian model averaging across models with *different* numbers of compartments using posterior model probabilities as weights - these being obtained from our reversible jump sampler - as well as trying hazards with more standard forms (such as the Weibull or Gamma distributions) for which conjugate priors are available to ease computation. We shall see how these may then be incorporated into our model averaging procedure as well and whether our predictive power is thus increased.

4.2 Initial observations

Here, we consider the initial observations we have made regarding the multiple compartment environment. First, observe that the hazard function is now very similar to the hazard of Chapter 3's single compartment setting, but with an adjustment in the form of a summation to account for there being M compartments, rather than just one. Our variables and parameters are given in Table 4.1.

Notice that this table is very similar to the table of parameters (Table 3.1) given in Chapter 3 but that each parameter has acquired an index $i = 1, 2, \ldots, M$ indicating to which compartment the given parameter corresponds as each compartment will have one of each parameter T, τ, g and P_L (whence these being vectors of parameters in the expressions that follow). We now obtain the following form for the survivor function S_j of a particular dive as being

$$S_j(\boldsymbol{T}, \boldsymbol{\tau}, \boldsymbol{g}, \boldsymbol{P_L} | M, x_{1j}, x_{2j}, x_{3j}) = \exp\left(- \sum_{i=1}^{M} \int_0^\infty h_i(T_i, \tau_i, g_i, P_{Li}, t | x_{1j}, x_{2j}, x_{3j}) \, dt \right),$$

Identifier	Type	Definition
$\dot{Q}_i$	Parameter (but subsumed by τ)	The rate of blood flow to compartment i per unit of compartment volume (min^{-1})
α_{ti}	Parameter (but subsumed by τ)	Tissue-nitrogen solubility coefficient for compartment i (ml/Pa)
T_i	Parameter	Amount of nitrogen that compartment i can safely tolerate (Pa)
τ_i	Parameter	Rate of exchange of nitrogen into and out of compartment i (min)
g_i	Parameter	A dimensionless scaling parameter for the weight of the ith compartment within the hazard function
P_{Li}	Parameter	Kinetics switching parameter for compartment i (Pa)
x_1	Covariate	The saturation pressure (Pa)
x_2	Covariate	The pressure at the escape depth (Pa)
x_3	Covariate	The bottom time (minutes)
P_a	Function	The arterial nitrogen partial pressure (Pa)
P_b	Function	The compartment nitrogen burden (Pa)
P_g	Function (a function of P_b)	The compartment nitrogen tension (Pa)
n	Constant	The number of datapoints
P_m	Constant	The total partial pressures of all metabolic gases (Pa)
α_b	Constant	Blood-nitrogen solubility coefficient (ml/Pa)
t	Variable	The time after decompression begins (minutes)
R	Function	The supersaturation ratio
h	Function	The derived hazard function
L	Function	The derived likelihood
S	Function	The derived survivor function

Table 4.1: Functions, parameters, covariates, variables and constants used in Chapter 4, together with their definitions

with likelihood given by

$$L(\boldsymbol{T}, \boldsymbol{\tau}, \boldsymbol{g}, \boldsymbol{P_L}|M, n, \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}) = \prod_{j=1}^{n} S_j^{1-\delta_j}(1-S_j)^{\delta_j},$$

where $\delta_j = 1$ for the jth dive if DCS occurs and $\delta_j = 0$ if DCS does not occur. Note that h_i is the hazard function defined in Chapter 3, Equation 3.9 as

$$h_i(T_i, \tau_i, g_i, P_{Li}, t|x_1, x_2, x_3) = g_i \times \max\left\{\frac{P_b(\tau_i, P_{Li}, t|x_1, x_2, x_3) - P_a(t|x_2, x_3) - T_i + P_m}{P_a(t|x_2, x_3)}, 0\right\},$$

but now includes the index i as there is a different function for each compartment. Note that while it would be useful to interchange the order of summation and integration here, so as to avoid computing M integrals per evaluation of the survivor function rather than just one, this is not possible as the expression for h_i depends on the taking of a maximum on a compartment-by-compartment basis.

As in Chapter 3, estimation of the parameters takes place using an MCMC Metropolis-Hastings procedure with acceptance probability α_w given by

$$\alpha_w = \min\left\{\frac{L(\boldsymbol{\theta}^*)f(\boldsymbol{\theta}^*)}{L(\boldsymbol{\theta}^{(t)})f(\boldsymbol{\theta}^{(t)})}\left|\frac{\partial(\boldsymbol{\xi})}{\partial(\boldsymbol{\theta})}\right|, 1\right\}, \tag{4.1}$$

where $\boldsymbol{\xi}$ is a vector of the parameters on their original scales while $\boldsymbol{\theta}$ is a vector containing all of the parameters on the scales to be used in the procedure (i.e. on the log scale). The factor $|\partial(\boldsymbol{\xi})/\partial(\boldsymbol{\theta})|$ here is the Jacobian of the transformation and represents the area factor required for our change of variables in moving from $\boldsymbol{\xi}$ to $\boldsymbol{\theta}$. Also, $f(\boldsymbol{\theta})$ in the above is the prior distribution for θ (of which more below).

We use a symmetric random walk Normal proposal q, whence q does not

appear in the above ratio (1) as it will be cancelled in the numerator and denominator of the acceptance probability due to the symmetry of q giving $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}) = q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^*)$. Note that we consider the aforementioned transformations of the parameters in question to ensure that the parameters themselves are all positive. This avoids our having to restrict our search to only the positive real line and instead allows us to search the whole of $\mathbb{R}^{4M}$ (there being four parameters per compartment), which is considerably more efficient.

As reasoned in Section 3.4, we again consider diffuse priors for the parameters on their original scales so we set $i = 1, 2, \ldots, M$, $T_i \sim U(0, 10^6)$, $g_i \sim U(0, 1)$, $P_{Li} \sim U(0, 10^6)$. We employ three different priors for the τ_i as we expect there to be a 'short' compartment, a 'medium length' compartment and a 'long' compartment. To that end, we set $\tau_i \sim U(0, 10^{i+4})$. We make one further restriction on the *joint* prior of our parameters. This is to require that $\tau_1 < 1.2\tau_2 < 1.2\tau_3 < \cdots < 1.2\tau_M$. This ensures that there is at least some spacing between compartments so that the gas uptake rate in each compartment is at least 20% greater than that in the previous compartment. The reason for imposing this latter constraint is perhaps not immediately obvious. In fact, it is to help us to avoid the problem of *label switching*. If we did not impose such a constraint then it would be unclear (to the MCMC algorithm) which compartment was which and so in, for instance, the two compartment case, it would not be clear which would be the 'fast' and which the 'slow' compartment. This could well lead to the algorithm switching seemingly randomly between the areas of the two required values of τ for each of the faster and slower compartments. By imposing our constraint (and requiring some separation between each τ_i), we avoid this problem as the labels on the compartments will be well defined.

4.3 A Simulation

Before deploying the method described above on the true dataset, it is sensible to consider a simulation to ensure that our method is working appropriately. To do this, we shall first require some simulated data and will need to select some parameter values for the model from which to simulate. We choose to simulate from a three compartment model as Thalmann et al. (1997) already developed maximum likelihood estimates for this model (for the original first QinetiQ dataset) and we therefore have a reasonable idea of sensible parameter estimates to use. To that end, unlike in the rest of the thesis, we do not separate the three different dive-types here into separate models, in order that we can directly compare with the maximum likelihood estimates obtained in the aforementioned paper. Note that while it would be ideal to carry out more than one simulation, these are quite time consuming to undertake and of course we could never simulate for all possible parameter values in any case, so it seems sensible to be guided some what by values that were previously found to be reasonable and see if these can be recovered by our method. The maximum likelihood estimates obtained therein were: $T_1 = 0, \tau_1 = 81.5393, g_1 = 0.0000708745 = 7.08745 \times 10^{-5}, P_{L_1} = \infty, T_2 = 0, \tau_2 = 3139.0449, g_2 = 0.0000018216 = 1.8216 \times 10^{-6}, P_{L_2} = 7440.8616, T_3 = 8780.26, \tau_3 = 29121.5146, g_3 = 0.000016902 = 1.6902 \times 10^{-5}, P_{L_3} = \infty$. Where they denoted a parameter value by ∞, this indicates that it took very high values and effectively the parameter had no effect (as it is a threshold parameter, and a high value indicates that the threshold is never reached).

In order to have reasonable simulated data, it will be necessary to draw from the full range of covariates (saturation and escape depths and dive-type) employed there. Furthermore, we shall limit the size of the simulated

dataset to reduce computation times. As the first QinetiQ dataset is comprised of approximately 4400 points, it would seem reasonable to take a simulated dataset of size approximately 880 (which is 20% of the size of the original dataset). To ensure coverage of the covariate ranges, we order the covariate value pairs first by saturation depth and then by escape depth and in each case take every 10th pair in each case for each dive-type. This will ensure that the requisite range is covered reasonably.

We must now choose suitable parameter values. To ensure that these are realistic, we shall choose values close to (but not exactly the same as) the maximum likelihood estimates mentioned above. In particular, we set $T_1 = 1000$, $\tau_1 = 70$, $g_1 = 7 \times 10^{-6}$, $P_{L_1} = 30000$, $T_2 = 1000$, $\tau_2 = 3000$, $g_2 = 2 \times 10^{-6}$, $P_{L_2} = 7000$, $T_3 = 20000$, $\tau_3 = 30000$, $g_3 = 1 \times 10^{-6}$, $P_{L_3} = 30000$.

Doing this, we obtain the results summarised in Table 4.2 and Figures 4.1-4.3. We see that in general the simulated results are quite reasonable, the posterior modes in general being at most two posterior standard errors from their set values, and generally rather less than that. The exception to this is the P_{L_i} parameters. To understand why this might be the case, note that these parameters are *threshold* parameters. That is, for any nitrogen pressure value in the ith compartment above this threshold, linear kinetics applies. However, it may be that linear kinetics is *never* appropriate for this compartment, in which case any suitably large threshold value above the maximum pressure reached will suffice. However, the value of the maximum nitrogen pressure reached depends in a complex way on the values of the other parameters for the compartment and also the particular dive profile and cannot be determined analytically. In their original paper, Thalmann et al. (1997) gave these thresholds a posteriori as being infinite where this seemed to be the case. However, this belies the possibility that they may not be and would not be suitable for our trans-dimensional method in Section 4.6.

Parameter	Approximate posterior mode (standard error)
T_1	850.23 (83.32)
τ_1	64.43 (25.43)
g_1	9.37×10^{-6} (2.21×10^{-6})
P_{L_1}	43739.47 (10601.33)
T_2	657.09 (299.17)
τ_2	3694.01 (735.01)
g_2	1.81×10^{-6} (4.90×10^{-7})
P_{L_2}	8953.10 (2557.07)
T_3	21073.39 (3579.59)
τ_3	32968.89 (4293.28)
g_3	9.77×10^{-7} (4.21×10^{-7})
P_{L_3}	55024.28 (7295.943)

Table 4.2: Estimated posterior modes (standard errors) of parameter distributions for simulated data

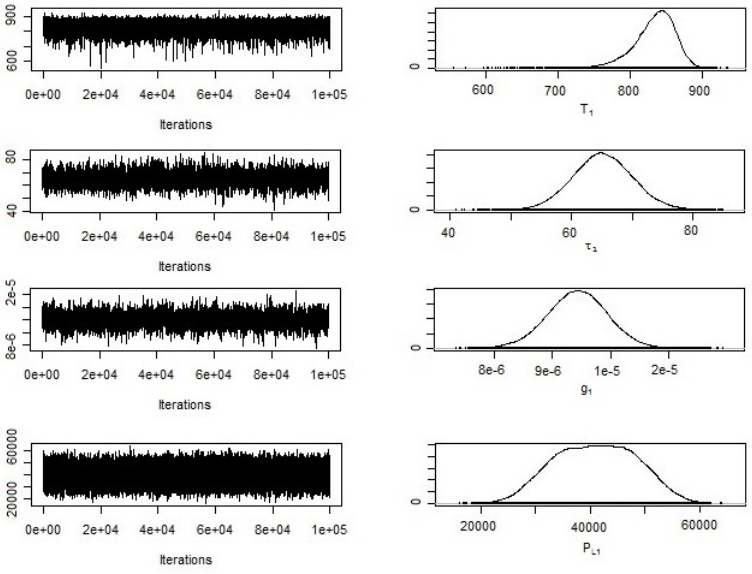

Figure 4.1: Trace plots and kernel density estimates for the first compartment using simulated data

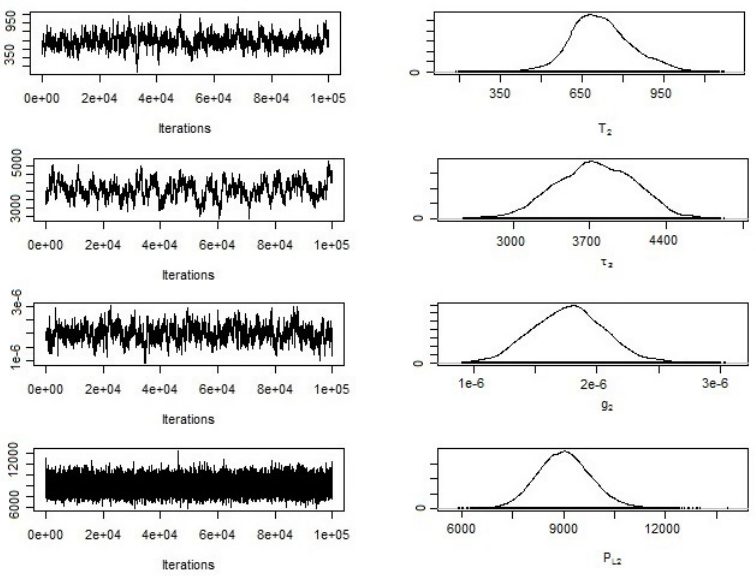

Figure 4.2: Trace plots and kernel density estimates for the second compartment using simulated data

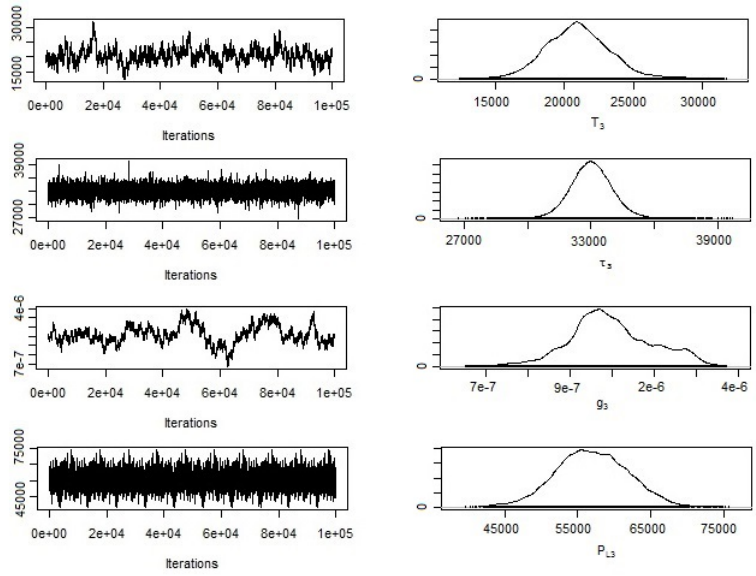

Figure 4.3: Trace plots and kernel density estimates for the third compartment using simulated data

4.4 Some Initial Results

In Figures 4.4-4.6, we give the trace plots and kernel density estimates for the parameter estimates in the three compartment case with bounce data to check that our algorithm works well with multiple compartments. We shall not proceed to evaluate multiple compartment models individually after this - rather, we shall proceed to using a transdimensional reversible jump algorithm in the next section and so we are not interested in the parameter estimates themselves here, but rather in examining whether the algorithm converges. 100,000 iterations were carried out and a burn-in of 10,000 iterations has been used.

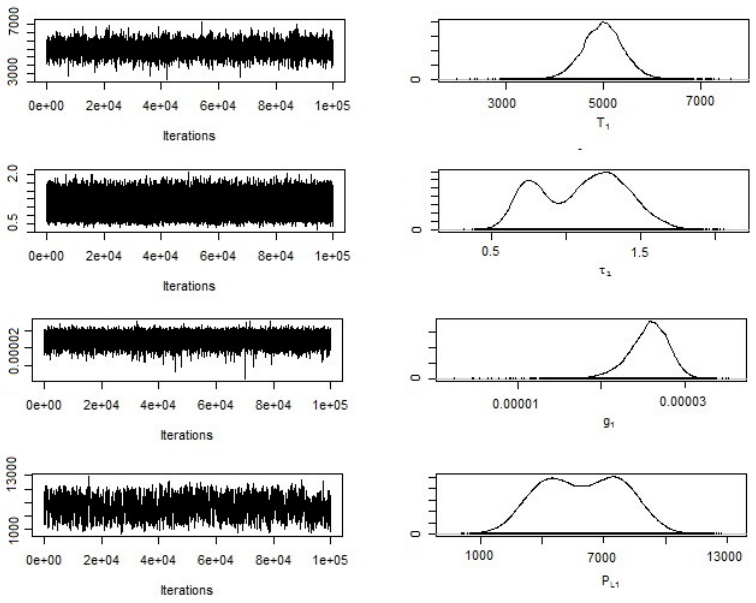

Figure 4.4: Trace plots and kernel density estimates for the first compartment using bounce data from the first QinetiQ dataset

We see in Figures 4.4-4.6 that although there is some reduction in mixing rate after the first compartment and the mixing in general is less good

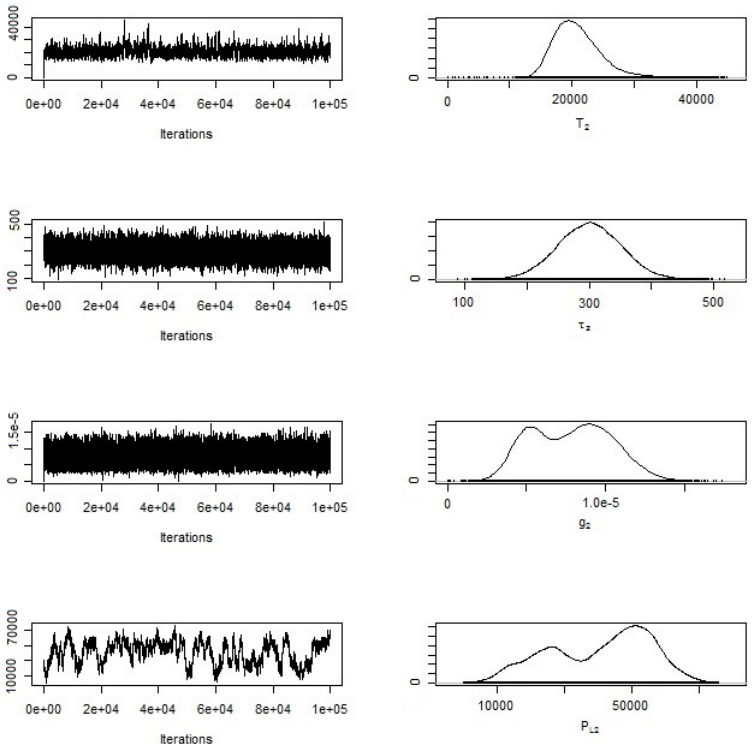

Figure 4.5: Trace plots and kernel density estimates for the second compartment using bounce data from the first QinetiQ dataset

than in our single compartment model or in our simulation, all the parameters for each compartment mix fairly well. This indicates that our algorithm is performing quite well and should indeed be apposite (after suitable modifications are made) for use in the transdimensional MCMC algorithm outlined in Section 4.4 below, though we should be aware that we are likely to require a good number of iterations for our MCMC procedure. In addition, there is definite evidence of bimodality here in some of the posterior distributions that was not found in the simulation run. Note that the AIC_c for this 3-compartment model is 452 for the saturation model, 447 for the sub-saturation data and 812 for the bounce data (all to the nearest integer). These figures all represent an improvement on those for

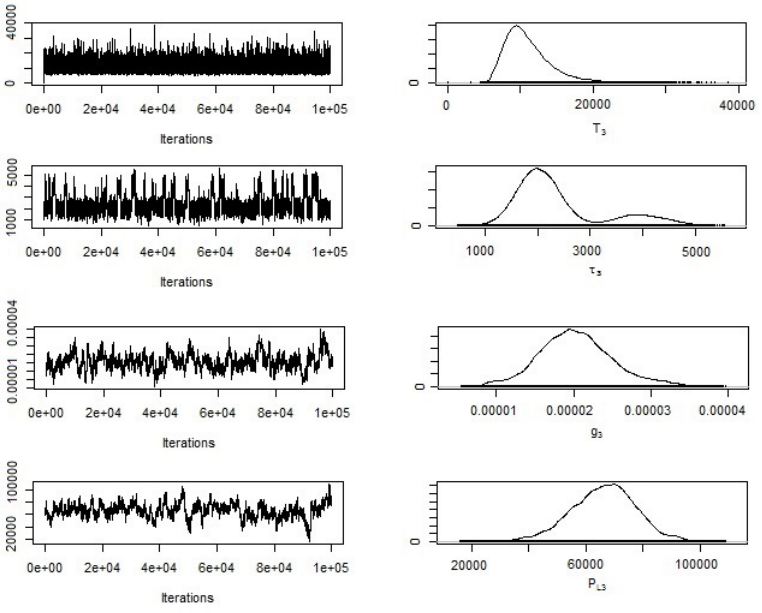

Figure 4.6: Trace plots and kernel density estimates for the third compartment using bounce data from the first QinetiQ dataset

the single compartment models found in Section 3.5.

4.5 Introduction to Reversible Jump MCMC

Essentially, the idea behind the Reversible Jump MCMC algorithm, as introduced by Green (1995), is one of dimension matching. This allows us to move between parameter spaces of different dimensions by eliminating variables according to some chosen algorithm, or by adding them, while all the time maintaining the same number of dimensions during the actual move process through the use of auxiliary variables. Thus, if we seek to move from a model of dimension m to one of dimension n then we must introduce (or eliminate) $|m - n|$ uniformly distributed random variables depending on whether m or n is the greater. Initially, Green suggested

birth and death type moves - moves where a parameter is simply introduced or removed (with the assistance of an auxiliary variable u for dimension matching). Subsequently, Richardson and Green (1997) suggested moves of a split and merge nature, where parameters are split apart and merged together, rather than simply being created and eliminated (i.e. one of the new parameters is created above the initial parameter value and the other is created below it). This has advantages in our case for the τ parameter, as moving to a new number of compartments is unlikely to create τ rate values that are wildly different from the previous model's values. Furthermore, we expect that the place of a single compartment will be taken by one 'faster' and one 'slower' compartment.

We now outline the algorithm proposed by Green, contextualising it for our particular problem.

Suppose that the chain is currently in a state $(k^{(t)}, \boldsymbol{\theta}^{(t)}_{k^{(t)}})$, where $\boldsymbol{\theta}^{(t)}_{k^{(t)}}$ is to be understood as a vector containing all of the parameters for a k-compartment model. Then, we first propose a new model k^*, say, with probability $v(k^{(t)}, k^*)$ where v is some pdf. We shall specify this function in Section 4.4.

Now, our choice of v leaves the possibility that $k^* = k^{(t)}$ so that we remain in the same model. In this case, we may use the MCMC procedure outlined in Section 4.2 for our move. If this is not the case and $k^* \neq k^{(t)}$ then we shall require a different approach.

Suppose, then, that $k^* \neq k^{(t)}$. We first generate a vector $\boldsymbol{u}$ from a proposal density $q(\boldsymbol{u}|\boldsymbol{\theta}_k, k, k^*)$. Next, set $(\boldsymbol{\theta}^*, \boldsymbol{u}^*) = w_{m,m^*}(\boldsymbol{\theta}, \boldsymbol{u})$, where w_{m,m^*} is a function whose inverse exists and which we shall specify in Section 4.6. To ensure that the function g_{m,m^*} exists, Dellaportas et al. (2002) note that it is sufficient that the *dimension matching* criterion $\dim(\boldsymbol{\theta}_k) + \dim(\boldsymbol{u}) = \dim(\boldsymbol{\theta}_{k^*}) + \dim(\boldsymbol{u}^*)$ hold and that, under this condition, $w_{m,m^*} = w^{-1}_{m^*,m}$.

Finally, we accept the proposed move to the new model k^* with probability

$$\alpha = \min\Bigg\{ \frac{f(\boldsymbol{y}|m^*, \boldsymbol{\theta}^*_{k^*}) f(\boldsymbol{\theta}^*_{k^*}|k^*) f(k^*) v(k^*, k) q(\boldsymbol{u}^*|\boldsymbol{\theta}_{k^{(t)}}, k^*, k)}{f(\boldsymbol{y}|m, \boldsymbol{\theta}_{k^{(t)}}) f(\boldsymbol{\theta}^{(t)}_{k^{(t)}}|k) f(k) v(k, k^*) q(\boldsymbol{u}|\boldsymbol{\theta}_{k^*}, k, k^*)} \times \left| \frac{\partial w(\boldsymbol{\beta}_{k^{(t)}}, \boldsymbol{u})}{\partial(\boldsymbol{\beta}_{k^{(t)}}, \boldsymbol{u})} \right|, 1 \Bigg\}, \tag{4.2}$$

where the final term here is the Jacobian of the transformation g_{m,m^*}.

In the next section, we shall elaborate on how to use this algorithm for our particular situation.

4.6 Reversible Jump algorithm for our compartmental model

We first need to define the function v as introduced in Section 4.5. We suppose that it is possible to jump to either one more compartments, remain at the same number of compartments or jump to one fewer compartment and that these possibilities are equiprobable. As this yields $v(k^{(t)}, k^*) = v(k^*, k^{(t)})$, no term in v will appear in our final acceptance probability, as it will be cancelled out, being the same in both the numerator and denominator.

We make use of a split and merge proposal as suggested in Section 4.5 above. In the case of a split move, consider the following possible scheme, where we move from model k to model $k+1$. We split the jth component into the components j and $j+1$ and give each parameter a second subscript so that g_{jk} represents the gain for parameter j under model k. Note that this will differ from the value of $g_{j(k+1)}$, which is the value of the gain for

parameter j under model $k+1$. Then, we consider the following proposal:

$$g_{jk} = \frac{1}{2}\left(g_{j(k+1)} + g_{(j+1)(k+1)}\right)$$
$$\tau_{jk} = \frac{1}{2}\left(\tau_{j(k+1)} + \tau_{(j+1)(k+1)}\right).$$

This will ensure that the new τ parameters are centred at the previous model's τ value, while allocating some proportion of the weight g, which was previously attached to just one compartment in the survivor function S, to each of the new compartments. In addition, in general one expects a higher gain value for shorter compartments (i.e. those with smaller τ values) as these will have a more immediate effect on more of the datapoints (given that some are non-saturation dives). It follows that a suitable split condition might be given by

$$T_{jk} = \frac{\tau_{j(k+1)}}{\tau_{jk}} T_{j(k+1)} + \frac{\tau_{(j+1)(k+1)}}{\tau_{jk}} T_{(j+1)(k+1)}.$$

Finally, there is no clear relation between P_L and any other parameters, so we take

$$P_{L_{jk}} = \frac{1}{2}\left(P_{L_{j(k+1)}} + P_{L_{(j+1)(k+1)}}\right)$$

to centre our new P_L values at the old P_L value.

Note that the reverse (merge) move is obtained deterministically from the above equations.

Now, to perform the split move outlined above, we generate auxiliary variables $u_{k(k+1)}$ as $u_1, u_2, u_3, u_4 \sim U(0,1)$. This accounts for q in the algorithm outlined in Section 4.3 and, as the distribution from which we have chosen to draw is symmetric and not dependent on the model parameters or model dimension, q will also not appear in our final acceptance probability as it will be cancelled in the numerator and denominator there. For our split

and merge proposal, therefore, we take

$$T_{j(k+1)} = u_1 T_{jk} \qquad T_{(j+1)(k+1)} = \frac{\tau_{jk} - \tau_{j(k+1)} u_1}{\tau_{jk} - \tau_{j(k+1)}} T_{jk} \tag{4.3}$$

$$\tau_{j(k+1)} = u_2 \tau_{jk} \qquad \tau_{(j+1)(k+1)} = (1 - u_2)\tau_{jk} \tag{4.4}$$

$$g_{j(k+1)} = u_3 g_{jk} \qquad g_{(j+1)(k+1)} = (1 - u_3) g_{jk} \tag{4.5}$$

$$P_{L_{j(k+1)}} = u_4 P_{L_{jk}} \qquad P_{L_{(j+1)(k+1)}} = (1 - u_4) P_{L_{jk}}. \tag{4.6}$$

The required Jacobian for this transformation is given by

$$\left| \frac{\partial w(\boldsymbol{\beta}_{k^{(t)}}, \boldsymbol{u})}{\partial(\boldsymbol{\beta}_{k^{(t)}}, \boldsymbol{u})} \right| = \begin{vmatrix} u_1 & \frac{\tau_{jk} - \tau_{j(k+1)} u_1}{\tau_{jk} - \tau_{j(k+1)}} & \dots & \dots & \dots & \dots & \dots & \dots \\ T_{jk} & \frac{-\tau_{j(k+1)} T_{jk}}{\tau_{jk} - \tau_{j(k+1)}} & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & u_2 & 1 - u_2 & \dots & \dots & \dots & \dots \\ 0 & 0 & \tau_{jk} & -\tau_{jk} & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & u_3 & 1 - u_3 & \dots & \dots \\ 0 & 0 & 0 & 0 & g_{jk} & -g_{jk} & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & u_4 & 1 - u_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & P_{L_{jk}} & -P_{L_{jk}} \end{vmatrix}.$$

Evaluating the Jacobian gives

$$\left| \frac{\partial w(\boldsymbol{\beta}_{k^{(t)}}, \boldsymbol{u})}{\partial(\boldsymbol{\beta}_{k^{(t)}}, \boldsymbol{u})} \right| = \frac{T_{jk} \tau_{jk}^2 g_{jk} P_{L_{jk}}}{\tau_{jk} - \tau_{j(k+1)}}.$$

Given the above considerations, and supposing that the probability of each model a priori is 1/8 (so that $f(k) = f(k^*) = 1/8$) and using the same diffuse uniform priors as in Section 4.3 for our parameters, we find that the acceptance probability (4.3) of moving between models is given by

$$\alpha_b = \min \left\{ \frac{T_{jk} \tau_{jk}^2 g_{jk} P_{L_{jk}} l(\boldsymbol{\theta}_{k+1})}{(\tau_{jk} - \tau_{j(k+1)}) l(\boldsymbol{\theta}_k)}, 1 \right\}, \tag{4.7}$$

where $l(\boldsymbol{\theta}_k)$ is the likelihood for the k-compartment model with parameters $\boldsymbol{\theta}_k$.

Having noted the required functions and inputs above, we can now use the algorithm given in Section 4.3. Thus, we first set an initial model, $k^{(0)}$ say, and initial state for the chain, $\boldsymbol{\theta}^{(0)}$, say. We then check whether there is a move between models. If there is, then we update $\boldsymbol{\theta}^{(0)}$ to $\boldsymbol{\theta}^{(1)}$ using the algorithm of Section 4.3. If there is not, then we update $\boldsymbol{\theta}^{(0)}$ to $\boldsymbol{\theta}^{(1)}$ using the method of Section 4.2, remaining within the same model. We now repeat the procedure but with the new model $k^{(1)}$ in place of model $k^{(0)}$ and $\boldsymbol{\theta}^{(1)}$ in place of $\boldsymbol{\theta}^{(0)}$. We continue this process to create our desired chain.

4.7 Another Simulation

Here, we make use of the same simulated dataset as was generated in Section 4.3. We set 8 as the maximum number of compartments allowed and 1 as the minimum and then run 8 chains (one starting at each of the possible numbers of compartments) for 100,000 iterations each, discarding the first 10,000 of each as burn-in. On averaging across the chains, this gives us the estimated posterior model probabilities in Table 4.3 for our simulated data. Here, we can see that the 3 compartment model is indeed selected as the most likely model, though there is also a fair amount of posterior weight on the 2 and (particularly) 4 compartment models. Having 8 chains also allows us to estimate the between-chains standard error, and assessments of this are also presented in Table 4.3. We see that there is (considered relatively) quite a wide range of variation - particularly for the very small and very large numbers of compartments - but that in terms of absolute variation this is not too large.

Number of compartments	Approximate posterior model probability (standard error)
1	0.02156 (0.03144)
2	0.12800 (0.05796)
3	0.47321 (0.12270)
4	0.28347 (0.06927)
5	0.04578 (0.05219)
6	0.01956 (0.02661)
7	0.02145 (0.03413)
8	0.00697 (0.00985)

Table 4.3: Estimated posterior model probabilities (between chains standard errors) for simulated data

4.8 Results

It is difficult to produce output plots for a transdimensional chain such as this one. Clearly, all parameters represent different things in different models, and so simply giving, say, the first set of parameters is unhelpful. What we can provide, however, is the posterior model probabilities, which are obtained by evaluating the proportion of time that the chain spends in each model state after the burn-in period is removed. We ran 100,000 iterations and removed the first 10,000 as burn-in. By using these posterior model probabilities, which are given in Table 4.2 below, together with the MCMC output produced by running our chain, which provides a (dependent) sample from our desired posterior distribution, we shall be able to use a model averaging procedure to obtain samples from the posterior predictive distribution of our model for predictive purposes.

We now give the table of posterior model probabilities for each dive-type. Notice that there is a slight increase in the weighting of more complex models with an increase in saturation pressure. This seems reasonable as on the longer ascents, we would expect different information at different points

in the dive. At the start of the ascent, it is the shorter, faster compartments that control the DCS risk. Meanwhile, towards the end of the dive as the surface is reached, the short compartments will have completed offgassing and the longer compartments will be the ones still retaining nitrogen. In the bounce-type dives, on the other hand, there will be no time for any slower compartments to become saturated with nitrogen, and so these are superfluous to requirements. We ran the chain 8 times (each chain taking approximately a week to run), starting at $1, 2, \ldots, 8$ compartments. We set 8 as the maximum number of compartments allowed and 1 as the minimum. Once all chains were run, we averaged the times spent in each model state over the 8 chains to obtain the results in Table 4.4. We also obtained between-chains standard errors and these are also reported in Table 4.4. We can see that in general there is (relatively) more variation in the 1, 2, 7 and 8 compartment models, as we found in the simulation. In any case, the absolute amount of time the chain spends in these states is small for all the chains.

In Sections 6.2 and 6.3, we shall consider how to use the posterior model probabilities in Table 4.4 together with the outputs from the chains themselves to predict the occurrence of DCS using a posterior predictive distribution.

Number of Compartments	Bounce	Sub-saturation	Saturation
1	0.00153 (0.02375)	0.00110 (0.00890)	0.00093 (0.00897)
2	0.02840 (0.04225)	0.02951 (0.01495)	0.00874 (0.00830)
3	0.34457 (0.16702)	0.30179 (0.20214)	0.31478 (0.15853)
4	0.41757 (0.11347)	0.38858 (0.16539)	0.33314 (0.12362)
5	0.18644 (0.12507)	0.23425 (0.08163)	0.28622 (0.09323)
6	0.01487 (0.02382)	0.03374 (0.02290)	0.03987 (0.02469)
7	0.00541 (0.00166)	0.00893 (0.00471)	0.01245 (0.02034)
8	0.00121 (0.00060)	0.00212 (0.01590)	0.00387 (0.01907)

Table 4.4: Posterior model probabilities (between chains standard errors) for Reversible Jump MCMC algorithm for the first QinetiQ dataset

Chapter 5

Bubble Score Based Methods

5.1 Introduction and Description of Data

This chapter concerns a different empirical model of the decompression process and subsequent prediction of DCS - as opposed to the more mechanistic, physical considerations of Chapter 3 - and makes use of the Second QinetiQ dataset (as outlined in Section 1.7). Recall from Section 1.9 that the data provided include: the identity of the subject in question (some subjects were used more than once) and whether the subject was observed to have any signs of DCS. In addition to this, the following key covariates were also included: the saturation depth used (i.e. the pressure of oxygen with which the subjects' bodies were saturated) and the depth from which subjects were subjected to escape conditions (where this was the type of dive undertaken). Finally, the covariates of key interest in the analysis that follows are the Kisman-Masurel (KM) bubble scores, given on a subjective ordinal scale from 0 to 4 in increments of one third (excluding one third itself, so that the scale is: 0,1-,1,1+,2-,2,2+,3-,3,3+,4-,4), indicating the approximate frequency and number of bubbles passing the subject's heart at the 25 timepoints (2, 5, 15, 30, 45, 60, 75, 90, 105, 120, 150, 180, 210, 240, 270, 300, 330, 360, 390, 420, 450, 480, 540, 600 and 1440 minutes)

after surfacing. It should be noted that some of these timepoints are extremely sparsely populated with bubble scores. For the purposes of our analysis, we recode these bubble scores as being on a scale of 1-12, with 1 corresponding to a KM score of 0 and 12 corresponding to a KM score of 4. Also provided was the Kisman Integrated Bubble Score (KISS), which is designed to be a weighted average of these various ordinal scores, accounting for the fact that the ordinal bubble score is in fact on a power-scale (bubble grade not being linear in bubble quantity), and aims to provide a continuous assessment of bubble severity, as noted by Tikuisis and Gerth (2002). More information on this follows in Section 5.2.

Unfortunately, there are a considerable number of missing values in the data set. Although the covariates are complete, so that we always know the saturation and escape depths of any dive, we do not know every bubble score for every timepoint. There are only very few observations present for some timepoints (notably at each of 210, 270, 330, 390, 450, 540, 600, 1440 minutes there are fewer than 25 observations of bubble scores). Initially, we will therefore discount these timepoints entirely as providing very little useful data and only serving to complicate our analysis and increase the computational time required considerably. For each of the remaining timepoints, there are at least 90 observations of bubble scores available and so these are liable to contain at least some usable information. In order to make the best use of the available information, we consider a method for imputing (i.e. attempting to infer through statistical methods) the missing values so as to provide one or more imputed datasets for further analysis. We outline below the method by which this is achieved (being of course only one such possible method), relying primarily on Anderson and Pemberton (1985). The method therein has the advantage of requiring only 1- and 2-way margins of the joint distribution of bubble scores across time and will also yield estimates of the correlation between pairs of bubble scores. This is useful to us as, despite the relative sparsity of bubble scores at some timepoints, the 1- and 2-way margins are still generally data-rich and thus

estimation using these margins will be easier than requiring that each datapoint have a complete set of bubble scores. We therefore make the best use of such pairs of data where they are available.

We shall initially focus on the timepoints at $5, 15, 30$ and 60 minutes as the last three of these are especially well observed, with the timepoint at 5 minutes missing around half of its observations (see Table 5.1 for further details). This would seem to provide a good basis for testing as we have observed the vast majority of the responses here. We will see later (in Section 5.4) that there are very sensible-looking correlations between the bubble scores at the various points, indicating that more distant bubble scores are less related to one another, while more proximate bubble scores are more closely associated. This would seem to be to be apt as the number of bubbles at nearby timepoints will reasonably be similar. We will then proceed to consider more timepoints in our second analysis, which is to say those at $5, 15, 30, 60, 90, 120, 180, 240, 300, 360, 420, 480$ minutes. As noted above, each of these 12 timepoints contains at least 90 observed bubble score observations, while the omitted 13 timepoints contain rather fewer.

While it is difficult to see the missingness pattern for 12 timepoints (given the sheer number of possible combinations), it is much easier to do so for our initial 4 timepoints and these are presented in Table 5.2. It is also useful to look at the proportion of missing data in each of the timepoints with a reasonable number of observations by DCS outcome. This is presented in Table 5.3. We can see here that while both the DCS and no-DCS groups have an increasing level of missingness as time progresses, there are distinctly more datapoints missing when DCS occurred than when it did not. This would seem to indicate that the data are Missing Not at Random (i.e. that whether or not the datapoint is missing depends on the value of the dependent variable) as it would seem to be more likely that data are missing when DCS occurs than when it does not. In addition, there would

seem to be a trend for an increasing proportion of missing data with increasing time. Despite the fact that the data are perhaps Missing Not at Random, Schafer and Graham (2002) indicate that it is not inappropriate to proceed with an imputation procedure, providing a Bayesian or likelihood based imputation method is used, and this is what we shall discuss in Section 5.3.

Finally, we shall aim to see whether there is a link between bubble scores at given timepoints and the occurrence or non-occurrence of DCS. We shall also consider whether KISS is indeed a good summary of bubble scores and whether this is any more useful as a predictor of DCS, or whether the information provided by saturation depth and escape depth dwarfs these effects. If this is the case, it may be that while bubble scores would be useful if we could in fact measure *all* the bubbles in the body, some bubbles may become trapped in various parts of the body, leading to their not being accounted for in the bubble scores, as this *only* considers bubbles passing the heart.

5.2 The Logistic Regression Model

We shall now seek to use a logistic regression model as outlined in Section 1.8 to model these data. In our case, the set of predictors will comprise what we shall later term continuous bubble concentrations at various timepoints. We will also consider models where we use escape and saturation depth as predictors together with KISS, defined in Jankowski, Nishi and Eaton (1997) as

$$KISS = \frac{100}{4^v(t_k - t_1)} \sum_{i=1}^{k-1} \frac{(t_{i+1} - t_i)(d_{i+1}^v + d_i^v)}{2}$$

where k is the highest timepoint index, t_j is the value in minutes of the jth timepoint, $v = 3$ is an adjustment to account for the non-linearity of the

	Proportion of missing data
2 minutes	91.18%
5 minutes	54.27%
15 minutes	7.18%
30 minutes	10.54%
45 minutes	88.64%
60 minutes	11.81%
75 minutes	90.45%
90 minutes	54.91%
105 minutes	91.36%
120 minutes	21.55%
150 minutes	89.55%
180 minutes	30.27%
210 minutes	98.18%
240 minutes	37.00%
270 minutes	99.45%
300 minutes	45.45%
330 minutes	99.45%
360 minutes	51.00%
390 minutes	99.73%
420 minutes	67.54%
450 minutes	99.73%
480 minutes	73.36%
540 minutes	98.91%
600 minutes	99.91%
1440 minutes	98.82%

Table 5.1: Proportion of missing data in the second QinetiQ dataset by timepoint

5 minutes	15 minutes	30 minutes	60 minutes	Proportion of Observations
1	1	1	1	37.55%
0	1	1	1	43.73%
1	0	1	1	1.36%
1	1	0	1	0.18%
1	1	1	0	2.64%
0	0	1	1	1.91%
0	1	0	1	2.91%
0	1	1	0	1.55%
1	0	0	1	0.18%
1	0	1	0	0.09%
1	1	0	0	1.45%
0	0	0	1	0.36%
0	0	1	0	0.64%
0	1	0	0	2.82%
1	0	0	0	2.27%
0	0	0	0	0.36%

Table 5.2: Missingness patterns for the 5, 15, 30 and 60 minute timepoints (0 indicates an observation is absent while 1 indicates it is present)

ordinal bubble score as a measure of the number of bubbles present and d_j is the ordinal bubble score at the jth timepoint. We will also consider a possible interaction between escape and saturation depth. To maximize the use of the available data and inferences that can be made from it, we will seek to impute the bubble scores that are missing in the dataset - a process which we now explain.

	Proportion of missing data (No DCS)	Proportion of missing data (DCS)
5 minutes	45.55%	51.89%
15 minutes	3.09%	24.32%
30 minutes	2.82%	45.95%
60 minutes	2.82%	53.51%
90 minutes	43.82%	65.95%
120 minutes	10.18%	67.57%
180 minutes	19.82%	62.16%
240 minutes	25.55%	68.11%
300 minutes	32.36%	77.84%
360 minutes	38.09%	76.76%
420 minutes	53.18%	85.41%
480 minutes	59.45%	82.70%

Table 5.3: Proportion of missing data in the second QinetiQ dataset by timepoint and DCS occurrence

5.3 Imputation Model

As mentioned in Section 1.10, we consider the idea that some suitable joint model underlies the data and that the observed ordinal bubble score at each timepoint is derived from a latent continuous variable. The model that we use here for the underlying variable is drawn from Anderson and Pemberton (1985) who consider a multivariate normal underlying joint distribution as, while we could use a distribution more complex than the multivariate normal distribution, we may draw quickly and efficiently from this distribution which, as we will see, will be an important factor. We denote the observed ordinal bubble score at timepoint i for the jth dive on our 1-12 scale by y_{ij}. The existence of an underlying continuous variable Z_i for each timepoint i, which we term the bubble concentration, would seem to be a reasonable model given that the number of bubbles passing the heart at a particular time, while not necessarily able to be categorised by listeners into anything more than grades, would in fact be a specific number of bubbles

on a considerably finer scale. The method, then, supposes that conditional on the covariates $\boldsymbol{x}_j$ for a particular dive j, the bubble concentrations are drawn from a multivariate normal distribution $\boldsymbol{Z}_j \sim N(\boldsymbol{\beta x}_j, \boldsymbol{R})$. Here, if we suppose that we have a total of m timepoints and p covariates (where in our case $p = 2$ as we are using escape depth and saturation depth as the two covariates), $\boldsymbol{\beta}$ is a $m \times p$ matrix of regression coefficients. Further, $\boldsymbol{R}$ is a suitable $m \times m$ covariance matrix describing the way in which the observations at different pairs of timepoints are related to one another. Note that this matrix must be symmetric (being a covariance matrix), so that in computing these correlations, we need not in fact estimate all m^2 elements of $\boldsymbol{R}$.

Now, let the number of possible levels of the rth timepoint's ordinal variable be denoted by k_r. In our case, we have $k_r = 12$ for each timepoint as there are 12 possible bubble grades for every timepoint. Suppose also that there are $k_r - 1$ cutpoints $\alpha_{1,i}, \alpha_{2,i}, \ldots, \alpha_{k_r-1,i}$ such that for each timepoint i and the values of the new continuous variable between the pairs of cutpoints determine to which ordinal bubble score our continuous variable corresponds. Suppose, for example, that at timepoint 1 our continuous bubble *concentration* lies in the interval $[\alpha_{5,1}, \alpha_{6,1}]$. Then this corresponds to an ordinal bubble score of 6 on our scale (equivalent to a KM score of 2). We also suppose that there are two additional 'cutpoints' $\alpha_0 = -\infty$ and $\alpha_{k_r} = \infty$, so that, in general, an ordinal bubble score of q on our scale corresponds to the bubble concentration lying in the interval $[\alpha_{q-1}, \alpha_q]$.

In order to identify the location and scale of Z, which are not generally unique, we must fix two parameters in each dimension. We choose to fix the cutpoint $\alpha_1 = 0$ for each timepoint, as well as setting the diagonal entries of $\mathbf{R}$ to 1, so that $\mathbf{R}$ is a correlation matrix. Note that $\mathbf{R}$ must be symmetric and have 1s along its leading diagonal (being a correlation matrix) so that, in computing these correlations, we need not in fact find all m^2 values, which is useful for our computations. To actually estimate

the remaining coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ we use maximum likelihood based on the 1-way margins of the joint distribution of $\mathbf{y}$. It follows that for the ith timepoint of the jth dive with covariates $\mathbf{x}_j$, the probability that an observation of ordinal score y_{ij} is observed is given by

$$P(y_{ij}) = \Phi(\alpha(i, y_{ij}) + \boldsymbol{\beta}(i)\boldsymbol{x}_j) - \Phi(\alpha(i, y_{ij} - 1) + \boldsymbol{\beta}(i)\boldsymbol{x}), \tag{5.1}$$

where as previously noted the observed KM bubble scores have been converted to lie on a 1-12 scale, (12 representing a KM score of 4.00 and 1 representing a KM score of 0.00), where Φ is the standard normal CDF and where $\boldsymbol{\beta}(i)$ represents the ith row of $\boldsymbol{\beta}$.

Using (5.1) for each of the n dives, taking logs and summing, we may now construct a log likelihood

$$l = \sum_{i=1}^{k_r} \sum_{j=1}^{n} \log P(y_{ij}) \tag{5.2}$$

and maximizing (5.2) leads us to maximum likelihood estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Having obtained these maximum likelihood estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we can now similarly find estimates for the components of $\boldsymbol{R}$ by constructing a similar log likelihood instead based on the 2-way margins (i.e. the joint probability that at the ith timepoint we have a bubble score of y_{ij} while at the kth timepoint we have a bubble score of y_{kj}). Using equation (5) of Anderson and Pemberton (1985), which states that

$$P(y_{ij}, y_{kj}) = \int_{\alpha(i,y_{ij}-1)+\boldsymbol{\beta}(i)\boldsymbol{x}_j}^{\alpha(i,y_{ij})+\boldsymbol{\beta}(i)\boldsymbol{x}_j} \int_{\alpha(k,y_{kj}-1)+\boldsymbol{\beta}(k)\boldsymbol{x}_j}^{\alpha(k,y_{kj})+\boldsymbol{\beta}(k)\boldsymbol{x}_j} f(\zeta, \eta, \rho_{ik})\, d\zeta\, d\eta,$$

where h is the standardised bivariate normal PDF given by

$$f(x, y, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 + y^2 - 2\rho xy}{2(1-\rho^2)}\right),$$

we may construct the new log likelihood

$$l_2(\boldsymbol{\rho}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{k_r} \sum_{\substack{k=1 \\ k \neq i}}^{k_r} \sum_{j=1}^{n} \log P(y_{ij}, y_{kj}). \tag{5.3}$$

Proceeding as before by maximization of this new likelihood (5.3) (which, it should be noted, is a conditional likelihood given the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$), we may obtain estimates of the components of $\boldsymbol{R}$, which are the correlations between bubble concentrations at different timepoints. Reasonably, we might expect these to be fairly high for neighbouring timepoints, as it is likely that bubble scores at nearby timepoints will be closely related. Having estimated $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (by maximizing the likelihood (5.2)) and $\boldsymbol{R}$ (by maximizing the likelihood (5.3) conditional on our estimates for $\boldsymbol{\alpha}$and $\boldsymbol{\beta}$), we must now consider how to impute our missing values. Splitting the data for the ith timepoint and jth dive with corresponding covariates $\boldsymbol{x}_j$ into observed $y_{obs,j}$ (where the original bubble score is observed) and unobserved $y_{unobs,j}$ (where the original bubble score is unobserved) sections, with corresponding bubble concentrations $z_{obs,j}$ and $z_{unobs,j}$. Thence, for each dive j where the ith bubble score is observed and takes value y_{ij}, we have that

$$Z_{ij}|y_{ij}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{x}_j \sim N(\boldsymbol{\beta}(i)\boldsymbol{x}_j, \boldsymbol{R})|z_{obs,j} \in [\alpha_{y_{ij}-1}, \alpha_{y_{ij}}], \tag{5.4}$$

where the above notation indicates that Z_{ij} conditional on $y_{ij}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{x}_j$ is distributed as a truncated normal distribution with mean $\boldsymbol{\beta}(i)\boldsymbol{x}_j$, variance $\boldsymbol{R}$ and truncated to the interval $[\alpha_{y_{ij}-1}, \alpha_{y_{ij}}]$.

We then wish to calculate the expectation

$$E[Z_{unobs,j}|y_{obs,j}, \boldsymbol{x}_j] \tag{5.5}$$

to form our mean imputation of bubble concentration for the unobserved concentrations at the given timepoint.

To do this, we use a Gibbs sampler, which involves the full conditional distributions of each datapoint. That is, for a particular datapoint (with given covariates of saturation and escape depth) then given the bubble score at every other timepoint, we may find the conditional distribution for the latent bubble concentration at the remaining timepoint and subsequently sample from this using a Gibbs sampler to give us an estimate of the latent bubble concentration on our new continuous scale. The desired conditional distribution will depend on whether we have observed the original bubble score at the given timepoint. If we have, then we draw from a truncated normal, truncated at the desired cutpoints (after adjustment for the covariates) with mean and variance being standard results as given in, for example, Gelman et al. (2003), for the required conditional distributions. We use a truncated normal distribution in this case as we know that the continuous bubble concentration must lie between a given pair of cutpoints. On the other hand, if we have a missing value initially then we must draw from a (standard, non-truncated) normal distribution, again with mean and variance $\boldsymbol{\beta}(i)\boldsymbol{x}_j$ and $\boldsymbol{R}$ respectively as given in (5.4) (but this time without the truncation requirement for the distribution) as we do not know the original datavalue and have no bounds for the possible value of the continuous variable. We then iterate the process, cycling through each of the timepoints for a given observation a number of times until we appear to have achieved convergence in our Gibbs sampler (this turns out to require only a few burn-in iterations). Having done this, we then average the Gibbs sampler outputs over all the post burn-in iterations and perform a mean imputation using (5.5). We may then optionally back-transform

using our cutpoints (adjusted for the covariates) onto the original ordinal KM scale to provide an ordinal imputation if this is desired. In either case an initial value for the continuous bubble concentration is required. In the case of a non-missing value, we use the midpoint of the cutpoints of the interval in which the continuous score must lie. In the case of a missing value, we use 0 as our initial value.

Note that while we could treat the above parameters as missing and then incorporate them into our Gibbs sampler, this would necessitate our running the sampler every time we wanted to find the continuous bubble score for a new dive - including when we wanted to make a prediction for new covariate values. As we shall see in Section 6.3, this would be highly undesirable in a time-critical rescue situation and this is why we have chosen the above method instead.

5.4 Results

In this section, we first consider the results of applying the model outlined in the previous section. We begin by considering results from using only the 4 timepoints at $5, 15, 30$ and 60 minutes. In this case, we obtain the following maximum likelihood estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and we also give the estimated correlation matrix $\boldsymbol{R}$ between the bubble concentrations at the different timepoints. Note that the first column of $\boldsymbol{\beta}$ corresponds to saturation depth while the second corresponds to escape depth. Also note that the models are only valid over the range in which we have data (i.e. for saturation depths between around 0 and 15m and escape depths between 0 and 300m). In the results for $\boldsymbol{\alpha}$ below, we omit the zeroth column (which is -1×10^{10} for all timepoints) and the final column (which is 1×10^{10} for all timepoints).

Thus, for $\boldsymbol{\alpha}$ we obtain the estimates

$$\hat{\boldsymbol{\alpha}} = \begin{pmatrix} 0 & 0.0002 & 0.412 & 0.430 & 0.430 & 0.871 & 0.902 & 1.204 & 1.881 & 2.532 & 2.772 \\ 0 & 0.113 & 0.716 & 0.820 & 0.865 & 1.639 & 1.649 & 2.018 & 2.659 & 3.427 & 3.612 \\ 0 & 0.024 & 0.643 & 0.712 & 0.722 & 1.489 & 1.514 & 1.897 & 2.714 & 3.559 & 3.798 \\ 0 & 0.055 & 0.821 & 0.849 & 0.870 & 1.671 & 1.690 & 2.047 & 2.868 & 3.769 & 4.085 \end{pmatrix}.$$

Next, we have

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} -0.100 & -0.00739 \\ -0.186 & -0.00885 \\ -0.222 & -0.00769 \\ -0.259 & -0.00622 \end{pmatrix}.$$

Finally, for our estimated correlation matrix, we have

$$\hat{\boldsymbol{R}} = \begin{pmatrix} 1.000 & 0.843 & 0.771 & 0.704 \\ 0.843 & 1.000 & 0.882 & 0.789 \\ 0.771 & 0.882 & 1.000 & 0.877 \\ 0.704 & 0.789 & 0.877 & 1.000 \end{pmatrix}.$$

We see here that the coefficients $\boldsymbol{\beta}$ are all negative. This indicates that when saturation is higher, or when escape occurs at deeper depths, we expect higher mean bubble concentrations. That is, it is easier for a given dive to fall into a higher category, so that, as we would certainly expect, dives that have higher saturation or escape depths are more likely to produce higher bubble scores. Also, our estimated correlation matrix shows that closely related timepoints have similar bubble scores in an almost conditionally independent manner (so that at each step further away we square the correlation to obtain the next nearest correlation). This would also seem to be a promising result as we would expect close timepoints to have similar bubble scores, with more distant timepoints having less similar scores.

Now, using these results to impute our missing values, while also converting the datavalues to continuous bubble concentrations, we find that the median bubble concentrations for the $5, 15, 30, 60$ timepoints after imputation are 2.027, 2.872, 2.496 and 2.477 respectively. Comparing with $\hat{\boldsymbol{\alpha}}$ above, we have that these correspond to values of 10, 12, 10 and 10 on our ordinal scale (or 3+, 4, 3+ and 3+ on the KM scale). Comparing with the median for the original unimputed data (ignoring missing values), where the medians are 3+, 3, 3, 3 respectively, it would seem that imputation has indicated that the bubble scores that are missing are, in general, higher than those that are observed. This implies that we do not have data that are Missing Completely at Random (MCAR). That is, the values of the *other* covariates may not be independent of whether the covariate values are missing or not. As higher bubble scores are more prevalent in dives with higher scores, an analysis based on only the complete cases may well lead to misleading results.

Having completed the imputation, we would ideally use the Bayesian framework to perform our logistic regression. However, as QinetiQ wishes to compare and integrate this work with their current work - which is undertaken under the maximum likelihood framework - our results here shall also be undertaken within that framework. In order not to lose our Bayesian perspective, however, we shall also outline in Section 5.7 the Bayesian approach that could be used to obtain these regression results and obtain DCS predictions.

We may use our imputed dataset and then apply the method of scoring to find the maximum likelihood estimates for a logistic regression of the DCS response on the continuous bubble concentration at each timepoint. Note that the null deviance for all the models below (with the exception of the KISS model as this contains fewer datapoints) is 900.21 on 985 degrees of freedom. This yields the results in Table 5.4. This model has a residual deviance of 571.54 on 981 degrees of freedom and an AIC_c of 581.54. We

	Estimate	Std. Error
Intercept	-5.2236	0.3592
5 minutes	-1.0981	0.2770
15 minutes	0.7617	0.4542
30 minutes	0.1055	0.5130
60 minutes	0.9500	0.2760

Table 5.4: Estimates for regression model after imputation using only bubble scores

find that the model can be improved by including saturation and escape depth as covariates. Doing this, we obtain the results in Table 5.5 (note that the final row represents the interaction effect between the saturation and escape depths): This model has a residual deviance of 505.81 on 978

	Estimate	Std. Error
Intercept	-6.7191	0.8335
5 minutes	0.3750	0.3522
15 minutes	-0.3170	0.4831
30 minutes	0.4623	0.5344
60 minutes	-0.0457	0.3219
Saturation Depth	0.4494	0.0906
Escape Depth	0.0086	0.0038
Saturation Depth:Escape Depth	-0.0010	0.0003

Table 5.5: Estimates for regression model after imputation using bubble scores, saturation depth, escape depth and their interaction

degrees of freedom and an AIC_c of 521.81. We see then that the AIC_c has dropped significantly with the inclusion of the three extra parameters and indeed none of the bubble concentrations are significant predictors when taken together. This would seem to indicate that any information they give is perhaps provided by a smaller subset of these timepoints. Before we consider selecting the 'best' subset of variables for predictive power, we first give an indication of how to predict from this model, should we wish to do so.

We now consider a worked example. Suppose we have a dive with bubble scores of $4, 3, 2.66, 2$ on the KM scale at timepoints 5,15,30 and 60 minutes respectively, with a saturation depth of 5m and an escape depth of 200m. The first thing we must do is convert our observed bubble scores to continuous bubble concentrations for use with our models. To achieve this, we can approximate using the midpoint of a suitable pair of the cutpoints $\hat{\boldsymbol{\alpha}}$, depending on which timepoint we are considering (with the exception of bubble scores of 0 or 4, for which we must adjust slightly, given the need for artificially large end cutpoints). Now, to convert our bubble scores to bubble concentrations, we consider each row of $\hat{\boldsymbol{\alpha}}$ in turn. We consider the second, third and fourth scores first. For the second score, which is 3 on the KM scale (or 9 on our 1-12 scale), we examine the second row of $\boldsymbol{\alpha}$ and take the midpoint of the interval $[\alpha_{2,8}, \alpha_{2,9}]$ as our bubble concentration estimate. This gives 2.339 as our estimate. For the third score, which is 2.66 on the KM scale (8 on our 1-12 scale), we look at the third row of $\hat{\boldsymbol{\alpha}}$ but now take the midpoint of the interval $[\alpha_{3,7}, \alpha_{3,8}]$ as our bubble concentration estimate, giving the estimate 1.706. Finally, for the fourth score, which is 2 on the KM scale (6 on our 1-12 scale), we look at the fourth row of $\hat{\boldsymbol{\alpha}}$ but now take the midpoint of the interval $[\alpha_{4,5}, \alpha_{4,6}]$ as our bubble concentration estimate, giving the estimate 1.2705. To obtain the estimate for the first timepoint, we must follow a slightly different procedure as the observed score is 4 - we cannot take the midpoint of $[\alpha_{1,11}, \alpha_{1,12}]$ as we recall that the latter is artificially large (1×10^{10}). Thus, instead, we simply add 0.3 to the cutpoint $\alpha_{1,11}$ (giving 3.072) as this is similar to the average distance between cutpoints. Note that this only provides a rough and ready estimate (as does using the midpoint of each pair of cutpoints in general) as usually all the continuous bubble concentrations would be imputed using a run of our Gibbs sampler that takes into account the bubble scores measured during the dive. Thus, we obtain the following (where the

coefficients have been rounded to 3 d.p.)

$$\begin{aligned} P(DCS) &= \text{invlogit}\big(-6.719 + 0.375 \times 3.072 - 0.317 \times 2.339 \\ &\qquad + 0.463 \times 1.706 - 0.046 \times 1.2705 + 0.449 \times 5 \\ &\qquad + 0.008 \times 200 - 0.001 \times 5 \times 200\big) \\ &\approx \text{invlogit}(-2.73) \\ &= \frac{e^{-2.73}}{1 + e^{-2.73}} \\ &= 0.0612, \end{aligned}$$

where the inverse logit function is given by $\text{invlogit}(x) = \exp(x)/(1 + \exp(x))$.

By following the procedure outlined in Collett (2002, p. 98-99), we may also find a confidence interval for this prediction. To do this, we must first obtain from R the overall standard error of the fitted value - s, say - and in this case, this is given by 0.6415. Collett then gives that if our fitted value on the logit scale is $\hat{y}$, say, then a suitable 95% confidence interval on the logit scale is given by $(\hat{y} - 1.96s, \hat{y} + 1.96s)$, which in this case evaluates to (-3.987,-1.473). Further, he asserts that this may be back transformed to the original scale to give a confidence interval for the fitted probability of DCS and in this case this evaluates to (0.0182,0.1865).

Next, using both forward and backward stepwise selection with decreasing AIC_c as our selection criterion, the model we obtain is shown in Table 5.6. This model has only a slightly larger residual deviance of 507.03 on 980 degrees of freedom and an AIC_c of 517.03. We have thus improved on our model by considering only those predictors that seem to encapsulate the most information.

We also have the results for the 12 timepoint model. In this case, we again omit for $\boldsymbol{\alpha}$ the zeroth column (which is -1×10^{10} for all timepoints) and

	Estimate	Std. Error	z-value	p-value
Intercept	-6.6663	0.8108	-8.222	$< 2 \times 10^{-16}$
30 minutes	0.4245	0.1357	3.128	0.0018
Saturation Depth	0.4131	0.0814	5.074	3.90×10^{-7}
Escape Depth	0.0083	0.0037	2.279	0.0227
Saturation Depth:Escape Depth	-0.0009	0.0003	-2.988	0.0028

Table 5.6: Estimates for regression model after imputation using bubble scores, saturation depth, escape depth and their interaction after model selection by AIC_c

the final column (which is 1×10^{10} for all timepoints). Here then, for $\hat{\boldsymbol{\alpha}}$ we have

$$\hat{\boldsymbol{\alpha}} = \begin{pmatrix} 0 & 0.0002 & 0.412 & 0.430 & 0.430 & 0.870 & 0.902 & 1.204 & 1.881 & 2.532 & 2.772 \\ 0 & 0.113 & 0.716 & 0.820 & 0.865 & 1.639 & 1.649 & 2.018 & 2.659 & 3.427 & 3.612 \\ 0 & 0.024 & 0.643 & 0.712 & 0.722 & 1.489 & 1.514 & 1.897 & 2.714 & 3.559 & 3.798 \\ 0 & 0.055 & 0.821 & 0.849 & 0.870 & 1.671 & 1.690 & 2.047 & 2.868 & 3.769 & 4.085 \\ 0 & 0.051 & 0.590 & 0.630 & 0.630 & 1.461 & 1.522 & 1.934 & 2.844 & 3.792 & 4.093 \\ 0 & 0.130 & 0.775 & 0.842 & 0.866 & 1.510 & 1.543 & 1.953 & 2.703 & 3.622 & 3.946 \\ 0 & 0.224 & 0.996 & 1.039 & 1.058 & 1.683 & 1.683 & 2.132 & 3.024 & 4.026 & 4.447 \\ 0 & 0.313 & 0.919 & 0.990 & 1.027 & 1.671 & 1.686 & 2.207 & 3.148 & 4.048 & 4.730 \\ 0 & 0.257 & 0.797 & 0.926 & 0.954 & 1.760 & 1.777 & 2.349 & 3.302 & 4.077 & 4.594 \\ 0 & 0.251 & 0.805 & 0.938 & 0.972 & 1.841 & 1.854 & 2.404 & 3.294 & 3.867 & 4.813 \\ 0 & 0.132 & 0.630 & 0.749 & 0.775 & 1.748 & 1.801 & 2.192 & 3.334 & 3.859 & 4.380 \\ 0 & 0.156 & 0.794 & 0.979 & 1.013 & 1.929 & 1.973 & 2.387 & 3.178 & 3.554 & 4.909 \end{pmatrix}.$$

For $\hat{\boldsymbol{\beta}}$, we obtain

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} -0.100 & -0.00739 \\ -0.186 & -0.00885 \\ -0.222 & -0.00769 \\ -0.259 & -0.00622 \\ -0.281 & -0.00251 \\ -0.232 & -0.00407 \\ -0.217 & -0.00398 \\ -0.190 & -0.00368 \\ -0.170 & -0.00317 \\ -0.157 & -0.00236 \\ -0.103 & -0.00218 \\ -0.128 & -0.00140 \end{pmatrix}$$

while for the correlation matrix we have

$$\hat{\mathbf{R}} = \begin{pmatrix} 1.000 & 0.843 & 0.771 & 0.704 & 0.562 & 0.591 & 0.505 & 0.408 & 0.438 & 0.320 & 0.229 & 0.248 \\ 0.843 & 1.000 & 0.882 & 0.789 & 0.400 & 0.629 & 0.595 & 0.540 & 0.494 & 0.364 & 0.086 & 0.044 \\ 0.771 & 0.882 & 1.000 & 0.877 & 0.671 & 0.739 & 0.669 & 0.652 & 0.605 & 0.537 & 0.460 & 0.525 \\ 0.704 & 0.789 & 0.877 & 1.000 & 0.828 & 0.850 & 0.740 & 0.650 & 0.604 & 0.564 & 0.571 & 0.525 \\ 0.562 & 0.400 & 0.671 & 0.828 & 1.000 & 0.810 & 0.700 & 0.600 & 0.550 & 0.426 & 0.413 & 0.486 \\ 0.591 & 0.629 & 0.739 & 0.850 & 0.810 & 1.000 & 0.875 & 0.777 & 0.698 & 0.677 & 0.665 & 0.690 \\ 0.505 & 0.595 & 0.669 & 0.740 & 0.700 & 0.875 & 1.000 & 0.895 & 0.835 & 0.783 & 0.760 & 0.653 \\ 0.408 & 0.540 & 0.652 & 0.650 & 0.600 & 0.777 & 0.895 & 1.000 & 0.915 & 0.845 & 0.808 & 0.671 \\ 0.438 & 0.494 & 0.605 & 0.604 & 0.550 & 0.698 & 0.835 & 0.915 & 1.000 & 0.927 & 0.871 & 0.775 \\ 0.320 & 0.364 & 0.537 & 0.564 & 0.426 & 0.677 & 0.783 & 0.845 & 0.927 & 1.000 & 0.905 & 0.812 \\ 0.229 & 0.086 & 0.460 & 0.571 & 0.413 & 0.665 & 0.760 & 0.808 & 0.871 & 0.905 & 1.000 & 0.893 \\ 0.248 & 0.044 & 0.525 & 0.525 & 0.486 & 0.690 & 0.653 & 0.671 & 0.775 & 0.812 & 0.893 & 1.000 \end{pmatrix}$$

Here, we again see that all the $\hat{\boldsymbol{\beta}}$ coefficients are negative, which is as

expected. The correlation matrix here is not quite so intuitive as in the four-point example, though bubble scores at adjacent timepoints are again generally well correlated, so that this matrix would seem to indicate that nearby timepoints have similar bubble scores.

Having obtained these results, we may now use them to impute the missing values in the dataset, and convert the ordinal bubble scores to continuous bubble concentrations, being sure to adjust for the covariates. Unfortunately, it happens that this correlation matrix is, in fact, not positive definite, and so our imputation process will fail (as we will obtain negative conditional variances, which is clearly impossible). To circumvent this issue, we note that the matrix above is somewhat akin to the correlation matrix for an autocorrelated process (i.e. one where adjacent timepoints are correlated with some correlation coefficient ρ, timepoints separated by one timepoint are correlated with correlation coefficient ρ^2 and, in general, timepoints separated by k timepoints have correlation coefficient ρ^{k-1}). We may select such a matrix that is in fact positive definite and use this to approximate $\mathbf{R}$. We choose to do this by minimizing the sum of squares of the matrix $\mathbf{R} - \mathbf{R}^*$, where $\mathbf{R}^*$ is our new matrix, as a function of ρ. This

yields the following matrix $\mathbf{R}^*$:

$$\mathbf{R}^* = \begin{pmatrix}
1.000 & 0.894 & 0.798 & 0.713 & 0.637 & 0.570 & 0.509 & 0.455 & 0.406 & 0.363 & 0.324 & 0.290 \\
0.894 & 1.000 & 0.894 & 0.798 & 0.713 & 0.637 & 0.570 & 0.509 & 0.455 & 0.406 & 0.363 & 0.324 \\
0.798 & 0.894 & 1.000 & 0.894 & 0.798 & 0.713 & 0.637 & 0.570 & 0.509 & 0.455 & 0.406 & 0.363 \\
0.713 & 0.798 & 0.894 & 1.000 & 0.894 & 0.798 & 0.713 & 0.637 & 0.570 & 0.509 & 0.455 & 0.406 \\
0.637 & 0.713 & 0.798 & 0.894 & 1.000 & 0.894 & 0.798 & 0.713 & 0.637 & 0.570 & 0.509 & 0.455 \\
0.570 & 0.637 & 0.713 & 0.798 & 0.894 & 1.000 & 0.894 & 0.798 & 0.713 & 0.637 & 0.570 & 0.509 \\
0.509 & 0.570 & 0.637 & 0.713 & 0.798 & 0.894 & 1.000 & 0.894 & 0.798 & 0.713 & 0.637 & 0.570 \\
0.455 & 0.509 & 0.570 & 0.637 & 0.713 & 0.798 & 0.894 & 1.000 & 0.894 & 0.798 & 0.713 & 0.637 \\
0.406 & 0.455 & 0.509 & 0.570 & 0.637 & 0.713 & 0.798 & 0.894 & 1.000 & 0.894 & 0.798 & 0.713 \\
0.363 & 0.406 & 0.455 & 0.509 & 0.570 & 0.637 & 0.713 & 0.798 & 0.894 & 1.000 & 0.894 & 0.798 \\
0.324 & 0.363 & 0.406 & 0.455 & 0.509 & 0.570 & 0.637 & 0.713 & 0.798 & 0.894 & 1.000 & 0.894 \\
0.290 & 0.324 & 0.363 & 0.406 & 0.455 & 0.509 & 0.570 & 0.637 & 0.713 & 0.798 & 0.894 & 1.000
\end{pmatrix}$$

We see that the matrix $\mathbf{R}^*$, where $\rho = 0.894$ (which is the minimizer for $\mathbf{R} - \mathbf{R}^*$), is fairly similar to $\mathbf{R}$ but is not identical to it. However, it would seem to be a reasonable approximation and is indeed a positive definite matrix, as we require to continue.

We now use 2,000 iterations of the Gibbs sampler, with our matrix $\mathbf{R}^*$, taking the first 100 as a burn-in and discarding them. We then impute the mean of the remaining 1,900 iterations as the desired continuous bubble concentration. We use a specific point as an example here. Notably, we take the final datapoint. This has missing values at $5, 90, 300, 360, 420, 480$ minutes, with a saturation depth of 5 and escape depth of 0 (as this is not an escape type dive). We see from the histogram in Figure 5.1, which concerns the imputation for the bubble concentration at the first timepoint at 2 minutes, that there is a continuous bubble concentration spread of around 0.53 to 2.85, with the majority of the mass occurring between about

0.9 and 2.5. Comparing with $\hat{\boldsymbol{\alpha}}$ above, this leads us to a probable value of 8, 9 or 10 for our ordinal bubble score (3-, 3 or 3+ on the KM scale), or possibly 4 or 8 (1+ and 3- respectively on the KM scale). This would seem reasonable given that the first observed score, which occurs at 15 minutes, is 3 on the KM scale. Using the mean of our Gibbs sampler after removing the burn-in iterations, we obtain a continuous concentration of 1.64, which back-transforms to a score of 9 on the ordinal scale (3 on the KM scale), which seems in order given that the score at 5 minutes is 8 (or 3 on the KM scale).

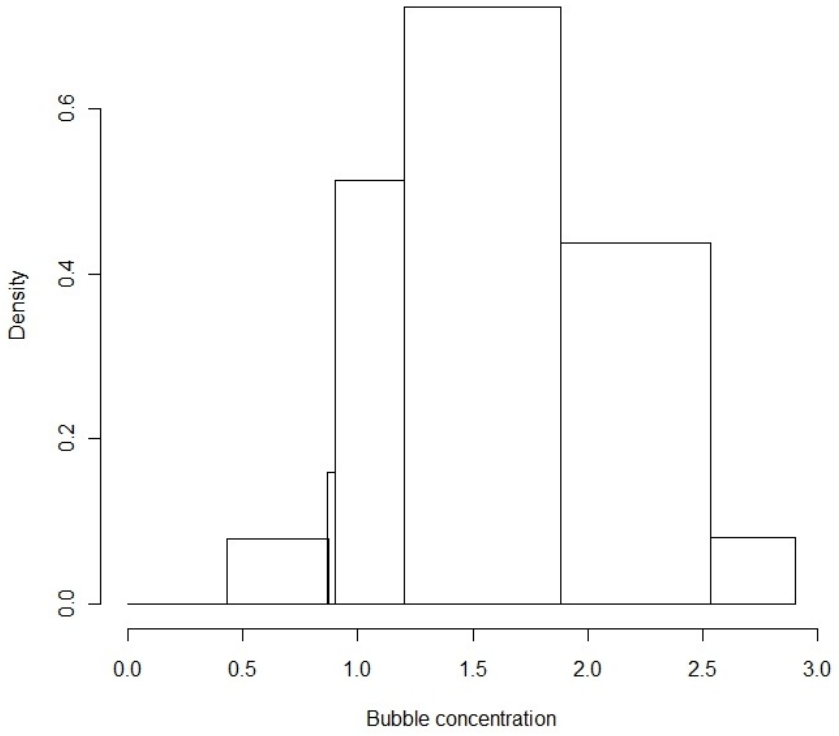

Figure 5.1: Histogram of imputation for 5 minute timepoint of dive 986

Having completed the imputation, we may again use our imputed dataset to perform a logistic regression of the DCS response on the continuous bubble concentration at each timepoint, though we should note that we

are using the data twice here (albeit for two rather different purposes, but this cannot be avoided). This yields the results in Table 5.7. This

	Estimate	Std. Error
Intercept	-4.813	0.397
5 minutes	-0.444	0.322
15 minutes	0.295	0.479
30 minutes	0.162	0.518
60 minutes	1.023	0.436
90 minutes	0.341	0.383
120 minutes	-1.201	0.428
180 minutes	0.207	0.441
240 minutes	0.428	0.443
300 minutes	-0.915	0.490
360 minutes	1.441	0.521
420 minutes	-2.296	0.588
480 minutes	1.767	0.515

Table 5.7: Estimates for regression model after imputation using only bubble scores

model has a residual deviance of 533.49 on 973 degrees of freedom and an AIC_c of 559.49. This time, we see that now there is considerable overlap in the information provided by the bubble concentrations at the various timepoints, causing a lack of significance of any of the bubble concentrations as predictors. However, the AIC_c is considerably lower than when we were just using 4 timepoints, so it is clear that a good deal of information is imparted by the later timepoints (perhaps another reason for the success of KISS, rather than the previous individual first few bubble scores), though our model here is not as good as the model using only KISS.

Suppose we now use saturation and escape depth as additional predictors. Recall that in the 4 timepoint case this provided a considerable improvement. This model, whose parameters are given in Table 5.8, has a residual deviance of 487.88 on 970 degrees of freedom with an AIC_c of 519.88. Here

	Estimate	Std. Error
Intercept	-6.756	0.858
5 minutes	0.382	0.358
15 minutes	-0.443	0.503
30 minutes	0.440	0.542
60 minutes	1.169	0.474
90 minutes	-1.258	0.538
120 minutes	-0.263	0.474
180 minutes	-0.033	0.445
240 minutes	0.654	0.465
300 minutes	-0.705	0.511
360 minutes	0.710	0.563
420 minutes	-0.654	0.619
480 minutes	0.523	0.500
Saturation Depth	0.466	0.096
Escape Depth	0.006	0.004
Saturation Depth:Escape Depth	-0.001	0.000

Table 5.8: Estimates for regression model after imputation using bubble scores, saturation depth, escape depth and their interaction

we see that the effect on the AIC_c of predictors is smaller than it was previously, but is still significant. This clearly indicates that using more timepoints has improved the amount of information obtained using more of the bubble profile, as we might expect. Suppose we now use stepwise (backward) selection on the above model. Then we obtain the results in Table 5.9. This model has a residual deviance of 493.77 on 970 degrees of freedom with an AIC_c of 507.77 - clearly an improvement on the full model above. As we can see, only a few of the timepoints are selected, so that there is an indication that some of the timepoints provide overlapping information.

For comparison, suppose we only include the effect of saturation depth, escape depth and the interaction of these two factors. Then, we have

	Estimate	Std. Error	z-value	p-value
Intercept	-6.7892	0.8172	-8.308	$< 2 \times 10^{-16}$
60 minutes	1.4884	0.3360	4.431	9.40×10^{-6}
90 minutes	-1.4819	0.3945	-3.756	0.0002
240 minutes	0.3606	0.1986	1.815	0.0695
Saturation Depth	0.4658	0.0840	5.545	2.94×10^{-8}
Escape Depth	0.0057	0.0038	1.515	0.1298
Saturation Depth:Escape Depth	-0.0009	0.0003	-2.952	0.0032

Table 5.9: Estimates for regression model after imputation using bubble scores, saturation depth, escape depth and their interaction after model selection by AIC_c

the results in Table 5.7. This model has a residual deviance of 517.24

	Estimate	Std. Error	z-value	p-value
Intercept	-7.018	0.8570	-8.189	2.64×10^{-16}
Saturation Depth	0.5481	0.0756	7.256	3.99×10^{-13}
Escape Depth	0.0138	0.0035	3.961	7.45×10^{-5}
Saturation Depth:Escape Depth	-0.0012	0.0003	-3.746	0.0002

Table 5.10: Estimates for regression model after imputation using only saturation depth, escape depth and their interaction

on 982 degrees of freedom with an AIC_c 525.24. This is clearly greater than the AIC_c for the previous model, whence we may conclude that the bubble concentrations do provide some useful information over and above the escape and saturation depths.

For practical reasons, we also consider the model using only the five minute timepoint together with the saturation and escape depths. We do this because we wish to ascertain whether the 5-minute timepoint is a useful observation to collect - it is easily and rapidly obtainable and may thus be useful in the event of an emergency. This yields the results in Table 5.8. This model has a residual deviance of 517.25 on 982 degrees of freedom and an AIC_c of 525.25. In fact, using backward selection we find that the

	Estimate	Std. Error	z-value	p-value
Intercept	-5.0080	0.4472	-11.199	$< 2 \times 10^{-16}$
5 minutes	0.5580	0.1367	4.084	4.43×10^{-5}
Saturation Depth	0.3033	0.0353	8.592	$< 2 \times 10^{-16}$
Escape Depth	-0.001227	0.0016	-0.7680	0.443

Table 5.11: Estimates for regression model after imputation using only saturation depth, escape depth and the 5-minute bubble score

escape depth term is not required here and the AIC_c may be improved by eliminating it to yield the model whose parameters are given in Table 5.12 instead. This model has a residual deviance of 517.84 on 983 degrees of free-

	Estimate	Std. Error	z-value	p-value
Intercept	-5.2123	0.3660	-14.240	$< 2 \times 10^{-16}$
5 minutes	0.4979	0.1113	4.473	7.72×10^{-6}
Saturation Depth	0.3198	0.0284	11.274	$< 2 \times 10^{-16}$

Table 5.12: Estimates for regression model after imputation with AIC_c selection, using only saturation depth and the 5-minute bubble score

dom and an AIC_c 523.84. This is better even than the model incorporating the escape depth and saturation depth together with their interaction (a model with an AIC_c of 525.24) and is certainly better than that model without the relevant interaction (which has an AIC_c of 540.88). Thus, it seems that the 5 minute timepoint does in fact provide valuable information - information that could well be useful in practical circumstances if easily obtained.

Suppose we now check the effect of using KISS rather than the imputed bubble concentrations. This requires us to first back-transform our imputed bubble concentrations onto the original ordinal scale before using these figures to calculate the KISS. Doing this, we obtain the results in Table 5.13. This model has a residual deviance of 634.28 on 984 degrees of freedom

	Estimate	Std. Error	z-value	p-value
Intercept	-3.8214	0.2195	-17.41	$< 2 \times 10^{-16}$
KISS	0.0627	0.0046	13.52	$< 2 \times 10^{-16}$

Table 5.13: Estimates for regression model after imputation using only KISS

with an AIC_c of 638.28.

We see that KISS is a less effective predictor than our model that uses a set of bubble concentrations together with the escape and saturation depths, and is less effective even than the model that uses only the bubble concentrations. It seems then that it would be helpful to find some modified version of KISS that is able to incorporate our continuous bubble concentrations, rather than requiring that the ordinal KM scale be used, as we might expect some form of KISS-type score that averages over the various timepoints and takes account of the non-linearity of bubble quantity with respect to bubble concentration to provide a more useful predictor.

Thus, it seems that perhaps it would be useful in future work to find some more flexible version of KISS that incorporates the ability to use continuous bubble concentrations, as this may be able to provide more information as a predictor than we would obtain from just using the saturation and escape depths alone, or any particular bubble score. Further, use of an integrated score would allow us to recoup some degrees of freedom that are lost if we make use of the bubble concentrations at individual timepoints.

Now, using the imputed KISS scores together with the saturation depths, we obtain Table 5.14. This model has a residual deviance of 515.49 on 981 degrees of freedom with an AIC_c of 525.49. This model is, however, not noticeably better than the model excluding KISS (and we see that the KISS regression coefficient has a rather high p-value), so we might conclude that there is no information to be gained by using the bubble scores in addition

	Estimate	Std. Error	z-value	p-value
Intercept	-6.8674	0.8531	-8.050	8.29×10^{-16}
KISS	0.0098	0.0074	1.325	0.185
Saturation Depth	0.5050	0.0812	6.216	5.10×10^{-10}
Escape Depth	0.0127	0.0035	3.591	0.0003
Saturation Depth:Escape Depth	-0.0012	0.0003	-3.597	0.0003

Table 5.14: Estimates for regression model after imputation using KISS, saturation depth, escape depth and their interaction

to the saturation and escape depths where available. However, it would, as noted previously, be interesting to see whether a more flexible version of KISS might provide a yet more substantial amount of information from the available bubble scores than does the current version, particularly if we could use our continuous concentrations to this end. Moreover, it is interesting to note that if, rather than simply using our imputed KISS scores, we use the 'true' (observed) KISS score where available and only elsewhere use the imputed scores (we label this variable KISS2), we obtain the model given in Table 5.15:

	Estimate	Std. Error	z-value	p-value
Intercept	-6.5940	0.8032	-8.210	$< 2 \times 10^{-16}$
KISS2	0.0430	0.0058	7.431	1.08×10^{-13}
Saturation Depth	0.3567	0.0741	4.812	1.49×10^{-6}
Escape Depth	0.0070	0.0034	2.073	0.0382
Saturation Depth:Escape Depth	-0.0008	0.0003	-2.692	0.0071

Table 5.15: Estimates for regression model after imputation using KISS2, saturation depth, escape depth and their interaction

This model has a residual deviance of 459.81 on 981 degrees of freedom with an AIC_c of 469.81. This is clearly better than any of our previous models. Thus, it seems that the observed KISS values provide a better source of information than the imputed KISS scores. To see why this

might be the case, we examine those datapoints for which there is a large difference between the observed KISS scores and those calculated using the imputed data. It emerges that those points with a lower observed KISS than imputed KISS are almost universally missing the observation at the 5 minute timepoint. Meanwhile, those points with a higher observed KISS than imputed KISS are generally those dives that have a great many missing values after the first two or three timepoints and have large KM scores for the first couple of timepoints. In such dives, the KISS may well be 100 or close to it, while the imputed scores, reflecting the general tendency for bubble scores to decrease over time, have a rather lower KISS as the imputed scores become smaller as time progresses (reflecting the correlation structure in the data). Given both these observations, and the fact that using the observed KISS where available seems to provide superior results, it would seem that perhaps the bubble scores at the earlier timepoints provide more information than do the bubble scores at later timepoints. It may well then be instructive to create a new version of KISS giving enhanced weight to these points, in an effort to reflect the additional information provided here.

5.5 Applying Multiple Imputation to the Bubble Dataset

We can now proceed to apply the multiple imputation procedure of Section 1.10 to our bubble dataset, taking $m = 40$ to be our number of imputations as we have under 70% missing data and so should obtain a suitably robust result. As it is of particular interest, we consider the regression given by

$$\text{logit}(P(DCS)) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_{12} x_1 x_2 + \gamma_3 x_3,$$

where x_1 is the saturation depth, x_2 is the escape depth, x_3 is the imputed variable on the concentration continuous scale, and the γs are regression coefficients. In our case, we are particularly interested in the γ_3 coefficient. Recall that this model has estimated coefficients as given in Table 5.16.

	Estimate	Std. Error	z-value	P-value
Intercept	-6.9270	0.8727	-7.938	2.05×10^{-15}
5 minutes	0.2380	0.0686	3.469	0.0005
Saturation Depth	0.5000	0.0801	6.127	8.96×10^{-10}
Escape Depth	0.0070	0.0038	1.867	0.0619
Saturation Depth:Escape Depth	-0.0010	0.0003	-3.235	0.0012

Table 5.16: Estimates for regression model after imputation using saturation depth, escape depth and their interaction together with the 5 minute bubble score

On creating 40 imputed datasets, we obtain 40 values for γ_3 all ranging between 0.22 and 0.24 with standard errors around 0.07. Using Rubin's rules given above, we have $Q = 0.2327, \overline{U} = 0.07234, B = 5.118623 \times 10^{-6}, T = 0.07234337$, so that q is very large indeed (we have $q \approx 7.5 \times 10^9$) and we may effectively compare with a normal distribution rather than a t-distribution. In this case, then, we do have evidence that our regression coefficient is significantly different from zero as

$$\frac{0.2327 - 0}{\sqrt{0.07234}} \approx 3.217,$$

which gives a p-value of 0.00145.

Clearly, then, it is better to obtain the bubble score at 5 minutes to assist shipboard medics in predicting occurrence of DCS than it is to go without it and, given that it is such an easy reading to obtain, is well worth the time to acquire as it will certainly enhance their triage and treatment objectives.

5.6 Some Validation

In order to validate the above results to some extent, we split the (imputed) second QinetiQ dataset into a training dataset (comprising 80% of the original dataset, chosen at random) and a test dataset (comprising the remaining 20% of the original dataset). We then refit the model with the same covariates as Table 5.16 using only the training dataset and then use the results to make predictions for the test dataset.

We again use the pseudo-R_2, R_2^T, defined in Section 3.6 and, when considering this restricted training dataset, we would again hope to obtain similar parameter estimates as we did using the whole dataset, as well as a similar R_T^2 value for the predictions on the test dataset.

Fitting the model to the training dataset yields the parameter estimates in Table 5.17 below. We see that, as hoped, these parameter estimates are quite similar to those obtained in Table 5.16 for the whole of the second QinetiQ dataset (and certainly within two standard errors of the previous estimates, though do note that these standard errors are now larger, as we would expect, given that we have a smaller dataset). In addition, the R_T^2 value for the test dataset using these parameter values is 0.4311. The value obtained for the whole dataset is 0.4863, and this similarity is again reassuring.

5.7 Bayesian Logistic Regression

In Section 5.4, we used maximum likelihood methods to enable QinetiQ to make comparisons directly with their own models. However, it would of course be possible to use a Bayesian framework to fit our logistic regression,

	Estimate	Std. Error	z-value	P-value
Intercept	-6.3781	1.4920	4.275	1.91×10^{-5}
5 minutes	0.2670	0.0927	2.880	0.0040
Saturation Depth	0.4174	0.1242	3.361	0.0008
Escape Depth	0.0070	0.0076	1.867	0.0619
Saturation Depth:Escape Depth	-0.0012	0.0006	-2.18	0.0292

Table 5.17: Estimates for regression model after imputation using saturation depth, escape depth and their interaction together with the 5 minute bubble score for the training dataset

and this would avoid losing the Bayesian coherence of our work. To do this, we first note that the likelihood for the logistic regression with p covariates $x_1, x_2, \ldots, x_p$ is given by

$$\prod_{i=1}^{n} \left(\frac{e^{\eta}}{1+e^{\eta}} \right)^{y_i} \left(1 - \frac{e^{\eta}}{1+e^{\eta}} \right)^{1-y_i} \tag{5.6}$$

where $\eta = b_0 + b_1 x_1 + \cdots + b_p x_p$ is the linear predictor, $y_i = 1$ if DCS occurs on the ith dive and $y_i = 0$ otherwise and $b_0, b_1, \ldots, b_p$ are some real parameters. If desired, interaction parameters can also be included.

Taking the product of the likelihood 5.6 with a suitable prior, we can (using a suitable MCMC sampler), obtain samples from the posterior distribution of the regression parameters.

As an example of how this might be achieved, and in order to compare our results with the maximum likelihood results obtained previously, we shall follow this procedure for our final model above (from Table 5.16). That is, we shall fit a model incorporating the 5 minute bubble score, as well as saturation depth, escape depth and the interaction of the last two of these, so that our linear predictor is $\eta = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_3 x_3$, where x_1 is the saturation pressure, x_2 the escape depth and x_3 the imputed bubble score. Then, we have 5 regression parameters to deal with, including our

Parameter	Approximate posterior mode (Standard error)
b_0 (Intercept)	-7.1463 (0.8561)
b_1 (Saturation Depth)	0.5181 (0.0784)
b_2 (Escape Depth)	0.0100 (0.0038)
b_{12} (Saturation Depth:Escape Depth)	-0.0011 (0.00033)
b_3 (5 minutes)	0.2625 (0.1406)

Table 5.18: Estimated posterior modes (standard errors) of regression parameters

intercept. As we have no prior knowledge about possible parameter values, we shall use a diffuse uniform prior, $U(-100, 100)$, for each of the regression parameters.

Using the rjags package in R to sample from the desired chain 100,000 times and discarding the first 10,000 as burn-in, we obtain the results given in Table 5.18 and Figure 5.2. We can see that the chains mix quite well and that the posterior modes presented here are very similar to the maximum likelihood estimates noted in Table 5.16 above, which is reassuring.

5.8 Discussion

In summary, this chapter has concerned use of bubble scores to predict DCS, as well as the imputation of those bubble scores to ensure comprehensive use of the given information, trying to avoid the complete discarding of missing data. To that end, we have outlined an imputation method by which we may impute the missing values in a dataset, using the mean of a Gibbs sampler that also converts the observed ordinal bubble data into continuous bubble concentrations.

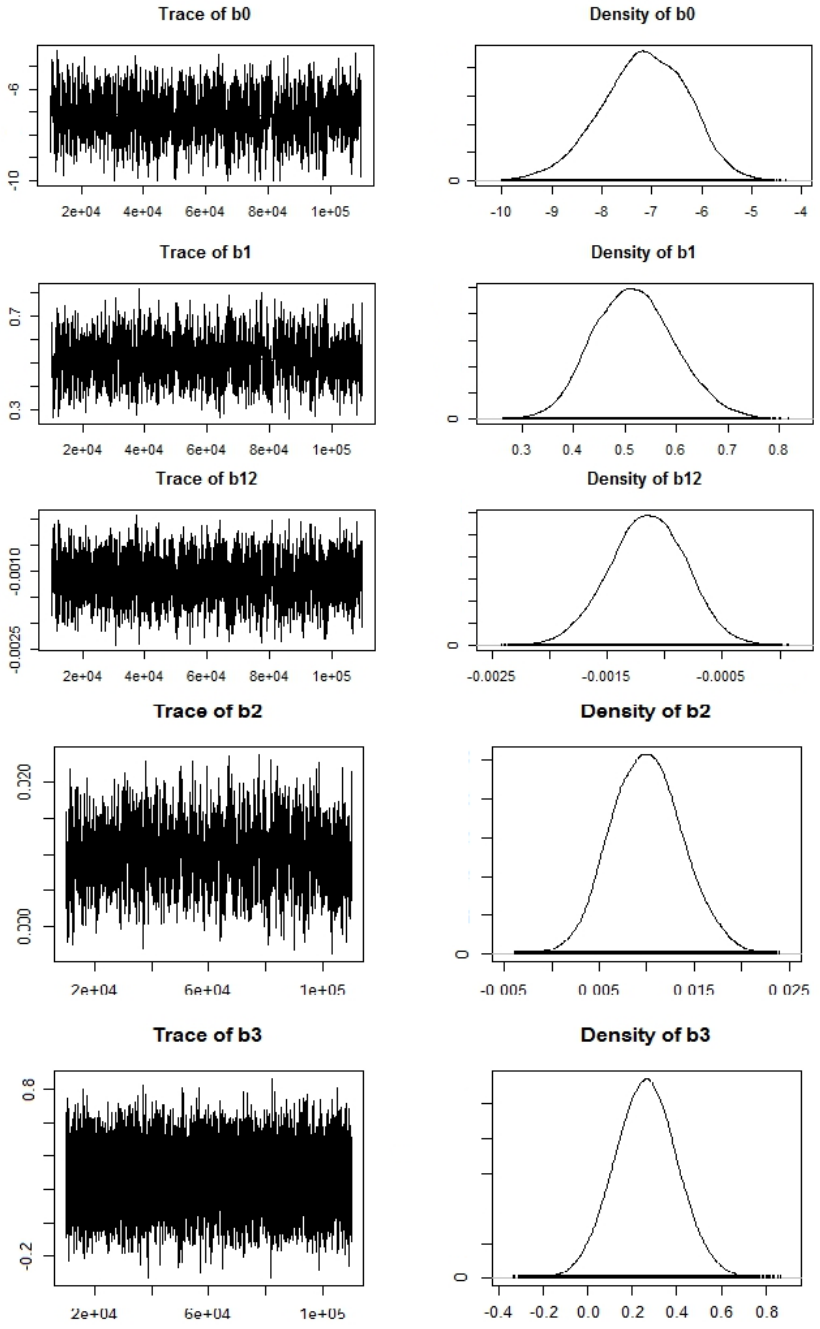

Figure 5.2: Trace plots and kernel density estimates for Bayesian logistic regression model

Using these results, we found that using more timepoints seems to give considerably better results on performing logistic regression of occurrence of DCS on bubble score, and that KISS, using as it does all the available timepoints, gives a yet better result, with the proviso that imputed KISS is used only where the original KISS is unavailable. We would suggest that it might in future be helpful to devise a version of KISS that can use the continuous bubble concentrations to provide an even better overall summary of the data and to give enhanced weight to the first few datapoints. This would seem to be helpful, given the additional information on DCS probability that these points seem to provide. However, if this KISS is to use imputed data, we should also form some measure of its uncertainty and incorporate this into its subsequent usage as a predictor for DCS.

In future work, we could usefully refit the model of Chapter 3 using the Second QinetiQ dataset as employed in this chapter, and we could then directly compare the predictive power of the models in Chapters 3 and 4. This would provide further information on the predictive effectiveness of each of the models, together with an indication as to which of the two might a more fruitful avenue for further investigation. However, should the bubble-score based method prove to be more effective, this does not mean that we should necessarily dismiss Chapter 3's model but may rather be an indication that the single-compartment model is insufficient and that an extension to multiple compartments is required for full predictive efficacy.

Chapter 6

Combining Models, Model Averaging and Predicting Risk

6.1 Introduction

Having considered a variety of models in Chapters 3, 4 and 5, we now seek to bring our previous work together here. Recall that we have two primary aims. Our first aim is to provide easily accessible advice for submariners trapped in a damaged submarine who must determine whether they should attempt to escape from the submarine - a choice which is only sensible if their DCS risk is not too high should they choose to do so, or if they will otherwise be killed anyway - or, alternatively, wait for a rescue sub to arrive. The latter is far preferable from a DCS risk point of view but may not be viable from a practical one. We therefore seek to provide a submarine commander with advice to better weigh up their options in this regard. In particular, we shall seek to make some account for the importance of the length of time for which saturation at the given depth has been experienced - that is, whether we are effectively dealing with a saturation dive, a subsaturation dive or a bounce dive. The first will be the case where the submarine has been flooded to a greater or lesser extent for a long period

before an escape attempt is made, the second where there has been flooding for a short period prior to consideration of escape, and the latter where there is no internal submarine flooding at all. In this instance (i.e. within the stricken submarine), no computational facilities can be relied upon. To this end, we shall attempt to improve on the tool that already exists, which is a 2D contour plot of saturation depth against escape depth with DCS probability as the response (Loveman et al. (2014)). We shall use the results that we have obtained in Chapters 3 and 4 to outline how to produce a superior visual aid for use by submariners in this situation.

Having recalled our first aim, we now recall our second: to provide improved practicable applicable models for the prediction of DCS in submariners *after escape* from a damaged submarine. Our purpose now is not to provide aid to the submarine commander, but rather to provide information to any rescuers. Upon rescuing the submariners, the crew of the rescue ship must assess each submariner for their risk of contracting DCS in order to decide whom, if anyone, to place in their limited available on-board recompression chamber treatment space or whether to advise the submariners to wait for additional rescue resources to arrive before attempting escape. To this end, we shall not be able to use any technique that requires vast amount of computational effort to formulate our predictions given the limited time and intrinsic pressures of an emergency situation. In particular, we should be able to run our predictive model for each member of a disabled submarine's crew in the space of around five minutes. Given that a standard naval submarine crew is a little under 100 hands, with the largest submarines having just under 200, this means that our predictions should not take more than about a second each to deliver. Furthermore, the computational resources available to make any predictions on board a rescue ship will be limited. While a small computer may be available, it would be ideal if any calculations could be completed on (suitably reprogrammed) watch-size diving computers (or perhaps via a phone app) that are readily available and are easily carried by members of the rescue operation. Note that since

rescue has now occurred, we will now be able to leverage our results from Chapter 5 as well as those from Chapters 3 and 4, as the rescuers will be able to make an assessment of the divers' KM bubble scores five minutes after surfacing, and this should be taken by the rescue team themselves after bringing survivors to the surface. Given that this data is highly significant in the prediction of cases of DCS, they will be able to use this invaluable information in addition to whatever other dive depth and time details they have to hand.

Given the above considerations, we do not want to actually perform any MCMC (and particularly not any RJMCMC) at this stage. All computations are to have been completed in advance so that, rather than performing any likelihood evaluations, we need only use predetermined models and parameter values. On the other hand, it will certainly be feasible to perform a small number of low-dimension integrations as this is well within the realm of possibility for computationally low-powered devices. Thus, our models from Chapters 3 and 4 remain useful. While we noted that computation times there were high, this is because we had to compute several thousand integrals per likelihood evaluation. Here, however, if we have a k compartment model, we need only compute at most $1+(1+2)+(1+2+3)+\cdots+(1+\cdots+k)$ separate integrals per submariner. Even if we make use of all possible models from 1-compartment up to and including an 8-compartment model in some way, then we need only compute a maximum of $(1+(1+2)+\cdots+(1+\cdots+8))\times 200 = 24000$, which is computationally equivalent to only three likelihood evaluations of the 8-compartment model and is easily achievable! Furthermore, our results from Chapter 5 rely only on logistic regression, predictions from which are easily and quickly calculated. Indeed, this was why we took a non-Bayesian approach there; while we could have used a Bayesian procedure, this would have entailed drawing from a posterior predictive distribution for a new observation or running a Gibbs sampler to include this 'new observation' as we went along running our imputation and doing so would have required

Number of compartments	Label	Bounce	Sub-saturation	Saturation
3	p_3	0.37450	0.33240	0.32445
4	p_4	0.41757	0.38858	0.33314
5	p_5	0.20793	0.27904	0.34241

Table 6.1: Table of reweighted posterior model probabilities

running a lengthy sampler which is not something we have the time or computational power to do for each submariner in this time-critical situation.

6.2 A submariner's aid

Our first aim of helping the submariners to determine whether or not to escape from their damaged vessel is to be accomplished using just our results from Chapters 3 and 4 (as of course no bubble score is available until the post-ascent stage). To this end, we focus on using our results from Chapter 4 and, in particular, using a Model Averaging procedure with model weights as found in performing our RJMCMC.

Suppose, then, that we use the results obtained in Chapter 4 and recorded in Table 4.2 where we used our RJMCMC algorithm to obtain posterior model probabilities for the $1, 2, \ldots, 8$-compartment models. Notice that the vast majority of the posterior model probability distribution is concentrated on the 3, 4 and 5 compartment models. To that end, to avoid unnecessary computation, we reallocate the probability from the 1, 2, 6, 7 and 8 compartment models to the 3 and 5 compartment models by adding the weight of the 1 and 2 compartment models to the 3 compartment model and from the 6, 7 and 8 compartment models to the 5 compartment model (these respectively being the closest compartment values to those of the removed compartments). Then, the model probabilities assigned to these models are as in Table 6.1.

Now, as suggested by Kass and Rafferty (1995) we may use Bayesian Model Averaging, to obtain the pdf for the probability of DCS not occurring as being given by

$$p(y|\boldsymbol{\theta}, x_1, x_2, x_3) = \sum_{k=3}^{5} p_k \left(\exp\left(-\sum_{i=1}^{k} \int_0^{\infty} h_i(t|\boldsymbol{\theta_i}, x_1, x_2, x_3)\, dt \right) \right), \tag{6.1}$$

where h is our hazard function as obtained in Section 4.2.

In the following section, we shall want to save further computational effort and insert the modal posterior parameter values here in order to consequently obtain the estimated probability of DCS occurring in the ascended submariners. However, this is not necessary at this stage as our computational time is not limited here. To that end, we would like to obtain the posterior predictive distribution in order to account for our uncertainty in the estimates of the parameter values, as this will effectively average out the prediction over these possible parameter values. To that end, we note that the posterior predictive distribution for a new observation y_{new} will be given by

$$p(y_{new}) = \int_{\boldsymbol{\theta}} p(y_{new}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|x_1, x_2, x_3)\, d\boldsymbol{\theta}, \tag{6.2}$$

where θ is the whole set of parameters. In particular, to use (6.2) with (6.1), we need to interchange the order of summation and integration to obtain

$$p(y_{new}) = \sum_{k=3}^{5} p_k \int_{\boldsymbol{\theta_k}} \left(\exp\left(-\sum_{i=1}^{k} \int_0^{\infty} h_i(t|x_1, x_2, x_3)\, dt \right) \right) p(\boldsymbol{\theta_k}|\boldsymbol{x})\, d\boldsymbol{\theta_k}, \tag{6.3}$$

where $\boldsymbol{\theta_k}$ is the set of parameters required for model k.

Now, (6.3) may not seem much more amenable to computation than its predecessor statement. However, recall that we already have samples from

$p(\boldsymbol{\theta_k}|\boldsymbol{x})$, as these are simply our samples from the RJMCMC runs in Chapter 4. It follows that we can use these (after a suitable burn-in period has been removed from the samples, of course). Then, supposing that at step t of our MCMC iteration our parameters are $\boldsymbol{\theta}_{\boldsymbol{k}}^{(t)}$, we may then simulate from $p(y_{new}|\boldsymbol{\theta}_{\boldsymbol{k}}^{(t)})$ by drawing from a Bernoulli distribution with parameter given by $p(y, \boldsymbol{\theta}_{\boldsymbol{k}}^{(t)})$. Having done this, we may collate these results to estimate the probability of DCS occurring for a new observation, while also taking account of the uncertainty in our parameter values. By doing this for a range of covariate values, we can form a grid of posterior predictive probabilities for these parameter values and subsequently create an enhanced 2D contour plot.

As an example, we provide such a plot in Figure 6.1 for saturation dives followed by escape. While plots of this nature take quite some time to produce and are highly computationally intensive, they could be prepared in advance by personnel knowing the approximate depths and exposure times likely to be involved in any operations to be undertaken. Note that we limit the plot to a maximum depth of 700m as below this depth it is highly likely that the submarine itself will be crushed.

6.3 Helping the rescuers

We now move on to consider our second problem - helping the rescuers to predict which submariners are liable to suffer DCS, so as to be able to provide the requisite treatment more effectively to those who are in most need. Recall that it can take several hours following surfacing for DCS to develop, but these decisions must be made within minutes of divers' surfacing, so we must make use of all the information we can obtain. To that end, we now use not only our results from Chapters 3 and 4, but also those from Chapter 5 as the rescuers will be able to measure the bubble

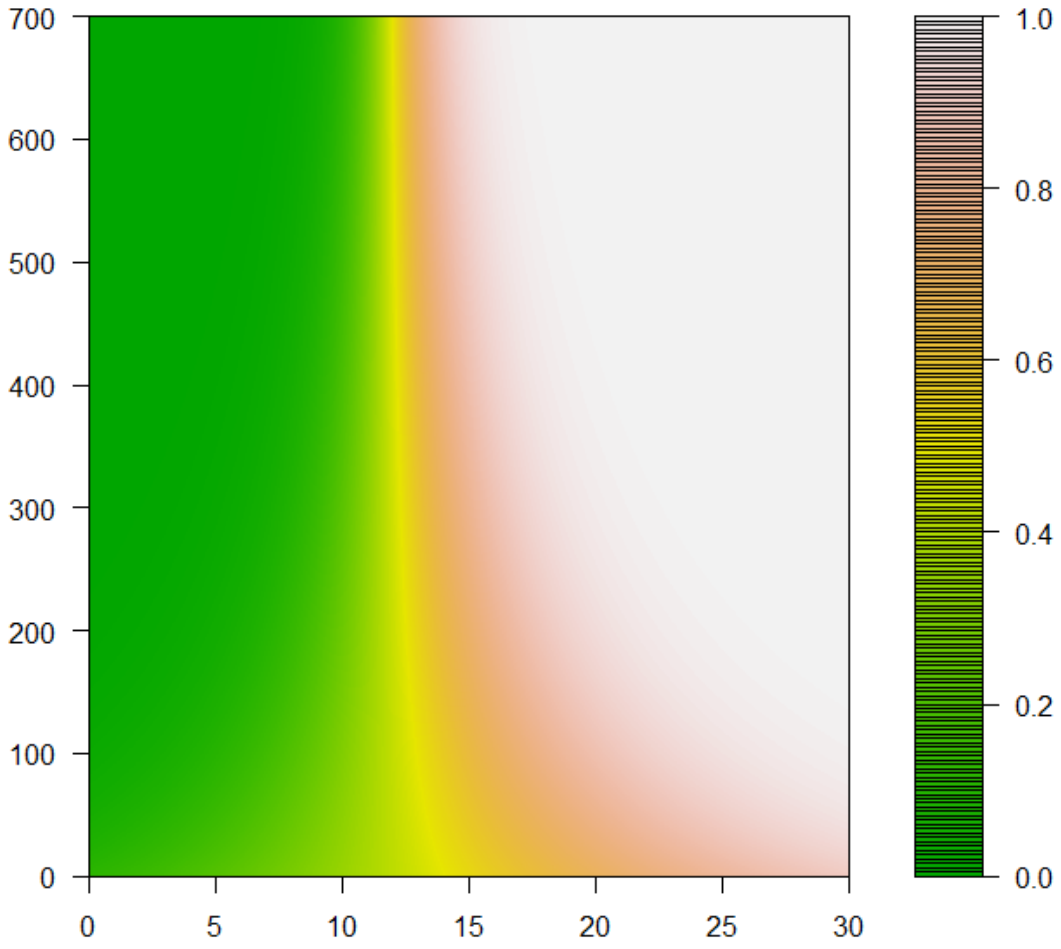

Figure 6.1: Contour plot showing estimated DCS probabilities for a range of saturation depths (x-axis) and escape depths (y-axis)

scores of those submariners whom they rescue. To do this, however, we must find some way to combine the two types of model and, in order to achieve this, we use the idea of a mixture of the two possible approaches.

First, note that to save computation time here, we use a point estimate of the parameter values for our compartmental models, rather than making use of the posterior predictive distribution for a new observation as we did in the previous subsection. We do, however, continue to account for model uncertainty by using the posterior model probabilities obtained in Chapter 4, rather than simply choosing a 'best model'. For our point estimates of the desired compartmental model parameters, we use the posterior mode obtained by maximizing the marginal kernel density estimates of our posterior distribution, as outlined in Chapter 3 (with the estimates used being as given in Chapter 4).

Supposing, then, that we have the posterior model probabilities of our 3, 4 and 5 compartment models as given in the previous subsection, with θ_k again corresponding to the parameters of the k-compartment model, we formulate the probability y of DCS *occurring* as being

$$f(y|\boldsymbol{\theta}, p, x_1, x_2, x_3) = pp_B + (1-p)p_M \tag{6.4}$$

where p is our 'mixture' probability which will weight the importance of each of the two types of model, p_B is the DCS probability obtained from the logistic regression model considered in Section 5.6 and p_M is the DCS probability obtained from our averaged model from Section 6.2. We shall suppose that there are n datapoints and take

$$L(p) = \prod_{i=1}^{n} f(y_i)^{\delta_i}(1 - f(y_i))^{1-\delta_i}$$

to be our likelihood function, where $\delta_i = 0$ if DCS does not occur on the ith dive and $\delta_i = 1$ if DCS does occur on that dive. We further suppose that

Parameter	Approximate posterior mode (Standard error)
β_0	-65.034 (4.214)
β_1	1.637×10^{-6} (2.056×10^{-7})
β_2	-8.078×10^{-6} (6.310×10^{-7})
β_3	-5.973 (1.672)

Table 6.2: Estimated posterior modes (standard errors) of mixture parameter distributions for saturation dives

$\text{logit}(p)$ may depend linearly on the maximum depth, x_1, the saturation depth, x_2 and the 5-minute imputed bubble score, x_3. It follows that $p = \text{logit}^{-1}(\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3)$ for some hyperparameters $\beta_0, \beta_1, \beta_2, \beta_3$. We may now run an MCMC algorithm to estimate these parameters. Note that since $\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3$ may take any real value while p takes a value in the interval $[0, 1]$ as required, given that it is a probability. As previously, we sample each component separately and use a normal proposal as well as a non-informative uniform prior on each parameter- this time $U(-200, 200)$ - so that the acceptance probability, α, for our move for β_0, for example, becomes

$$\alpha = \min\left\{\frac{L(\beta_0^{(*)}|\beta_1^{(t)}, \beta_2^{(t)}, \beta_3^{(t)}, y)}{L(\beta_0^{(t)}|\beta_1^{(t)}, \beta_2^{(t)}, \beta_3^{(t)}, y)}, 1\right\},$$

where superscript (t) indicates the current, tth iterated value and superscript $(*)$ indicates our new proposed value for β_0.

Our results are summarised in the Table 6.2, which gives the approximate posterior modes of the distributions of the parameters in question and a trace plot is given in Figure 6.2. Note that as the bubble data do not contain information on dive type, we must select a particular set of parameters from those found in Chapter 4 to use. To this end, we use the saturation parameters as QinetiQ have stated that the vast majority of the dives for this dataset were undertaken in saturation conditions (Loveman, Personal Communication, 2012).

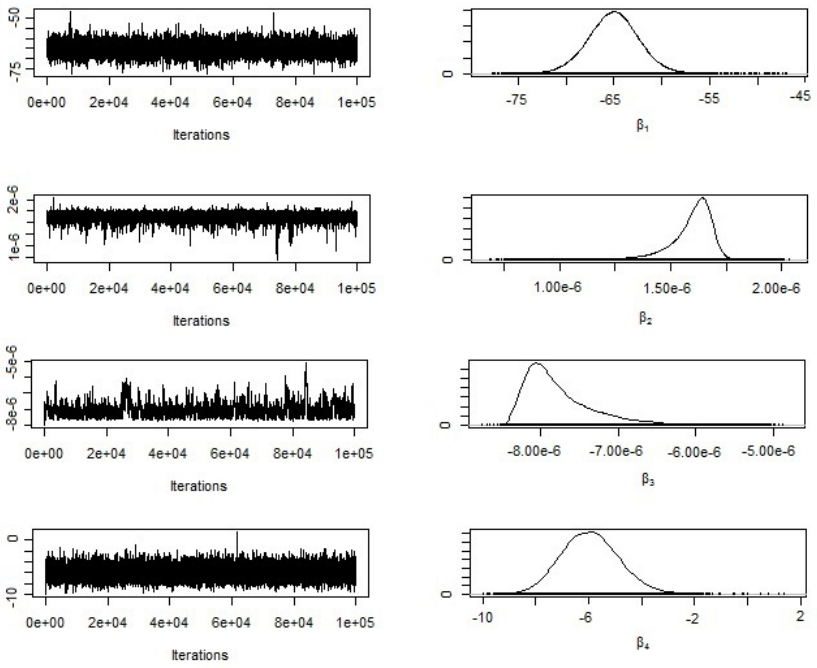

Figure 6.2: Trace plot for $\boldsymbol{\beta}$ hyperparameters

Note that our method does not add any additional computational time at the point of rescue - all necessary MCMC sampling is done here, in advance. All we need do in order to estimate the probability of DCS for a given dive profile and bubble score, is to input the Saturation Depth, Escape Depth and Bottom Time of the dive, together with the bubble score, so as to obtain a suitable point estimate of DCS probability. Note, however, that Bottom Time will likely have to be estimated for a sub-saturation dive, though a 'standard' bottom time of 12 hours is used here where no further information is available.

To see how our new combined model performs, it is helpful to recall the pseudo-R^2, R_2^T, suggested by Tjur (2009) that we defined in Section 3.6, found by calculating the average predicted probability for each of the three

possible outcomes (DCS, no DCS, again counting marginal DCS as a no DCS outcome) and taking the absolute difference between them.

Now, if we set $p = 0$ and produce a prediction for each datapoint, we can see the effect of considering only the averaged compartmental model. This gives an R_T^2 of 0.4119. On the other hand, setting $p = 1$ and again making predictions for each datapoint, we can now see the effect of considering only the bubble-score based method. This gives an R_T^2 of 0.4863. Finally, by setting p equal to the suggested inverse logit function of x_1, x_2, x_3 above, we can see the ultimate effect of, and improvement created by, combining our models. In this case, we obtain an R_T^2 of 0.6155. Meanwhile, the R_T^2 for the original QinetiQ logistic regression in Section 3.2 is 0.3417. We see, then, that our combined model makes a good improvement on the original QinetiQ logistic regression model and a reasonable improvement on the separate models presented here, without too much additional computational effort being needed.

6.4 An Overview

Throughout this and preceding chapters, we have made various approximations to a fully Bayesian solution. In this Section, we shall aim to outline what those have been and how a fully Bayesian solution to our problem might fit together.

Firstly, in Chapter 4, we would ideally have liked to have found the exact marginal likelihood for each model under consideration (i.e. the models with $1, 2, \ldots, 8$ compartments). However, the models under consideration are clearly analytically intractable and we therefore had to consider some means by which to approximate the desired marginal likelihoods. To that end, we could have chosen either a within-model approach (involving the

evaluation of a series of numerical approximations to suitable integrals) or a between-models, transdimensional approach. This latter approach was selected as approximating the integrals involved in the computing of each marginal likelihood directly - even with a numerical approximation - would have been difficult, given the complexity of the likelihood function in question. The transdimensional approach, while still requiring likelihood evaluations, did not require as many as would have been needed by evaluating the marginal likelihood of a model with each of $1, 2, \ldots, 8$ compartments separately.

Secondly, in Chapter 5, Section 5.3, rather than performing a mean imputation, we could have used our Gibbs sampler to directly sample the desired values for imputation by treating these as missing values. However, as mentioned in Section 6.1 above, this would have necessitated the running of the Gibbs sampler for each prediction we wished to make, rather than running it initially and then making use of the results obtained during the time-critical rescue period.

Finally, in Section 6.3 above, we have made use of two sets of approximations: in the first instance, we have used point estimates (posterior modes) for the compartmental models used to obtain our estimated DCS probabilities, rather than making use of a posterior predictive distribution after integrating out the relevant parameters. This allows us to compute the desired survival probabilities rapidly, again in what would be a time-critical situation, rather than having to numerically evaluate multiple integrals of the form (6.2). In the second instance, we have similarly approximated the posterior parameter distributions for the regression models of Chapter 5 by their posterior modal estimates. Overall, then, a fully Bayesian solution would necessitate avoiding both these approximations and that noted above that was used in Section 5.3 (since the imputed bubble score is used in (6.4)). Firstly, we should have to use our Gibbs sampler to draw our imputed bubble score, rather than approximating this by using a

mean imputation. Secondly, we should need to find the posterior predictive distribution for the DCS probability of a new dive, not by using point estimates of the posterior density, but rather by marginalising suitably over all the parameters in question. That is, we should have to integrate out the parameters $\boldsymbol{\theta}_k$ for *each* of our k compartmental models (whether that be the reduced number we used in Sections 6.2 and 6.3 or the full range of $1, 2, \ldots, 8$ suggested in Section 4.7), over the regression parameters obtained in Section 5.7 (since we would need to use our Bayesian regression framework, rather than the maximum likelihood framework used in Section 5.4) and also over the hyperparameters $\boldsymbol{\beta}$ introduced in Section 6.3. Clearly, however, this would represent a considerably higher computational burden than making use of the approximations we have chosen and could not be done in a time-critical situation.

Chapter 7

Future Work

In this thesis, we have considered both 'black box' logistic regression models and physically based Thalmann-type models for the modelling of the probability of the occurrence of DCS both with and without information being available about post-surfacing bubble scores. In doing this, we have supposed that the compartments in question are not physical, identifiable locations in the body that correspond with particular tissue types, but rather are simply useful modelling tools, not corresponding with particular tissue types. However, it has been proposed in Bühlmann (1984), for instance, that with 16 compartments these could directly correspond to parts of the body. If this were the case, perhaps it would be possible to take measurements in different parts of the body and we would subsequently be able to form a model in which compartments were directly interpretable as being associated with specific body parts or tissue types. A further reason for doing this is that, although we have considered only the incidence of DCS, the first QinetiQ dataset also includes information on the type of DCS that occurred. For example, depending on where the majority of bubbles are lodged, there can be neurological symptoms, breathing difficulties, heart trouble or joint issues. If we were able to assign compartments to body parts in a one-one fashion, it might be possible to make inferences about not only whether DCS might occur, but also what type of DCS would be

most likely to occur. This is of interest to medics as some types of DCS (notably those forming in the brain and lungs) require more immediate treatment than do those occurring in the joints or extremities, say. In addition, we have effectively considered DCS as a binary outcome (with a third possible 'marginal' outcome in a few cases). It may also be possible by, for example, requesting that experimenters use their expert judgement, to ask them to assign a DCS severity level in a manner similar to the KM bubble scores considered in Chapter 5. These ordinal data would provide us with more information than a simple binary outcome, and would allow us to better predict those most serious, problematic DCS cases.

Due to time and computational constraints, we have only been able to focus our attention on a single physical decompression model - that of Thalmann et al. (1997) in which a supersaturation approach to hazard build-up over time is considered. However, as noted in Section 2.3, there are several models that directly incorporate the formation of bubble dynamics into their models - though often *without* considering levels of supersaturation, but merely considering absolute levels of inert gas in the compartments at a given time. We could seek to extend our methods in Chapter 4 and Chapter 6 by allowing the addition of a further model that directly incorporates these dynamics, and subsequently including this information in our hazard function and then into our likelihood. We would then be mixing *three* types of model in Chapter 6. Clearly, this would lead to considerably increased computational requirements as these bubble modelling processes involve the use of partial differential equations whose solutions cannot be found exactly. This would require the use of a numerical approximation scheme for their solution, rather than being able to use methods such as those outlined in Chapter 3 for the Thalmann et al. model. Even more promising would be the possibility of using the very recent model suggested in Goldman and Solano-Altamirano (2015), which accounts not only for bubble formation but also the possibility of interchange of gas being permitted *between* the compartments of the model. This occurrence is something not accounted

for by the model that we selected, nor by any of the other models described in Section 2.3, all of which are 'perfusion limited', which is to say no diffusion is permitted between the compartments in these models. However, the Goldman-Solano-Altamirano model is an order of magnitude more complicated, requiring not only the solution of partial differential equations, but *coupled* partial differential equations, where the coupling occurs due to the diffusion between the compartments. To evaluate the required likelihood many hundreds of thousands of times, as is required to use our methods in Chapter 4, would simply not be viable. However, we could consider making use of *emulation*. In this process, details of which can be found in, for example, Conti and O'Hagan (2010), one attempts to avoid having to evaluate the likelihood by calculating it at a carefully selected set of points and then extrapolating what the computer code would output by use of a statistical emulator, the latter often being based on a Gaussian process. As Gaussian processes are quick to evaluate, we can effectively find approximations to computationally expensive likelihood values by using our cheap emulator. Of course, this comes at the cost of additional error and uncertainty, as we are now approximating the true output of our computer program. However, it would allow us to approximate the output Goldman-Solano-Altamirano method for use in our MCMC and RJMCMC procedures.

As noted in Section 3.6, we could further investigate the use of time-to-event data by allowing the parameters in each compartment to vary before and after DCS occurs (in those cases where it does occur), in order to obtain a better model fit. However, this would immediately double the size of our parameter space and likelihood evaluation already takes several seconds, with the MCMC code requiring several days to run several hundred thousand iterations. To decrease this computation time, we could attempt to *parallelize* our MCMC code. There are two possible ways to do this: in the first instance, we could perform an embarrassing parallelization in which we simply have several instances of the code running at the same

time. On the other hand, we could also utilise the massively parallel structure of a system such as a GPU (Graphics Processing Unit), rather than a CPU, to parallelize a single chain. A GPU can perform many hundreds of identical computations simultaneously. For details of how this works see, for example, Kirk and Hwu (2010). Such a system would be an excellent choice for, say, simultaneously checking whether each of the parameters in our RJMCMC compartment is to be updated during the within-model phase, then feeding this back to the main CPU. The CPU then performs the single step of checking whether there is to be an update to the number of compartments, before passing new data back to the GPU if required for the next update of the various parameters. By doing this, we can actually calculate many results at the same time and considerably reduce required computation time, possibly allowing us the leeway to explore either particularly large numbers of more closely specified compartments that directly correspond to particular body parts (as suggested in the first paragraph above), or by allowing us to explore different, more highly parameterized versions of our hazard functions. For example, we could allow the parameters in each compartment to become some function of the covariates.

A final possibility to extend the work would be to consider the effect of different types of dive. In this thesis, we have only considered individual dives in isolation and all these dives have been undertaken using air as the breathing gas. However, we could, for example, consider examining dives undertaken with a different breathing gas (such as Heliox - a mixture of Helium and Oxygen), or we could consider 'repetitive type' dives where there is more than one descent and subsequent ascent and decompression. In these cases, bubble score is qualitatively observed not to follow a similar pattern with time as with the other non-repetitive type dives (Loveman, Personal Communication, 2010). There are a not insignificant number of dives of this type that have already been undertaken in otherwise suitable escape-type conditions. Thus, if possible, we feel that it would be beneficial

to make use of this information. This might provide useful data and lead to more informed practice in a tactical situation.

Chapter 8

References

Akaike H. *A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6): 716-723, 1974.*

Anderson JA, Pemberton JD. *The grouped continuous model for multivariate ordered categorical variables and covariate adjustment. Biometrics, 41(4):875-885, 1985.*

Ando T. *Bayesian Model Selection and Statistical Modelling. Chapman and Hall, 2010.*

Boycott AE, Damant GCC, Haldane JS. *The prevention of compressed air illness. Journal of Hygiene, 8(3):342-443, 1908.*

Browne WJ. *MCMC algorithms for constrained variance matrices. Computational Statistics and Data Analysis, 50(7):1655-1677, 2006.*

Brubakk AO, Neuman TS. (ed.) *Bennett and Elliott's Physiology and Medicine of Diving, 5th edition. Saunders Ltd., 2002.*

Bühlmann AA. *Decompression-Decompression Sickness. Springer-Verlag, 1984.*

Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd edition. Springer-Verlag, 2002.*

Burnham KP, Anderson DR. *Multimodel inference: understanding AIC and BIC in model selection. Sociological Methods and Research 33: 261-304, 2004.*

van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, Rubin DB. *Fully conditional specification in multivariate imputation. Journal of Statistical Computation and Simulation, 76(12):10491064, 2006.*

Collett D. *Modelling Binary Data, Second Edition. Chapman and Hall, 2002.*

Conti S, O'Hagan A. *Bayesian emulation of complex multi-output and dynamic computer models. Journal of Statistical Planning and Inference 140: 640-651, 2010.*

Corless RM, Gonnet GH, Hare DEG, Jeffrey DJ, Knuth DE. *On*

the Lambert W function. Advances in Computational Mathematics, 5:329-359, 1996.

Cowles MK, Carlin BP. *Markov Chain Monte Carlo convergence diagnostics: a comparative review. Journal of the American Statistical Association, 91(434):883-904, 1996.*

Davison AC. *Statistical Models. Cambridge University Press, 2008.*

Dellaportas P, Forster JJ, Ntzoufras I. *On Bayesian model and variable selection using MCMC. Statistics and Computing, 12:27-36, 2002.*

Gelman A. *Prior distributions for variance parameters in hierarchical models. Bayesian Analysis, 1(3): 515-533, 2006.*

Gelman A, Carlin J, Stern H, Rubin D. *Bayesian Data Analysis, 2nd Edition. Chapman and Hall/CRC, 2003.*

Gerth WA, Vann RD. *Probabilistic gas and bubble dynamics models of decompression sickness occurrence in air and nitrogen-oxygen diving. Journal of Undersea and Hyperbaric Medicine, 24:275-292, 1997.*

Goldman S. *A new class of biophysical models for predicting the probability of decompression sickness in scuba diving. Journal of Applied Physiology, Articles 103, 484-493, 2007.*

Goldman S, Solano-Altamirano JM. *Decompression sickness in breath-hold diving and its probable connection to the growth and dissolution of small arterial gas emboli. Mathematical Biosciences, 262:1-9, 2015.*

Graham JW, Olchowski AE, Gilreath TD. *How many imputations are really needed? Some practical clarifications of Multiple Imputation theory. Prevention Science, 8:206-213, 2007.*

Green PJ. *Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. Biometrika, 82(4):711-732, 1995.*

Hastings WK. *Monte Carlo sampling methods using Markov Chains and their applications. Biometrika, 57(1):97-109, 1970.*

Heller R, von Schrötter H, Mager W. *Luftdruck-Erkrankungen mit besonderer Berücksichtigung der sogenannten Caissonkrankheit. Hölder, 1900.*

Hosmer DW, Lemeshow S. *Goodness-of-fit tests for the multiple logistic regression model. Communications in Statistics: Theory and Methods, Part A, 9: 10431069, 1980.*

Howle LE, Weber PW, Vann RD, Campbell MC. *Marginal DCS events: their relation to decompression and use in DCS models. Journal of Applied Physiology 107: 15391547, 2009.*

Howle LE, Weber PW, Vann RD. *A parameter estimation system with computational advantages for fitting probabilistic decompression models to empirical data. http://rubicon-foundation.org/dspace/handle/123456789/5066 (last accessed January 2012), 2007.*

Hurvich CM, Tsai CL. *Regression and time series model selection in small samples. Biometrika, 76:297-307, 1989.*

Jankowski LW, Nishi RY, Eaton DJ, et al. *Exercise during decompression reduces the amount of venous gas emboli. Journal of Undersea and Hyperbaric Medicine. 24:59-66, 1997.*

Kass RE, Rafferty AE. *Bayes factors. Journal of the American Statistical Association, 90:773-795, 1995.*

Kirk DB, Hwu WW. *Programming Massively Parallel Processors: A Hands-on Approach. Morgan Kaufmann, 2010.*

Lee, P. *Bayesian Statistics: An Introduction, 3rd edition. Hodder, 2004.*

Link WA, Eaton MJ. *On thinning of chains in MCMC. Methods in Ecology and Evolution. 3: 112-115, 2012.*

Loveman GAM, Seddon FM, Thacker JC, Stansfield MR, Jurd KM. *Submarine tower escape decompression sickness risk estimation. Journal of Undersea and Hyperbaric Medicine. 41(4):315-329, 2014.*

McQuarrie ADR, Tsai CL. *Regression and Time Series Model Selection. World Scientific, 1998.*

Metropolis N et al. *Equations of state calculations by fast computing machines. Journal of Chemical Physics, 21(6):1087-1092, 1953.*

Nelder JA, Wedderburn RWM. *Generalized Linear Models. Journal of the Royal Statistical Society. Series A (General), 135(3):370-384, 1972.*

O'Hagan A, Forster JJ. *Bayesian Inference, 2nd edition, volume 2B of Kendall's Advanced Theory of Statistics. Arnold, 2004.*

Powell MJD. *Approximation Theory and Methods. Cambridge University Press, 1981.*

Richardson S, Green PJ. *On Bayesian analysis of mixtures with an unknown number of components. Journal of the Royal Statistical Society. Series B (Methodological), 59(4):731-792, 1997.*

Robert CP, Casella G. *Monte Carlo Statistical Methods, 2nd Edition. Springer, 2004.*

Rubin DB. *Multiple Imputation for nonresponse in surveys. Wiley, 1987.*

Schafer JL, Graham JW. *Missing Data: Our View of the State of the Art. Psychological Methods, 7(2):147-177, 2002.*

Schwarz G. *Estimating the dimension of a model. The Annals of Statistics, 6(2):461-464, 1978.*

Silverman BW. *Density Estimation for Statistics and Data Analysis. Chapman and Hall, 1986.*

Thalmann ED. *Phase II testing of decompression algorithms for use in the US Navy underwater decompression computer. US Navy Experimental Diving Unit, rep. 1-84, 1984.*

Thalmann ED, Parker EC, Survanshi SS, Weathersby PK. *Improved probabilistic decompression model risk predictions using linear-exponential kinetics. Undersea and Hyperbaric Medicine Journal, 24(4):255-274, 1997.*

Tikuisis P. *Modeling the observations of in vivo bubble formation with hydrophobic crevices. Undersea Biomedical Research, 13:165-180, 1986.*

Tikuisis P, Gerth WA. *Chapter 10, Bennett and Elliotts' Physiology and Medicine of Diving ed. Brubakk AO and Neuman TS, 5th edition. Saunders Ltd. 2002.*

Tjur T. *Coefficients of determination in logistic regression modelsA new proposal: The coefficient of discrimination. The American Statistician, 63:366-372, 2009.*

Vann RD. (ed) *The Physiological Basis of Decompression. 38th Undersea and Hyperbaric Medical Society Workshop. UHMS Publication Number 75(Phys) 6-1-89. Undersea and Hyperbaric Medical Society, 1989.*

Wand MP, Jones MC. *Kernel Smoothing. Chapman and Hall, 1995.*

Waters S et al. *Parameter estimation in decompression sickness. http://www.maths-in-medicine.org/uk/2007/decompression-sickness/report.pdf (last accessed January 2012), 2007.*

Weathersby PK, Homer LD, Flynn ET. *On the likelihood of decompression sickness. Journal of Applied Physiology: Respiratory and Environmental Exercise Physiology, 57:815-825, 1984.*

Wienke BR. *Reduced Gradient Bubble Model (RGBM). International Journal of Biomedical Computation, 26:237-256, 1990.*

Wienke BR. *Technical Diving In Depth. Best, 2001.*

Wilkinson DJ. *Chapter 16, Handbook of Parallel Computing and Statistics ed. Kontoghiorghes EJ. Chapman and Hall/CRC, 2005.*

Workman RD. *Calculation of decompression schedules for nitrogen-oxygen and helium-oxygen dives. Navy Experimental Diving Unit, Research Report 6-65, 1965.*

Wu W, Jia F, Enders C. *A comparison of imputation strategies for ordinal missing data on Likert scale variables. Multivariate behavioral research 50 (5):484-503*

Yount DE. *On the elastic properties of the interfaces that stabilize gas cavitation nuclei. Journal of Colloid and Interface Science, 193(1):50-59, 1997.*

Yount DE, Hoffmann DC. *On the use of a bubble formation model to calculate diving tables. Aviation, Space and Environmental Medicine, 57(2): 149-156, 1986.*

Zellner A. *An Introduction to Bayesian Inference in Econometrics. Wiley, 1971.*