Markov chain Monte Carlo analyses of longitudinal biomedical magnetic resonance data

Thesis

Markov chain Monte Carlo analyses of longitudinal biomedical magnetic resonance data

Martin D. King

Department of Statistics
Faculty of Mathematics and Computing
The Open University

Submitted for the degree of Doctor of Philosophy
September 2004

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

Markov chain Monte Carlo simulation was used in an analysis of the data acquired in three longitudinal biomedical magnetic resonance studies. The first of these investigations uses a Bayesian nonlinear hierarchical random coefficients model to examine the longitudinal extracellular direct current (DC) potential and apparent diffusion coefficient (ADC) responses to focal ischaemia in the rat. The purpose is to perform a formal analysis of the temporal relationship between the two responses, and thus to examine the data for compatibility with a common latent (driving) process and, alternatively, the existence of an ADC threshold for anoxic depolarisation. The DC-potential and ADC transition parameter posterior probability distributions were generated, paying particular attention to the within-subject differences between the DC-potential and ADC transition characteristics. The results indicate that the DC-potential and ADC changes are not driven by a common latent process and, in addition, provide no evidence for a consistent ADC threshold associated with anoxic depolarisation.

The second analysis uses data acquired in a nuclear magnetic resonance spectroscopic study into the effects of intestinal ischaemia and subsequent reperfusion on liver metabolism in the rat. The purpose of the analysis is to examine the temporal relationship between energy status [inorganic phosphate to adenosine triphosphate ratio (PAR)] and the pH response, the former of which is an indicator of liver energy failure. The posterior distribution obtained for the PAR-pH onset time difference indicates that the pH response precedes the change in PAR, suggesting that intracellular acidosis cannot be ruled out as a contributing factor to the observed liver failure.

The third dataset was acquired in an electron spin resonance study of the Arrhenius behaviour of the rabbit muscle sarcoplasmic reticulum membrane. An MCMC Arrhenius plot changepoint analysis is used to estimate the order parameter 'transition' temperature.

Contents

Appendices

List of figures and tables

Figures

Tables

Acknowledgements

I am extremely grateful to my two external supervisors, Prof. David Hand and Prof. Martin Crowder for their support and guidance throughout the eight year period of my registration for an Open University research degree. Not only have I had the advantage of having two internationally renowned statisticians supervising my work, but each of my meetings has been an enjoyable and enlightening experience. I approached Prof. Hand when he was Head of the Department of Statistics at the Open University to enquire into the possibility of my embarking on a postgraduate course in statistics, motivated by the need to become more proficient in longitudinal data analysis. Prof. Crowder was asked to act as co-supervisor, and together they suggested that I adopt a Bayesian approach and make MCMC methods the focus of my work. I am most grateful for being guided in this direction.

During my registration as an Open University student I have been employed at the Institute of Child Health with Prof. David Gadian as my Head of Department. I wish to express my gratitude to him for his continued interest in the work outlined in this thesis, and for his critical contribution to the papers that have arisen from this work. The Institute of Child Health (ICH) works closely with the Great Ormond Street Hospital for Children (GOS), and I consider myself extremely fortunate to be working within this stimulating clinical research environment. The final chapter of this thesis provides some insight into the exciting prospects for applying MCMC methods to important problems in paediatric neurology; I feel privileged to have had the opportunity to work in this environment, surrounded by physicists and clinicians, each of whom are leaders in their respective fields. In particular I wish to thank the following colleagues for their friendship and numerous informative discussions: Prof. Alan Connelly, Dr. Fernando Calamante, Dr. Donald Tournier and Dr. Jo Perthen (MR physicists), Dr. Rod Scott and Dr. Brigitte Vollmer (clinicians), Dr. Mark Lythgoe (neuroscientist) and Sally Dowsett (departmental administrator). Numerous other colleagues, past and present, have contributed to making ICH/GOS a special workplace.

In addition, thanks are due to my internal supervisor, Prof. C. Jones and my third-party monitor, Dr. Karen Vines. Dr. Vines has considerable experience in MCMC methods, and I appreciate her comments on some of the work presented in this thesis. Finally, I wish to thank most sincerely my wife, Wan-Sen, and daughter, Suling, for their understanding and patience during the last eight years; they have been somewhat ignored.

1. INTRODUCTION

1.1 Motivation

Clinical magnetic resonance imaging (MRI) is an important diagnostic tool, especially in neurology. To take an example, MRI is claimed to have revolutionised the understanding of disease activity in multiple sclerosis (Arnold et al., 1997; Miller et al., 1998). Similarly MRI plays a central role in the assessment of a variety of other conditions including central nervous system tissue damage in cerebral ischaemia (stroke). Among the reasons for this prominent role is the superiority of MRI as a method for assessing disease progression, as compared with the alternatives. The diagnostic information provided by MRI can be supplemented by using magnetic resonance spectroscopy (MRS), which is a non-invasive method for performing quantitative biochemical analysis. Among the clinical applications for MRS are the assessment of traumatic brain injury (Vink and McIntosh, 1997), including damage due to birth asphyxia (Cady, 1997; Hope et al., 1984; Hope and Reynolds, 1985) and the diagnosis and management of brain tumours (Ross et al., 1997). In addition to their clinical role, MRI and MRS are widely used in biomedical research. A common feature of many biomedical magnetic resonance-based investigations is the multivariate and longitudinal nature of the acquired data. This applies to both clinical and laboratory studies. Numerous longitudinal magnetic resonance (MR) studies have been published but, in most instances sub-optimum or invalid statistical methods have been adopted. This observation provides the motivation for this thesis.

1.2 Background to the role of MRI and MRS in clinical diagnosis and biomedical research

Magnetic resonance is a non-invasive and non-destructive method for performing diagnostic imaging and biochemical analysis and, as such, is ideally suited to longitudinal studies. Soon after the discovery of the nuclear magnetic resonance (NMR) phenomenon (Bloch et al., 1946; Purcell et al., 1946), NMR spectroscopy became established as an important tool in analytical chemistry, mainly for chemical structure determination, although it is also used for quantitative analysis.

Over the last two decades it has developed into a biomedical research tool of major importance, mainly due to the versatility of the technique. In fact it fulfils two distinct roles. The first of these is an extension of its use in analytical chemistry, and it is now used for performing biochemical analyses within intact and living systems in the form of *in vivo* MR spectroscopy. Its major role is, however, as a clinical imaging technique. The development of MRI started in the 1970s when it was recognised that proton spin density (water) maps can be generated using the NMR method (Lauterbur, 1973). The basis of the method is the Fourier relationship between spin density and the NMR signal acquired in the presence of a magnetic field gradient (Callaghan, 1991, Chapters 1 and 3; see footnote in Section 1.3). This is, in effect, an NMR diffraction phenomenon (Mansfield and Grannell, 1975). MRI emerged as an important diagnostic imaging tool during the late 1980s. In addition to its role in clinical diagnosis it is widely used in biomedical research. Among the reasons for its present-day predominance in medical imaging are: it is a radiofrequency technique and, as such, is not associated with the risks involved with x-ray and other methods that employ ionising radiation; the superior quality of the resulting images when compared with alternative methods, especially in brain imaging; and the fact that MR images can be sensitised to a variety of biophysical processes. As a consequence of the last of these properties, MRI can be used in an immense variety of differing modalities, including conventional imaging modes, angiography (MRA), perfusion imaging and diffusion imaging (Gadian, 1995; King et al., 1997b; Thomas et al., 2000). Although the majority of these imaging methods are based on various types of water (proton) imaging, MR can be used to generate images of important tissue metabolites, thus combining the imaging and spectroscopic modalities (spectroscopic imaging; Gadian, 1995). As mentioned previously, one of the major advantages of MR as a laboratory biomedical research technique is the non-invasive, non-destructive nature of the method. This yields a huge advantage over conventional histology, for example, when a characterisation of progression in pathophysiology is a major objective.

The advantage of the MR method, as applied to biomedical laboratory studies, over the traditional techniques, is made clear by taking, as an example, animal studies designed to establish the

relationship between evolving pathology and the underlying biochemistry and physiology. The data used in Chapters 3 and 4 are typical of the results acquired in this kind of study, although the observation period might be extended to several hours, days, weeks or months, depending on the research question. The traditional approach in which separate histological and biochemical data are acquired requires large numbers of animals due to the necessarily destructive nature of the method. Thus it is not possible to follow individuals over time, and separate groups of animals are required for each observation occasion. Furthermore, separate animals might be required for the histological and biochemical components of the study, making within-subject comparisons impossible, even at a single time point. Among the considerable advantages associated with MRI/MRS is the non-invasive and harmless nature of the method, this facilitating a longitudinal investigative approach. This feature, together with the manner in which MRI can be used in a variety of modes to provide surrogate markers of pathology, in vivo (see, for example, King et al., 1997b, p195), effectively allows the acquisition of simultaneous biochemical, pathological and anatomical data. For this reason multivariate longitudinal studies are extremely common in biomedical MR research. Invariably, huge amounts of data are generated within a single study, and frequently the typical MR researcher is required to undertake analytical problems of considerable complexity. It is unfortunate that inferior statistical methods continue to be used by many MR researchers, with the result that many studies are severely compromised. Accordingly, the main emphasis in the statistical sections of this introduction is on methods for dealing with longitudinal data where the traditional approaches either fail, are not robust or might be dismissed as inappropriate, paying particular attention to the kind of statistical problems that arise in biomedical MR research.

The main part of this thesis describes the reanalysis of data obtained from two published multivariate longitudinal MR studies with the aim of addressing some important biomedical questions. In common with many longitudinal MR studies, one of these takes the form of a nonlinear random coefficients modelling problem, while the other requires some form of

multivariate changepoint analysis. Markov chain Monte Carlo (MCMC) simulation is the method of choice in both cases, and this provided the motivation for this work.

Recent developments in MCMC simulation-based statistical analysis has been accompanied by a resurgence in the Bayesian inferential method with which MCMC is associated. A huge amount has been written on the contentious debate regarding the relative merits of the frequentist and Bayesian approaches. None of this is reproduced in this thesis, for the following two reasons. Firstly, the debate has become an irrelevance because the availability of both software and computing power now enables the analyst to turn to the most appropriate method for any given problem. For example, random coefficients modelling is associated with a variety of well-known difficulties, the traditional reliance on asymptotic approximation being a major concern. Nonlinear random coefficient models are especially problematic. The possibility of avoiding the need for approximation by resorting to MCMC simulation has an obvious appeal. Secondly, clinical researchers in general, and biomedical MR researchers in particular, are faced with a huge range of rather challenging statistical problems. Accordingly, the focus of the introduction to this thesis is restricted to those topics that are of direct relevance to the types of problem and issues that have arisen in the clinical environment in which the candidate works. The majority are practical problems, but there are a few instances in which some apparently theoretical issue has huge clinical implications. The manner in which p-values are used by clinicians is an important example, and Appendix D is devoted to an appraisal of the use of p-values as evidence. The diverse nature of current MR imaging modes has been emphasised in this section because it underlies the multivariate nature of the data acquired in a typical MR study. The following section includes a very brief introduction to the magnetic resonance phenomenon, sufficient to provide some insight into the mechanism by which so much information can be acquired using a single technique.

1.3 The nuclear magnetic resonance phenomenon and MR imaging. A brief summary

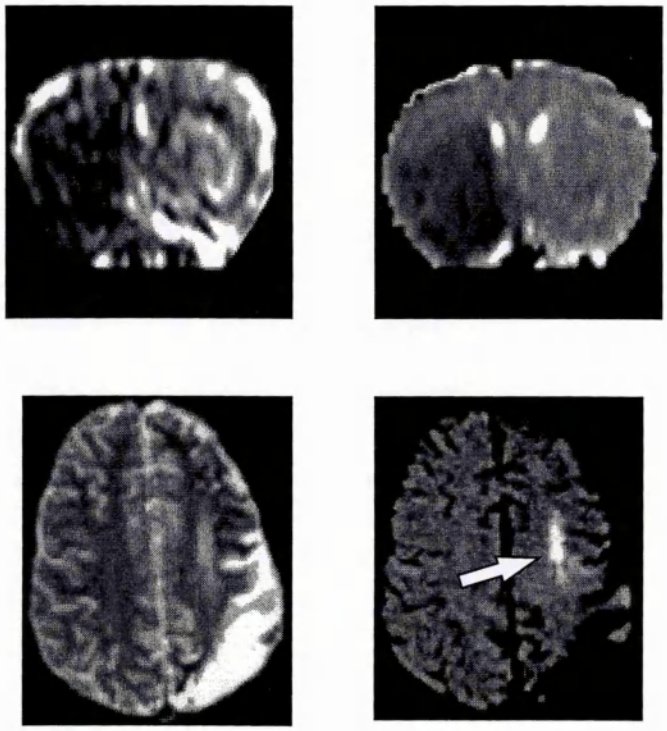
Many nuclei, including the ^{1}H (proton) and ^{31}P nucleus, possess spin angular momentum and an associated nuclear magnetic moment. The application of an external magnetic field causes a quantisation of the energy states associated with these magnetic moments (Carrington and McLachlan, 1967). Both the proton and the ^{31}P nucleus have spin quantum number 1/2, giving rise to two quantised energy levels. Essentially, the nuclear moments can adopt only one of two allowed spatial orientations, i.e., either parallel or anti-parallel to the direction of the applied field. The application of a second, perpendicular radiofrequency magnetic field stimulates transitions between these energy states and this is the origin of the magnetic resonance phenomenon. In MR imaging this radiofrequency excitation is applied as short pulses. These induce a degree of coherence within the nuclear spin system giving rise to a bulk magnetisation that precesses about the static magnetic field. Inductive coupling to a receiver coil generates an electric current and subsequent signal within the receiver system. This signal is acquired in the presence of various time-varying magnetic field gradients, thus generating a magnetic field having a strength that is spatially dependent. These gradients cause a spatial dependence in the resonance condition and give rise to a Fourier relationship[1] between proton density (usually tissue water density) and the NMR signal, thus generating the required spatial dependence in the MR data (Callaghan, 1991).

---

[1] $\rho(\mathbf{r}) = \iiint S(\mathbf{k}) \exp[-i2\pi \mathbf{k}.\mathbf{r} d\mathbf{k}]$, where $S(\mathbf{k})$ is the signal amplitude in k-space, and $\mathbf{k} = (2\pi)^{-1}\gamma \mathbf{G}t$. $\rho(\mathbf{r})$ is the spin density (water density), expressed as a function of $\mathbf{r}$, the nuclear spin co-ordinates. γ is the gyromagnetic ratio of the proton, $\mathbf{G}$ is the (pulsed) magnetic field gradient, and t is time.

As stated previously MR imaging is extremely versatile and is used in a variety of modes. This includes a mode of imaging designed to be sensitive to spin relaxation processes. The spin relaxation rate determines the rapidity with which the spin system returns to equilibrium after radiofrequency excitation. Tissue differences in relaxation rates [the so-called T_1 (longitudinal) and T_2 (transverse) relaxation rates] are used to generate MR contrast and it is this contrast that gives rise to anatomical structure within a conventional MR image (Gadian, 1995). MR images can also be sensitised to a variety of other physical processes, including flow, which is used in MR angiography, and diffusion, as implemented in various forms of diffusion-weighted imaging (King et al., 1997b). An analysis of the data obtained from an MR diffusion study of the response to transient cerebral ischaemia forms a major part of this thesis. To put this in context, Fig. 1.1 shows two sets of images, one acquired in a laboratory diffusion-imaging study of the effects of middle cerebral artery occlusion in the rat, the other acquired in a clinical examination of a 12-year-old child, 2 days after an ischaemic event. The two images shown on the right are diffusion-weighted, while the lower left image is T_2-weighted. Diffusion-weighting causes a reduction in signal-to-noise ratio and a consequent loss of grey-matter to white-matter contrast, with the result that the two diffusion-weighted images lack the structural/anatomical detail of a conventional T_2-weighted image. Nevertheless the diffusion-weighted images provide useful information on the extent of the damage caused by the ischaemic insult. It might be noted in passing that T_2- and diffusion-weighted images contain little direct information on the pathophysiology of the tissue response to ischaemia. q-Space imaging is a diffusion-imaging variant that uses magnetic field gradients in a manner designed to provide structural information at the microscopic (cellular) level (Cory and Garroway, 1990; King et al., 1994) . Localised q-space imaging was developed as an experimental technique for *probing* the microstructural changes that occur during cerebral ischaemia (King et al., 1997a). This technique requires extremely large magnetic field gradients, however, and cannot be implemented on a clinical scanner. Thus, standard diffusion-weighted imaging remains an important diagnostic tool in the management of stroke patients, although it cannot provide direct information on the microstructural changes that occur in pathology, and the mechanistic basis of the image intensity changes that occur in cerebral ischaemia remain to be established.

Fig. 1.1. **MR central nervous system imaging and cerebral ischaemia**



The upper two images were acquired from a rat 4 hours after the onset of an ischaemic insult produced by surgical occlusion of the left middle cerebral artery. A perfusion image (also referred to as a *cerebral blood flow map*) is shown on the left, and a diffusion [trace(D)] image on the right. The perfusion image was obtained using an arterial spin tagging technique (Calamante et al., 1999). The trace(D) image was generated by equating the intensity of each pixel to the trace of the diffusion tensor (see King et al., 1997, for details). The main feature of the perfusion image is the reduction in blood flow in the left hemisphere relative to that in the normal contralateral right hemisphere. The region of reduced intensity in the left hemisphere of the trace(D) image indicates an area of cytotoxic oedema caused by the ischaemic insult. In a clinical setting this tissue would become infarcted. (Images provided by Dr. M. Lythgoe, Institute of Child Health, UCL, London.)

Fig.1.1. continued

The two images in the lower row were acquired from a 12-year-old child 2 days after an ischaemic event. An axial T_2-weighted echo planar image is shown on the left and a diffusion-weighted image on the right. There is a clear distinction between an acute lesion and pre-existing infarcted tissue (chronic lesion). The chronic lesion appears as an area of high intensity in the T_2-weighted image and low intensity in the diffusion-weighted image, while the acute lesion appears as an area of high intensity in the diffusion-weighted image (indicated by the arrow). (Images from the Great Ormond Street Hospital for Children, provided by Dr. F. Calamante, Institute of Child Health, UCL, London.)

1.4 Longitudinal data analysis and related topics. An overview.

A common problem with the classical approach to longitudinal data analysis is the frequent need to force the problem into a format that is not ideal. This might occur at either the design stage or analytical stage in order to conform to some statistical model for which an analysis has been developed. The sums of squares based ANOVA approach to experimental data analysis provides an example. The important principles of experimental design were established in the 1920s and 1930s. The ANOVA calculation was central to the approach and, at that time, computational feasibility was a major consideration. The recommended procedure was to adopt some standard experimental or study design, one for which an analysis had been laid out. A number of texts were exclusively concerned with these designs, the monograph *Experimental Designs* by Cochran and Cox (1957) being particularly well-known. An analysis of variance is reasonably straightforward provided the study or experiment conforms to one or other of the standard plans (details are given in Cochran and Cox, 1957, and Cox, 1992). In contrast, a departure from these standard designs often caused major problems for the analyst. In many cases the failure to conform was unintentional. Unintended incomplete data was, for example, a major problem. Data imbalance due to missing observations is associated with a number of difficulties. Non-uniqueness in the sums of squares introduces a major complication, and the sums of squares may not be related in any simple way to the hypotheses of interest (Rawlings, 1988, Chapter 16; Littell et al., 1991, Section 4.3). Furthermore, some key effects may not be estimable. The latter problem is not, of course, restricted to the ANOVA calculation. An insufficiency of data may result in a condition in which various effects cannot be separated, regardless of the analysis that is adopted. In this case the model in question is intrinsically under-identified. Nevertheless, some of the computational problems associated with missing data disappear when the sums of squares analysis is abandoned. Of course, the problem of possible bias remains if data are missing due to some non-random process.

Improvements in computing power were accompanied by a movement away from the sums of squares ANOVA calculation. The resulting methods are rather less restrictive and the emphasis on adopting a standard design has receded. Thus the prominent role of published treatment plans, as detailed by Cohran and Cox (1957) and Jones and Kenward (1989), for example, has given way to a more flexible approach. Mead (1988, p8), for example, suggested that 'if the set of treatments fits simply with the structure of units using a standard design, then the experiment is already designed.' He recommends, however, that 'the natural pattern of units and treatments should not be deformed in a Procrustean fashion to fit a standard design', and that the available resources (units) should be used in an optimal way, not constrained to achieve computational simplicity (Mead, 1988, pages 8 and 154 et seq.). A major advantage associated with the more modern, likelihood-based modelling approach to statistical inference is the elimination of the need to coerce the problem into a sums of squares/ANOVA format. Consequently, hypotheses of interest can be specified in a more natural and direct manner.

1.4.1 A summary of common non-Bayesian approaches to multilevel and longitudinal data analysis

It is not uncommon to find biomedical research papers in which ordinary least squares (OLS) is used in the analysis of longitudinal and multilevel data. The BLUE property of least-squares estimation depends, however, on the validity of the assumption that the errors are identically and independently distributed (abbreviated *iid*; Rawlings, 1988, p237), and this assumption is not realistic in many longitudinal/multilevel contexts. Correlated errors result in a loss of efficiency in the OLS parameter estimates, *i.e.*, the OLS estimates are no longer minimum variance, although the properties of unbiasedness and consistency are retained (Goldstein, 1987, p18 et seq. and p28 et seq.; Diggle, 2002, Section 4.3). In addition the OLS estimate of the covariance matrix is incorrect, in general, and all measures of precision biased, yielding invalid tests of significance (Goldstein, 1987, p18 et seq. and p28 et seq.; Diggle, 2002, Section 4.3). The direction of the bias in the variance estimates depends on the sign of the autocorrelation function (Theil, 1971, p254 et

seq.; Zeger and Liang, 1992, p1826). A variety of alternative frequentist approaches to longitudinal data analysis have been outlined with differing objectives in mind. These range from simple methods aimed at researchers wanting quick but acceptable methods and, at the other extreme, rigorous methods requiring a considerable degree of statistical sophistication. A number of *repeated measures* review papers have been published including Everitt (1995), Cnaan et al. (1997), Omar et al. (1999) and Sullivan et al. (1999). The texts by Crowder and Hand (1990), Hand and Crowder (1996) and Jones (1993) are among those that provide a comprehensive coverage. A brief summary of the important methods is given in the following pages.

1.4.1.1 Summary measures

The summary measures approach has been recommended as a procedure for those researchers requiring an easily applied method of analysis (Matthews et al., 1990; Everitt, 1995; Omar et al., 1999; Crowder and Hand, 1990, p11 et seq.). Essentially, a summary parameter is chosen, one that captures the principal characteristics of the response or a particular feature of interest. The summary parameter is then treated as a raw observation in a subsequent univariate analysis using some simple textbook method, ANOVA for example. Refinements to the approach have been suggested including a weighted analysis (Matthews, 1993; Omar et al., 1999). The latter procedure is, however, far from straightforward. This is, perhaps, inevitable when a simple method is refined to deal with inadequacies in the original approach. The loss of simplicity defeats the original objective and a more rigorous approach might be adopted at the outset.

1.4.1.2 Repeated measures ANOVA

An alternative approach to longitudinal data analysis is based on a modified ANOVA calculation (referred to as mixed-model ANOVA). It is widely used in biomedical research. While readily applied to studies involving a small number of observation occasions, with no missing data, the method has a number of disadvantages. These include the well-known non-uniqueness in the sums of squares associated with unbalanced data and the problem that arises if subjects differ in their observation times, either due to missing data or because the design does not stipulate well-defined

observation occasions. Various *types of sums of squares* have been defined for the unbalanced case (Rawlings, 1988, Chapter16; Littell et al., 1991). Non-uniqueness in the sums of squares causes a considerable degree of complication for the analyst, however. Although imputation provides an alternative approach to dealing with missing observations (Schafer, 1999; Barnard and Meng, 1999), the problem can be avoided by adopting a mixed-model regression approach. This is among the strong arguments in favour of the latter.

In any discussion on the relative merits of the ANOVA approach to repeated measures (RM) data analysis, it is useful to distinguish two types of RM study. The first type might be described as "intrinsically longitudinal" because the time dependent behaviour is the primary focus of the study. Growth curve studies fall into this category. This is in contrast to the second type in which information on the effect of one or more treatments or factors is sought and an RM design is chosen, although it is not obligatory. In the latter case, time is of no interest in its own right; instead it is a nuisance factor, but one which must be dealt with properly. Many cross-over experiments fall into this latter category. This distinction between types of RM study is important in the context of the ANOVA-based methods of analyses because, although the factorial nature of the cross-over type of design fits naturally into the ANOVA format (despite a potential violation of the fundamental ANOVA assumptions), this is not true of growth curve data. Forcing growth curve data into an ANOVA structure can be extremely detrimental to the analysis and the mixed-model regression approach is preferred.

1.4.1.3 Mixed-model regression

The theory that underlies the likelihood based approach to mixed-model regression has been outlined in numerous texts including Crowder and Hand (1990), Jones (1993), Hand and Crowder (1996), and Littell et al. (1996). Laird and Ware (1982) is regarded as the seminal publication, while Jennrich and Schluchter (1986) have provided a useful tutorial paper on the Laird-Ware model.

Following Laird and Ware (1982) and Littell et al. (1996) the mixed model is written as

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \varepsilon \qquad [1.1]$$

$$Var\begin{bmatrix}\gamma \\ \varepsilon\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}\end{bmatrix}, \qquad [1.2]$$

where **X** and **Z** are the fixed- and random-effect design matrices, β is a vector of fixed-effect coefficients, γ a vector of random-effect coefficients and ε the vector of within-subject random errors. **G** is the random-coefficients covariance matrix (between-subject variation) and **R** the within-subjects covariance matrix.

The variance of y is given by

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}} + \mathbf{R} \qquad [1.3]$$

and the parameter covariance matrix is

$$\mathbf{C} = \begin{bmatrix}\mathbf{X}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\end{bmatrix}^{-}, \qquad [1.4]$$

where the final superscript indicates the generalised inverse. It has been shown (McLean and Sanders, 1988; see Littell et al., 1996, p500-1 for additional information) that the fixed-effect covariance matrix is given by $\hat{\mathbf{V}}_{\beta} = (\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}^{-1}\mathbf{X})^{-}$. In practice, **R** and **G** are unknown, and must be estimated using, for example, maximum likelihood (ML) or restricted maximum likelihood (REML) methods. The relevant log likelihood functions are (Littell et al., 1996, p498)

$$\text{ML: } l(\mathbf{G},\mathbf{R}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{n}{2}\log \mathbf{r}^{T}\mathbf{V}^{-1}\mathbf{r} - \frac{n}{2}[1 + \log(2\pi / n)] \qquad [1.5]$$

$$\text{REML: } l(\mathbf{G},\mathbf{R}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log\left|\mathbf{X}^{T}\mathbf{V}^{-1}\mathbf{X}\right| - \frac{n-p}{2}\log \mathbf{r}^{T}\mathbf{V}^{-1}\mathbf{r} - \frac{n-p}{2}\left\{1 + \log\left[2\pi / (n-p)\right]\right\}. \qquad [1.6]$$

These are used as the objective function in some standard minimisation routine (Fletcher, 1980) to obtain the required maximum. Estimates of β and γ are obtained by solving the mixed model equations, giving

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}^{-1}\mathbf{y} \qquad [1.7]$$

$$\hat{\boldsymbol{\gamma}} = \hat{\mathbf{G}}\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \qquad [1.8]$$

Inspection of the second of these equations reveals the interesting property whereby the random coefficient estimates display *shrinkage towards the origin*, the degree of shrinkage depending on the relative magnitude of the between-subject and within-subject variation. The resulting estimates are a weighted combination of the individual least-squares estimate and zero. Laird and Ware (1982, p 966-7) describe these random coefficient estimates as Empirical Bayes.

Given the linear hypothesis $\mathbf{Lb} = \mathbf{h}$, a traditional test is based on the Wald statistic

$$W = (\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{h})^{\mathrm{T}}(\mathbf{L}\mathbf{V}_{asy}\mathbf{L}^{\mathrm{T}})^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{h}) \qquad [1.9]$$

where $\mathbf{V}_{asy}$ is the asymptotic covariance matrix given by the inverse of the second derivative of the likelihood with respect to the model parameters. W is distributed as χ^2. Clearly, the application of the Wald statistic to small sample problems is dubious, and a number of alternative tests have been suggested. The Wald statistic divided by the rank of the design matrix (i.e., the numerator degrees of freedom), for example, has an F-distribution (Berk, 1987). A variety of other conversion schemes have been suggested, the details of which depend on the design structure (Schluchter and Elashoff, 1990; Dixon and Merdian, 1992).

A related procedure is to form an F-ratio using (Littell et al., 1996, p502)

$$F = \frac{(\mathbf{L}\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\boldsymbol{\gamma}}\end{bmatrix} - \mathbf{h})^{\mathrm{T}}(\mathbf{L}^{\mathrm{T}}\hat{\mathbf{C}}\mathbf{L})^{-1}(\mathbf{L}\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\boldsymbol{\gamma}}\end{bmatrix} - \mathbf{h})}{\mathrm{rank}(\mathbf{L})}. \qquad [1.10]$$

But the question regarding the denominator degrees of freedom remains. One approach is to use a value given by the number of observations minus the rank of the design matrix, which is the standard ANOVA definition. Other alternatives have been suggested, however (see, for example, Littell et al., 1996).

An interesting feature of the Laird-Ware mixed-model regression method is the manner in which it embraces both Bayesian and frequentist concepts. On the one hand p-value calculation and comparison against a critical value is a fundamentally frequentist approach while *shrinkage towards the mean* is an essentially Bayesian property. Laird and Ware (1982) describe the latter behaviour as Empirical Bayes.

1.4.1.4 Multilevel models

A category of mixed-effect regression models consisting of a nested or hierarchical structure of random effects is referred to as the multilevel model. Multilevel modelling is well-known to educational research statisticians, mainly due to the work of Goldstein and co-workers and the resulting computer program, *MLn*, which is designed specifically for performing hierarchical random-effects modelling calculations (see, for example, Goldstein et al., 1994; Goldstein, 1995). The focus of the program is maximum likelihood estimation based on iterative generalised least-squares (IGLS) (Goldstein et al., 1994, pages 1645 and 1654), although recent releases of *MLn* have been expanded to include a variety of developments, including Gibbs sampling.

1.4.1.5 Generalised estimating equations

Two distinct likelihood-based approaches to longitudinal data analysis have been developed. The preceding sections have focused on methods that involve an explicit modelling of between-subject heterogeneity in the regression parameters. This includes the Laird-Ware mixed-model in which subject-specific effects are assumed to conform to some parametric distribution. Zeger et al. (1988) refer to this approach as a subject-specific (SS) analysis. The alternative method involves a modelling of the population-average response as a function of covariates, without explicitly accounting for between-subject heterogeneity. The resulting regression coefficients apply to the population only, and provide no information about individuals. These models are referred to as population averaged (PA) models or, alternatively, marginal models. The focus is fixed-effect parameter estimation. Additional parameters are required to define the variance structure, but these are regarded as nuisance variables of no interest in their own right. An attractive feature of

PA modelling is that the only requirement for consistent inference is the validity of the link function (Zeger et al., 1988, p1059). Zeger et al. (1988) outline a generalised estimating equations (GEE) approach to both SS and PA modelling. The main objective is to extend the quasi-likelihood generalised linear modelling technique to applications involving correlated data. [Quasi-likelihood methods apply when the likelihood cannot be specified fully. Furthermore, the assumption that the model belongs to the exponential family of generalised linear models can be relaxed (Garthwaite et al., 1995, p266 et seq.; Armitage and Colton, 1998, p1626 et seq.).] An important characteristic of the GEE approach is the need to specify a structure for the so-called working correlation matrix, although validity of the standard error estimates does not depend on the selected structure matching the true correlation structure (Burton et al., 1998, and references therein). Robustness in the standard errors is among the attractive features of the method[1]. The question arises regarding the applicability of PA models. Clearly, subject-specific modelling is required when individuals are the focus of the study. In some instances, however, the objective is to acquire data on average treatment effects or average covariate effects (Omar et al., 1999), in which case PA methods can be adopted.

1.4.2 Bayesian inference

1.4.2.1 Bayes' theorem and inverse probability. Introduction

Bayes' theorem enables an analyst to make probabilistic statements about model parameters, θ, conditional on the acquired data, $\mathbf{y}$, as follows. The joint probability $p(\theta,\mathbf{y})$ is written as a product of the prior probability, $p(\theta)$, and the sampling distribution, $p(\mathbf{y}|\theta)$, i.e.,

$$p(\theta,\mathbf{y}) = p(\theta)p(\mathbf{y}|\theta), \qquad [1.11]$$

---

[1] Appendix A provides a table of GEE parameter estimates and standard errors obtained in an analysis of the RATS growth curve data of Gelfand et al. (1990) using SAS PROC GENMOD with several alternative working correlation matrix structures. Mixed model regression and MCMC results are included for comparison.

from which it follows that

$$p(\theta|\mathbf{y}) = \frac{p(\theta,\mathbf{y})}{p(\mathbf{y})} = \frac{p(\theta)p(\mathbf{y}|\theta)}{p(\mathbf{y})}. \qquad [1.12]$$

Box and Tiao (1973, p5) state that $p(\mathbf{y}|\theta)$ is a probability density of any hypothetical data, $\mathbf{y}$, before data acquisition. Written in this form the sampling distribution $p(\mathbf{y}|\theta)$ is a density function of $\mathbf{y}$ for fixed θ. But following data collection y is fixed and known. Then, assuming the purpose of the investigation is to estimate θ, it is usual to regard $p(\mathbf{y}|\theta)$ as a function of θ for fixed $\mathbf{y}$. $p(\mathbf{y}|\theta)$ is then referred to as the likelihood, a conceptual change that is often emphasised by using the notation $l(\theta|\mathbf{y})$. Thus we arrive at Bayes' theorem, which states that posterior probability is proportional to the product of prior probability and likelihood, i.e.,

$$p(\theta|\mathbf{y}) \propto l(\theta|\mathbf{y})p(\theta). \qquad [1.13]$$

Bayes and Laplace are credited as having worked independently to arrived at this relationship for calculating inverse probability. The important point is that a reversal of conditioning is achieved, thus providing the information that is sought by the researcher in most applications, *i.e.*, the probability distribution of some parameter, for example, given the information provided by an experiment. The latter takes the form of the likelihood of the observed data, conditional on some model or hypothesis. Bayes' theorem encapsulates the process of learning. In this context it is important to distinguish subjective and mathematical probability (see, for example, Box and Tiao, 1973, p14). This issue is discussed at some length in Appendix D, but the main point is that subjective probability relates to degrees of belief, and the Bayesian approach provides a mathematical expression of the researcher's degree of belief with respect to some proposition. Box and Tiao (1973, p20) state that Bayesian statistical inference is concerned with the state of nature in probability terms. This information is provided by the posterior distribution. To summarise, $p(\theta|\mathbf{y})$ tells the researcher what can be inferred about the parameter, θ, given the data, $\mathbf{y}$, and relevant prior knowledge, as represented by the distribution $p(\theta)$. Newly acquired data, $\mathbf{y}$, modifies prior belief through the likelihood function, $l(\theta|\mathbf{y})$, the latter of which expresses the information content of the data relating to θ.

1.4.2.2 Comments on the Bayesian method and some properties of Bayesian models

A number of general points are worth noting regarding Bayesian inference. Firstly, the well-known criticism directed at the Bayesian method is the need to specify a prior distribution, which is mandatory. This appears contradictory to the premise that all scientific investigation should be objective, with no place for subjectivity. Critics suggest that, in some cases, the problem surrounding the arbitrary nature of the adopted prior is overwhelming. A host of important questions arise. Box and Tiao (1973, p9) address the issue, noting that some prior knowledge enters into all inferential systems. They suggest that this is overlooked on many occasions, and give the application of sampling theory as an example in which prior belief that a given distribution is exactly normal might be invoked, together with the belief that each observation is independent and has the same variance. Subsequent residuals analysis may show one or other of these assumptions to be invalid. They point out that among the attractive features of the Bayesian method, compared with the sampling approach, is the relative ease with which the prior assumptions, for example, the normality assumption, can be relaxed. Nevertheless, in many studies an overriding requirement for objectivity calls for the use of so-called noninformative priors, many statisticians making the point that in any scientific investigation 'the data must speak for themselves'. Clinical trials are an interesting exception, and the recommendation is to formulate several prior distributions representing differing clinical opinion, including extreme forms of prior belief (Spiegelhalter et al., 1994). This permits an assessment of the degree to which a given trial result is robust to differences in prior belief, information that carries considerable clinical importance.

Parameter shrinkage

Parameter shrinkage is a well-known Bayesian property. For example, in the simple case in which both the prior and likelihood have Gaussian distributions

$$\begin{aligned} \theta &\sim N(\mu, \tau^2) \\ y|\theta &\sim N(\theta, \sigma^2) \end{aligned} \qquad [1.14]$$

the marginal distribution of y

$$m(y) = \int l(\theta|y)p(\theta)d\theta \qquad [1.15]$$

is $N(\mu, \sigma^2 + \tau^2)$ and the posterior distribution, $p(\theta|y)$, is Gaussian with mean and variance given by

$$\begin{aligned} E(\theta|y) &= B\mu + (1-B)y \\ \text{var}(\theta|y) &= (1-B)\sigma^2 \end{aligned} \qquad [1.16]$$

respectively (Carlin and Louis, 2000, p11; Lee, 1989, p38-39; Gelman et al., 1995, p43), where $B = \sigma^2 / (\sigma^2 + \tau^2), 0 \le B \le 1$. The posterior mean is a weighted average of the prior mean and the estimate provided by the data. Accordingly, the observed estimate shrinks towards the prior mean, the amount of shrinkage depending on the relative precision of the prior and observed data. Bayesian shrinkage provides a compromise between bias and variance (Carlin and Louis, 2000, p11).

Marginal distributions

In many applications interest centres on a subset of model parameters, while the remaining elements are regarded as nuisance variables. Formally, given $\theta = (\phi, \varphi)$, with interest restricted to ϕ, the posterior distribution for ϕ is obtained from the joint posterior distribution by marginalization using

$$p(\phi|y) = \int p(\phi, \varphi|y)d\varphi \qquad [1.17]$$

Prediction

Inference concerning unknown observables is especially relevant to clinical decision making and diagnosis. Gelman et al. (1995, p8 et seq.) provide the following details. In the absence of any data the distribution of the unknown observable, y, is

$$p(\mathbf{y}) = \int p(\mathbf{y},\theta)d\theta = \int p(\mathbf{y}|\theta)p(\theta)d\theta. \quad [1.18]$$

After data acquisition the prediction for an unknown observable, $\tilde{\mathbf{y}}$, is given by the posterior predictive distribution, which takes the form

$$p(\tilde{\mathbf{y}}|\mathbf{y}) = \int p(\tilde{\mathbf{y}},\theta|\mathbf{y})d\theta = \int p(\tilde{\mathbf{y}}|\theta,\mathbf{y})p(\theta|\mathbf{y})d\theta = \int p(\tilde{\mathbf{y}}|\theta)p(\theta|\mathbf{y})d\theta, \quad [1.19]$$

where, as explained by Gelman et al. (1995, p9), the last equation follows from the conditional independence of $\tilde{\mathbf{y}}$ and $\mathbf{y}$, given θ.

1.4.2.3 Posterior distribution summary methods

The variance of the population posterior distribution of a given parameter vector combines the uncertainty in the parameter estimates with the true underlying natural variability in the parameters. An important issue that arises in any Bayesian analyses, but especially when implemented using MCMC, is the manner in which this distribution might be characterised. One approach is to use a Bayesian analogue of the frequentist point and interval estimates. In large sample applications the normal approximation applies, and the mode plus or minus two standard errors can be used to characterise the posterior distribution. In small sample analyses, the normal approximation does not necessarily hold and the choice of point estimate remains an issue for the analyst. Carlin and Louis (1996, p42 et seq.) provide some useful recommendations and suggest that for asymmetric posteriors the median might be used, since it lies between the mode and mean. The posterior mode is equal to the maximum likelihood estimate when the prior distribution is flat and is referred to as the *generalised maximum likelihood* (Carlin and Louis, 1996, p39). A variety of approaches to specifying posterior uncertainty have been suggested including the *central posterior interval*, which is the range of values lying within the $100(\alpha/2)\%$ posterior probability limits. An alternative summary is provided by the *highest posterior density* (HPD), which is the region containing $100(1-\alpha)\%$ of the posterior probability, subject to the condition that the density

within the interval is everywhere higher than outside. The HPD is identical to the central posterior interval if the former is symmetric and unimodal, but is narrower otherwise (Carlin and Louis, 1996, p44). Finally, a comment on terminology is required, namely that Lindley (1965, p15) uses the expression *Bayesian confidence interval* for posterior intervals with some specified coverage, while Edwards et al. (1963, p213) use the term *credible interval* for any interval containing 95% of the posterior density, noting the existence of an unlimited number of credible intervals for any specified probability.

1.4.3 Hierarchical and random effect models. Empirical Bayes

Useful introductions to hierarchical models and empirical Bayes are given in Gelman et al. (1995, Chapter 5) and Carlin and Louis (1996, Chapter 3). The terms *hierarchical* and *random-coefficients* are used synonymously on occasions although, in the modelling context, the term *hierarchical* refers to the nested structure that arises naturally in many random coefficient applications. These models apply to problems in which information is available at several levels within a nested structure of observational units and arise automatically, therefore, in the longitudinal data modelling context. An alternative use of the term *hierarchical* is made in connection with the Bayesian treatment of random coefficient models. Excepting those applications requiring informative priors in order to represent existing opinion, the usual demand for scientific objectivity necessitates the use of uninformative priors. A common approach to the specification of uninformative priors is to defer the point at which values are assigned to the distribution parameters. This is achieved through the introduction of a hierarchy of priors (Gelman et al., 1995, Chapter 5; Carlin and Louis, 1996, Chapters 2 and 3). Thus the prior distribution parameters at one stage are specified in terms of a higher level of prior distributions. Several levels can be introduced giving rise to an hierarchical model (Lindley and Smith, 1972; Carlin and Louis, 1996, Section 2.1). The terms *hyperprior* and *hyperparameter* are used in this context. A simple example is considered in Section 1.6.4.3 which deals with various computational issues surrounding hierarchical models.

An extensive literature exists on random-coefficient and hierarchical models, and this subsection focuses on a few topics of particular relevance to nonlinear random-effect models since this is especially relevant to the V_{DC}-ADC study that forms the major component of this thesis. In addition, a few topics of particular relevance to the clinical environment in which the candidate works have been included.

Nonlinear random effect models

A brief outline of the Laird-Ware two stage model is given in Section 1.4.1.3, the main feature of which is the separation of the fixed- and random-effects and their associated design matrices. In general, this separation is not possible when the response function is nonlinear. Nonlinear random-effect models play a major role in pharmacokinetic research and a huge effort has been made over a period of many decades in the search for useful analytical methods. The text by Davidian and Giltinan (1995) provides detailed coverage of the earlier methods, but most of this work has been superseded by MCMC. Hand and Crowder, (1996, Chapter 8) outline an alternative approach based on Gaussian estimation. Roe (1997) provides an assessment of the better-known algorithms through a comparative study of seven nonlinear random-effects modelling programs.

Mager and Göller (1997) give a useful introduction to the problems that typically arise in modelling toxicokinetic (pharamokinetic) data. The nonlinear problem can be expressed as

$$\mathbf{y}_i = g(t, \mathbf{X}_i, \boldsymbol{\beta}_i) + \boldsymbol{\varepsilon}_i \qquad [1.20]$$

where t is time, $\mathbf{X}_i$ is a matrix of covariates for the ith subject and $\boldsymbol{\beta}_i$ is a vector-valued random variable distributed about the population vector $\boldsymbol{\beta}$, i.e., $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i$. The precision of the parameter estimates provided by the traditional methods is difficult to evaluate. Similarly, an evaluation of the variance of any nonlinear function of the primary model parameters is problematic. Referring to the earlier approaches, Mager and Göller (1997) state that biased estimation is the rule and not the exception (even asymptotically) and the extent of the bias can be difficult to evaluate. But there was no alternative prior to the development of sampling methods, and researchers were

forced to rely on approximate asymptotic inferences and solutions. This was a particular problem for researchers involved in small sample studies.

Pharmacokinetic and toxicokinetic studies are a crucial component of drug development and, given the analytical difficulties referred to in the preceding paragraph, this provided the motivation for the development of improved statistical techniques. Sampling approaches to hierarchical modelling and longitudinal data analysis received particular attention. The resulting methods have a direct application to other areas of clinical research and beyond. Typical of the information sought in a pharmacokinetic study is the probability distribution of key characteristics, rates of metabolism or excretion, for example, and the manner in which these depend on various covariates, including body mass, age and possibly gender. Pharmacokinetic behaviour is typically modelled as a mixture of exponentials, and a common feature of these models is chronic ill-conditioning. Although sparse sampling of the longitudinal response might be unavoidable due to practical or clinical limitations, the model itself may be fundamentally under-identified, given a realistic physiological model. Both scenarios are common in pharmacokinetic research, and this causes major analytical problems. The fact that various key parameters are impossible to estimate from the data alone can be inescapable. Among the solutions adopted in earlier work was the application of crude constraints by assigning fixed values to the offending variables. But this leads to inaccurate estimates and underestimated uncertainty. A superior approach is to formulate the model in terms of parameters that have physical meaning, with informative priors assigned, as appropriate, using published physiological data (Gelman et al., 1996a). Hierarchical modelling lies at the centre of much of this work (see, for example, Wakefield, 1996), since individual-specific and/or subgroup-specific parameter estimation is an important aspect of pharmacokinetics, and hierarchical models provide the required estimates (Wakefield 1996; Gelman et al., 1996a).

Gelman et al. (1995, p119) make the point that, in most cases, the formulation of a given hierarchical problem in terms of a nonhierarchical model is inappropriate because the resulting

increase in parameter number causes the model to overfit the data in the sense of producing models with poor prognostic capability, despite the good fit obtained with the original data. This remark is aimed, presumably, at models in which units within an hierarchical structure (e.g., individuals in a longitudinal study, or centres within a multi-centre clinical trial) are modelled using fixed-effect terms. The key to hierarchical modelling is the removal of the effects of severe overparameterisation through the distributional constraint imposed on each of the unit-specific parameters within the hierarchical structure. An additional detail to note in relation to hierarchical modelling is the distinction between subject-level and observation-level covariates (Cnaan et al., 1997) the former of which remain constant over repeated observation within a unit while the latter may vary. Congdon (2001, Section 7.6) provides an introduction to time-varying coefficient models and linear dynamic modelling.

An additional issue that arises in hierarchical random-coefficients modelling is the method used to specify the final stage parameters. Given an hierarchy of priors with unknown parameters at each stage, the hierarchy must be terminated at some stage and the parameters given specific values (Lindley and Smith, 1972). One approach to this assignment, referred to as *empirical Bayes* (EB), uses the observed data to generate values for the final stage parameters. For example, the final stage parameters can be set at those values that maximize the marginal likelihood. An illustration of the distinction between a fully Bayesian analysis and empirical Bayes is provided by Smith et al. (1995) who compared various Bayesian approaches to a random-effects meta-analysis of trial data on the effectiveness of antibiotic treatment in respiratory tract infection. Using the notation r_i^C for the number of respiratory tract infections among n_i^C subjects in the control group within the ith trial, and assuming probability p_i^C of developing infection, with an equivalent notation for the treatment group, they compare the two analytical approaches using the model

$$
\begin{aligned}
r_i^C &\sim Bin(p_i^C, n_i^C) \\
r_i^T &\sim Bin(p_i^T, n_i^T) \\
\text{logit}(p_i^C) &= \mu_i - \delta_i / 2 \\
\text{logit}(p_i^T) &= \mu_i + \delta_i / 2 \\
\delta_i &\sim N(d, \sigma^2)
\end{aligned}
\qquad [1.21]
$$

where $\delta_i = \text{logit}(p_i^T) - \text{logit}(p_i^C)$ is the true treatment effect on a log-odds scale. They describe a process for assigning values to the prior distribution parameters in the fully Bayesian treatment, including the use of conservative prior guesses for the population and individual trial treatment effects. Similarly, parameters relating to the prior distribution of the between-trial variance (σ^2) are obtained by guesswork. In contrast an empirical Bayes analysis might use d and σ^2 estimated by moment-matching.

An important feature of hierarchical models is the distinction between *types* of posterior distribution that might be of interest. Thus, given a model characterised in terms of parameters (θ) and hyperparameters (ϕ), interest might focus on the distribution of future observations ($\tilde{\mathbf{y}}$), conditional on an existing θ_j or, alternatively, the distribution of future observations conditional on future values for θ_j ($\tilde{\theta}_j$). The required posterior predicted values can be obtained by drawing appropriate samples, the latter requiring an initial draw from $p(\phi)$ to generate $\tilde{\theta}_j$, followed by a subsequent draw from $\tilde{\theta}_j$ to generate $\tilde{\mathbf{y}}$ (Gelman et al., 1995, p127). An additional point of practical importance relates to the distinction between fixed- versus random-effect attributes and the manner in which these are assigned to each term in the model under consideration. Smith et al. (1995) discuss this issue in relation to meta-analysis, but it is central to the kind of brain image analysis in which the candidate is engaged. Referring to the meta-analysis example, the attribute assigned to *centres* depends on the manner in which centres enter into the trial and whether

interest is restricted to the individual centres recruited into the study (fixed-effect[1]) or, alternatively, to some wider population of which the study centres are a random sample (random-effect attribute[1]). Similar considerations apply to image analysis. A brain MR image consists of pixels nested within anatomically defined regions. Thus, given a set of serial images for each individual within a group, the research question might involve comparisons among a small number of distinct, predefined anatomical regions, with pixels nested within regions, in which case it is clear that regions should be treated as a fixed effect. Pixels within a homogenous region should be specified as random. Alternatively, regions-of-interest might be selected, but not based on anatomical structure. For example, a set of arbitrary grey-matter regions (or white-matter regions) might be selected for study. Then, in the absence of a defined anatomical basis for selection, and depending on the selection criteria, regions might be given a random-effects attribute. Smith et al. (1995) compared the difference in results obtained in their meta-analysis by changing the *trial* attribute from *random-effect* to *fixed-effect*, and found that the trial-specific treatment-effect 95% intervals obtained with the random-effects attribute were narrower than those generated using the fixed-effects model, and were drawn towards the overall mean. In contrast, the interval for the population treatment effect was wider (Fig. 4 in Smith et al., 1995).

---

[1] A fixed effect attribute is assigned when the units under consideration (centres in the trial context) are predefined and all inferences are restricted to these units with no reference to some broader population. Furthermore, the same fixed-effect units can be selected, in principle, in a replicate study. In this sense their inclusion is reproducible. Given a fixed effect attribute, the location differences among n individual levels/units of the effect in question (centres, for example) are incorporated into the model via n-1 fixed-effect design matrix columns (dummy variables), together with the associated regression model coefficients. In contrast, a set of n random coefficients is required under the random-effects model, but these are constrained by the assumed distribution. These coefficients enter into the model via n random-effect design matrix columns and associated variance component.

A caveat is required at this point. Although the units selected for a given study might be the focus of interest, as opposed to some wider population, a random attribute is sometimes adopted in order to impose the constraints provided by the assumed distribution. This approach can be used to advantage given data containing outlier/spurious observations. Furthermore, given a large number of units (eg., a large number of centres within a multicentre study) the number of fixed-effect parameters can become excessive under a fixed-effect model, and a considerable reduction in number can be achieved by using a random coefficients specification. In this context Gelman et al. (1995, p119) comment on the improved predictive performance that can sometimes be achieved by adopting a hierarchical random-effects model due to the resulting reduction in fixed-effect parameter number and random-coefficient distribution constraints.

1.5 Bayesian computation

A key component of Bayesian statistical inference is posterior interval evaluation or some other form of posterior characterisation. Often interest is restricted to a subset of the model parameters, and the required marginal posterior distribution is obtained from the joint posterior distribution by integrating over the remaining parameters. Thus, given

$$p(\theta|\mathbf{y}) = \frac{l(\theta|\mathbf{y})p(\theta)}{\int l(\theta|\mathbf{y})p(\theta)d\theta}, \qquad [1.22]$$

the analytical problem might be specified using the partitioning $\theta = (\theta_1, \theta_2)$, in which interest focuses on the parameter vector θ_1, exclusively. θ_2 is a vector containing the remaining parameters, often referred to as nuisance parameters. The required marginal distribution is obtained from the joint posterior by integrating over θ_2, using factorisation if this provides a simplification of the problem. Thus the marginal distribution $p(\theta_1|\mathbf{y})$ is given by

$$\begin{aligned} p(\theta_1|\mathbf{y}) &= \int p(\theta_1, \theta_2|\mathbf{y})d\theta_2 \\ &= \int p(\theta_1|\theta_2, \mathbf{y})p(\theta_2|\mathbf{y})d\theta_2. \end{aligned} \qquad [1.23]$$

In some applications the integration can be simplified by using appropriate conjugate priors, chosen to produce a posterior distribution that can be evaluated analytically. Simple linear models

with normal likelihoods are one type of problem in which this can be achieved. But in many practical applications, the analyst is faced with the evaluation of complex, multidimensional integrals, often of high order. Furthermore, it is usually necessary to perform additional calculations to generate posterior distribution summary information, quantiles, for example, and this requires additional integration. Given the analytical intractability of the integration in many applications a numerical approach was, for many years, the only alternative. Gaussian quadrature (Press et al., 1988, p131) was among the numerical techniques used to perform the integration, but this does not work well when applied to multidimensional integrals of high order (more than about 10 parameters). Although Monte Carlo integration is not new and offered an alternative to the standard numerical methods (Hammersley and Handscomb, 1964) CPU speed was, until recently, a limiting factor. For many years these practical difficulties were a major obstacle and prevented the implementation of various theoretical developments. For example, Lindley and Smith (1972) provided a detailed treatment of Bayesian hierarchical models and yet computational limitations remained an important practical issue for more than a decade. This is illustrated by the work of Racine et al. (1986) who discuss the application of Bayesian statistical models to various problems in pharmaceutics. Their paper includes a treatment of hierarchical nonlinear pharmacokinetic models, and refers to the need for efficient routines for performing the numerical integrations (Racine et al., 1986, p117). Given these almost insurmountable difficulties, the problem was, for many years, circumvented by adopting computationally less demanding approaches to providing posterior summary information. In particular, algorithms were developed for finding posterior modes and these were used in conjunction with various approximation methods for obtaining dispersion estimates. The EM algorithm and normal (asymptotic) approximation were especially important in this context and a very brief outline of these methods is given in the following pages. However, these techniques have been superseded to a large extent by modern Monte Carlo simulation algorithms. The latter effectively allow a complete characterisation of the posterior distribution. A major reason for the current movement in favour of Bayesian statistical inference is that the CPU intensive simulation methods that facilitate its implementation have become feasible due to recent developments in computer technology. Traditional numerical integration no longer

plays a prominent role and is not covered in any detail in this brief account of Bayesian computational methods. Although the same might be said of the EM algorithm and other methods for finding marginal posterior modes, these techniques and the associated normal approximation methods are not entirely redundant. These topics are, therefore, included (Section 1.5.2 and 1.5.3). The following overview includes very little mathematics because the aim is to provide a brief account of Bayesian computational methods with an historical perspective and an emphasis on the underlying concepts.

1.5.1 Numerical integration

The term *numerical integration* encompasses both stochastic simulation and the more traditional algorithmic integration methods (Simpson's rule and Gaussian quadrature, for example). For many years Gaussian quadrature and related traditional methods were all that remained when asymptotic approximation was inapplicable or insufficient in accuracy. Numerous texts have been published with sections devoted to these methods (Gerald, 1978; Conte and de Boor, 1981; Press et al., 1988) and this material is not reproduced in this introduction. As stated above, traditional numerical integration is no longer prominent among the methods used for Bayesian computation.

1.5.2 Normal approximation

An approach to Bayesian computation that was widely used prior to the availability of present day computing power was based on the combined use of some mode-finding algorithm and asymptotic approximation. Despite the development of iterative MCMC, these older methods are not entirely redundant since it has been suggested that posterior approximation should be used to generate the set of overdispersed starting points required for multiple chain MCMC simulation (Gelman and Rubin, 1992b, p459-460; Gelman et al., 1995, pages 322 and 335; see Section 1.6.1 for additional details). Furthermore, normal approximation is used by some analysts to check MCMC results for gross error (Gelman et al., 1995, p322). Among the methods that have been employed to perform the first-stage posterior mode identification are the standard minimisation techniques, including

Newton-Raphson (Fletcher, 1980), and conditional maximisation (Gelman et al., 1995, p272). The latter refers to a stepwise ascent procedure in which each parameter is adjusted in turn within each step of an iterative sequence. The EM algorithm is an alternative and widely used mode-finding technique. Subsequent to mode identification some asymptotic method, normal approximation or Laplace's method, for example, can be used to obtain an analytical approximation to the posterior distribution in the region of the mode. Finally, this approximation is used to generate the required joint posterior distribution summary statistics. The latter might take the form of posterior intervals or standard errors. Alternatively, the resulting normal approximation can be used in a subsequent simulation to generate a posterior sample. To illustrate this procedure, Gelman et al. (1995, Section 9.8) give an hierarchical normal model example in which a factorisation of the joint posterior is used, together with normal approximation and subsequent sampling, to generate samples from the joint posterior distribution and thence to calculate summary medians and other quantiles.

The validity of any posterior asymptotic approximation depends, in part, on the likelihood, which must take the form of a narrow peak relative to the prior (Carlin and Louis, 1996, p142). This requirement will be met given a sufficient number of observations, thus ensuring that the prior is relatively flat over the region in which the likelihood is non-negligible. Then, subject to various regularity conditions, the Bayesian Central Limit Theorem applies (Carlin and Louis, 1996, p142), the essence of which is that the posterior distribution can be approximated by a multivariate normal distribution with mean equal to the posterior mode and covariance matrix estimate given by the Hessian matrix (i.e., the matrix of log posterior density second partial derivatives with respect to the model parameters). This quadratic approximation to curvature is based on a Taylor series expansion. Specifically $p(\theta|\mathbf{y}) \approx N(\hat{\theta},[\mathbf{I}(\theta)]^{-1})$, with $\mathbf{I}(\theta)$ equal to minus the Hessian of the log posterior evaluated at the mode (Gelman et al., 1995, p95; Carlin and Louis, 1996, p 142), where the matrix of second partial derivatives has elements

$$-\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log p(\theta|\mathbf{y})\right]_{\theta=\hat{\theta}}. \qquad [1.24]$$

This is referred to as the generalised observed Fisher information matrix, generalised in the sense that it depends on the prior in addition to the likelihood. Under the large sample assumption the prior is locally uniform, however, suggesting an alternative approach in which the prior is ignored completely. The maximum likelihood estimate can then be taken as an approximate posterior mode, while the conventional Fisher information matrix provides an approximation to the posterior covariance matrix (Gelman et al., 1995, p101; Carlin and Louis, 1996, p143). With this modification the calculation reduces to the standard maximum likelihood method that is widely used in a number of classical applications including mixed model regression analysis (Section 1.4.1.3), generalised linear modelling (Dobson, 1990, Chapter 4) and nonlinear modelling (Jones, 1993, Section 7.1). These are well established areas of statistics in which some standard minimisation routine, Newton-Raphson or the method of scoring, for example, is used to obtain the maximum likelihood and covariance matrix estimates. Finally, it has been suggested (Gelman et al., 1995, p275) that the multivariate Student-*t* distribution might be employed to extend the applicability of asymptotic approximation to small sample problems. In addition, asymptotic approximation can be incorporated into a two-stage procedure in which the approximate distribution is refined via importance resampling (Gelman et al., 1995, p291 et seq.). Asymptotic approximation combined with importance resampling (the latter also referred to as sampling-importance resampling (SIR)) has been suggested as a method for drawing the sample of overdispersed starting values required for parallel chain MCMC. Although this is rather labour intensive and might appear excessive, Gelman and Rubin (1992b) demonstrate the benefits afforded by this approach in an application involving a multimodal posterior distribution. A brief overview of SIR is given in Section 1.5.4.

Asymptotic approximation refinements. Laplace's method for integral approximation

As mentioned above, one of a number of techniques can be adopted to improve the results obtained with normal approximation including importance resampling. Laplace's method provides an alternative refinement.

Normal approximation is expected to be poor if applied to skewed distributions. It is, however, a first-order Taylor series expansion, and an obvious refinement is the incorporation of higher order terms. The Laplace approximation adopts this strategy and includes second order terms. Historically, it was used mainly for performing multivariate integration when Simpson's rule and Gaussian quadrature were not successful (Gelman et al., 1995, p306). Additional details are given in Gelman et al. (1995, p306) and Carlin and Louis (1996, p145). Laplace's method replaces numerical integration with noniterative numerical differentiation, which is one of the main advantages, but this comes at the cost of unknown accuracy. Inaccuracies are expected to arise when dealing with samples of limited size and multidimensional problems of high order. Concerns regarding poor accuracy in large multidimensional calculations leads to the conclusion that Laplace's approximation cannot be recommended in most random coefficients modelling applications (Carlin and Louis, 1996, p149-150). It has been suggested, however, that the method produces estimates of better accuracy than posterior simulation for several types of model (Raftery, 1996, p170, quoting S. Rosenkranz, Ph.D. dissertation, Dept. Biostatistics, University of Washington, 1992).

1.5.3 A basic description of the EM algorithm

Although the EM algorithm is particularly well known for its application to *missing data* problems, it is used extensively as a mode-finding method. It is, in effect, an iterative optimisation (minimisation/maximisation) routine. The main feature of the EM method is the manner in which the observed data are augmented with latent data. This augmentation is performed in such a way that an initially intractable maximisation is converted into a simpler stepwise maximisation calculation. Lee(1989), Gelman et al. (1995) and Jackman (2000) all provide introductory accounts, the latter two of which relate the EM algorithm to the data augmentation method. Each iteration of the algorithm consists of two steps, i.e., an expectation (E) step and a maximisation (M) step. Specifically, given a decomposition of the data vector $\mathbf{y} = (\mathbf{y}_{obs}, \mathbf{y}_{aug})$ and the conditional probability density function of the augmented data, $f(\mathbf{y}_{aug}|\mathbf{y}_{obs}, \phi)$, where ϕ is a vector of current

values for the model parameters, the E-step requires an evaluation of $\int L(\phi|\mathbf{y}) f(\mathbf{y}_{aug}|\mathbf{y}_{obs},\phi) d\mathbf{y}_{aug}$. The subsequent M-step performs a maximisation of ϕ using the expected log-likelihood obtained in the E-step, thus generating an updated estimate of ϕ (Garthwaite et al., 1995, p48). Gelman et al. (1995) provide a concise but informative account of the EM method with a focus on its role as a marginal posterior mode-finding algorithm. Thus given $\theta=(\gamma, \phi)$ and the problem of approximating the marginal posterior density $p(\phi|\mathbf{y})$ in the region of the mode, the EM algorithm is used when direct maximisation of $p(\phi|\mathbf{y})$ is difficult but maximisation of $p(\phi|\gamma,\mathbf{y})$ is relatively easy. The EM algorithm is as follows (Gelman et al., 1995, p278):

1. Generate a set of starting parameter values

2. E step: Estimate $E_{old}[\log(p(\gamma,\phi|\mathbf{y}))] = \int \log(p(\gamma,\phi|\mathbf{y}))p(\gamma|\phi^{t-1},\mathbf{y})d\gamma$

   M step: Find the value of ϕ^t that maximises $E_{old}[\log(p(\gamma,\phi|\mathbf{y}))]$.

Lee (1997, p255-7) has outlined an implementation of the EM algorithm, as applied to the simple hierarchical normal model

$$\begin{aligned} x_{ij} &\sim N(\theta_i,\sigma^2) \\ \theta_i &\sim N(\mu,\psi) \end{aligned} \qquad [1.25]$$

with prior $p(\mu,\sigma^2,\psi) \propto 1/\sigma^2$. The M-step yields values for the set of hyperparameters $(\phi^{(t+1)} = (\mu^{(t+1)},\sigma^{2(t+1)}\psi^{(t+1)}))$ via a maximisation of the log posterior density which contains the augmented data $(\mathbf{x},\gamma^{(t)})$, where $\gamma^{(t)} = (\theta_1^{(t)},\theta_2^{(t)},...)$. $\gamma^{(t)}$ is among the parameters estimated at the E-step. In this particular example, closed form expressions are available at both the E- and M-steps.

1.5.4 Posterior simulation. An overview of noniterative Monte Carlo methods

The previous section outlines an approach to posterior interval evaluation, based on asymptotic approximation, designed to overcome the problems that arise due to the intractable form of the integrals encountered in most practical problems. But these methods do not always produce results

of sufficient accuracy (Gelman et al., 1995, p320; Cowles and Carlin, 1996, p883). Monte Carlo (MC) simulation offers an alternative to analytical approximation[1]. An additional advantage of the simulation approach is the relative ease with which various assumptions can be relaxed. For example, data with outliers can be modelled by replacing the usual normality distribution with a *t* distribution having some low value for the degrees of freedom. The degrees of freedom parameter can be included in the vector of model parameters (Gelman et al., 1995, p357; Congdon, 2001, p25).

MC simulation has been used for decades as a method for evaluating integrals (Hammersley and Handscomb, 1964). Rejection and importance sampling, together with *crude hit-or-miss* Monte Carlo are well-known methods (Morgan, 1984, Chapters 5 and 7). The idea underlying the Monte Carlo simulation method, as applied to Bayesian statistical inference, is to generate a sample from the joint posterior distribution or, alternatively, some marginal posterior distribution of interest. The resulting sample is used to generate the required statistics. In its basic form this might be achieved by direct sampling from the joint posterior distribution, in which case marginal statistics are generated from the resulting sample simply by ignoring the remaining parameters (Gelman et al., 1995, p65). Direct sampling from the joint posterior is not often performed in practice, however, since in many cases a considerable simplification can be achieved through factorisation. In fact, direct sampling from the joint posterior distribution might not be feasible, in which case some form of simplification is essential. For example, referring to [1.23], a sample from the marginal posterior distribution $p(\theta_1|y)$ can be generated by first drawing θ_2 from its marginal

---

[1] Although importance resampling can be used to refine the results obtained using analytical approximation, satisfactory accuracy is not guaranteed (Gelman et al., 1995, p320). Current practice is to use Monte Carlo simulation at the outset, thus avoiding analytical approximation entirely.

posterior distribution $p(\theta_2|\mathbf{y})$, and then drawing θ_1 from its conditional posterior distribution given θ_2. Sometimes direct simulation methods might be applied to the parts, otherwise some form of indirect sampling is used. Once the sample has been generated, quantiles can be derived from the ordered sample, as required. One of the many advantages of the simulation approach is the ease with which statistics can be obtained for complicated parameter functions, since the primary posterior sample can be used to generate a sample for any function of interest. Thus, given a set of draws $\theta^{(1)},\theta^{(2)},...,\theta^{(N)}$ from the posterior distribution, the mean, median, variance and/or posterior intervals can be estimated for virtually any posterior function of interest, $h(\theta)$, simply by using $h(\theta^{(1)}),h(\theta^{(2)}),...,h(\theta^{(N)})$. The sample can also be used to generate posterior predictive simulations of unobserved outcomes (Gelman et al., 1995, pages 301 and 329) as outlined in Sections 1.4.2.2 and 1.4.3. In those situations in which direct sampling is feasible it offers the advantage that it is an exact small-sample procedure, unlike ML and normal approximation, which are based on asymptotic theory. This assertion assumes a simulation size sufficient to ensure negligible simulation error. A striking illustration of the importance of the simulation approach to Bayesian inference is provided by the well-known *rats growth-curve* data (Gelfand et al., 1990). A linear random coefficients treatment of these data has been outlined in a number of papers and texts (Gelfand et al., 1990; Carlin and Louis, 1996, p166; Spiegelhalter et al., 1995b). This model is exceedingly simple conceptually, and yet, given the 30 individual growth profiles and a typical Bayesian random coefficients structure, a total of 66 model parameters is required. This number renders numerical approximation or numerical integration inapplicable. In fact, noniterative MC is expected to be somewhat problematic (Carlin and Louis, 1996, p168) leaving MCMC as the remaining viable alternative.

As implied in the previous paragraph, a number of categories of Monte Carlo simulation methods for Bayesian computation have been developed. These can be divided into two broad categories, namely the traditional noniterative methods and the iterative methods. Rejection sampling and importance sampling are well-known noniterative methods (Morgan, 1984) while the Metropolis-

Hastings algorithm, data augmentation method and Gibbs sampler are commonly used iterative techniques (Carlin and Louis, 1996). These three iterative sampling algorithms all belong to the Markov chain Monte Carlo integration class of methods. Simulation offers an advantage over earlier posterior summary methods (based on posterior modes) in providing a complete description of the posterior distribution, and both types of simulation, i.e., iterative and noniterative, are used for this purpose. Iterative simulation is now dominant among Bayesian computational methods due, in part, to a period of intense research activity during which various methodological improvements were established (Bernardo et al., 1992; Gilks et al., 1996a; see Section 1.6). A brief description of the commonly used algorithms is given in the following sections.

In contrast to MCMC, classical noniterative sampling produces independent non-Markov samples, assuming that correlation is not introduced deliberately. Broadly speaking, the noniterative methods can be sub-classified as *direct* or *indirect*, although no definitive categories exist. For example, Gelman et al. (1995) consider rejection sampling as a class of direct-sampling algorithm, while Carlin and Louis (1996) reserve the term *direct* for those methods in which an immediate draw from the target distribution is performed. Accordingly, they categorise rejection and importance sampling as indirect methods.

Noniterative direct sampling

As indicated in the preceding paragraph, noniterative direct sampling methods employ an immediate draw of samples from a target distribution. The basic, direct sampling approach to integration, as applied to Bayesian computation is conceptually simple. Suppose we seek an evaluation of the integral $\int f(\theta)h(\theta)d(\theta)$ and we can generate $\theta_1,...,\theta_N \sim h(\theta)$ by direct sampling. Then the required estimate is given by $\frac{1}{N}\sum_{j=1}^{N} f(\theta_j)$. A specific but simple illustration is provided by the normal model $y_i \sim N(\mu,\sigma^2)$, $i=1,...,n$, with prior $\pi(\mu,\sigma) = 1/\sigma$. The joint posterior distribution, $p(\mu,\sigma^2|y)$, is given by

$$\mu|\sigma^2,\mathbf{y} \sim N(\bar{y},\sigma^2/n)$$
$$\sigma^2|\mathbf{y} \sim K\chi^{-2}_{n-1} = K.IG\left(\frac{n-1}{2},2\right) \qquad [1.26]$$

where $K=\sum_{i=1}^{n}(y_i-\bar{y})^2$ and $IG(.)$ is the inverse gamma distribution (Carlin and Louis, 1996, p151; Lee, 1989, Section 2.12). A sample from the joint posterior distribution can be generated by making a draw $\sigma_j^2|\mathbf{y}$, and then drawing $\mu_j \sim N(\bar{y},\sigma_j^2/n)$. This process is replicated for $j=1,...,N$. Given this sample, an estimate of any function of the parameters and associated statistics can be generated, as required. For example, given some $f(\theta)$ of interest, an estimate of the posterior probability of the interval $(a<f(\theta)<b)|\mathbf{y}$ is provided by $\hat{p}=(\#f(\theta_j)\in(a,b))/N$ with associated binomial standard error estimate $\sqrt{\hat{p}(1-\hat{p})/N}$ (Carlin and Louis, 1996, p151).

Direct sampling has an obvious appeal since, as stated in the preceding paragraph, the accuracy of the simulation is determined solely by the sample size. Thus the latter can be calculated at the outset, which contrasts with MCMC simulation, where a lack of independence among the samples causes a reduction in accuracy. A useful tool in the implementation of direct sampling is the inverse cumulative density function method, in which a draw from $p(\theta|\mathbf{y})$ is made by using the inverse cumulative distribution, based on a lookup table implementation, if required (Morgan, 1984, Section 5.2; Gelman, et al., 1995, Sections 1.8, 3.7 and 10.2). Piecewise linear approximation (trapezoidal approximation) can be used as a refinement to the lookup results, if necessary.

Noniterative indirect sampling

In many applications it is not possible to perform direct sampling, with or without factorisation, and some other form of sampling is required. A number of noniterative indirect methods are available, including rejection sampling, importance sampling and the weighted bootstrap (also referred to as sampling importance-resampling (SIR)). In common with noniterative direct

sampling, the concept of convergence does not apply here. Instead, a sample of some prespecified size is drawn, at which point sampling is terminated.

Rejection sampling

Rejection sampling is used for generating a random sample when immediate sampling of the target distribution is not possible. It is conceptually simple and documented in numerous texts including Morgan (1984), Gelman et al. (1995) and Carlin and Louis (1996). Following Carlin and Louis (1996, p154), suppose a random sampling of the posterior distribution

$$h(\theta) = \frac{L(\theta)\pi(\theta)}{\int L(\theta)\pi(\theta)d\theta} \qquad [1.27]$$

is required. Further, suppose that an easily sampled envelope function $g(\theta)$ is available satisfying $L(\theta)\pi(\theta) < Mg(\theta)$ where $M > 0$ is a constant. The algorithm is

(1) Generate $\theta_j \sim g(\theta)$

(2) Generate $U \sim$ Uniform(0,1)

(3) if $MUg(\theta_j) < L(\theta_j)\pi(\theta_j)$, accept θ_j, otherwise reject θ_j and return to (1).

Formal treatments showing that the resulting sample is random with distribution $h(\theta)$ are available, although an inspection of Fig. 5.2 in Carlin and Louis (1996) or Fig. 10.1 in Gelman et al. (1995) suggests that this result is intuitive. One approach to defining the envelope function is to use piecewise trapezoidal approximation (Gelman et al., 1995, p305). An interesting application is provided by Wolfinger and co-workers, who developed a rejection sampling approach to Bayesian linear mixed model analysis for incorporation into a statistical package (Wolfinger and Rosner, 1996; SAS Institute, 1999). The main objective was the provision of a routine capable of producing reliable statistics without user intervention. Thus MCMC was ruled out due to the need to demonstrate convergence, which requires user input. An independence chain algorithm (Wolfinger and Kass, 2000) was added to subsequent releases of the software (SAS Institute, 1999). Rejection sampling is also used in the BUGS package to sample full conditional distributions not available in analytical form [Gilks, 1992; Gilks et al., 1994 (p176), 1995].

Importance sampling (Gelman et al., 1995, p307; Carlin and Louis, 1996, p153)

Suppose that an approximation to the posterior expectation

$$E(f(\theta)|\mathbf{y}) = \frac{\int f(\theta)L(\theta)\pi(\theta)d\theta}{\int L(\theta)\pi(\theta)d\theta} \qquad [1.28]$$

is sought and the density $g(\theta)$ is an easily sampled approximation to $cL(\theta)\pi(\theta)$. A sampled approximation can then be generated using

$$E(f(\theta)|\mathbf{y}) \approx \frac{\frac{1}{N}\sum_{j=1}^{N} f(\theta_j)w(\theta_j)}{\frac{1}{N}\sum_{j=1}^{N} w(\theta_j)} \qquad [1.29]$$

where $w(\theta) = L(\theta)\pi(\theta) / g(\theta)$ is a weight function. Importance sampling is among the methods used to compute marginal posterior densities (Gelman et al., 1995, p308). A related method is the sampling-importance resampling (SIR) algorithm that is described in the following subsection.

Sampling-importance resampling (weighted bootstrap)

Smith and Gelfand (1992) provide an introduction to SIR, describing it as a weighted-bootstrap. SIR provides a mechanism for sampling from $h(\theta) = f(\theta) / \int f(\theta)d\theta$ when the bounding constant required for rejection sampling is not available. Thus, referring to the problem in which a sample from

$$h(\theta) = \frac{L(\theta)\pi(\theta)}{\int L(\theta)\pi(\theta)d\theta} \qquad [1.30]$$

is required, the algorithm proceeds as follows (Smith and Gelfand, 1992; Carlin and Louis, 1996, p157):

(1) Generate a sample from an approximating density $g(\theta)$

(2) compute $w_i = \frac{L(\theta_i)\pi(\theta_i)}{g(\theta_i)}$ and $q_i = \frac{w_i}{\sum_{j=1}^{N} w_j}, i = 1,\ldots,N$

(3) Draw θ^* from $\{\theta_1,...,\theta_N\}$ with mass q_i on θ_i.

Smith and Gelfand (1992) point out that this is a bootstrapping algorithm, but with weights given by q_i. Furthermore, they show that the resulting sample has an approximate $h(\theta)$ distribution. Gelman and co-workers (Gelman et al., 1995, p339; Gelman and Rubin, 1992b, pages 460 and 471) recommend using a combined EM-SIR approach to generating the overdispersed set of starting values required for parallel chain MCMC simulation, the suggestion being that in some applications the refinement provided by SIR can be important. Finally, Gelman et al. (1995, p301) state that posterior simulation can be attempted by combining SIR with an approximation based on joint posterior modes. This, they suggest, can be applied to models of some complexity, thus providing an alternative to MCMC, although success is not guaranteed.

1.5.5 Posterior simulation. Markov chain Monte Carlo. The Metropolis, Metropolis-Hastings and Gibbs sampler algorithms

In principle, an ideal simulation uses independent samples drawn directly from the target distribution, since this is the most efficient. The accuracy of the simulation is determined immediately by the sample size, a property which is shared with the noniterative indirect methods[1]. This is an appealing feature and contrasts with MCMC, in which a lack of independence among the samples causes a reduction in precision. Direct sampling is not always feasible, however, and the indirect methods can be problematic when dealing with complex models. Consequently, MCMC is conceived as being the easiest approach to a wide variety of problems, including hierarchical analysis (Geyer, 1992; Gelman et al., 1995, p320). Geyer (1992, p473) states that, in contrast to independence sampling, MCMC can always be made to work. The basic idea behind MCMC-based estimation is to perform a random exploration of parameter space

---

[1] The algorithms of Section 1.5.4 are described as noniterative Monte Carlo methods in the sense that the concept of convergence does not apply. Instead, a sample of some predetermined size is generated, after which the algorithm stops.

which, in the Bayesian context, is a random exploration of the target posterior distribution. The aim is to create a Markov chain whose stationary distribution is identical to the specified target distribution. In practice, the simulation needs to be of sufficient length to ensure that the resulting chain is close to the stationary distribution. Once this is achieved the MCMC output is used to calculate the required statistics. Serial correlation within the sequence of samples is a major concern, since this reduces the rate at which convergence is met and thus has a detrimental effect on efficiency. Nevertheless, this can be overcome by increasing the length of the MCMC output, and MCMC provides exact results, in the sense that analytical/numerical approximation is not required. In principle estimates can be calculated with any specified precision, CPU time being the limiting factor. An additional advantage is the removal of some constraints regarding the types of prior that might be adopted. For example, a reliance on conjugate priors is eliminated. It might be noted that although MCMC is best known for its role in Bayesian analysis and has been described as a posterior simulation method (Raftery, 1996, p167), it is also used in various frequentist calculations, including missing-data analyses (Cowles and Carlin, 1996, p883; SAS Institute, 2001, Chapter 9).

The Gibbs sampler and the Metropolis/Metropolis-Hastings algorithms are the two best known Markov chain simulation methods. As outlined below, full conditional distributions are required to implement the Gibbs sampler and this can cause considerable difficulties since, in some applications, standard, closed form full conditionals are not available for sampling. The Metropolis/Metropolis-Hastings algorithms provide a useful alternative if sampling from the full-conditionals is problematic. On the other hand, the Metropolis algorithm uses a transition kernel to generate the Markov chain, and a successful implementation of the algorithm can be very dependent on the choice of kernel. Accordingly, it has been suggested that implementation of the Metropolis/Metropolis-Hastings methods is expected to be difficult in many mixed model (random coefficient model) applications due to the difficulty in finding a suitable transition kernel, given the large multidimensional nature of the parameter space associated with these models (Gelfand et al., 1995, p479-80).

The Metropolis and Metropolis-Hastings algorithms

The Metropolis algorithm was used in chemical physics and in image processing for many years before it was widely used by statisticians. Gelman et al. (1995, p323), Carlin and Louis (1996, p173) and Gilks et al. (1996b, p9) each provide an account of the algorithm, as applied to Bayesian computation. The main feature is the manner in which the algorithm uses draws from a distribution that is distinct from the target distribution, but modifies the output to achieve the target stationary distribution. This concept is explained below.

The Metropolis algorithm for generating a Markov chain is as follows:

1. Sample a candidate point θ^* from a jumping distribution $J_t(\theta^*|\theta^{t-1})$ where θ^{t-1} is the current point. $J_t(\theta^*|\theta^{t-1})$ must be symmetric in the sense that $J_t(\theta_a|\theta_b) = J_t(\theta_b|\theta_a)$.

2. Calculate the density ratio

$$r = \frac{p(\theta^*|\mathbf{y})}{p(\theta^{t-1}|\mathbf{y})} \qquad [1.31]$$

3. Set

$$\theta^t = \begin{cases} \theta^* \text{ with probability } \min(r,1) \\ \theta^{t-1} \text{ otherwise.} \end{cases} \qquad [1.32]$$

The requirement for symmetry in the transition kernel is important; a multivariate normal distribution centred at the current position is commonly used.

The Metropolis-Hastings algorithm, which is outlined by Gelman et al. (1995, p325), Carlin and Louis, (1996, p176) and Gilks et al. (1996b, p5), is a generalisation of the Metropolis method in which the requirement for a symmetric transition kernel is dropped. The acceptance/rejection ratio, r, modified to compensate for this asymmetry, is

$$\frac{p(\theta^*|\mathbf{y})J_t(\theta^{t-1}|\theta^*)}{p(\theta^{t-1}|\mathbf{y})J_t(\theta^*|\theta^{t-1})}. \qquad [1.33]$$

The successful implementation of both the Metropolis and Metropolis-Hastings algorithms requires a suitably scaled transition kernel in order to ensure reasonable efficiency. This matter is discussed in some detail in Sections 1.6.2 and 1.6.3.

A modification to the Metropolis algorithm in which $J_t(\theta^*)$ is used as the jumping distribution (i.e., samples are drawn independently of the current location) together with the acceptance rule $\min\left(1, \frac{p(\theta^*|y)J_t(\theta^{t-1})}{p(\theta^{t-1}|y)J_t(\theta^*)}\right)$ yields an independence sampler (Gilks et al., 1996b, p9; Chib and Greenberg, 1995, p330). Thus, in contrast to the random walk process that is generated with a transition kernel centred on the current position, independent samples are generated.

An appreciation of the manner in which the Metropolis and Metropolis-Hastings algorithms generate chains with the required target distribution as the stationary distribution can be acquired at various levels of mathematical sophistication. Formal treatments are given in Roberts (1996) and Tierney (1996). Informal treatments showing that the Metropolis algorithm converges to the target distribution are provided by both Carlin and Louis (1996, p174) and Gelman et al. (1995, p325). Each of the latter consists of two parts, the first of which shows that the sequence is a Markov chain with a unique stationary distribution. The second stage shows that, given the symmetric form of $J_t(.|.)$, this stationary distribution is equal to the target distribution. An informal proof showing that the Metropolis-Hastings algorithm converges to the target distribution is provided by Gilks et al. (1996b, p7). A central component of each treatment is the demonstration that both algorithms generate reversible chains. It is also shown that chain reversibility is a necessary condition for obtaining a statistical distribution equal to the target distribution (Chib and Greenberg , 1995, p328, equation 3; Gilks et al., 1996b, p7, equation 1.5; Carlin and Louis, 1996, p174). An examination of the reversibility condition provides an interesting insight into the mechanism through which the Metropolis/Metropolis-Hastings algorithm generates the required sampled distribution. This is the subject of the following subsection.

Chain reversibility. Insights into the Metropolis and Metropolis-Hastings algorithms

Several articles provide intriguing insights into these two algorithms, the account given by Chib and Greenberg (1995) being especially informative. Central to understanding these algorithms is chain reversibility. Dealing first with the more general Metropolis-Hastings case, reversibility is expressed by the equality

$$p(\theta^{t-1})P_{MH}(\theta^t|\theta^{t-1}) = p(\theta^t)P_{MH}(\theta^{t-1}|\theta^t) \qquad [1.34]$$

where

$$P_{MH}(\theta^t|\theta^{t-1}) = J_t(\theta^t|\theta^{t-1})r(\theta^{t-1},\theta^t). \qquad [1.35]$$

$r(\theta^{t-1},\theta^t)$ is the acceptance rate given the current position, θ^{t-1}, and candidate position, θ^t. If we suppose the condition

$$p(\theta^{t-1})J_t(\theta^t|\theta^{t-1}) > p(\theta^t)J_t(\theta^{t-1}|\theta^t) \qquad [1.36]$$

then, if it were not for the adjustment provided by the Metropolis-Hastings acceptance-rejection rule, the process would move from θ^{t-1} to θ^t too often. Thus a rule of the form $r(\theta^t,\theta^{t-1}) = 1$, $r(\theta^{t-1},\theta^t) < 1$ is required to meet the reversibility condition. Substituting [1.35] into [1.34] gives

$$p(\theta^{t-1})J_t(\theta^t|\theta^{t-1})r(\theta^{t-1},\theta^t) = p(\theta^t)J_t(\theta^{t-1}|\theta^t)r(\theta^t,\theta^{t-1}) \qquad [1.37]$$

which, in turn, gives the acceptance-rejection ratio

$$r(\theta^{t-1},\theta^t) = \frac{p(\theta^t)J_t(\theta^{t-1}|\theta^t)}{p(\theta^{t-1})J_t(\theta^t|\theta^{t-1})}. \qquad [1.38]$$

In the special case in which the transition kernel is symmetric (i.e., $J_t(\theta^t|\theta^{t-1}) = J_t(\theta^{t-1}|\theta^t)$) this reduces to the Metropolis acceptance-rejection ratio $p(\theta^t)/p(\theta^{t-1})$. In this case a candidate position of higher density, i.e., position with $p(\theta^t) > p(\theta^{t-1})$, is accepted on every occasion, while a candidate position of lower density is accepted with probability $p(\theta^t)/p(\theta^{t-1}) < 1$. Gelman et al. (1995, p324) describe the Metropolis algorithm as a stochastic type of stepwise mode finding process in which steps to a position of increased density are accepted always, while steps to a position of reduced density are accepted with a specified probability. In this sense the Metropolis

algorithm is a class of simulated annealing method (Press et al., 1988, p343). The same comments apply to the Metropolis-Hastings algorithm.

A number of additional points warrant attention. Firstly, the acceptance-rejection ratio can be calculated without knowing the normalisation constant associated with $p(\theta)$, since the latter appears in both the numerator and denominator. The choice of transition kernel is important, and common among the candidates is a multivariate normal or *t* distribution centred on the current position (Gelman et al., 1995, p324; Chib and Greenberg, 1995, p300; Carlin and Louis, 1996, p175). This produces a random walk within the posterior distribution. Chib and Greenberg (1995, p330) discuss some other alternatives, including the independence sampler variant mentioned previously. Finally, Geyer (1992, p474) states that the Metropolis-Hastings algorithm encompasses a huge family of methods. Consequently, although any one method might prove inadequate for a given problem, there are numerous alternatives. Reference is made to the advantages of adopting hybrid methods in which a variety of updating schemes are mixed within a given simulation.

Gibbs sampling

Many review articles and textbooks include accounts of the Gibbs sampler, a number of which (Casella and George, 1992; Carlin and Louis, 1996, p159 et seq.; Lee, 1997, Chapter 9; Jackman, 2000) adopt a comparative approach in which the Gibbs sampler is related to other methods including the *data augmentation* approach. Gelman et al. (1995, p326) provide the following summary. The parameter vector is divided into components $\theta = (\theta_1,...,\theta_d)$. At each iteration the algorithm cycles through the components θ_i, i=1,..,d, drawing a sample from the conditional distribution. Specifically, at the tth iteration, θ_j^t is sampled from the conditional distribution $p(\theta_j|\theta_{-j}^{t-1},\mathbf{y})$, where $\theta_{-j}^{t-1} = (\theta_1^t,...,\theta_{j-1}^t,\theta_{j+1}^{t-1},...,\theta_d^{t-1})$ is the vector containing ordered components of θ with θ_j removed. The required conditional distributions are available in closed form for a large class of hierarchical models with conjugate priors and hyperpriors (Carlin and Louis, 1996,

pages 165 and 171), and implementation is relatively straightforward in these cases. But in some applications conjugate priors do not exist, while in other cases non-conjugate priors might be chosen by design. A number of approaches have been developed for dealing with these more challenging problems (Gilks, 1992; Gilks et al., 1995).

A final point of interest is the relationship between the Gibbs sampler and the Metropolis-Hastings algorithm. The former is a special case of the latter (Gelman, et al., 1995, p328; Chib and Greenberg, 1995, p327; Jackman, 2000, p384). As stated above, the Metropolis-Hastings acceptance-rejection rule dictates that a candidate point is accepted with probability $r < 1$ if the transition is associated with a reduced density. The Gibbs transition kernel is given by

$$J_{j,t}(\theta^* | \theta^{t-1}) = p(\theta_j^* | \theta_{-j}^{t-1}, \mathbf{y}) \qquad [1.39]$$

for the jth component within the tth iteration, with $\theta_{-j}^* = \theta_{-j}^{t-1}$ by definition. Given this form of transition kernel, the Metropolis-Hastings acceptance-rejection rule yields an acceptance probability of unity (Gelman et al., 1995, p328; Tanner, 1996, p181). Thus the Gibbs variant of the Metropolis-Hastings algorithm updates each component in parameter space sequentially using a transition kernel given by the conditional density and with an acceptance rate of 100%.

1.6 MCMC implementation and related practical matters

1.6.1 Introduction and general issues

This section provides an introduction to various practical issues that arise in any MCMC simulation analysis, and includes a number of recommendations that have been made regarding implementation (Chapter 11 in Gelman et al., 1995, provides a useful summary). Among these are various procedures that have been proposed for performing an initial assessment of the MCMC output. Convergence is the major issue and, at some stage, the analyst needs to determine the number of samples required to achieve this. A related question is the sample size required to attain some specified precision in various target statistics. Both depend on the degree of autocorrelation in the MCMC output. While the number of samples required with an independence sampler is

determined directly by the required precision, this direct relationship does not hold in MCMC simulation. A high degree of correlation between consecutive iterations results in poor movement over parameter space (described as poor mixing) and a consequent reduced precision in the derived statistics. Thus, correlation reduces the effective sample size and has an adverse effect on convergence. It is an inescapable fact that the number of samples required cannot be determined until the degree of autocorrelation is known, and this information is available only after running the simulation. This section discusses briefly an informal approach to assessing convergence together with various procedures that have been suggested for overcoming poor performance and convergence failure. A few of the more formal convergence tests are outlined in Section 1.7.

A number of questions arise at the start of any MCMC simulation analysis, including the relative merits of using a single long chain in favour of a multiple parallel chains implementation (Geyer 1992, p474; Gelman, 1996, p132; Raftery and Lewis, 1996, p128). Given the fact that invalid inference can result from using chains of insufficient length, it has been suggested that a single long chain simulation is preferred (Geyer, 1992). Nevertheless, the diagnostic value of running multiple parallel chains is generally accepted. If a multiple chain implementation is chosen, the number of chains and their starting positions must be specified.

Many analysts argue in favour of multiple chain simulation, based on experience. A single chain may appear to have converged, as judged by visual inspection and formal convergence tests, but the chain may, in fact, be stuck in some restricted region of the target distribution. The detection of convergence failure is not, therefore, guaranteed. This kind of behaviour is more readily uncovered by examining multiple chains. Gelman et al. (1995, p330) argue that the distribution of each of the individual chains must be close to the distribution of the mixed chains before an individual chain is taken as an approximation to the target distribution. To illustrate the dangers associated with using a single MCMC chain, Gelman (1996, p132 et seq., see Fig 8.1) show two chains generated as part of a pharmacokinetic MCMC hierarchical model simulation analysis. An inspection of either chain in isolation gives the impression of convergence, while the two chains

are, in fact, well separated in parameter space. This example provides a compelling illustration of the benefits of using parallel chain trace plots as a mechanism for assessing convergence. In an hierarchical model analysis the random coefficients should be included in this assessment, irrespective of the fact that population statistics might be the sole focus of the study. Nevertheless, a substantial computational cost is incurred by running multiple chains and, in the past, some researchers have suggested that in all but the most important applications a single chain simulation might suffice, recognising that misleading parameter estimates might be obtained on occasions (Raftery and Lewis, 1996, p128). Undiagnosed convergence failure in a single chain was considered an acceptable risk when computational demands were an overriding consideration. Substantial improvements in computing power have occurred, however, since this early work on MCMC simulation, and computational demands are less of a consideration. Consequently, the benefits of the multiple chain approach have become generally accepted. The following guidelines have been proposed for a typical MCMC simulation analysis (based on Gelman et al., 1995, pages 322 and 330 et seq.; Gelman, 1996): (1) run several (three or more) parallel chains with overdispersed starting points derived from asymptotic approximation based on posterior modes; (2) discard the burn-in samples (see below) and use the resulting parallel chains to obtain an initial assessment of convergence, based on the degree of overlap between the chains; (3) if visual inspection suggests that the MCMC output is satisfactory, perform a formal assessment of convergence; (4) assuming satisfactory convergence, compare inferences derived from the MCMC output with those based on the approximate density generated in Step 1.

In those cases in which a visual inspection of the MCMC output indicates convergence failure in the form of poor coverage of parameter space, formal convergence tests become redundant. At this point the analyst might attempt to overcome the problem by using longer simulations runs, thus compensating for the inefficiency that arises due to correlation between successive samples. This approach might yield adequate coverage within reasonable CPU-time limits. Alternatively, the performance might be sufficiently poor that a change in approach is required (see Section 1.6.4). If, on the other hand, a visual inspection of the parameter chains reveals no signs of convergence

failure and the formal convergence test results are satisfactory, one can proceed with the analysis. The usual practice is to mix the chains (after discarding the burn-in section) and use the resulting sample to determine the required statistics.

In order to ensure that a given set of parallel chains is guaranteed to reveal convergence failure it is essential that the set of starting parameter values are dispersed widely in parameter space. But, usually, little is known about the target distribution in the initial stages of the analysis. One of a number of approaches can be adopted for generating the required overdispersed parameter values. As stated previously, Gelman et al. (1995, p322) suggest a formal approach using approximation based on posterior modes (see Sections 1.5.2 and 1.5.3). Samples can be drawn from the resulting distribution using SIR, for example. This procedure clearly requires a considerable amount of computation, and Carlin (roundtable discussion in Kass et al., 1998, p96) outlines a simpler approach which, he suggests, works well in practice. Given some approximation to the target distribution and assuming, for example, that five parallel chains are planned, the *j*th chain starting value for the *i*th parameter is equated to $\mu_i + (j-3)\sigma_i$, $j = 1,...,5$, where μ_i and σ_i are the estimated mean and standard deviation of the *i*th parameter, respectively. In an hierarchical model analysis the results obtained by fitting individual data is one approach to acquiring information about the target distribution (Gelman et al., 1995, p271).

Given the overdispersed nature of the parallel chain starting positions, it is necessary to discard the initial portion of each parallel chain, i.e., the so-called *burn-in* period (Geyer, 1992; Gelman et al., 1995, p330; Raftery and Lewis, 1996). A sufficient portion must be discarded to ensure that the truncated chain satisfies the stationarity condition. In this context Carlin and Louis (1996, p185) suggest a pragmatic definition of convergence, namely the condition under which the MCMC output *can be safely thought of as coming from the true stationary distribution of the Markov chain for all t>T.* Some researchers suggest that only the second half of each parallel chain should be retained (Gelman et al., 1995, p322), but this is an extravagant use of CPU time. Alternatively, the Raftery and Lewis (1992) calculation can be used to determine the number of samples that

might be discarded as *burn-in* (Section 1.7.3). Finally, given a marked degree of correlation, it is common practice to discard the majority of MCMC samples in order to obtain a reduced set of approximately independent samples (Geyer, 1992; Gelman et al., 1995, p330; Gelman, 1996). A simple approach is to retain one in every k samples, after discarding the burn-in section. This process is referred to as thinning. (Section 1.7.3 describes the Raftery-Lewis method for obtaining a suitable thinning ratio.) Alternatively, samples may be drawn at random from the MCMC output (Gelman et al., 1996a, p1406). It has been noted, however, that thinning is not essential, even if a considerable degree of autocorrelation exists within the MCMC output, since the only two advantages to be gained from thinning the sample are a reduction in data storage and a decrease in the computational demands of the post-MCMC processing phase of the analysis. Carlin and Louis (1996, p195-6) are more assertive in their opinion and state that thinning cannot be recommended because it is accompanied by an increase in sample mean estimator variance.

1.6.2 Efficient Metropolis jumping rules

The naïve application of the Metropolis algorithm can result in extremely slow mixing and inefficient estimation (Gelman et al., 1996b, p599). Poor mixing occurs for two reasons. Firstly, transitions that are too short must result in a slow movement over the target distribution. Conversely, transitions that move consistently into regions of low probability are rejected on most occasions, causing the chain to remain at a fixed position for many iterations. Intuition suggests that both extremes can be avoided by an appropriate adjustment of the transition kernel. Thus, a good jumping kernel generates transitions which cover reasonable distances in parameter space, subject to the requirement that candidate points are not rejected too frequently. This suggests that performance might be monitored either through an examination of the distance associated with each transition, or by using the acceptance frequency. The question of efficiency has been addressed by several researchers. Among the more formal treatments, Gelman et al. (1996b) have used both eigensystem calculations and limiting diffusion theory (limiting in the sense $d \to \infty$, where d is the dimension of the parameter space) to determine the optimum scaling for a d-dimensional multivariate normal (MVN) transition kernel. These studies have led to a number of

recommendations. Gelman and co-workers (Gelman et al., 1995, p334; Gelman, 1996, p141; Gelman et al., 1996b) suggest that draws should be made from a symmetric normal kernel with the same shape as the target distribution, as characterised by its covariance matrix (Σ), but scaled by a factor of approximately $\left(2.4/\sqrt{d}\right)$, giving the transition kernel $N\left(\theta^*|\theta^{t-1}, 2.4^2\Sigma/d\right)$. This assumes that the target has an approximately MVN distribution, which can be achieved through transformation, if necessary. Their limiting diffusion analysis indicates that optimum scaling gives a transition frequency of the order of 0.23, and accompanying simulations indicate that these results hold for dimensions as low as 6 (Gelman et al., 1996b, p603). A number of researchers have suggested an adaptive approach. For example, the following general procedure has been suggested for setting up an efficient Metropolis simulation (Gelman et al., 1995, p335). (1) Obtain an estimate of the target covariance matrix, (2) run a number of parallel simulations, using a scaling factor of $2.4/\sqrt{d}$, (3) use the resulting MCMC output to update the Metropolis jumping kernel, and finally (4) adjust the scaling of the kernel to obtain an acceptance rate in the range 0.44, for one-dimensional problems, to 0.23 for multidimensional problems of high order. The rationale for step (3) is the conservative nature of the covariance matrix estimate provided by the MCMC output, assuming a set of parallel chains with overdispersed start points (Gelman, 1996, p141). Transition kernel scaling can be built into a real-time adaptive scheme in which the acceptance rate is monitored and the scaling adjusted at intervals to keep the acceptance rate close to the target value. It should be noted, however, that caution must be exercised when using this kind of adaptive modification of the transition kernel since, by definition, the behaviour in one portion of the chain depends on previous iterations. This can compromise the stationarity of the resulting chain (Gilks and Roberts, 1996, p99; Gelman et al., 1996b, p606). An obvious solution is to use an adaptive algorithm to tune the transition kernel for a given target distribution. Once this has been achieved, the existing MCMC output can be discarded and the simulation restarted using a fixed kernel. Among the related procedures is one in which preliminary chains are used as a basis for model reparameterisation (Gelman et al., 1996b, p606; Gilks and Roberts, 1996, p97-99).

For example, this approach can be used to create a target distribution that is approximately spherical (Gelman et al., 1996b, p606).

1.6.3 Convergence issues and poor performance

The application of untuned MCMC simulation algorithms to complicated models invariably leads to poor performance due to extremely slow movement over the target distribution. One solution to the problem is to use longer chains since, assuming contiguity in parameter space, a chain will visit every region given sufficient time. There is a limit, however, beyond which it is not realistic to contemplate an increase in chain length. Thus it has been suggested that while simulations lasting several hours are tolerable, longer runs become impracticable (Gilks and Roberts, 1996). Faced with the situation in which an algorithm has not converged after a large number of iterations, some analysts discourage the brute force solution based on ever increasing chain lengths in the hope of obtaining convergence eventually. Depending on the nature of the problem, this can be futile (Brooks and Roberts discussion contribution to Polson, 1996). Ameliorative action is suggested, based on additional exploratory work, and aimed at solving the problem through model reparameterisation. Successful reparameterisation requires an understanding of the underlying causes of poor convergence, and this is the subject of this subsection. Two common causes of poor mixing are strong correlations between the model parameters and weak parameter identifiability (Gelfand et al., 1995), both of which can cause the likelihood and posterior surfaces to be poorly behaved due to extreme irregularity. This affects both the Gibbs sampler and the Metropolis algorithm (Gilks and Roberts, 1996, p90). In the case of the Gibbs sampler, a substantial degree of cross-parameter correlation tends to concentrate the full conditional densities close to the current position (Gilks and Roberts, 1996, p90), with the result that the majority of iterations generate very small steps. Several researchers have discussed this problem in relation to the simple MVN_k, constant correlation (ρ) target distribution. Polson (1996, p305) provides theoretical results indicating that in the case k=10 and ρ=0.9 Gibbs run lengths as large as 1.8×10^8 iterations are required to achieve convergence ($>10^{11}$ iterations for k=100 and ρ=0.9 using the Metropolis algorithm; Polson, 1996, Table 3), although some practitioners might regard these as

excessive and accept a reduction in accuracy. But it is clear that this thin, cigar-like distribution is a very poor candidate for Gibbs sampling and requires modification. Similarly, cross-parameter correlation has a detrimental effect on the Metropolis algorithm unless the transition kernel matches closely the unknown target distribution. The following explanation is intuitive (based on Gilks and Roberts, 1996, p90). The effect of neglecting the target correlation structure is most marked when a candidate point is generated some distance from the current position since the associated probability will be small relative to the current probability on most occasions, and the candidate will be rejected. The majority of accepted transitions will, therefore, lie close to their preceding positions, and the chain will be characterised by segments in which there is no movement, connected by a single small step. Under these conditions movement in parameter space will be impeded. It has been pointed out that these difficulties are related to the numerical problems that arise in multiple linear and nonlinear regression, given a marked cross-parameter correlation (Hills and Smith, 1992, p228; Gilks and Roberts, 1996, p92). One solution is to use a transformed parameterisation that is close to satisfying the ideal condition in which all cross-correlations are zero. In the linear case this can be achieved by using Gram-Schmidt orthogonalization (Strang, 1980), for example (Gilks and Roberts, 1996, p93; see Section 1.6.4.1). An alternative approach to improving the Metropolis algorithm is to use a transition kernel that matches closely the target distribution since, in a conjugate sense, tailoring the transition kernel is equivalent to reparameterisation. In fact it has been suggested (Carlin and Louis, 1996, p188) that a full MVN Metropolis approach provides a method for dealing with problems in which large cross-correlations are an inevitable consequence of poor identifiability.

A few miscellaneous comments are appropriate at this stage. A number of modifications to the basic Metropolis algorithm are possible including the so-called Metropolis within Gibbs variant, in which each parameter is updated individually (Raftery and Louis, 1996, p121). It is suggested that this variant of Metropolis should be easier to automate. This algorithm is not a proper Gibbs sampler, however, since it does not use full-conditionals. The name is, therefore, misleading. An

additional point of interest is the suggestion (Gilks and Roberts, 1996, p91) that a uniform transition kernel oriented along the major axis is among the strategies that might be used for dealing with prolate target distributions.

Given the disastrous deterioration in efficiency that can occur when a marked level of correlation exists among the parameters, it is important to understand the mechanisms through which cross-parameter correlation arises. It is useful to adopt a threefold classification, namely (1) cross-correlation that is attributable to the manner in which the model is specified and for which reparameterisation provides the solution, (2) correlation that arises because one or more parameters is poorly identified, given the available data, and (3) cross-correlation caused by overparameterisation, perhaps due to model specification error and/or inherent lack of identifiability. The sparse data problem that is common in pharmacokinetics (Wakefield, 1996; Gelman et al., 1996a, p1410; Roe, 1997) provides an example of the second of these three types. In this context the term *sparse* indicates that additional data could, in principle, provide a good identification of all parameters, even if the acquisition of these data is impracticable.

Poor identifiability of the second type often arises when vague prior distributions are specified and the information content of the data is insufficient to allow effective 'separation' of the parameters. Lack of identifiability due to data inadequacy also occurs in conventional fitting, and is a well-known cause of problems in linear regression, together with slow convergence in nonlinear regression, maximisation/minimisation, iteratively reweighted least squares and Fisher scoring applications. In the Bayesian setting nonidentifiable parameters have identical posterior and prior distributions (Kass et al., 1998, p97). A number of researchers have commented that poor identifiability and overparameterisation are a particular problem with large models (Carlin and Louis, 1996, p187), since these characteristics can be hard to detect when confronted with a complicated model. On the other hand, it has been noted that in some instances nonidentifiability arises naturally and is harmless. Mixture models are an example in which nonidentifiability is a

natural consequence of the nature of the problem and it has been suggested that in these situations any attempt to impose constraints to achieve identifiability is likely to be harmful (Kass et al., 1998, p97).

1.6.3.1 Random effects models

It is stated in the preceding subsection that cross-parameter correlation can be extremely harmful to MCMC simulation efficiency, and that this is related to the multicollinearity issue that arises in multiple regression, implying that some simple ameliorative action is available in the form of a suitable reparameterisation or transformation. The situation is, however, rather more complicated when dealing with random-effect models (hierarchical models) MCMC simulation. These models typically contain a large number of parameters and, although this does not result automatically in inadequate performance (Gilks and Roberts, 1996, p93), mixing can be extremely poor. This problem has received considerable attention. Gilks and Roberts (1996) and Gelfand et al. (1996) both discuss parameter correlation in the context of the simple random effects model

$$\begin{aligned} y_{ij} &= \mu + \alpha_i + \varepsilon_{ij} \\ \alpha_i &\sim N(0,\sigma_\alpha^2) \\ \varepsilon_{ij} &\sim N(0,\sigma_e^2) \end{aligned} \qquad [1.40]$$

$i=1,...,m, j=1,...,n$. The posterior cross-correlations for this model are given by (Gilks and Roberts, 1996)

$$\rho_{\mu,\alpha_i} = -\left\{1+\frac{m\sigma_e^2}{n\sigma_\alpha^2}\right\}^{-\frac{1}{2}} \qquad [1.41]$$

$$\rho_{\alpha_i,\alpha_j} = \left\{1+\frac{m\sigma_e^2}{n\sigma_\alpha^2}\right\}^{-1}. \qquad [1.42]$$

Thus the relative size of the variance components is important, and large posterior cross-correlations and consequent poor mixing will be avoided if σ_e^2 / n is large in relation to σ_α^2 / m. This result might appear counter-intuitive since a large number of random effects, m, or,

alternatively, a small random effects variance, σ_α^2, improves mixing, while a large number of observations per random effect, n, or small observational variance, σ_e^2, has a detrimental effect. Gilks and Roberts (1996) emphasise the point that mixing deteriorates as the data become increasingly informative. On a related matter, it has been noted that convergence failure can occur in MCMC hierarchical modelling when a chain enters a region of parameter space with a small random effect variance, since the chain tends to stall at this point (Gilks and Roberts, 1996, p96; Raftery and Lewis, 1996, p128).

1.6.4 Methods for improving performance. Transformation, centering and hybrid methods

A considerable research effort has been devoted to the development of methods for improving performance in MCMC applications. A number of review and summary papers are available, including Gilks and Roberts (1996), Gelman et al. (1996b) and Kass et al. (1998). Central to the performance issue is the detrimental effect of high correlations (Carlin and Louis, 2000, p162; see Sections 1.6.1 and 1.6.3) and a variety of methods have been suggested for speeding up convergence, including the 'feedback' or so-called adaptive processes in which the MCMC output generated at a given point in time is used to alter some aspect of the simulation process. Adaptive MCMC simulation is mentioned in Section 1.6.2 in relation to Metropolis kernel updating. An alternative strategy is adaptive model reparameterisation based on previously obtained Metropolis output, with the objective of achieving a near spherical target distribution (Gelman et al., 1996b, p606). At the same time a near optimal transition rate can be obtained by rescaling the transition kernel. A related approach to model reparameterisation has been suggested in Gibbs sampling (Hills and Smith, 1992, Section 3). A number of researchers (Carlin and Louis, 2000, p162; Gelman et al., 1996b, quoting work by Gelfand and Sahu; Gilks and Roberts, 1996, p99) have noted the need for caution regarding adaptive MCMC simulation since the resulting output is not, in general, a stationary Markov process. Nevertheless, it has been demonstrated that regular

updating schemes can yield stationary chains (Carlin and Louis, 2000, p162, quoting Mira and Sargent, 2000).

1.6.4.1 Reparameterisation methods

The preceding paragraph refers to adaptive reparameterisation methods. In essence, these methods are empirical. In some cases, however, convergence can be helped at the outset through a rational reparameterisation of the physical model, effective independence among the parameters being the objective (Gelman et al., 1995, p334). A number of approaches have been outlined, some of which are discussed below. Related to reparameterisation is a hierarchical blocking approach to acceleration (Carlin and Louis, 2000, p163) in which the model parameters are updated in blocks, each consisting of a set of highly correlated parameters.

The numerical approach

Ill-conditioning due to lack of independence is a well-known problem in several areas of numerical analysis, and a number of approaches to accelerating convergence in linear models MCMC have their origin in the numerical methods discipline. Among these is the application of orthogonal transformations/rotations and parameter scaling, the former of which is aimed at achieving approximate posterior independence among the model parameters (Gilks and Roberts, 1996, p93; Kass et al., 1998, p95-6). For example, an initial Gram-Schmitt orthogonalization (Strang, 1980) of the entire parameter space can be performed (Gilks and Roberts, 1996, p93), which can be conceived as a numerical alternative to the analytical reparameterisation mentioned in the preceding paragraph. It has been suggested, however, that this kind of algorithmic transformation is problematic in high-dimensional problems (Kass et al., 1998, p95). Furthermore, orthogonal transformation and related methods are restricted to linear problems. Nonlinear problems are more difficult (Gilks and Roberts, 1996, p96), although Hills and Smith (1992, p228) discuss a reparameterisation based on residuals.

An alternative procedure founded on orthogonal transformations is an adaptive algorithm in which subsidiary Metropolis iterations are used within a Gibbs algorithm, combined with regular parameter space transformation (Carlin and Louis, 2000, pages 160 and 170, citing Müller, 1991). Each update is based on the covariance matrix estimate provided by the current MCMC output, but with the refinement that a Cholesky decomposition is performed to obtain an orthogonal basis for the parameter space. The computational cost associated with approximating the covariance matrix with sufficient accuracy has been noted, however, and this limits the feasibility of this approach in high-dimensional applications. Overrelaxation (Carlin and Louis, 2000, p162) is another approach to convergence acceleration that has its origins in the numerical methods field, in the sense that it is related to the Gauss-Seidel/overrelaxation algorithms (Strang, 1980) for solving linear equations. It achieves an improvement in performance by ensuring that each new sample is negatively correlated with the previous one, thereby suppressing the random walk generated by the basic Gibbs sampling and Metropolis algorithms (Kass et al., 1998, p97; Neal, 1995).

As stated previously, a considerable effort has been made to improve the efficiency of the basic Gibbs sampling and Metropolis algorithms. This subsection has provided a brief overview of some of this work, much of which centres on model reparameterisation. It is appropriate to conclude by reiterating the point that an alternative strategy can be adopted in Metropolis MCMC, namely the updating method in which the transition kernel is tailored to the target distribution, as described in Section 1.6.2.

1.6.4.2 Hybrid and related algorithms

A variety of hybrid schemes or mixture algorithms have been suggested for obtaining an improved sampling of parameter space, the central feature of which is algorithm switching. For example, Metropolis MCMC can be implemented by using a set of alternative transition kernels which are used in a cycle (Carlin and Louis, 2000, p159-160). In addition, hybrid algorithms can be used to

force an occasional jump across parameter space, thus reducing autocorrelations while, apparently, preserving convergence to the stationary state (Carlin and Louis, 2000, p159; Gilks and Roberts, 1996, p110). A related issue that has received some attention is the need for algorithms capable of forcing jumps between poorly connected regions of parameter space (Kass et al., 1998, p96; Gilks and Roberts, 1996, p102). Other mixing strategies include random swapping between different parameterisations (Gilks and Roberts, 1996, p110). An associated technique is a simulated tempering approach (Kass et al., 1998, p96; Gilks and Roberts, 1996, p104; both provide references to simulated tempering in MCMC) for generating random interchanges between a set of MCMC chains.

1.6.4.3 Random coefficients model hierarchical centering

As stated previously poor mixing is a frequent occurrence in MCMC simulation work. Although a given problem might be inherently pathological due to lack of identifiability, the choice of model parameterisation can dramatically affect the shape of the surface. Hierarchical random coefficient models are known to be particularly problematic in this respect due to the high posterior autocorrelations and cross-parameter correlations involving the random coefficients. A lack of information about the variance components has been implicated as an additional cause of slow mixing (Gilks and Roberts, 1996, p96). In particular a problem arises when the MCMC sampler enters a region of parameter space in which one of more of the posterior random effect variances is small. MCMC algorithms sometimes fail in this situation because these regions tend to trap the MCMC output, effectively preventing a proper sampling of the target distribution (Raftery and Lewis, 1996, p128; Gilks and Roberts, 1996, p96). This occurs, however, only if the variance component prior distribution has nonzero probability close to zero[1], and an acceptable solution

---

[1] DuMouchel and Waternaux (1992, discussion contribution to Morris and Normand, 1992), Spiegelhalter (1995a, Section 9.2) and others have discussed the related impropriety issue that arises in connection with random effect variances, pointing out that $\sigma_a^2 = 0$ can be supported by a

might be a change in prior specification to incorporate an appropriate lower bound that excludes zero (Gilks and Roberts, 1996, p96). The rationale for imposing this constraint is the argument that a given variance should not be small if a random effects attribute is deemed necessary at the outset. This problem aside, in many cases a marked improvement in performance can be achieved by reducing the level of posterior correlations through a reparameterisation of the random coefficients. One approach is to use so-called hierarchical centering in which a strictly hierarchical structure is adopted among a set of random effects, each of which is centred about the mean defined by the next level in the hierarchy. Gelfand and co-workers have provided a thorough treatment of hierarchical centering, as applied to both normal linear models (Gelfand et al., 1995) and generalised linear models (Gelfand et al., 1996). A brief outline of the former is given below. Vines et al. (1996) have proposed an alternative scheme, based on the observation that random coefficient models are typically overparameterised. This arises because the random coefficients are usually expressed as deviations from a population mean, resulting in an inevitable strong posterior correlation between each random coefficient and the associated mean. Accordingly, the authors suggest a change in parameterisation to one in which each random effect is expressed as a deviation from its sample mean. This reparameterisation assumes, however, that the purpose of the analysis is an evaluation of the random effects relative to their overall mean as opposed to the absolute random effect and grand mean estimates (Vines et al., 1996, p341-2). Gilks and Roberts (1996, p95) refer to this method as reparameterisation by sweeping. Hierarchical centering is better known and has become established as a useful approach to the parameterisation of linear random coefficient models. It is unfortunate that no general approach has been devised for nonlinear models (Gilks and Roberts, 1996, p96).

---

non-negligible likelihood, leading to an improper posterior, if allowed under the variance component prior distribution (e.g., a prior of the form $p(\sigma_\alpha^2) \propto 1/\sigma_\alpha^2$).

Hierarchical centering details

Hierarchical centering is the name given to the reparameterisation scheme in which random effects are centred about their respective population means (Kass et al., 1998, p95; Gilks and Roberts, 1996, p94). Gelfand et al. (1995) provide a useful introduction using the simple random effects model

$$\begin{aligned} &\text{Level 1: } Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk} \\ &\text{Level 2: } \varepsilon_{ijk} \sim N(0,\sigma_e^2),\ \beta_{ij} \sim N(0,\sigma_\beta^2),\ \alpha_i \sim N(0,\sigma_\alpha^2),\ \mu \sim N(\mu_0,\sigma_\mu^2). \end{aligned} \qquad [1.43]$$

Hierarchical centering yields the alternative parameterisation

$$\begin{aligned} &\text{Level 1: } Y_{ijk} = \rho_{ij} + \varepsilon_{ijk} \\ &\text{Level 2: } \rho_{ij} = \mu + \alpha_i + \beta_{ij} = \eta_i + \beta_{ij} \\ &\text{Level 3: } \eta_i = \mu + \alpha_i \\ &\text{Level 4: } \varepsilon_{ijk} \sim N(0,\sigma_e^2),\ \beta_{ij} \sim N(0,\sigma_\beta^2),\ \alpha_i \sim N(0,\sigma_\alpha^2),\ \mu \sim N(\mu_0,\sigma_\mu^2). \end{aligned} \qquad [1.44]$$

The first model consists of two hierarchical levels while the second has four. The important feature of the centred parameterisation is the manner in which, at each level, the random coefficient term is centred with respect to the next level in the hierarchy. Thus the units η_i are centred about μ, the units ρ_{ij} are centred about η_i and the units Y_{ijk} are centred about ρ_{ij}. It should be noted that the two models have identical sample space dimensionality. The potential for improved performance afforded by hierarchical centering has been shown through an examination of the simpler two-level model given previously (equation [1.40])[1]. After centering the model becomes $y_{ij} = \eta_i + \varepsilon_{ij}$, $\eta_i \sim N(\mu,\sigma_\alpha^2)$, $\varepsilon_{ij} \sim N(0,\sigma_e^2)$, and the resulting posterior correlations are (Gilks and Roberts, 1996, p94; Gelfand et al., 1996)

$$\rho_{\mu,n_i} = \left\{1 + \frac{mn\sigma_\alpha^2}{\sigma_e^2}\right\}^{-\frac{1}{2}} \qquad [1.45]$$

---

[1] Hierarchical centering is examined in Appendix C. In particular, Fig. C.1 provides an example of the gain in performance that can be achieved by adopting an hierarchical centred parameterisation, relative to the performance of a non-centred model.

$$\rho_{n_i,n_j} = \left\{1 + \frac{mn\sigma_\alpha^2}{\sigma_e^2}\right\}^{-1}. \qquad [1.46]$$

These equations show that, in contrast to the properties of the non-centred model, mixing improves as σ_α^2, m and n increase. Thus a reversal of the relationship between the magnitude of σ_α^2 and the posterior correlation behaviour is achieved (compare [1.45] and [1.46] with [1.41] and [1.42]). Among the models examined by Gelfand et al. (1995) are those that are not naturally hierarhical, such as the 2-way additive model

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}. \qquad [1.47]$$

Simulation results indicate that centering the effect with the larger posterior variance can lead to improved performance (Gelfand et al., 1995, p486). Accordingly the general recommendation is to centre effects having large posterior variances relative to the residual variance, together with all effects lower in the hierarchy (Gelfand et al., 1995, p486; Gilks and Roberts, 1996, p94).

1.7 Convergence assessment. Formal tests of convergence

Convergence is an issue of fundamental importance to MCMC statistical inference. Given a Markov chain, $X(t_1),...,X(t_n)$, convergence is met at the point t_i provided the MCMC output is stationary for all $t > t_i$. Thus an MCMC algorithm is described as having converged when its output behaves as coming from a stationary distribution. An informal definition of the stationary condition is the absence of any systematic change in the mean or variance, after removing any strictly periodic variation (Chatfield, 1989, p10). More formally, a time series is strictly stationary if the joint distribution of $X(t_1),...,X(t_n)$ is the same as the joint distribution of $X(t_1+\tau)$, ..., $X(t_n+\tau)$ for all $t_1,...,t_n,\tau$ (Chatfield, 1989, p28). The main point to note is that, due to the stochastic nature of the problem, the goal is to achieve convergence to a distribution as opposed to convergence to a point. Given the central role of MCMC-simulation in the work presented in this thesis it is necessary to define a Markov process and some related concepts, starting with the definition of a *process* as any function of time, stochastic or deterministic, that specifies the instantaneous state of a system (Gillespie, 1992, page xi). An important subclass of processes are

stochastic and *memoryless*, meaning that the value $X(t)$ determines the probabilistic behaviour of $X(t+\tau)$, uniquely. Thus, given a memoryless stochastic process, any probability assignments $p(X(t+\tau))$ based on the value $X(t)$ cannot be improved by using information $X(t'), t' < t$ (Gillespie, 1992, page xii). The process forgets past values. Memoryless stochastic processes are referred to as Markov process after the Russian mathematician A.A. Markov (1856-1922).

It is clear from the preceding overview of the MCMC literature that the convergence issue is central to the MCMC method, as applied to Bayesian statistical inference. An assessment of output convergence is a necessity because the statistics derived from the MCMC output are invalid if it is not achieved. Furthermore, the magnitude of the discrepancy (bias in both the parameter and precision estimates) between the simulation results and the target statistics is unknown. Some degree of serial correlation is expected within the chains generated by a Markov process, and the chain length required to reach convergence increases with increasing autocorrelation. The latter is unknown at the start of the analysis. It is, therefore, incumbent on the analyst to demonstrate convergence, either formally or by using some informal procedure to show that the posterior distribution is sampled adequately. Care must be taken to ensure that sufficient accuracy and precision are both achieved. Serial correlation rules out any simple estimation of simulation error, but estimates can be obtained, post simulation, by using the time-series methods described below.

Rigorous mathematical treatments have provided theoretical results on the relative convergence rates of various competing algorithms and bounds on the number of iterations needed to ensure that a distribution is obtained within some specified limit of a given true stationary distribution (Carlin and Louis, 1996, pages 186 and 189; Carlin and Louis, 2000, pages 174 and 176). It appears, however, that theoretical results are seldom used in practice and that an alternative convergence diagnostic approach is used in the majority of applied analyses. The question that must be addressed is the point at which a given MCMC output can be terminated and treated as representative of the required stationary distribution. The aim is to determine the extent to which

inferences based on a given Markov chain differ from the intended target results. An informal but extremely useful approach to acquiring this information is to run several parallel chains started at remote positions within parameter space (overdispersed starting points). Overlaid parallel chain trace plots provide an immediate visual indication of convergence problems since this results in a failure to obtain good overlap over the entire length of the set of chains in one or more parameters. Poor mixing among the chains indicates inadequate sampling of the posterior distribution. This behaviour arises when a high degree of serial correlation causes the sampler to move slowly within a local region of the target distribution. Assuming the level of serial correlation is insufficient to cause overwhelming problems, convergence is achieved at some point given a sufficient number of iterations. It is then necessary to discard the initial nonstationary portion of each parallel chain. This section of the chain is referred to as the burn-in period, and the question arises as to the number of samples that should be discarded. Various diagnostic procedures have been developed to address the two key questions, i.e., the number of iterations to discard as burn-in, and the chain length required for adequate sampling of the posterior distribution. Both depend on the level of correlation.

Types of convergence diagnostic

Convergence diagnostic methods can be distinguished according to several key features (Carlin and Louis, 1996, p191; Cowles and Carlin, 1996, p884). Some are designed to produce diagnostics based on a single MCMC chain while others require at least two parallel chains. Some apply to univariate MCMC output while others can be used with the full joint posterior sample. Some diagnostics are restricted in their application to chains produced using the Gibbs sampler, some work with a specific subset of MCMC schemes, while others can be applied to any MCMC output. Some methods are quantitative and produce a summary diagnostic parameter, while others are qualitative in the sense of producing graphical output designed for visual inspection. Finally, most focus on estimating bias, while other methods provide estimates of precision. The following description is restricted to three particularly well-known diagnostics. The Gelman-Rubin diagnostic was selected as an example parallel-chain test while the Geweke method was selected

as a single chain diagnostic. Finally, the Raftery-Lewis diagnostic was selected as a commonly used method for assessing precision. All three methods are restricted to univariate MCMC output, but with no constraint on the type of algorithm used to generate this output.

Software

A number of software routines are available for performing MCMC convergence diagnoses including CODA (Convergence Diagnosis and Output Analysis software) and BOA (Bayesian Output Analysis Program). These two programs consist of a set of menu-driven S-Plus functions, also available in an R language implementation (the latter referred to as R-CODA). In addition to providing a comprehensive set of diagnostics, both programs include various statistical and graphics tools. CODA is supplied as a supplement to the BUGS/winBUGS Gibbs sampler program.

1.7.1 The Gelman and Rubin diagnostic

The Gelman-Rubin diagnostic is concerned with bias and requires at least two univariate parallel chains with overdispersed starting points (Gelman and Rubin, 1992a, 1992b; Gelman et al., 1995, p331; Gelman, 1996; Cowles and Carlin, 1996; Brooks and Roberts, 1998). It is based on an ANOVA-type estimation of the within- and between-chain variance, and uses this to provide an index of convergence. Specifically, given J parallel chains of length n, the Gelman-Rubin convergence diagnostic is based on the so-called scale reduction factor ($\hat{R}$) which is defined by

$$\sqrt{\hat{R}} = \sqrt{\frac{\hat{\text{var}}^{+}(\psi|\mathbf{y})}{W}} \qquad [1.48]$$

where

$$\hat{\text{var}}^{+}(\psi|\mathbf{y}) = \frac{n-1}{n}W + \frac{1}{n}B, \qquad [1.49]$$

with

$$B = \frac{n}{J-1}\sum_{j=1}^{J}\left(\overline{\psi}_{.j} - \overline{\psi}_{..}\right)^2, \quad \overline{\psi}_{.j} = \frac{1}{n}\sum_{i=1}^{n}\psi_{ij}, \quad \overline{\psi}_{..} = \frac{1}{J}\sum_{j=1}^{J}\psi_{.j},$$
$$W = \frac{1}{J}\sum_{j=1}^{J}s_j^2, \; s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\psi_{ij} - \overline{\psi}_{.j}\right)^2. \qquad [1.50]$$

Thus, $\hat{\text{var}}^{+}(\psi|\text{y})$ is a weighted average estimate of the marginal posterior variance of the estimand in question involving the between-chain variance (B) and the within-chain variance, (W). Gelman and Rubin (1992, p463) show that this weighted average, which uses an ANOVA-like partitioning of mean-squares, provides an unbiased estimate of the target distribution variance, provided the stationarity condition is met.

Given a set of parallel chains with an overdispersed distribution of starting points, the observation that the resulting average within-chain variance (W) does not exceed the between-chain estimate of the variance in the chain means, B/n, by a substantial amount indicates that the parallel chains are far removed from a common distribution. Under these conditions, $\hat{\text{var}}^{+}(\psi|\text{y})$ overestimates the marginal posterior variance, $\text{var}(\psi|\text{y})$, while W is an underestimate and the scale reduction factor provides a measure of the magnitude of the extent to which the scale of the current distribution might shrink if sampling were continued indefinitely ($\sqrt{\hat{R}} \to 1 \text{ as } n \to \infty$). To summarise, the method is based on an ANOVA-like construction of an over- and underestimate of the variance of the target distribution, with the property that the two estimates become approximately equal at convergence, but not before. The Gelman-Rubin diagnostic can be applied to any type of MCMC output. The approach is univariate, but can be used with -2log(posterior density) as a mechanism for examining the convergence of the joint density (Cowles and Carlin, 1996, p885). The recommendation is to examine all model parameters and other quantities of interest, including parameter functions (see, for example, Gelman et al., 1995, p331; Gelman, 1996). Each should have a near normal distribution for the method to work well. Transformation is, therefore, recommended using logits, for example, in the case of estimands bounded by zero

and unity. The Gelman-Rubin diagnostic is critically dependent on the identification of a set of overdispersed starting points, a feature that Carlin and Louis (1996, p193) find dubious.

Some analysts apply the Gelman-Rubin diagnostic to the second half of each of a set of parallel chains and calculate the scale reduction parameter 50% and 97.5% quantiles (Gelman et al., 1995, p332; Best et al., 1995, p19; Cowles and Carlin, 1996, p885). The recommended criterion for convergence is that both quantiles are close to unity. If satisfied, the second half of each chain is used to form a composite sample from which the required statistics are generated. This is, of course, extremely wasteful in terms of CPU resources, and alternative approaches have been adopted based on Gelman-Rubin plots.

Gelman-Rubin plots

The CODA and BOA packages both produce shrink factor plots using the following protocol. Each chain is divided into a number of segments, the first consisting of samples 1 to 50 (or some larger number when dealing with long chains), the second consisting of samples 1 to 50+n, the third consisting of samples 1 to 50+2n, etc. The Gelman-Rubin shrink factor median and 97.5% quantile is computed for the second half of each segment and plotted against the maximum iteration number of the segment. Convergence is indicated if the plots obtained for all model parameters stabilize around a value of unity at some point. Since the diagnostic is based on the second half of each chain, convergence can be assumed to have occurred at the midpoint of the first segment giving a shrink factor median and 97.5% quantile close to unity.

1.7.2 Geweke's spectral density diagnostic

The Geweke convergence diagnostic is based on a time series (signal processing) approach to making a comparison of the means of two portions of a single chain (Geweke, 1992; Brooks and Roberts, 1998; Cowles and Carlin, 1996). The two segments, which may be unequal in length,

should have similar means if the chain is stationary. The comparison uses a spectral density estimate of the asymptotic variance of the parameter in question. Accordingly, the Geweke *Z*-score is defined as

$$Z_n = \frac{(\hat{\theta}_n^A - \hat{\theta}_n^B)}{\sqrt{\frac{1}{n_a}\hat{S}_\theta^A(0) + \frac{1}{n_b}\hat{S}_\theta^B(0)}} \qquad [1.51]$$

where $\hat{S}_\theta^k(0)$ is the spectral density variance estimate provided by the kth segment, and n_k is the segment length. Z_n should have an approximate standard normal distribution given a stationary chain. An excess of *Z*-scores outside the 95% confidence interval of the standard normal distribution indicates that the chain is not fully convergent over its entire length. The form of the Geweke *Z*-score follows from the Wiener-Khintchine Fourier relationship between the normalised power spectral density, $S_{norm}(\omega)$, and the autocorrelation function, $\rho(\tau)$, together with the relationship between $\mathrm{var}[\overline{X}]$ and $\rho(\tau)$ (the latter is given in Priestley, 1981, equation 5.3.5). The Wiener-Khintchine Fourier relationship is

$$S_{norm}(\omega) = |F_{norm}(\omega)|^2 = \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty}\rho(\tau)e^{-j\omega\tau} \qquad [1.52]$$

where $F_{norm}(\omega)$ is the normalised spectrum of the process, $X(t)$. Thus, it follows that

$$\mathrm{var}[\overline{X}] \approx (2\pi\sigma^2 / N)S_{norm}(0) \qquad [1.53]$$

as shown by Priestley (1981, equation 5.3.7) where $\sigma^2 = \mathrm{var}[X_t]$. The Geweke diagnostic can be applied to any MCMC chain, but among its limitations is a sensitivity to the manner in which the two segments (spectral windows) are specified (Cowles and Carlin, 1996, p886).

Geweke diagnostic plots

CODA and BOA both include a routine for plotting the Geweke *Z*-score based on the following procedure. Given a chain consisting of N samples, a number of chain segments are formed, the

first consisting of all N samples, the second consisting of the last N-n samples, the third consisting of the last N-$2n$ samples, etc. The Geweke Z-score is calculated for each segment and plotted against the sample number at which the segment starts. An abundance of Z-scores outside the range +/-1.96 suggests convergence failure. Large Z-scores restricted to early chain segments are attributed to burn-in and these samples discarded.

Methods for assessing simulation error

The preceding convergence diagnostics are concerned solely with the estimation of bias in the MCMC estimates. Although bias might be the main worry in the first instance, the precision of the MCMC sample statistics must be determined at some stage. Given the posterior estimate

$$\hat{E}(\lambda|\mathbf{y}) = \hat{\lambda}_N = \frac{1}{N}\sum_{t=1}^{N}\lambda^{(t)} \qquad [1.54]$$

for some parameter or function of interest (λ), as obtained by combining chains after discarding burn-in iterations, a naïve estimate of the simulation standard error can be calculated using

$$\hat{\text{var}}(\hat{\lambda}_N) = \frac{1}{N(N-1)}\sum_{t=1}^{N}(\lambda^{(t)} - \hat{\lambda}_N)^2. \qquad [1.55]$$

But this estimate ignores autocorrelation in the sample and is, therefore, anti-conservative. An improved estimate is obtained by using the effective sample size $N/\kappa(\lambda)$ (Kass et al., 1998, p99; Carlin and Louis, 2000, p171), where $\kappa(\lambda)$ is the autocorrelation time given by

$$\kappa(\lambda) = 1 + 2\sum_{k=1}^{\infty}\rho_k(\lambda) \qquad [1.56]$$

and $\rho_k(\lambda)$ is the autocorrelation at lag k for the parameter of interest.

A number of alternative approaches to standard error estimation are documented, including the use of a sample obtained by taking the kth iteration from each of m independent chains, or retaining

every *k*th sample from a given chain, with *k* sufficiently large to achieve approximate independence (Carlin and Louis, 1996, p195; Carlin and Louis, 2000, p171). The former procedure is extremely wasteful of CPU time since the proportion of retained MCMC output is extremely small. The latter procedure is referred to as thinning and has been discussed in Section 1.6.1. Yet another method, based on *batching*, divides each MCMC chain into *m* segments (batches) of sufficient length to ensure that the correlation between segments is negligible. A subsequent calculation using the segment means (B_i) yields (Carlin and Louis, 2000, p172)

$$\hat{\text{var}}_{batch}(\hat{\lambda}_N) = \frac{1}{m(m-1)} \sum_{i=1}^{m} (B_i - \hat{\lambda}_N)^2 . \qquad [1.57]$$

CODA provides the following three standard error estimates: (1) the naïve estimate provided by the sample standard deviation divided by the root of the number of iterations; (2) a time-series estimate given by the root of the spectral density variance estimate divided by the sample size; and (3) a batch estimate calculated as outlined above (Best et al., 1995, p14). The degree to which these estimates are dissimilar depends, in part, on the degree of autocorrelation in the sample. The Raftery-Lewis analysis, which is described in the following subsection, provides an alternative method for assessing simulation accuracy.

1.7.3 The Raftery and Lewis convergence rate test

The following outline is based on Raftery and Lewis (1992, 1996), Cowles and Carlin (1996, p885) and Brooks and Roberts (1998). The Raftery and Lewis (1992) diagnostic is used to detect convergence to a stationary distribution in a single chain and to obtain bounds on the accuracy of the estimated quantiles. An estimate is provided of the number of iterations (*N*) required to obtain a given quantile to some specified accuracy, taking into account the correlation between samples. In addition, the number of iterations displaying burn-in behaviour is indicated together with the thinning number, *k*. The latter refers to the maximum proportion of samples (1:*k*) that may be retained in order to achieve independence within the chain, given an MCMC process with non-

negligible autocorrelation. Standard implementations of the Raftery-Lewis diagnostic provide two additional parameters, namely N_{min} and I. N_{min} is the minimum number of iterations required to achieve the specified accuracy given independent samples. I is a measure of the increase in chain length required to compensate for dependence in the sample ($I = N / N_{min}$). N_{min} is based on the binomial variance (see, for example, Raftery and Lewis, 1996), leading to the counterintuitive result that the number of iterations required to estimate quantiles near the median to a specified accuracy is greater than the number required for a quantile in the tails. To be specific, suppose the posterior quantile estimate q=Prob(λ<u|Y) is required to within $\pm r$ units with probability s. The required standard error is given by r/z, where $z = \Phi^{-1}\left(\frac{s+1}{2}\right)$, and $\Phi^{-1}(.)$ is the standard normal cumulative distribution function. Thus, assuming independence among the samples, the required sample number follows from the binomial variance and is given by $N_{min} = q(1-q)z^2 / r^2$. An example calculation is given by Raftery and Lewis (1996), in which they suppose that the quantile q_l=0.025 is required to within ± 0.0125 units with probability 0.95. The required sample number (N_{min}), assuming independence, is approximately 600. Taken together with an identical specification for q_u=0.975, it then follows that a true posterior interval coverage of between 0.925 and 0.975 will be achieved in 95% of applications.

Briefly, the remaining parameters provided by the Raftery-Lewis procedure are calculated as follows. The thinning parameter, k, is determined by comparing a 1st-order Markov model against a 2nd-order model, using the Bayes information criterion (BIC) to determine the smallest value of k for which the 1st-order model is preferred [Raftery and Lewis, 1992 (p765), 1996 (p116); Brooks and Roberts, 1998, (p331)]. Finally, Markov chain theory is used to determine the number of initial (burn-in) iterations to discard in order to obtain a chain that is within some specified limit of its estimated stationary distribution (Raftery and Lewis, 1992; Cowles and Carlin, 1996, p885; Raftery and Lewis, 1996, p117).

1.7.4 Autocorrelation and cross-correlation

As stated previously, autocorrelation is a major issue in MCMC simulation-based statistical inference because it has a direct impact on efficiency. Similarly, parameter cross-correlation has a detrimental effect on the Gibbs sampler (see Sections 1.6.3 and 1.6.4), and can affect the efficiency of the Metropolis algorithm if the transition kernel is not well matched to the target distribution. Some statisticians might argue that autocorrelation and cross-correlation are no longer a major concern in these days of cheap computing power, since poor efficiency can be overcome by increasing the sample number, as required to achieve a given accuracy. In contrast, others suggest that some attention should be paid to this issue because a marked level of correlation indicates that some form of reparameterisation is required (Best et al., 1995, p26-7; Gelfand et al., 1995, 1996; Carlin and Louis, 2000, p183). Accordingly, CODA and BOA provide options for generating auto-correlation tables and plots, together with cross-correlation data. As stated previously, *thinning* is suggested by some analysts as a mechanism for reducing the level of autocorrelation. Gelman (1996, p140) points out, however, that thinning offers no advantages except for reduced storage and computational costs (see Section 1.6.1 for additional information).

1.8 MCMC in practice. A summary of recommendations and comments

In an ideal situation the analyst would know how well the simulated distribution approximates the target distribution. In practice, however, the latter will always be unknown . This is a fundamental problem and it follows that no diagnostic procedure can prove convergence. The analyst is forced to adopt a pragmatic approach. Fortunately, an abundance of practical experience has been accumulated over the past decade or so, and a number of MCMC texts and review papers have been published in which the major practical issues are discussed. On the whole a broad consensus exists regarding good practice, although, of course, some differences remain regarding details.

Experience bears out the benefits of using a set of initially overdispersed parallel chains as a simple mechanism for detecting convergence problems. The essential problem with the single chain approach is that single-chain diagnostic tests may fail to detect significant convergence failure [Gelman and Rubin, 1992a, 1992b (pages 458 and 471); Gelman, 1996 (p132-3); Raftery and Lewis, 1996 (p128)]. Thus, convergence cannot be assumed solely because a single-chain diagnostic test appears satisfactory. The importance of performing a visual inspection of parallel chain trace plots has been emphasised by a number of analysts. Interestingly, one participant in a 'roundtable discussion on MCMC practice' does not use any formal test of convergence, but relies entirely on autocorrelation estimates and a visual inspection of trace plots, including the log(posterior density) trace plot (contribution from Neal to the discussion in Kass et al., 1998, p95). A final point discussed by Kass et al. (1998, p95) is the situation in which all parameters of interest exhibit convergence while one or more of the remaining parameters fail to meet the convergence criteria. Clearly, any temptation to use the MCMC output without satisfying convergence on every parameter should be resisted because the true distribution of any individual parameter must be uncertain until the entire target distribution is sampled adequately.

The following general recommendations have been compiled from a variety of sources, including Gelman et al. (1995, p322), Carlin and Louis (1996, p196), Gelman (1996) and Kass et al. (1998).

- Run a number of chains starting at points taken from a distribution believed to be over dispersed with respect to the stationary distribution, covering, for example, ±3 standard deviations from the prior mean. The starting points might be chosen systematically rather than at random (Carlin and Louis, 1996, p196; See Section 1.6.1 for additional information).
- Visually inspect overlaid chain trace plots for each parameter.
- Investigate cross-correlations among parameters suspected of being nearly confounded.

- When working with hierarchical models, examine all fixed effect parameters, some variance components and a few well-chosen random effects (say two individuals at opposite extremes).
- Use several alternative convergence diagnostic tests rather than any one method, although some analysts might focus on a particular diagnostic in the first instance. For example, both Gelman (contribution to the discussion in Kass et al., 1998, p94-5) and Carlin and Louis (1996, p196) pay attention to the Gelman-Rubin shrink factor during the initial stages of an analysis. Gelman aims for a value less than 1.2 in every parameter.

Finally, the desirability of using several alternative MCMC algorithms has been noted, although this might be impractical.

2. METHODS

2.1 Experimental methods

Brief descriptions of the experimental and data acquisition methods are given in each of the relevant chapters as follows: the DC-potential/ADC study methods in Chapter 3 (Section 3.2.1), (the same data are used in the ADC threshold analysis of Chapter 4); the electron spin resonance Arrhenius changepoint study methods in Chapter 5 (Section 5.2.1); the intestinal ischaemia-reperfusion study methods in Chapter 5 (Section 5.2.2). These details are provided mainly in the form of references to published experimental methods, with the exception of the unpublished electron spin resonance work, which is described in greater detail.

2.2 Statistical methods

A complete description of the statistical model used in the DC-potential/ADC latent variable study of Chapter 3 is given in Section 3.2.2. Section 4.2.2 provides additional details relating specifically to the ADC threshold analysis of Chapter 4. The electron spin resonance Arrhenius changepoint and the intestinal ischaemia-reperfusion changepoint statistical models are described in Section 5.3.

The DC-potential/ADC latent variable and ADC threshold simulation analyses were both performed using Markov chains generated using the Metropolis algorithm (see Section 1.5.5). These simulations were performed using software written in SAS IML (SAS Institute, 1990). With the exception of convergence testing, all subsequent analyses were performed using SAS procedures (SAS Institute, Cary, NC., USA.). In particular SAS/GRAPH was used to generate the graphical output (overlaid chain plots, trajectory plots, etc.) that are an important component of the work. Markov chain convergence testing was performed as outlined below (Section 2.2.2). The same basic approach was used in the changepoint Metropolis MCMC simulation analyses of Chapter 5. For comparison Gibbs sampler (winBUGS) MCMC changepoint analyses were also performed. Additional details regarding the changepoint MCMC methods are given in Section 5.3.

2.2.1 Metropolis transition kernel. Initial estimate and adaptive updating procedure

DC-potential/ADC analyses

A key to the successful implementation of the Metropolis algorithm, especially when applied to problems containing a large number of parameters, is the choice of transition kernel, since a considerable gain in efficiency is achieved by using a kernel that matches the target distribution (see Sections 1.6.2 and 1.6.4). Specifically, given a multivariate normal (MVN) target distribution (a transformation can be used to achieve an approximate MVN target, if necessary), a general rule states that near optimum efficiency (i.e., near optimum performance in terms of the rate of convergence to a stationary distribution) is obtained by using a MVN transition kernel with covariance $c^2\Sigma$, $c \approx 2.4/\sqrt{d}$, where Σ is the covariance matrix of the target distribution, and d is the number of parameters. This gives an acceptance rate of about 0.23 for problems with $d > 5$ (Section 1.6.2). Thus, a starting estimate of Σ is required in order to calculate the transition kernel. In addition, this estimate is used to generate a set of overdispersed starting points for the set of parallel chains. The procedure used to obtain an estimate for Σ and to generate a suitable transition kernel is outlined in the following paragraphs. The MCMC simulations of Chapters 3 and 4 [i.e., the DC-potential/ADC (V_{DC}-ADC) response analyses] were found to be particularly sensitive to the choice of transition kernel, mainly because the number of parameters is large (134). The following description refers specifically to the procedure used in that work. The changepoint problem of Chapter 5 is relatively easy in terms of transition kernel initialisation. The same general procedure was adopted, however, as described at the end of this section. Referring to the V_{DC}-ADC simulation analysis, the random coefficients were held at fixed values obtained by fitting each individual curve (or set to reasonable values, if an individual nonlinear regression calculation failed to converge). Initial fixed-effect parameter variance estimates were then derived from bivariate likelihood plots, taking parameter pairs selected on the basis of their potential for

correlation. Accordingly, the transition rate and associated transition-time parameters were taken in pairs and the transition scaling and location parameters were treated in pairs. These estimates were used to construct a starting diagonal transition kernel which, in turn, was used to perform a preliminary simulation. Three parallel chains were generated, two of which were started at positions in parameter space well removed from the first chain. The latter was started using individual-specific least-squares parameter estimates and other approximations to the posterior mode. At this stage, the same transition kernel was used for the random-effect and fixed-effect parameters. The sample obtained by combining the three parallel chains was used to calculate a target distribution covariance matrix estimate, which was used, in turn, to calculate an updated transition kernel. The equality of the fixed-effect and random-effect parameter transition kernels was lifted at this second stage, since the first stage parallel chains permit the estimation of separate fixed- and random-effect transition kernels. This iterative transition kernel updating procedure (referred to as *manual updating* in the following paragraphs) was repeated, as required, to obtain a satisfactory transition kernel. At each stage overlaid trace plots were produced for each parameter and used to assess performance based, at this stage, on visual inspection alone. A failure to obtain a good overlap over the entire length of the set of overlaid trace plots was taken to indicate convergence failure caused by inadequacy in the transition kernel. After several iterations satisfactory overlaid trace plots were obtained.

In addition to the manual updating scheme outlined in the preceding paragraph, a within-chain transition kernel updating procedure was also incorporated into the algorithm. Within-chain updating was based on a covariance matrix calculated using a preceding segment of the current chain. This gives rise to a within-chain adaptive algorithm in which transition kernel updates are performed at intervals using a moving window approach. Despite the use of this within-chain, moving-window updating procedure, good performance was not achieved without first going through several cycles of the manual parallel chain transition kernel updating sequence. The manual updating procedure is particularly effective because the extra dispersion reflected in the set of partially offset chains obtained for the parameters that fail to converge to a common

distribution serves to increase the estimated variance, relative to the variance estimate provided by any of the chains in isolation. This depends, of course, on the provision of a set of suitably overdispersed parallel chain starting values. Thus the dispersed nature of the parallel chains obtained for the subset of problem parameters causes their variance estimates to become inflated, while the variance estimates obtained for the remaining parameters, ie., those exhibiting good overlap among their parallel chains, will be relatively unaffected by the update. The resulting increase in the subset of transition kernel elements causes a more rapid movement over the corresponding parameter space during the next-stage simulation run.

Although this manual updating procedure was successful in the sense that a given set of problem parameter chains always showed a markedly improved movement through parameter space, as reflected in the overlaid chain trace plots obtained subsequent to the transition kernel update, in the initial stages this was accompanied by the introduction of poor coverage in one or more of the remaining, previously well-behaved parameters. Thus, some parameter chains found to be satisfactory at a given stage, as indicated by their overlaid trace plots, exhibited a marked deterioration in chain overlap at some subsequent stage in the manual updating sequence. Eventually, however, a transition kernel was generated that produced satisfactory overlaid trace plots for each of the 145 parameters in the model. At this stage the manual updating procedure was terminated. Then, using the transition kernel generated at the last stage of the cycle of manual updates, three new parallel simulations were performed, each generating a parameter vector chain consisting of 9.6×10^6 samples. Run-time thinning was applied at this stage (thinning ratio 1:100), giving 96,000 stored samples per chain. One chain was started with parameter values close to the median while the other two chains were started using an overdispersed value (approximately equal to the median plus-or-minus 3 standard deviations) for each parameter. Within-chain transition kernel updating was continued, initially, but the resulting changes were found to be negligible at this point. All updating was therefore terminated and the simulation continued. Given the negligible changes generated by the within-chain updates at this stage, the associated MCMC output was retained, together with all subsequent output. The resulting set of three parallel chains

was subjected to several diagnostic tests of convergence, as outlined in the following section. Guided by the results provided by the Raftery-Lewis procedure, a 3000-sample burn-in section was subsequently removed from the start of each chain, after which the three chains were combined and used to generate the required parameter medians and quantiles (2.5% and 97.5%).

Changepoint analyses

The method used to establish a transition kernel in the changepoint simulation analyses of Chapter 5 was similar to that described above, with the exception that a single manual transition kernel update was sufficient to generate a satisfactory kernel, as determined by a visual inspection of the resulting overlaid chain plots.

2.2.2 Convergence assessment

As stated above, a preliminary assessment of convergence was made through a visual inspection of overlaid trace plots. By starting the three parallel chains at different points in parameter space, two of which were overdispersed in every parameter, a failure to achieve a good coverage of parameter space was easily identified in the form of poor mixing of the parallel chains obtained for one or more parameters and a consequent failure to achieve good overlap over the entire length of the set of chains. The sequence of manual adaptive simulations, described above, was continued until this visual assessment indicated good coverage of parameter space. Then, using an overdispersed set of starting parameter values the simulation was restarted and a final set of parallel chains generated and used as input to the Geweke, Raftery-Lewis and Gelman-Rubin convergence tests, as implemented in the CODA (Convergence Diagnosis and Output Analysis Software for Gibbs sampling output) and BOA (Bayesian Output Analysis) programs (Best et al., 1995). A brief description of these diagnostic tests is provided in Section 1.7, and a subset of results obtained in the DC-potential/ADC analysis is given in Section 3.3. In the initial stages of this work these analyses were performed on a Unix system using an S-Plus implementation of CODA (Version 0.4, obtained from URL http://www.mrc-bsu.cam.ac.uk). Subsequent calculations

were performed on a PC using R-CODA (obtained form URL http://www-fis.iarc.fr) and BOA (obtained from URL http://www.pmeh.uiowa.edu) running under the R language.

3. THE TEMPORAL RELATIONSHIP BETWEEN THE ADC AND DC-POTENTIAL RESPONSES TO TRANSIENT FOCAL ISCHAEMIA IN THE RAT

3.1 Introduction

3.1.1 Cerebral ischaemia[1]

Ischaemia is the term given to the condition in which tissue blood flow is reduced to an abnormal level, too low to support normal cell function. Ischaemia can be transient or persistent. Furthermore, cerebral ischaemia can be global, as occurs in cardiac arrest, or focal if, for example, it is caused by embolic occlusion of a small arterial vessel. Associated conditions are hypoxia and anoxia, the former of which indicates an inadequate supply of oxygen while the latter is the term given to the state in which there is a complete absence of oxygen. Impaired blood flow of sufficient severity and duration is accompanied by a disruption of cell membrane function and cell energy failure. The resulting ischaemic damage, which is reversible in the initial stages, is selective at both the regional and cellular level (Abe et al. (1995) and Griesemer et al. (2002) both provide references to the relevant literature).

The brain cannot store oxygen and has minimum reserves of energy yielding substrates. It is, therefore, critically dependent on a sustained delivery of glucose and oxygen via the blood. [Sokoloff (1976) and Siesjö (1981) both provide a brief overview of brain metabolism in relation to the effects of ischaemia/hypoxia.] Given a condition in which cerebral blood flow (CBF) starts to decline, the brain compensates initially by extracting a greater proportion of oxygen and glucose from the circulation, thus maintaining normal metabolism and function. But there is a

---

[1] A brief review on cerebral ischaemia with an MRI emphasis is provided by Jacobs and Brant-Zawadzki (1992).

critical CBF threshold below which it becomes impossible to support normal activity through an increase in oxygen and glucose extraction. Neurons in this condition become quiescent, although ion homeostasis is maintained. At this stage the subject becomes unconscious if the insult is global, or suffers selective loss of function if the deficit is focal. Beyond this limit any additional reduction in CBF causes a massive movement of ions between various tissue compartments and this is accompanied by cell death.

A host of biochemical and physiological events are associated with cerebral ischaemia, including brain oedema. As outlined in Section 1.2, MRI is, in essence, a water imaging technique. It is, therefore, sensitive to oedema, although the precise relationship between the various MR responses (relaxation rate and diffusion image changes) and oedema is not established. The present longitudinal study of the apparent diffusion coefficient (ADC)[1] and DC-potential (V_{DC}) responses to focal ischaemia was undertaken to explore a particular aspect of this relationship.

The oedematous response to ischaemia consists of two distinct phases. In the initial phase, membrane ion pump failure occurs together with a consequent increase in intracellular water and loss of cell volume regulation. This is referred to as cellular or cytotoxic oedema. If this continues the blood-brain barrier becomes compromised and this allows a redistribution of plasma proteins and an accumulation of water via the remaining patent vessels giving rise to a second phase oedematous response referred to as vasogenic oedema. The latter occurs over a period of hours.

---

[1] The term apparent diffusion coefficient was adopted in recognition of the fact that absolute values cannot be extracted from bounded-system pulsed field gradient data in the absence of additional information.

3.1.2 MR diffusion imaging

The intrinsic sensitivity of the NMR signal to diffusive displacement (Hahn, 1950; Carr and Purcell, 1954) became established soon after the NMR phenomenon was first observed. This sensitivity to diffusion, which arises due to inhomogeneity in the magnetic field within the sample, is a nuisance in many applications. For example, it gives rise to a fundamental lower limit to the resolution that can be achieved in MR microscopic imaging (i.e., high resolution k-space imaging, Callaghan, 1991, p201). On the other hand, the deliberate application of a constant magnetic field gradient yields an increased sensitivity to diffusion, and NMR soon became established as a method for the precise measurement of diffusion in liquids (McCall et al., 1959, 1963). There is, however, a lower limit to the range of diffusion coefficients that can be measured with useful precision using the original steady-gradient method (Callaghan, 1991, p162). This lower limit was reduced considerably through a modification in which a pulsed magnetic field gradient was employed (Stejskal and Tanner, 1965; Tanner and Stejskal, 1968), and the resulting pulsed-gradient spin-echo technique soon became established as a method for measuring both diffusion and coherent displacement in liquids, colloids and porous systems (Tanner and Stejskal, 1968; Callaghan, 1991, Chapters 6 to 8).

The measurement of tissue perfusion status (i.e., blood flow at the microscopic, cellular level) is a fundamentally important component of the neurological assessment that is carried out routinely in connection with a number of pathologies. Although a variety of techniques have been developed, none are entirely satisfactory (Bell, 1984; Calamante et al., 1999; Thomas et al., 2000). Soon after MRI became established in neurology, it was suggested that pulsed magnetic field gradient spin-echo imaging might be used to measure perfusion, based on the argument that perfusion can be regarded as a pseudo-diffusive process (LeBihan et al., 1986). This expectation was never realised, however, mainly because the calculations required to extract perfusion information from diffusion-weighted (DW) images are fundamentally ill-conditioned, and cannot produce useful parameter estimates (King et al., 1992). Nevertheless, it is fortuitous that pulsed-gradient spin-echo/stimulated-echo images were found to exhibit an unexpected sensitivity to one or more of the

tissue changes that occur following an ischaemic insult (Moseley et al., 1990a, 1990b). Within a short period of time diffusion-weighted imaging was shown to provide diagnostic information in a variety of neurological conditions (Doran et al., 1990). Furthermore, diffusion behaviour within bounded systems carries structural information (Callaghan, 1991, Chapter 7) and a so-called q-space variant of the pulsed field gradient stimulated-echo technique was developed with a view to obtaining structural data, in vivo, as it relates to the tissue response to a perfusion deficit (King et al., 1994, 1997a). Thus, although the original proposal to use pulsed field gradient imaging as a surrogate for perfusion imaging was never realised, pulsed field gradient imaging [commonly referred to as diffusion imaging or ADC mapping] is a central component of the neurological assessment of stroke patients. Furthermore, it is widely used in experimental stroke research, and is becoming established as a method for investigating other central nervous system pathologies (King et al., 1997b; Thomas et al., 2000).

Although DW imaging is used extensively in both clinical diagnosis and experimental stroke research, a consensus view of the processes involved in the DW image sensitivity to stroke remains elusive, despite almost two decades of research. Numerous investigations into the extracellular DC-potential response to ischaemia have also been performed and its temporal behaviour is well documented. Furthermore, the cellular electrophysiology of the ischaemic response is well established. A study was therefore undertaken to examine the temporal relationship between the extracellular DC-potential and ADC changes that occur during a transient ischaemic episode, since this is expected to lead to an improved understanding of the biophysics that underlies the ADC response. This chapter outlines an MCMC simulation analysis of the resulting DC-potential and ADC data.

3.1.3 Cerebral ischaemia biophysics

Several hypotheses have been advanced to account for the appearance of ischaemic lesions in DW images. Latour et al. (1994b) and Norris et al. (1994) both suggested, for example, that DW image changes occur as a result of reduced extracellular water diffusivity. This is attributed to a

reduction in extracellular volume driven, in turn, by the cell swelling that occurs when water moves into the intracellular space. (Hoehn-Berlage et al. (1995) provide references documenting the extracellular space changes that occur in ischaemia.) An increase in the tortuosity of the extracellular space has also been implicated (Latour et al., 1994a; van der Toorn et al., 1996). (Helmer et al. (1995) provide an introduction to tortuosity and related measures in biological systems.) It should be noted, however, that Krizaj et al. (1996) have examined the sensitivity of water compartmentation and extracellular tortuosity to changes in osmolarity and have shown that the resulting changes in volume fraction are not accompanied by substantial changes in tortuosity (see Fig. 6 in Krizaj et al., 1996). They argue, therefore, that a change in tortuosity may not be pivotal in the DW imaging response to ischaemia, and suggest that the extra- and intracellular volume changes are themselves the primary driving force, except under conditions of extreme tortuosity. Similarly, Anderson et al. (1996) have argued that extracellular volume changes are a principal cause of the observed shifts in ADC. This conclusion is consistent with the well-documented decrease in extracellular volume that occurs in association with anoxic depolarisation (see, for example, Hansen and Olsen, 1980, and Hansen, 1985; anoxic depolarisation is the name given to the rapid and persistent negative extracellular DC-potential shift seen in response to ischaemia/anoxia). Alternative explanations for diffusion changes in ischaemia include a reduction in cell membrane permeability (Helpern et al., 1992) and a decrease in intracellular diffusivity, the latter of which is based on intracellular metabolite (Wick et al., 1995) and ^{133}Cs diffusion measurements (Neil et al., 1996). A tight coupling between the ADC response and ATP depletion has been suggested (Busza et al., 1992).

Clearly, a general theory for the DW image changes that occur in stroke remains to be established. A key to making progress towards a better understanding of the biophysical chemistry underlying the DW image response is the acquisition of more detailed information concerning the temporal relationship between the diffusion changes and the primary biophysical and biochemical events. Numerous multivariate longitudinal studies have been performed, but relatively little progress has been made. In part, this might be due to limitations in the statistical methods that have been

employed. In particular there has been a tendency to ignore the fundamentally longitudinal nature of the problem. Thus many researchers have focussed on the relationship between two or more key variables but have not always performed a rigorous analysis of the temporal relationship between them. It is unfortunate that, in general, the MR research community appears unaware of the extensive literature that now exists on the statistical analysis of longitudinal data (see Section 1.4). Furthermore mean statistics do not provide a complete characterisation of the temporal response since important information is contained in the between-subject differences in temporal behaviour.

The main purpose of the present study was to determine whether a more detailed examination of previously published longitudinal data can provide new information concerning the relationship between the various pathophysiological changes that occur during an ischaemic insult. Given the central role attributed to tissue depolarisation in cerebral ischaemia [see, for example, Balestrino (1995) and references therein and the review papers by Hansen (1985), Somjen et al. (1990) and Lipton (1999)] together with the well established relationship between diffusion changes and tissue pathology, it was decided to focus on the DC-potential and ADC data obtained in a series of experiments, performed by Harris et al. (2000) in which simultaneous measurements were made over a period of time prior to, during and following an ischaemic insult. Extracellular glutamic acid measurements were also made but these data are not used in this analysis. It is shown that by adopting a more formal analytical approach, new information can be extracted from the data, information that is key to gaining an insight into the pathophysiology of cerebral ischaemia. Particular attention is paid to the within-subject temporal relationship between the DC-potential and ADC responses. This is central to the common latent process question. A latent variables model assumes that, among a set of observations, several are related directly to some common underlying process. Direct proportionality is usually assumed although some nonlinear relationship might apply in the current context. Indistinguishable response characteristics are a necessary condition for the involvement of a linear common latent process. It has been argued that the ADC is one of several MR observables that can be used as a surrogate marker of

pathophysiology. A formal analysis of the ADC response is a prerequisite to adopting this approach.

3.2 Methods

3.2.1 Experimental methods

Focal ischaemia was induced in rats by intraluminal occlusion of the middle cerebral artery as outlined previously (Roussel et al., 1994; Harris et al., 2000). DC-potential and MR data were acquired simultaneously and continuously during an observation period consisting of a pre-occlusion phase, followed by a post-occlusion phase and subsequent recovery period. Echo-planar spin-echo images were acquired with a temporal resolution of 15s, while the DC-potential sampling interval was 4s. All other experimental details, including the MR data acquisition and DC-potential recording details are given in Harris et al. (2000). One of the 11 animals was excluded from the present analysis because the observed ADC was unusually high during both the pre-occlusion and post-occlusion phases.

3.2.2 Data analysis

3.2.2.1 Image analysis

Region-of-interest diffusion-tensor trace values were calculated using the image processing procedure described by Harris et al. (2000). Measurements were taken from a 1.5mm^2 region within the striatum, immediately below the position of the DC-potential recording electrode, resulting in a separate longitudinal set of trace values for each animal.

3.2.2.2 Longitudinal data analysis

A Bayesian MCMC analysis was performed using the Metropolis algorithm, as described in Section 2.2. A two-phase simulation approach was used, the first phase of which was adaptive and used a transition kernel updating algorithm to 'train' the Metropolis transition kernel. This was followed by a second phase in which transition kernel updating was stopped. The majority of the

MCMC output generated in the first phase was discarded and the remaining output retained for convergence assessment and subsequent analysis. Central to this analysis are the individual-specific posterior probability distributions for the DC-potential and ADC transition rates and half-response times, these permitting a statistical assessment of the temporal relationship between the two responses. Thus, in contrast to the usual interest in population-average statistics, the main focus of the present analysis is the within-animal temporal relationship between the DC-potential and ADC responses, as characterised by the individual-specific transition parameter posterior medians and 95% intervals.

A formal description of the statistical model is given in the following paragraphs, a main feature of which is the logistic function used to model the ADC and DC-potential responses. A single logistic term was used for the ADC data, while a sum of two logistic terms was used for the DC-potential response. A full random coefficients model was adopted as described below. Uniform prior distributions were used for all parameters except for the variance components, for which priors of the form $p(\sigma) \propto \sigma^{-1}$ were adopted. The Metropolis algorithm was used to generate three parallel chains, each consisting of 9.6×10^6 samples. A run-time thinning ratio of 1:100 was employed to reduce the storage and subsequent processing demands of the analysis, giving rise to a final chain length of 9.6×10^4 samples. Convergence was assessed using the CODA implementation of the Geweke, Raftery-Lewis and Gelman-Rubin diagnostics (see Section 1.7 and 2.2). In addition, the parallel chains obtained for each model parameter were inspected for signs of convergence failure.

3.2.2.3 Statistical model

The following model was adopted:

$$y_{ij(adc)} = \beta_{1i} - \frac{\beta_{2i}}{1+\exp\left[\beta_{3i} + \beta_{4i} t_j\right]} + (\varepsilon_{ij(adc)})$$
$$y_{ij(dc)} = \beta_{5i} - \frac{\beta_{6i}}{1+\exp\left[\beta_{7i} + \beta_{8i} t_j\right]} - \frac{\beta_{9i}}{1+\exp[\beta_{10i} + \beta_{11i} t_j]} + \varepsilon_{ij(dc)} \quad [3.1]$$

where

$$\beta_{1i} = (\alpha_1 + \gamma_{1i}) + (\alpha_2 + \gamma_{2i})$$
$$\beta_{ki} = 2(\alpha_k + \gamma_{ki}), k = 2, 6, 9$$
$$\beta_{ki} = \exp(\alpha_k + \gamma_{ki}), k = 3, 7, 10$$
$$\beta_{4i} = \frac{\exp(\alpha_3 + \gamma_{3i})}{\alpha_4 + \gamma_{4i}}, \quad \beta_{8i} = \frac{\exp(\alpha_7 + \gamma_{7i})}{\alpha_8 + \gamma_{8i}}, \quad \beta_{11i} = \frac{\exp(\alpha_{10} + \gamma_{10i})}{\alpha_{11} + \gamma_{11i}}$$
$$\beta_{5i} = (\alpha_5 + \gamma_{5i}) + (\alpha_6 + \gamma_{6i}) + (\alpha_9 + \gamma_{9i})$$
$$\gamma_{ki} \sim \mathrm{N}(0, \sigma^2_{\gamma k}), \; k = 1,\ldots,11; \; \varepsilon_{ij(l)} \sim \mathrm{N}(0, \sigma^2_{e(l)}), \; l = adc, dc.$$

The subscripts *ij(dc)* and *ij(adc)* indicate the *j*th DC-potential and *j*th ADC observation in the *i*th animal, respectively. α_k, $k = 1,2,\ldots,11$ and γ_{ki}, $k = 1,2,\ldots,11$, are the fixed-effect and random-effect terms, respectively. The individual-specific half-response times are given by $-(\alpha_k + \gamma_{ki})$, $i = 1,\ldots,10$, $k = 4$ (ADC response), $k = 8$ (fast DC-potential response (DC_{fast})) and $k = 11$ (slow DC-potential response (DC_{slow})). The individual-specific normalised rates (rates of change at half-maximum response) were calculated using $\beta_{4i}/4$, $\beta_{8i}/4$ and $\beta_{11i}/4$ for the ADC, DC_{fast} and DC_{slow} transitions, respectively. This amounts to an individual-specific scaling of each of the three rate coefficients to a unitary transition, i.e., a transition between the initial and final asymptotic levels of one and zero, respectively, as shown in Fig. 3.20 and Figs. 3.24 to 3.26. This is necessary when testing for compatibility with a common latent process. The above reparameterisation of the logistic function (i.e., the transformation from the β-parameter space to the α-parameter space seen by the Metropolis algorithm) was adopted in order to reduce the level of parameter cross-correlation. In addition, parameter exponentiation confers a near log-normal distribution on each of the rate parameters, β_{ki}, $k = 4, 8, 11$. For the sake of programming simplicity, the MCMC simulation analysis was performed using a time-origin shift of 10min, giving $t_{shift} > 0$ at all observation occasions.

A likelihood of the form

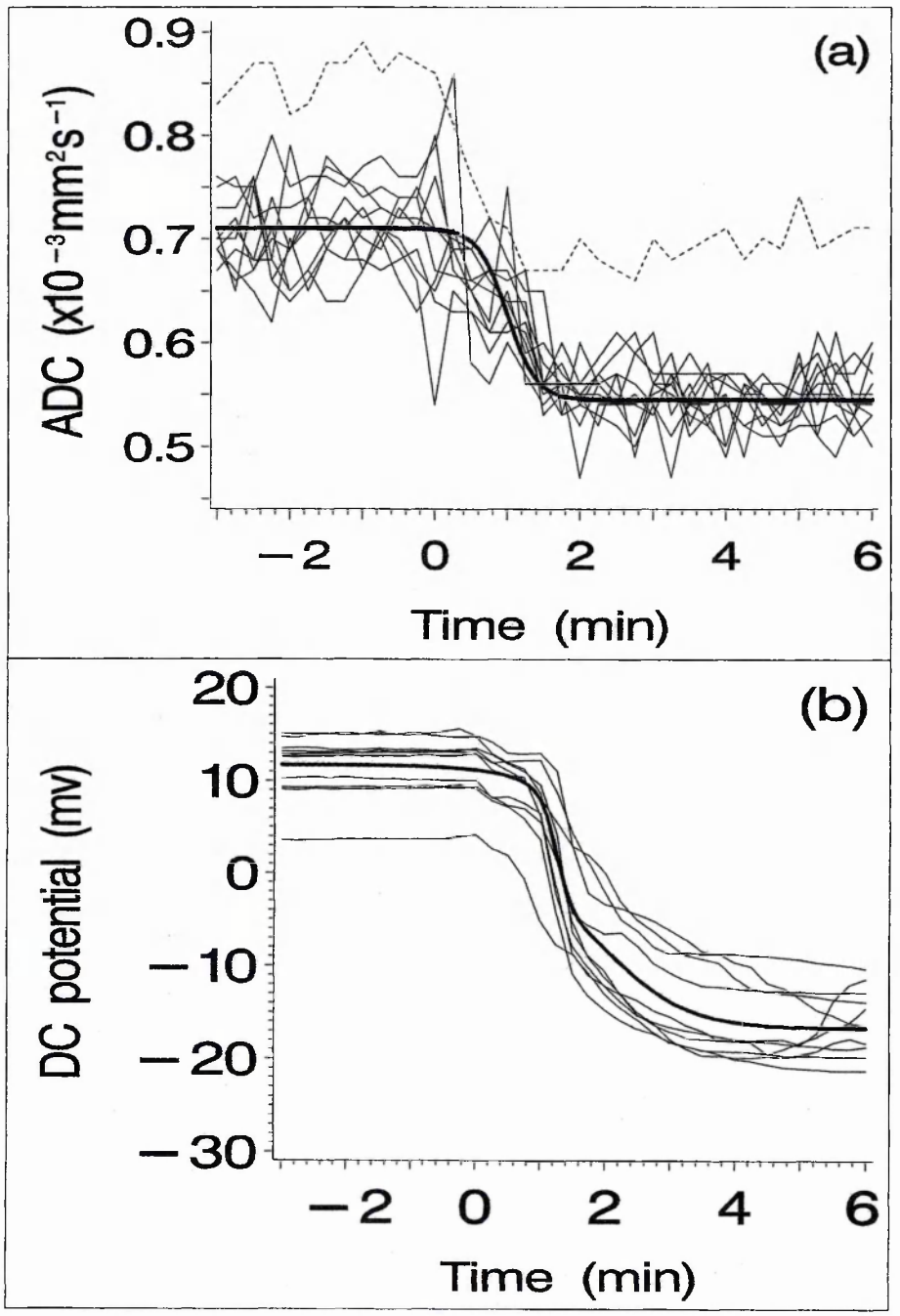
$$L(\alpha,\gamma_i,\sigma^2_{e(l)},\sigma^2_{\gamma k}) = \prod_{l=1}^{2}\prod_{i=1}^{n_s} N_{n_{i(l)}}(\mathbf{Y}_{i(l)}|f_l(\alpha,\gamma_i),\sigma^2_{e(l)}\mathbf{I}_{n_{i(l)}})\prod_{i}^{n_s} N_{11}(\gamma_i|0,\mathbf{V}_{11}) \qquad [3.2]$$

was adopted where α is the vector $[\alpha_1,\alpha_2,...,\alpha_{11}]^T$, and γ_i is the vector of random coefficients $[\gamma_{1i},\gamma_{2i},...,\gamma_{11i}]^T$ for the ith subject. $\mathbf{Y}_{i(l)}$ is the vector of observations for the lth response variable, l=(*dc*, *adc*), in the ith subject, $N_k(.|\mu,\Sigma)$ is a k-dimensional normal distribution with mean μ and covariance matrix Σ, $\mathbf{I}_k$ is the k-dimensional identity matrix, $\mathbf{V}_{11}$ is $\text{diag}(\sigma^2_{\gamma_k})$, k=1,2,...,11, and $f_l(\alpha,\gamma_i)$ is the nonlinear model (logistic function) given above for l=(*dc*, *adc*). n_s is the number of animals and $n_{i(l)}$ is the number of observations acquired for the lth response variable in the ith animal.

3.3 Results

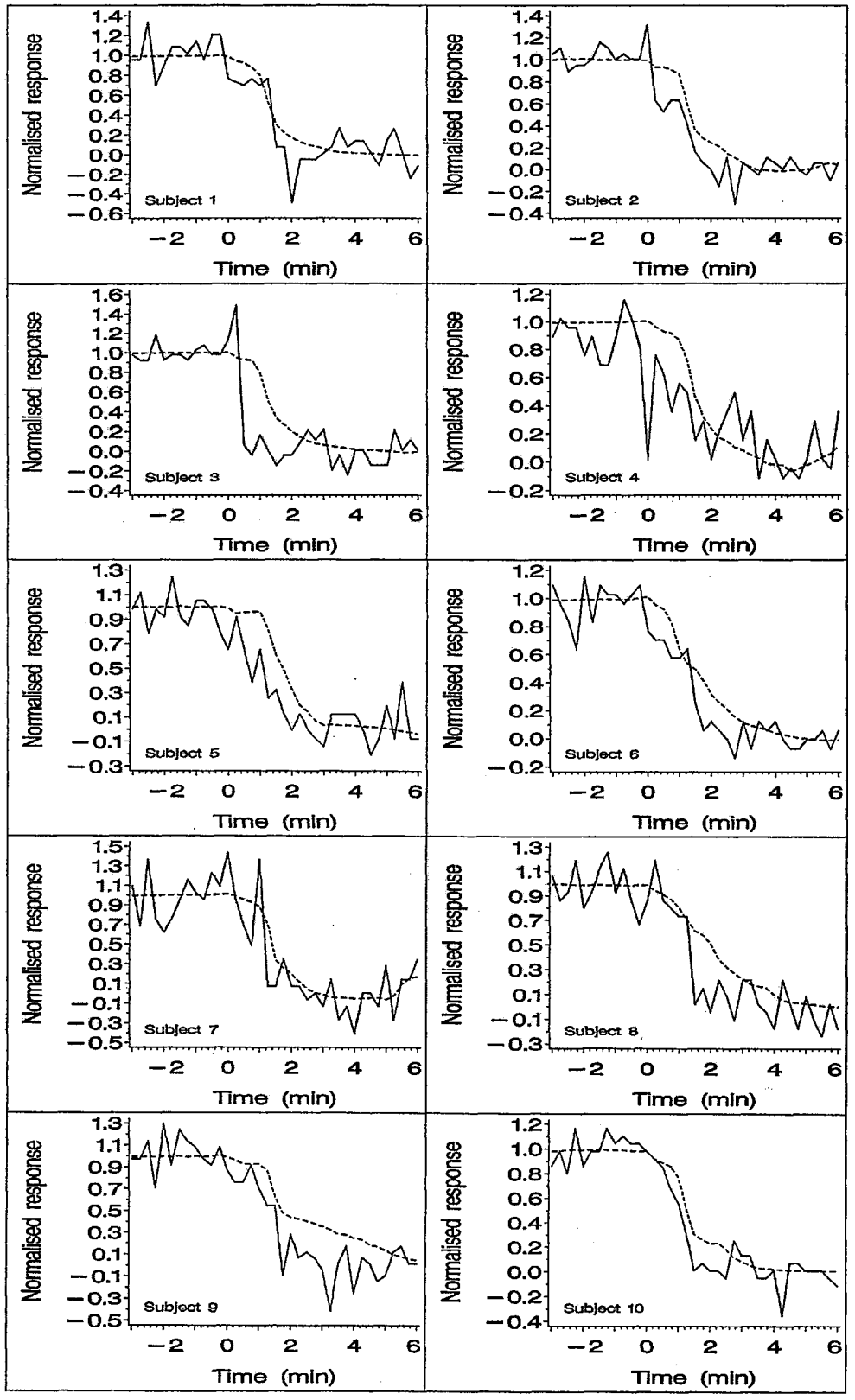
Fig. 3.1 shows the DC-potential and ADC responses acquired from each of the ten individual animals, together with the ADC data obtained from one animal that was excluded because of the unusually high ADC associated with both the pre- and post-occlusion phases. Fig. 3.2 shows the same data plotted to show the ten within-individual pair of ADC and DC-potential time courses. MCMC simulations were performed as described in Sections 2.2 and 3.2.2 and the resulting output assessed for convergence. In accordance with accepted procedure (Sections 1.6 and 1.7) formal convergence testing was preceded by a visual inspection of various overlaid chain trace plots. These plots were generated for all parameters in the model, including all random effects and variance components. Plots were also generated for the transition rate and half-response time parameters that were derived from the model and various other parameters of interest. The following subset of overlaid plots are shown: (1) the primary (i.e., Metropolis space) ADC fixed-effect parameters (α_1-α_4, Fig. 3.3); (2) the primary DC-potential fixed-effect parameters (α_5-α_{11}, Fig. 3.4); (3) var($\varepsilon_{ij(l)}$), l=ADC, DC (Fig. 3.3); (4) the derived DC-potential and ADC transition parameters (the transition rate parameter chains are displayed after log transformation; Fig. 3.5); (5) the subset of random coefficient chains γ_{ki}, i = 1,2,...,10; k = 3, 4, 7, 8, 10, 11, where i is the

Fig. 3.1. **ADC and DC-potential response profiles**



Individual ADC and DC-potential responses from ten animals are shown (thin lines) together with the population median response profiles superimposed (bold lines). The data acquired from an eleventh animal (ADC profile shown using a broken line) were excluded from the analysis because the ADC associated with both the pre- and post-occlusion phases was unusually high.

Fig. 3.2. **Normalised ADC and DC-potential response profiles**



Overlaid plots showing the normalised ADC (continuous irregular line) and DC-potential (broken line) data acquired from each of the ten animals.

Fig. 3.3. Overlaid Metropolis-space parameter chain plots are shown for each of the four fixed-effect ADC parameters and the two residual error variances [VC_1, $\log_e$(ADC error variance); VC_2, $\log_e$(DC-potential error variance)]. Two chains (yellow and red) were started using overdispersed values for all model parameters while the third chain (blue) was started at a position close to the expected posterior median. A run-time thinning ratio of 1:100 was applied giving a post thinning chain length of 96,000 iterations. For the sake of programming simplicity, the MCMC analysis was performed using a time-origin shift of 10min, hence the apparent discrepancy between the Metropolis-space trace plot shown for a_4 and the half-response time data shown in subsequent figures.

Fig. 3.4. Overlaid Metropolis-space parameter chain plots are shown for each of the seven fixed-effect DC-potential parameters. For the sake of programming simplicity, the MCMC analysis was performed using a time-origin shift of 10min, hence the apparent discrepancy between the Metropolis-space trace plots shown for a_8 and a_{11}, compared with the half-response time data shown in subsequent figures. Additional details are given in the legend to Fig. 3.3.

Fig. 3.3. **ADC fixed-effect parameter and residual error-variance overlaid chain plots**

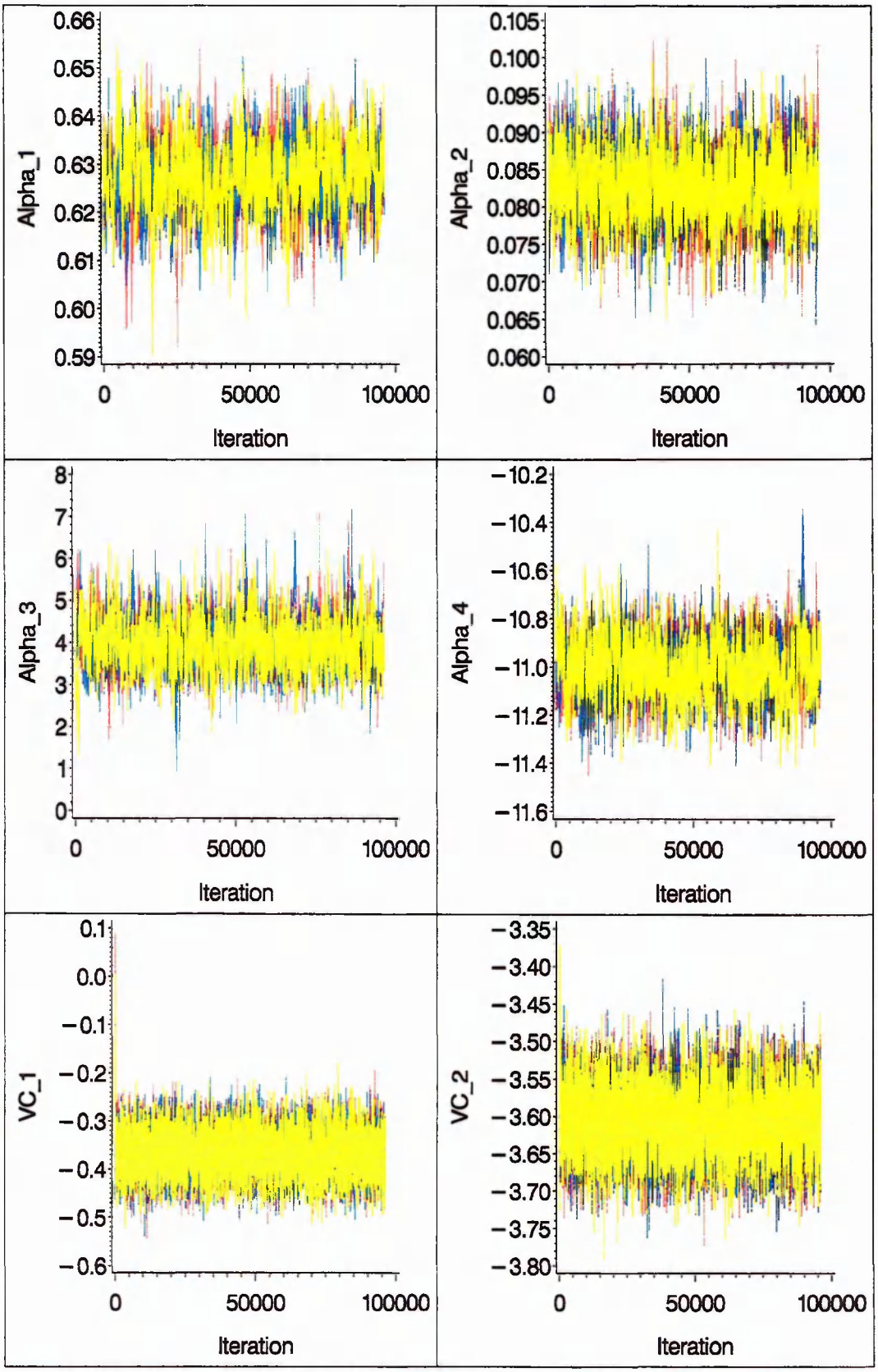


Figure legend on page 93.

Fig. 3.4. **DC-potential fixed-effect parameter overlaid chain plots**

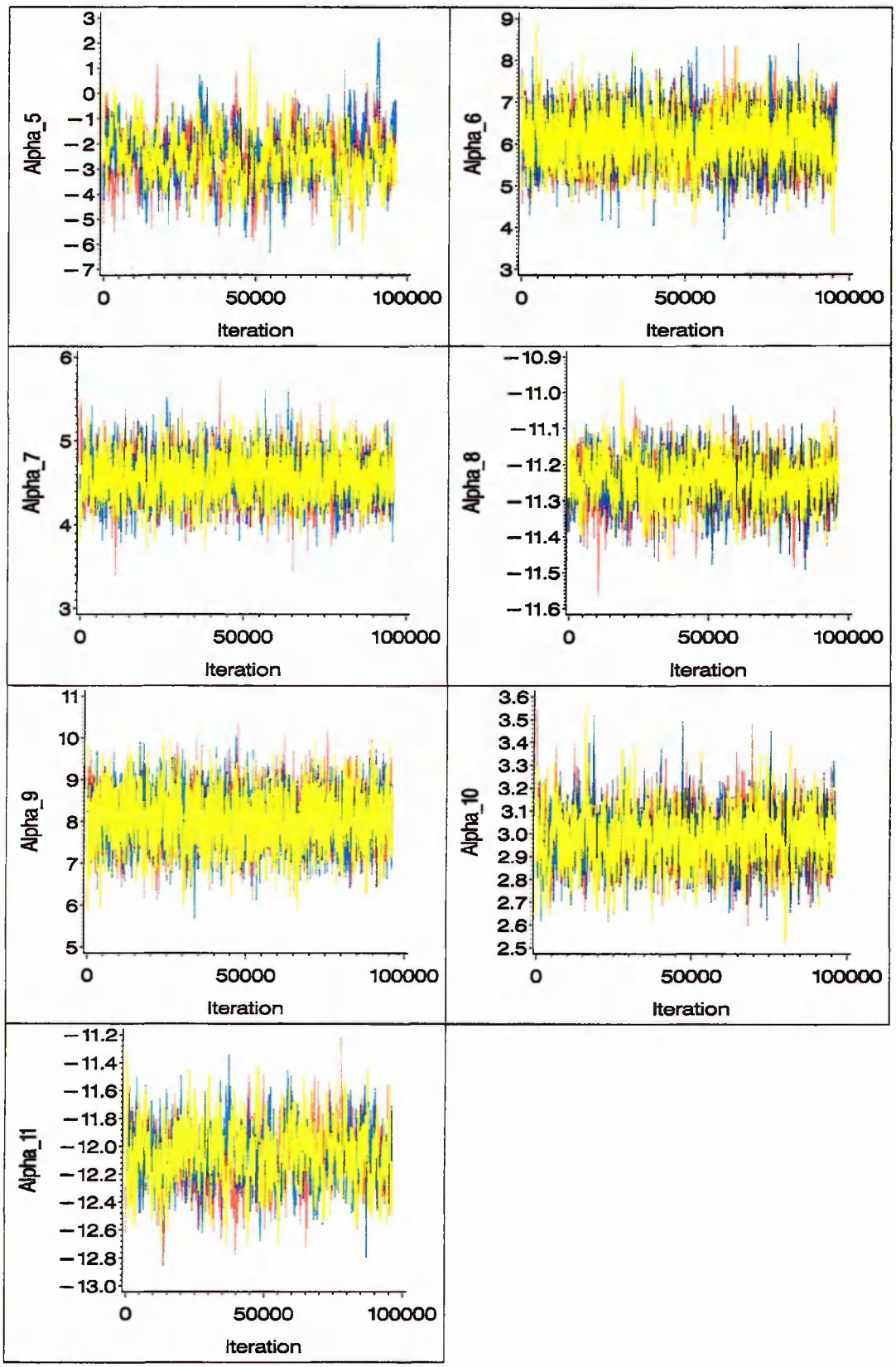
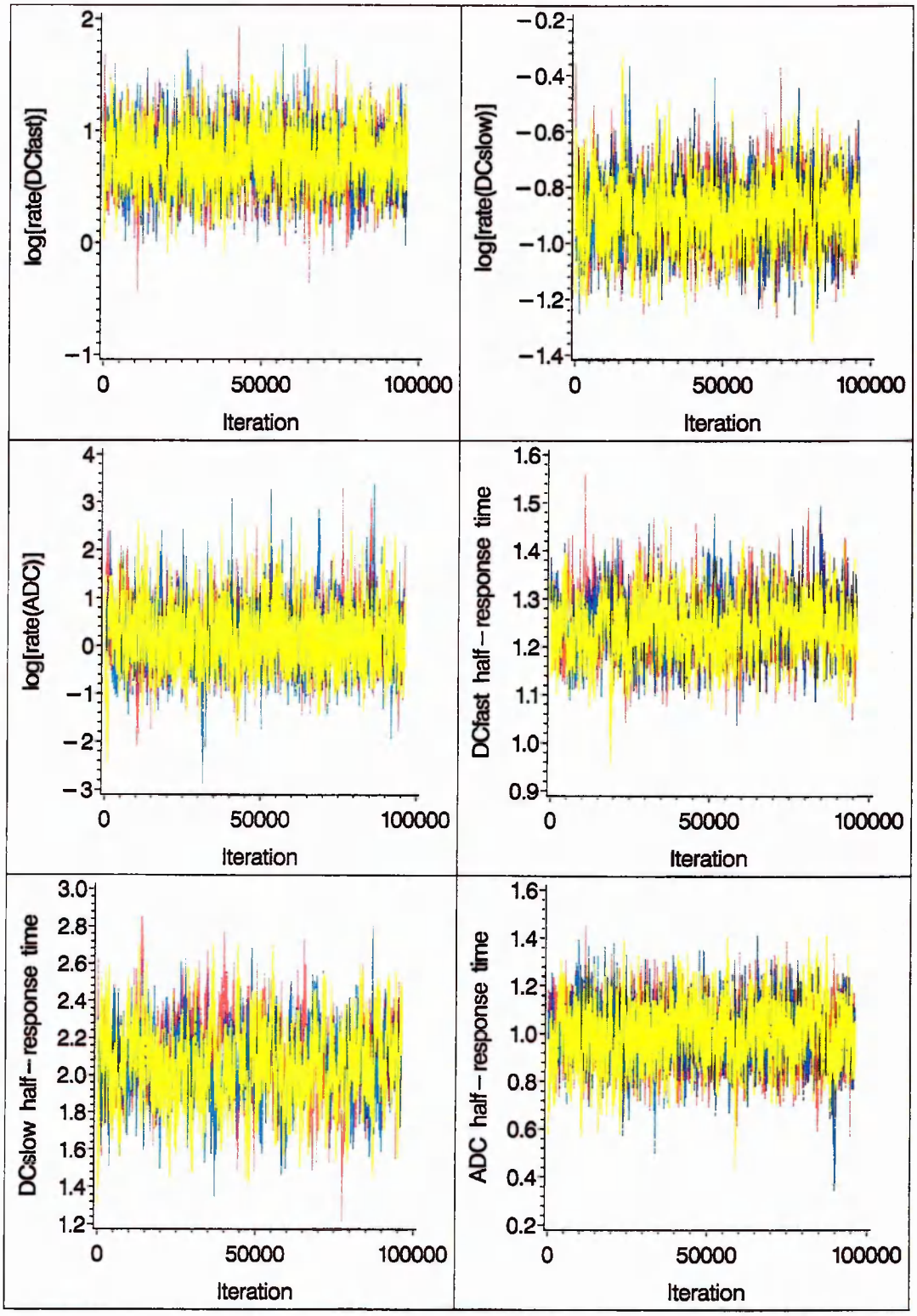


Figure legend on page 93.

Fig. 3.5. **Transition rate and half-response time overlaid chain plots**



Overlaid chain plots are shown for each of the three rate coefficients (plotted on a $\log_e(|\text{rate}|)$ scale) and for the three half-response time parameters that characterise the ADC and DC-potential transitions. Additional details are given in the legend to Fig. 3.3.

subject index (Figs. 3.6 to 3.11. This subset was selected because these random coefficients, together with their corresponding fixed-effect parameters, determine directly the individual-specific transition rate and half-response time estimates that are the focus of the analysis); (6) the log(posterior density) and a selected contrast of interest involving the difference between two individual-specific rate parameters (Fig. 3.12).

The rate of 'diffusion' over parameter space is somewhat tortuous in some instances. α_1 (Fig. 3.3) and α_5 (Figs. 3.4), for example, exhibit a substantial degree of autocorrelation despite using a run-time thinning ratio of 1:100. Nevertheless, the plots indicate an acceptable coverage of parameter space in every case. Furthermore, a good overlap of the three chains is achieved in every parameter, and the overlaid chain plots show no signs of convergence failure. Similarly, the log(posterior density) overlaid chain plot (Fig. 3.12) shows no signs of convergence problems. A variety of 2D trajectory plots were also produced, since these contain additional information. Again, the subset of six half-response time and rate-determining fixed-effect parameters are selected for display (Figs. 3.13). Although these plots indicate no major problems in the form of substantial chain separation, they do show several instances of individual chain migration into isolated regions of parameter space. Nevertheless, there is no indication of prolonged residence within these regions, suggesting that these occur due to a few excursions into regions of low probability. Histograms were produced for each fixed-effect parameter together with the six derived transition rate and half-response time parameters. A subset of these histograms is shown in Fig. 3.14. All have an acceptable shape and there are no instances in which the tails reveal sampling problems. Finally, visual inspection of various plots derived from the random coefficient

---

Figs. 3.6 to 3.11. Overlaid Metropolis-space random-coefficient chain plots are shown for each of the six parameters that characterise the ADC and DC-potential transitions. Figs. 3.6 and 3.7, ADC transition parameters; Figs. 3.8 and 3.9, fast DC-potential transition parameters; Figs. 3.10 and 3.11, slow DC-potential transition parameters. Additional details are given in the legend to Fig. 3.3.

Fig. 3.6. **Random-coefficient overlaid chain plots (γ_3)**

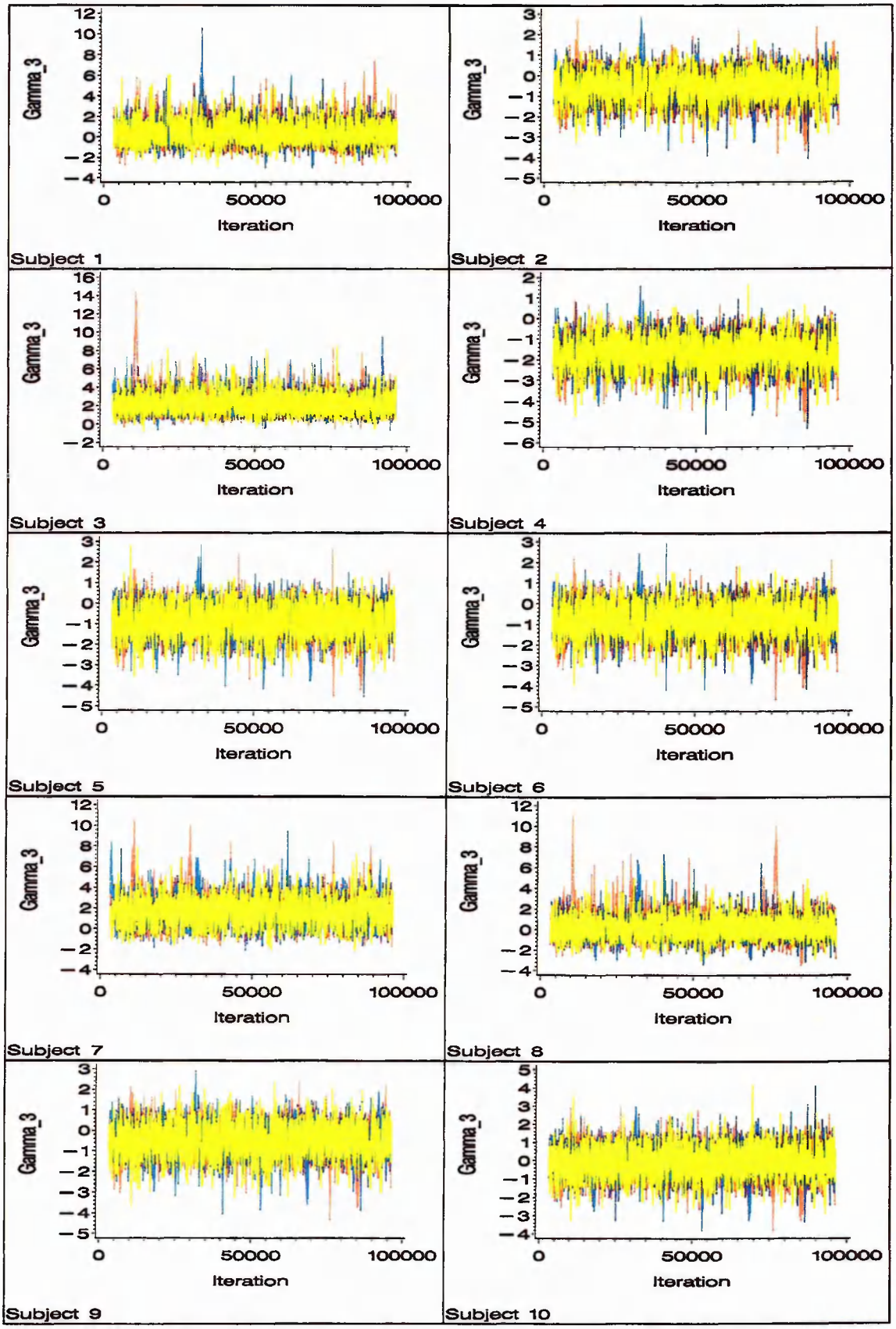


Figure legend on page 97.

Fig. 3.7. **Random-coefficient overlaid chain plots (γ_4)**

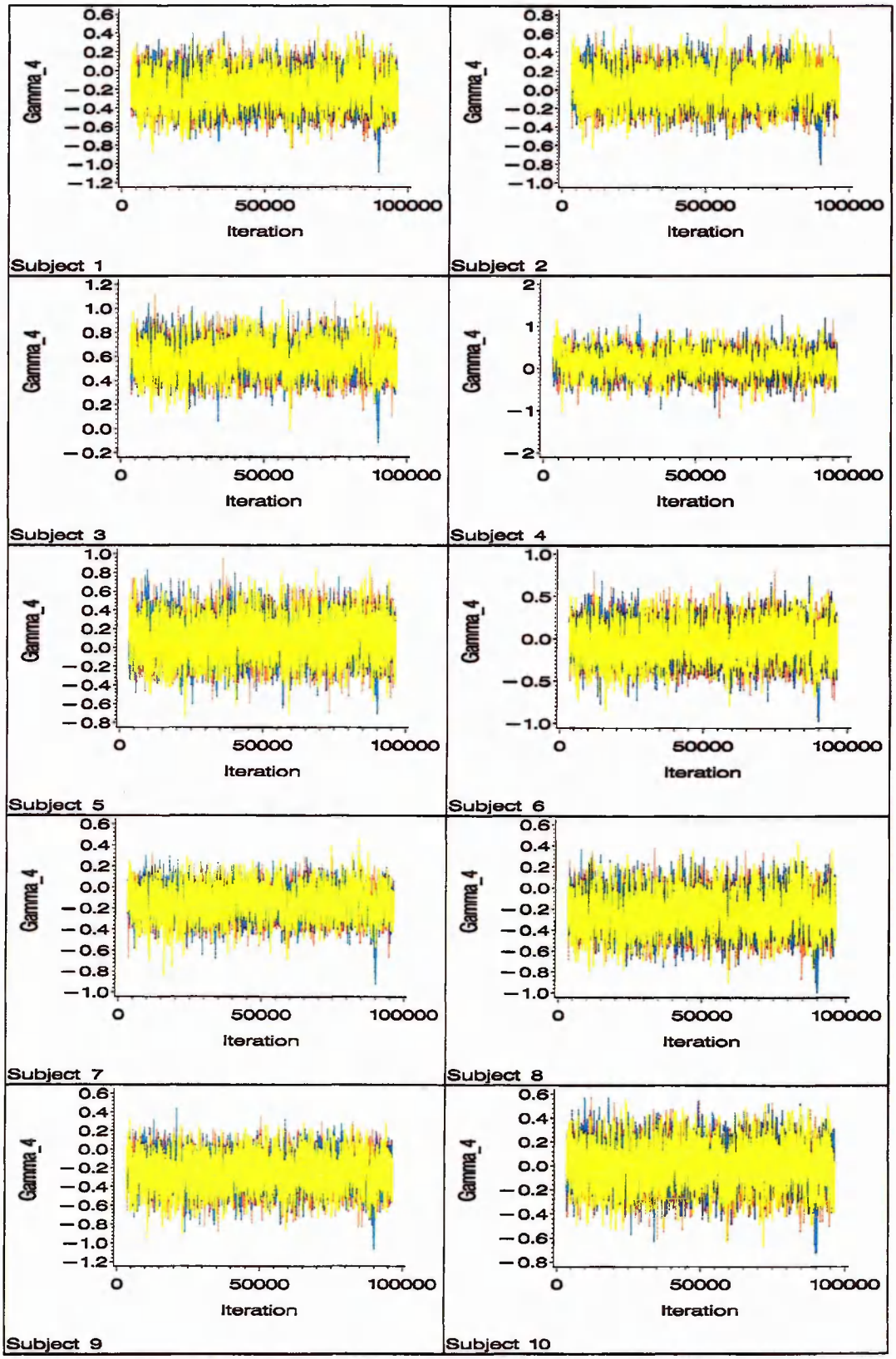


Figure legend on page 97.

Fig. 3.8. **Random-coefficient overlaid chain plots (γ_7)**

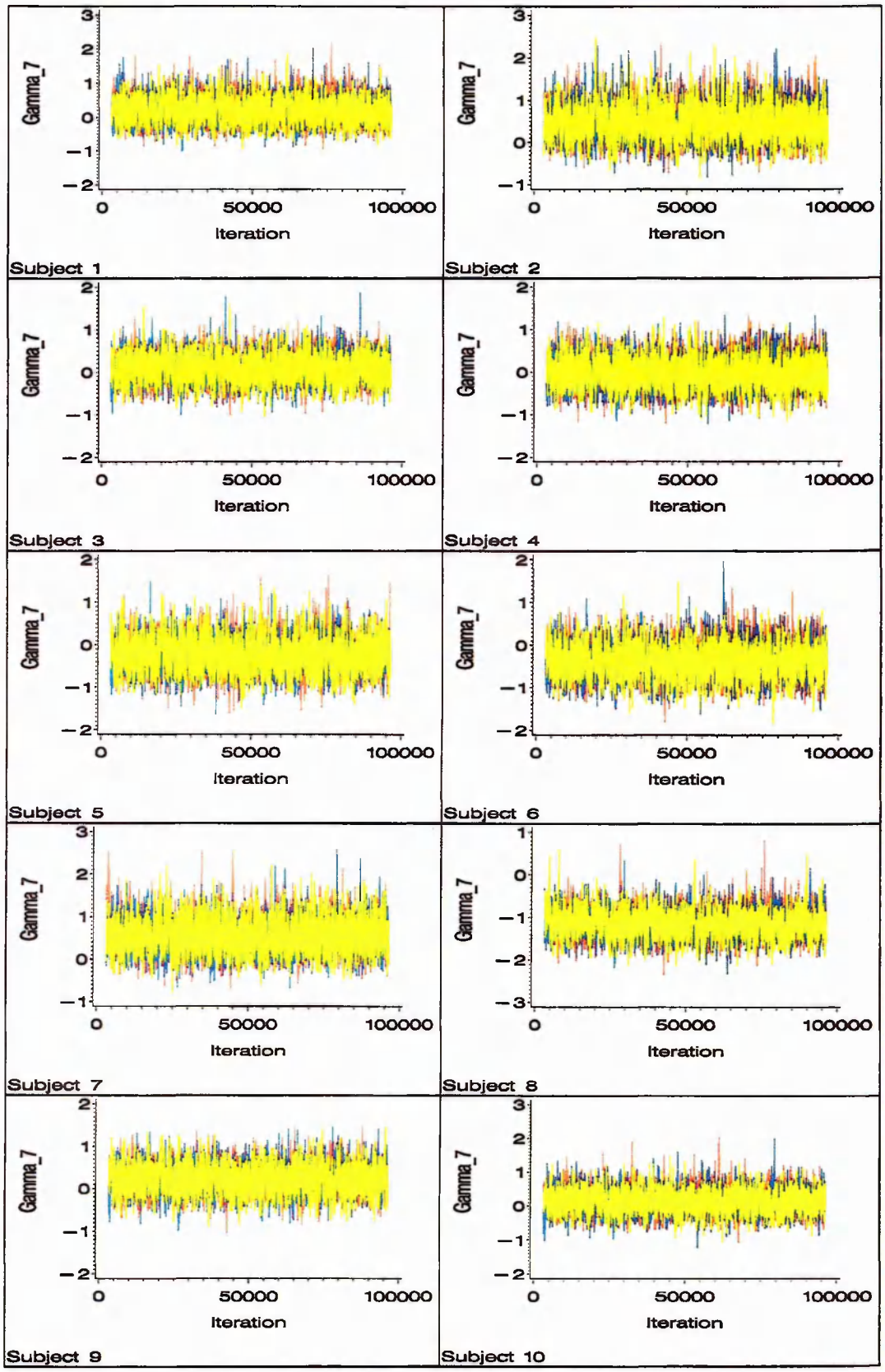


Figure legend on page 97.

Fig. 3.9. **Random-coefficient overlaid chain plots (γ_8)**

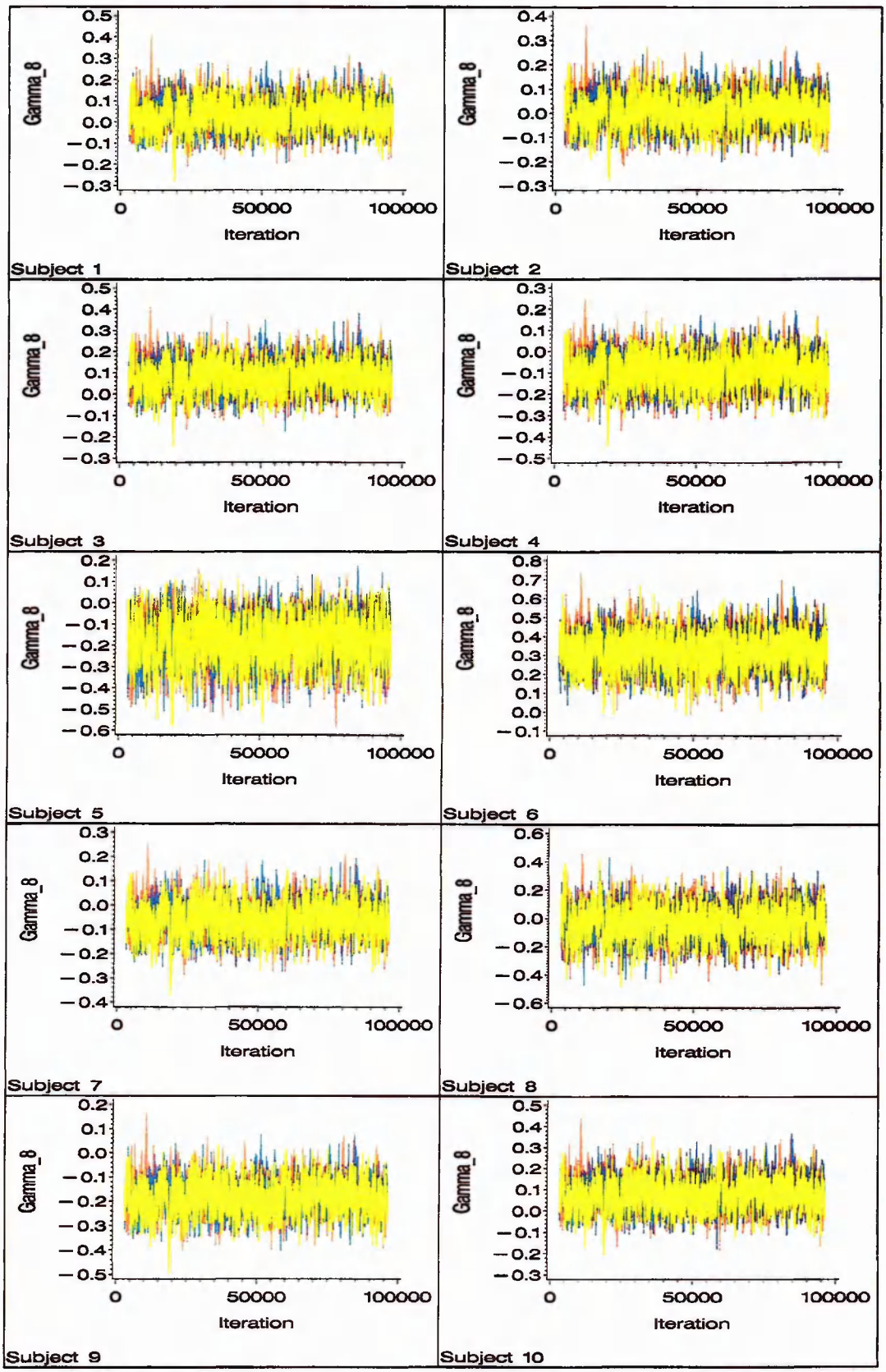


Figure legend on page 97.

Fig. 3.10. **Random-coefficient overlaid chain plots (γ_{10})**

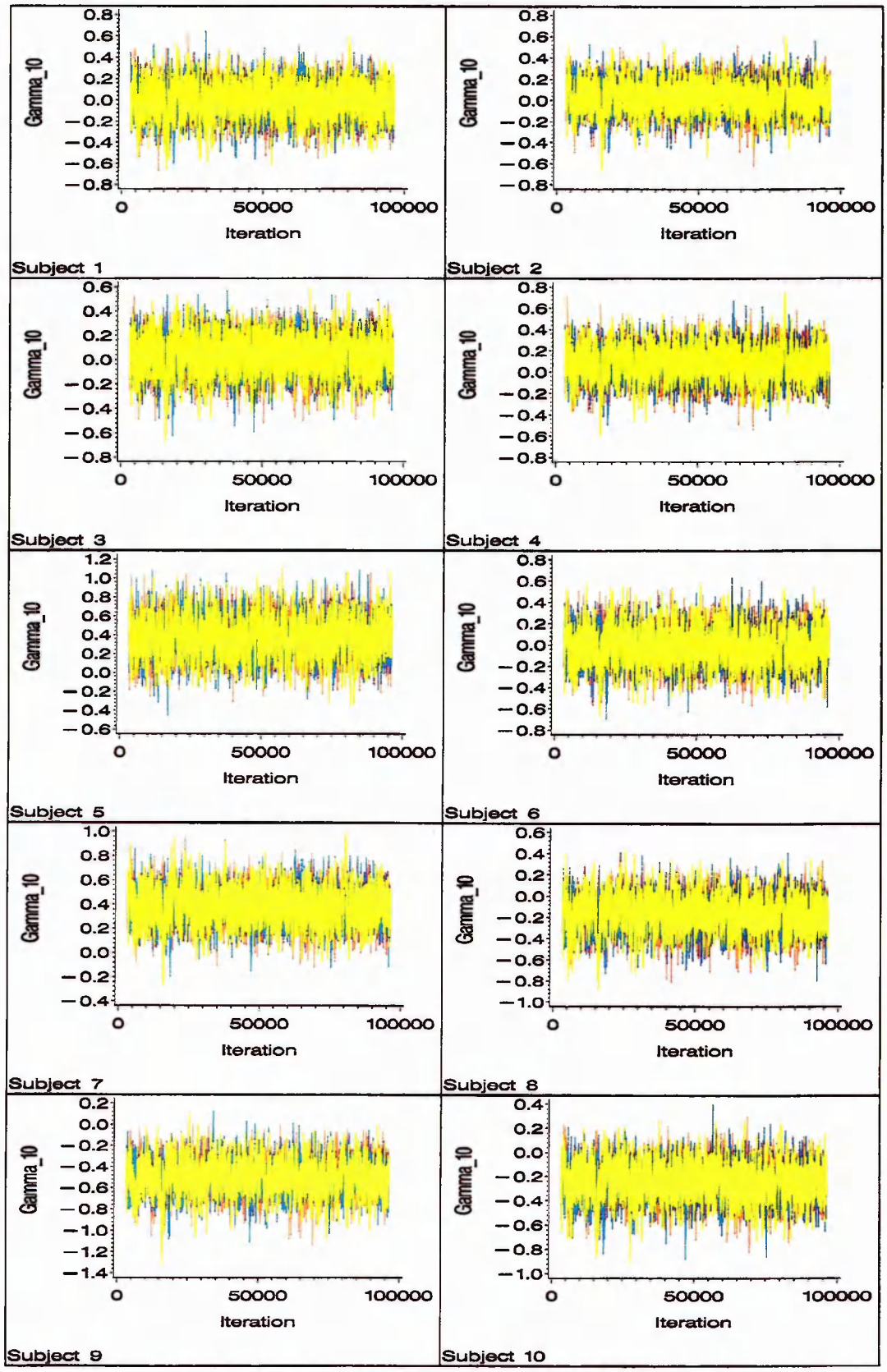


Figure legend on page 97.

Fig. 3.11. **Random-coefficient overlaid chain plots (γ_{11})**

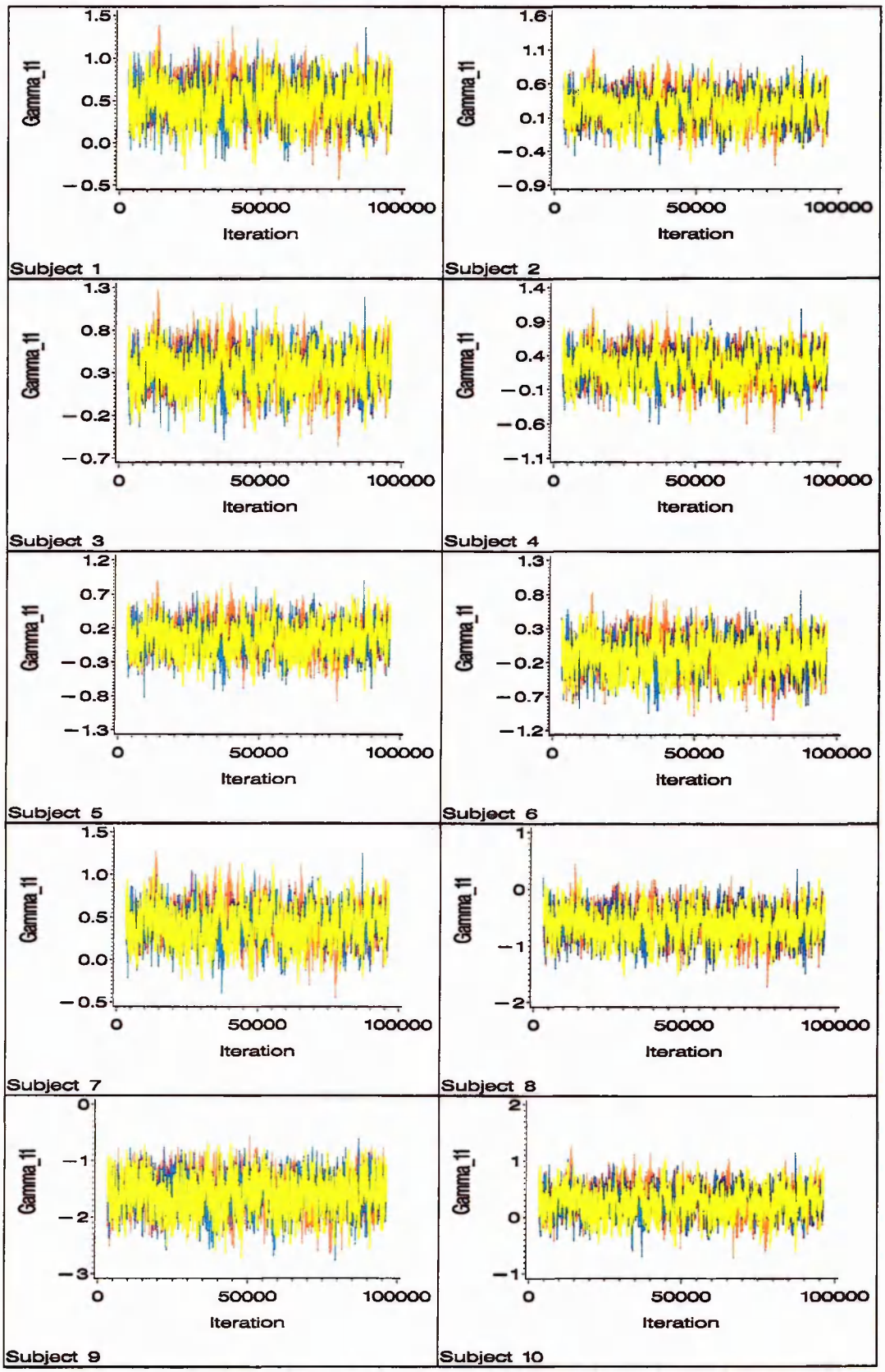
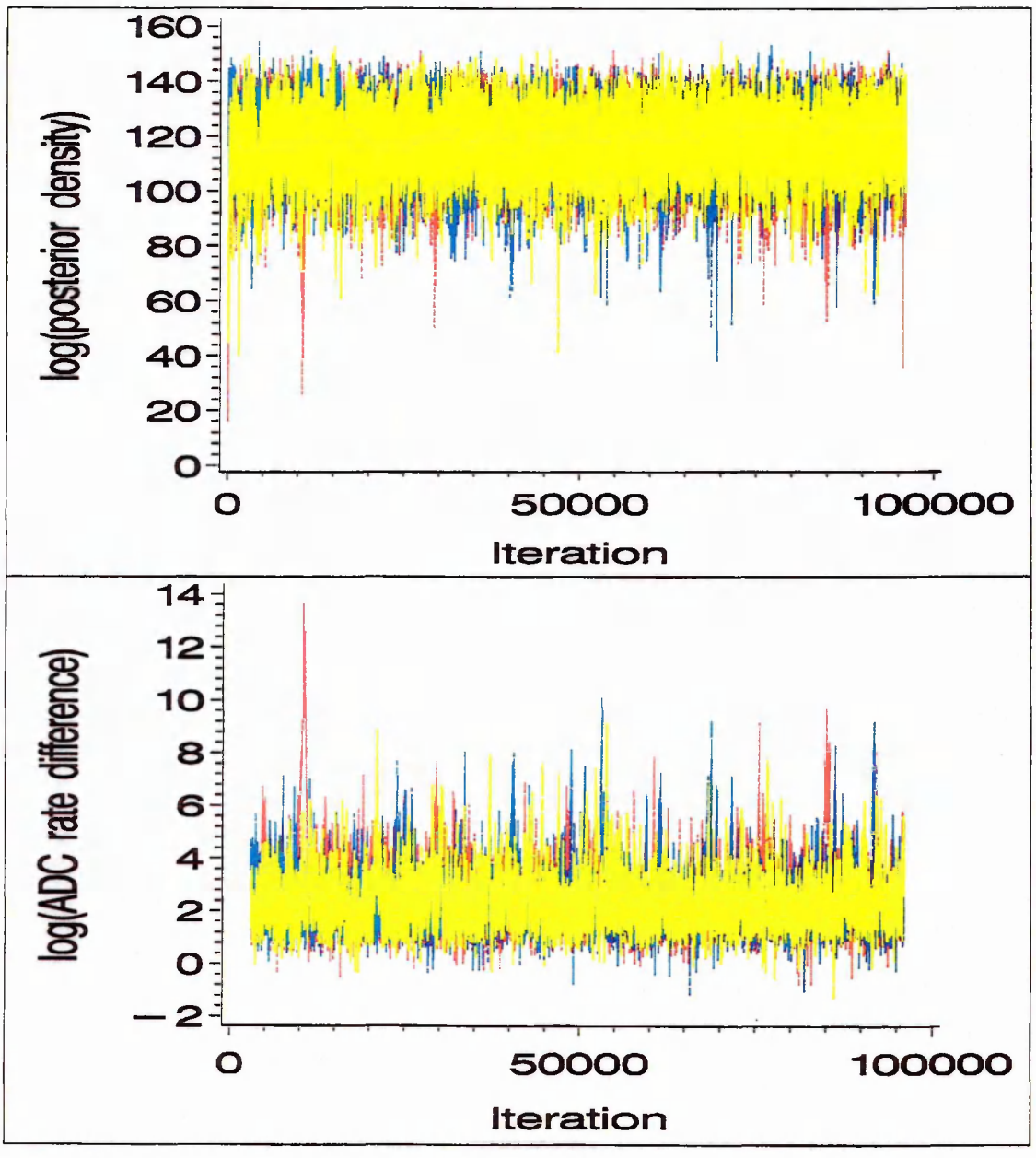


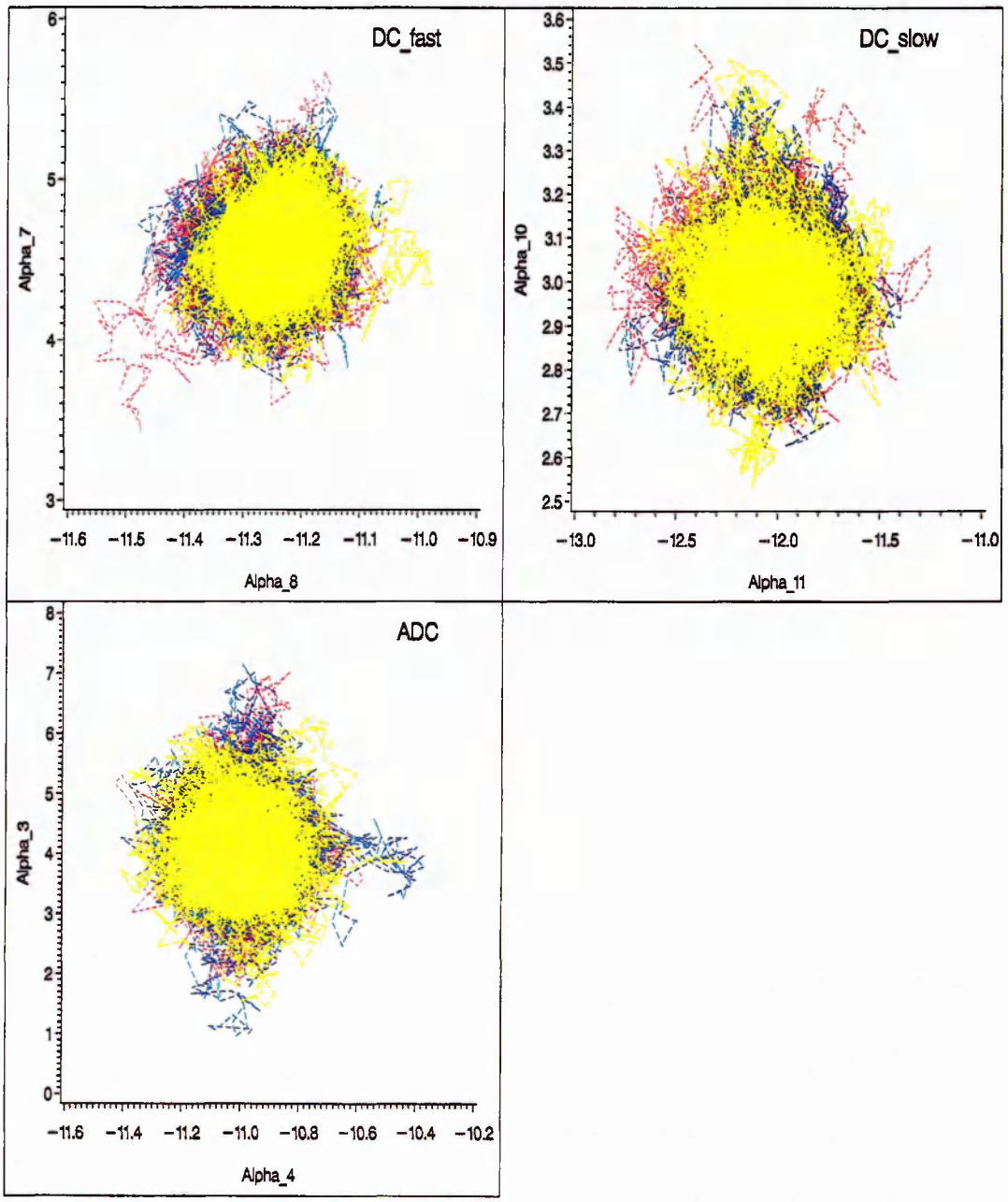
Figure legend on page 97.

Fig. 3.12. **Overlaid chain plots for the posterior density and a selected ADC transition rate difference**



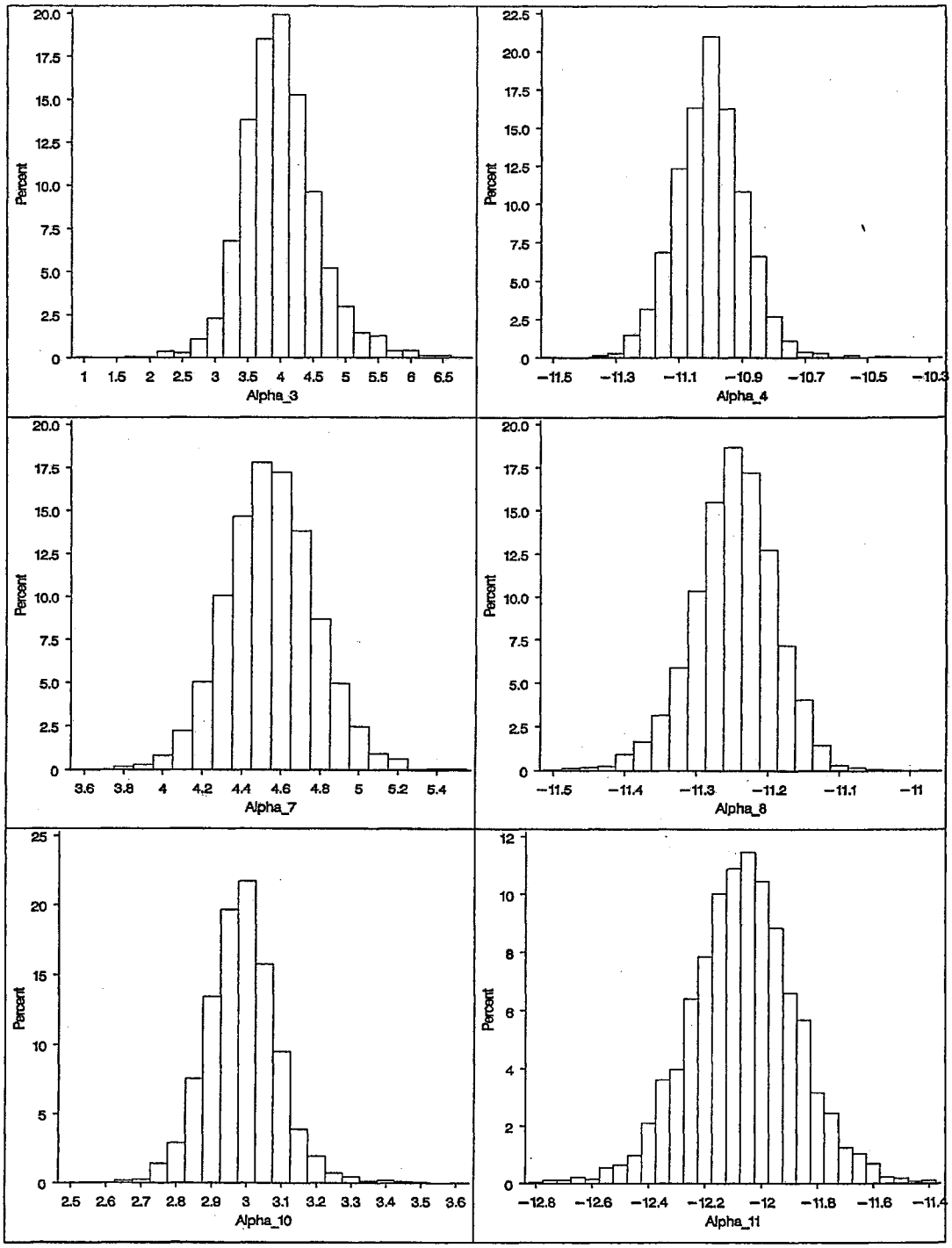
Upper panel, $\log_e$(posterior density) overlaid chain plot. Lower panel, overlaid chain plot of the natural logarithm of the difference between the ADC transition rates in the 2nd and 3rd animals.

Fig. 3.13. **Fixed-effect transition-rate and transition half-response time parameter overlaid chain trajectory plots**



Overlaid chain Metropolis-space trajectory plots are shown for each of the three pairs of fixed-effect transition-rate and half-response time parameters [DC_{fast} (a_7 and a_8), DC_{slow} (a_{10} and a_{11}) and ADC (a_3 and a_4)]. See legends to Figs. 3.3 and 3.4 for additional information.

Fig. 3.14. **Fixed-effect transition-rate and transition half-response time parameter histograms**



Histograms are shown for each of the three pairs of fixed-effect transition-rate and half-response time parameters [ADC (α_3 and α_4), DC_{fast} (α_7 and α_8) and DC_{slow} (α_{10} and α_{11})]. See legends to Figs. 3.3 and 3.4 for additional information.

chains indicate an adequate coverage of parameter space with no signs of convergence failure (a subset of overlaid chain plots is shown in Figs. 3.6 to 3.11). This is important because the random coefficients are central to the analysis of the within-individual temporal relationship between the DC-potential and ADC responses, and this is the major focus of the analysis presented in this chapter.

In summary, the overlaid chain plots, trajectory plots and histograms appear satisfactory and show no signs of convergence failure. Accordingly, a second phase assessment using formal convergence diagnostic tests was performed. This is required for two reasons. Firstly it is necessary to substantiate the assertion, based on visual inspection of the parameter chain plots, that convergence has been achieved within reasonable limits. A more formal convergence analysis provides a quantification of these limits . Secondly, it is important to obtain an estimate of the simulation error associated with the various parameters.

Three convergence test procedures were used in this study, namely the Gelman-Rubin shrink factor diagnostic, the Geweke time-series Z-score diagnostic, and the Raftery-Lewis diagnostic procedure. A number of parameters of interest were included in these convergence tests, but for illustrative purposes, a subset of five sets of results are included in this chapter. These are the log(posterior density), the three transition rates (diagnostic tests performed using log transformed rates) and the difference between two individual-specific ADC transition rate parameters. The last of these was selected because the main conclusion arising from this analysis follows from the observed pattern of between-animal ADC transition rate differences. It is essential, therefore, that these differences are estimated with sufficient accuracy. The Raftery-Lewis procedure provides an estimate of the accuracy of the associated quantiles.

Figs. 3.15 to 3.19 show, for each parameter, the three single chain Geweke Z-score plots, together with the Gelman-Rubin parallel chain shrink-factor plot. (To reduce the processing time required for these calculations, each MCMC output chain was subjected to additional thinning (1:100) prior

Fig. 3.15. **Log_e(posterior density) Geweke *Z*-score and Gelman-Rubin shrink factor plots**

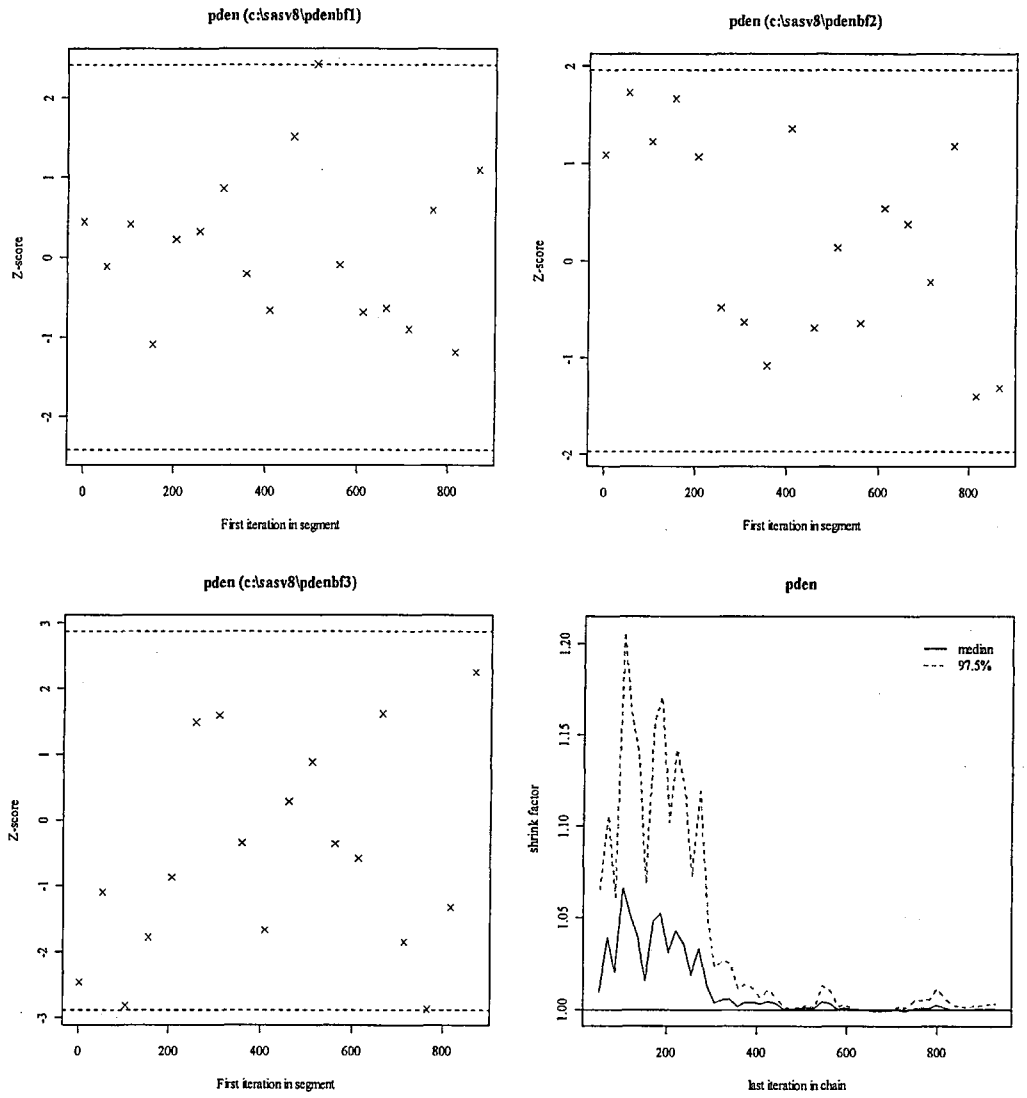
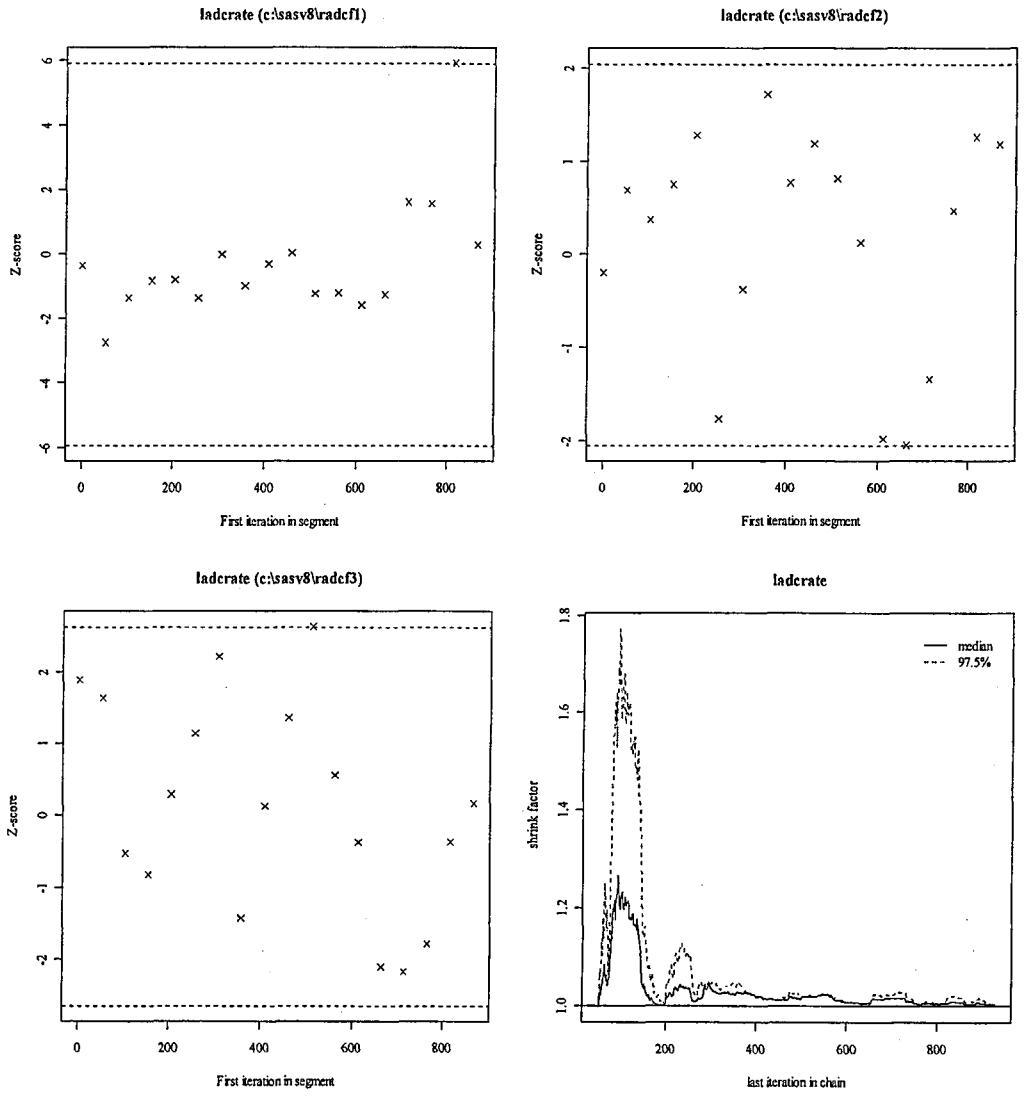
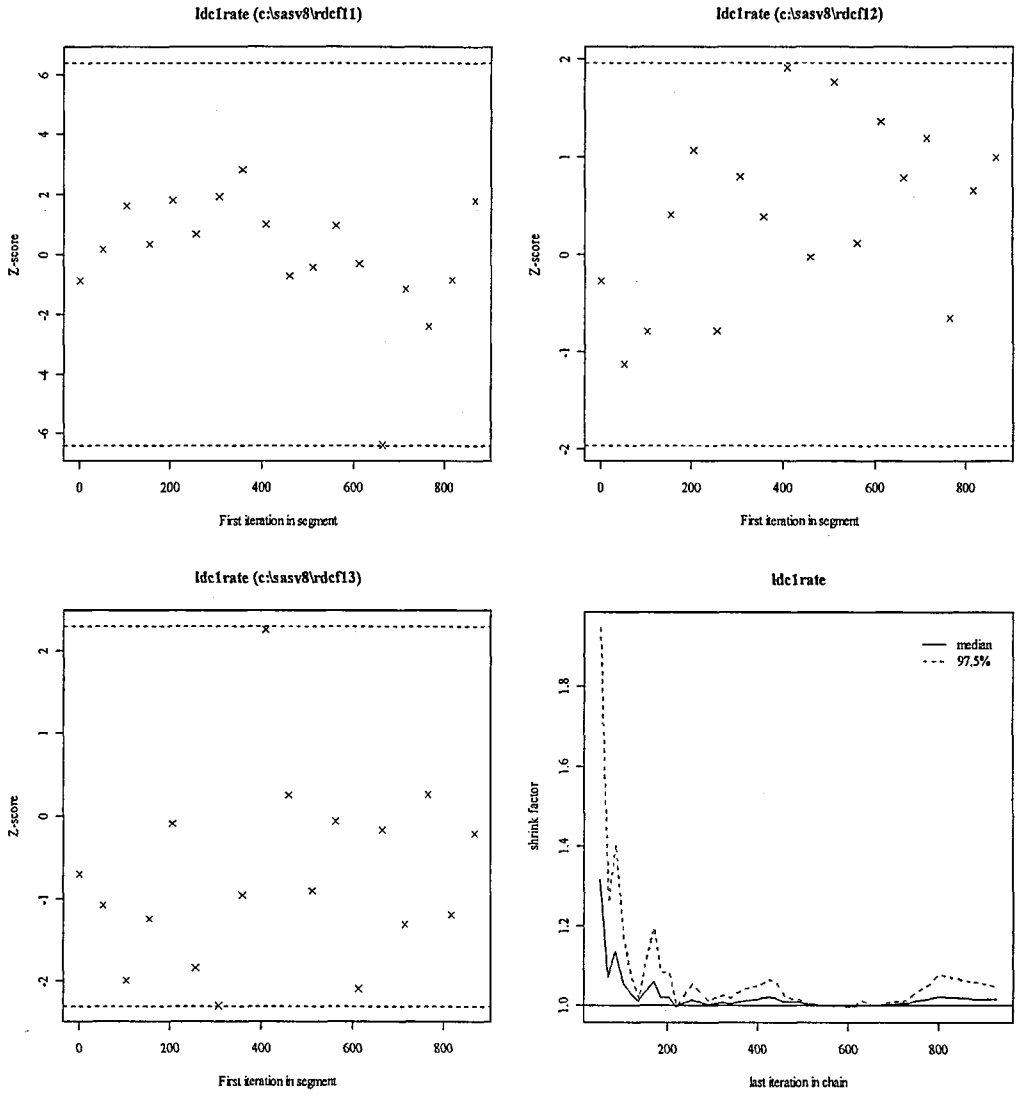


Fig. 3.16.**ADC transition rate Geweke *Z*-score and Gelman-Rubin shrink factor plots**



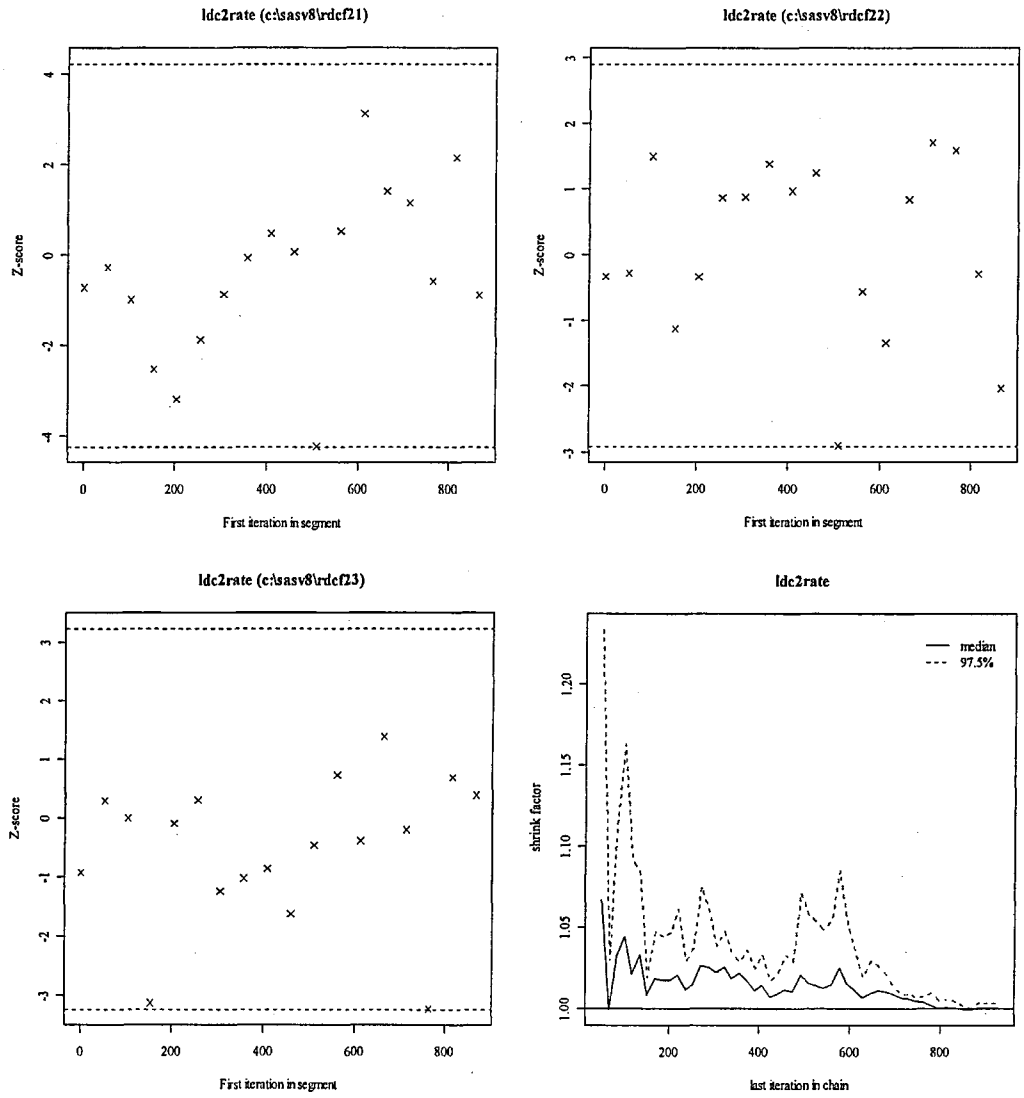
The figure shows the convergence diagnostic plots obtained for the $\log_e(|\text{ADC transition rate}|)$ chain generated from the MCMC output.

Fig. 3.17. **Fast DC-potential transition rate Geweke *Z*-score and Gelman-Rubin shrink factor plots**



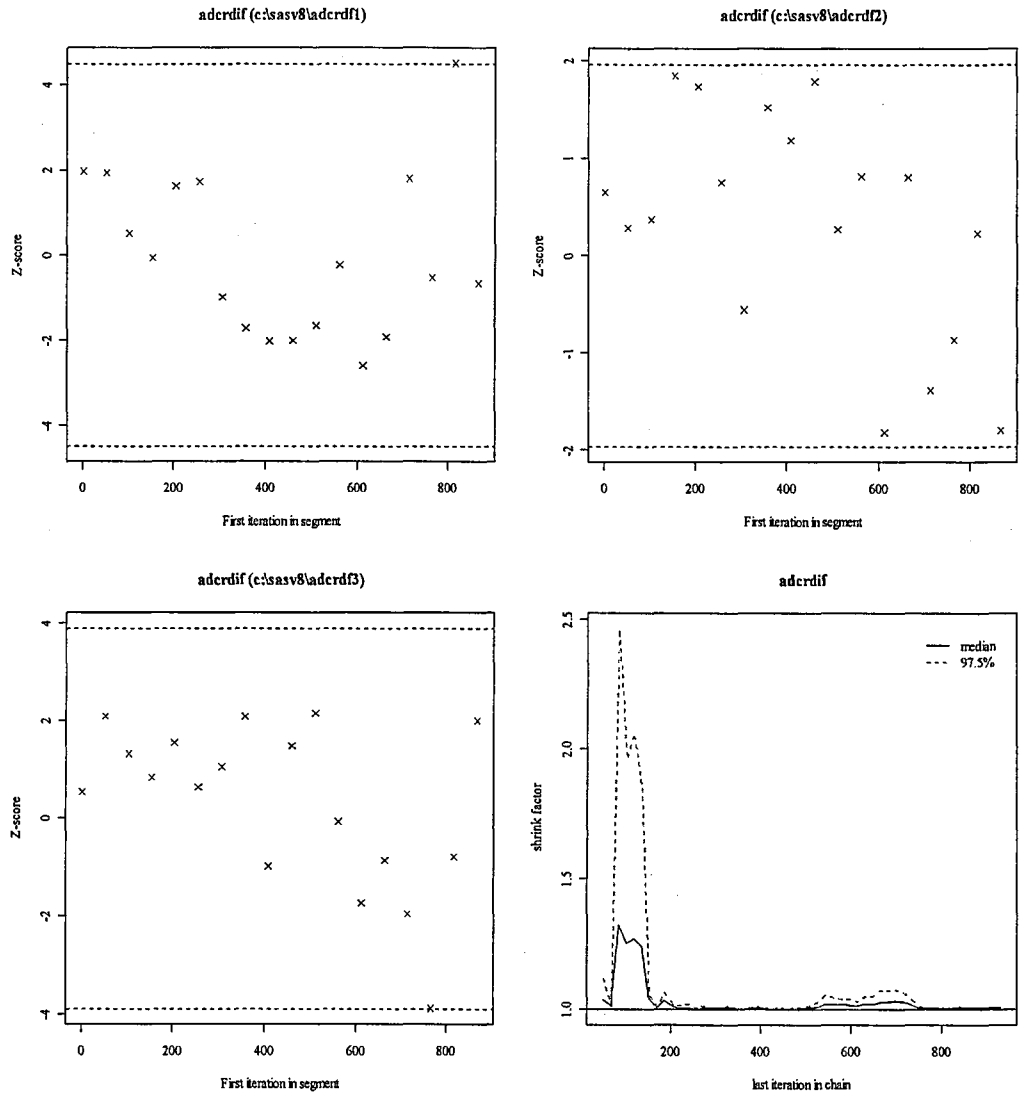
The figure shows the convergence diagnostic plots obtained for the $\log_e(|DC_{fast}$ transition rate|) chain generated from the MCMC output.

Fig. 3.18. **Slow DC-potential transition rate Geweke *Z*-score and Gelman-Rubin shrink factor plots**



The figure shows the convergence diagnostic plots obtained for the $\log_e(|DC_{slow}$ transition rate|) chain generated from the MCMC output.

Fig. 3.19. **Geweke *Z*-score and Gelman-Rubin shrink factor plots obtained for a selected ADC transition rate difference**



The figure shows the convergence diagnostic plots obtained for the natural logarithm of the difference between the ADC transition rate in the 2nd and 3rd animals, as generated from the MCMC output.

to processing, giving a final chain length of 930 samples, after discarding burn-in iterations.) The majority of Z-score points lie within the range -2<Z<2 as expected, assuming convergence, although an occasional Z-score outlier is evident. Similarly, each of the Gelman-Rubin shrink-factor plots indicates a value close to unity towards the end of each set of chains, suggesting that little improvement is achievable by increasing the number of iterations. Table 3.1 lists the Geweke Z-score obtained for each selected parameter and each chain, together with the Gelman-Rubin shrink factor and Raftery-Lewis estimate of the number of iterations, *N*, required to obtain the 0.025 quantile with an accuracy of 0.0125 with probability 0.95. These diagnostic results were judged adequate. Although the Raftery-Lewis *N* exceeds 930 in some cases (maximum *N*=1803), subsequent parameter estimates are based on the combined parallel chains giving an equivalent *N* of 2790. The second phase of the analysis was therefore undertaken using the existing output without additional chain extension.

The posterior fixed-effect transition-parameter medians (i.e., the estimated population transition-parameter medians) are listed in Table 3.2. Fig. 3.2 shows the observed DC-potential and ADC pair of response profiles obtained for each of the 10 animals, after normalisation[1] while Fig. 3.20 shows the corresponding individual-specific posterior median DC-potential and ADC response

---

[1] The main purpose of the present analysis was to examine the ADC and DC-potential responses for behaviour consistent with a common latent process. If a common process drives either of the two DC-potential component responses and the ADC response, then the ADC and corresponding DC-potential component must move between their two asymptotic states in synchrony. Thus it is necessary to scale each rate coefficient to the magnitude of their respective transitions when testing for compatibility with a common latent process. Very similar results were obtained, however, if this scaling was performed using the corresponding group medians.

Table 3.1. **V_{DC}-ADC MCMC simulation analysis. Convergence diagnostics**

	Geweke *Z*-score			GR shrink factor (97.5% quantile)	Raftery-Lewis *N*		
	Chain 1	Chain 2	Chain 3		Chain 1	Chain 2	Chain 3
Log(DC_1 rate)	-0.89	-0.27	-0.71	1.01 (1.05)	910	910	763
Log(DC_2 rate	-0.73	-0.34	-0.93	1.00 (1.00)	1090	994	1456
Log(ADCrate)	-0.36	-0.20	1.89	1.00 (1.01)	1616	1803	763
Log(rate diff)	1.96	0.65	0.52	1.00 (1.01)	642	600	600
Log(pden)	0.44	1.08	-2.47	1.00 (1.00)			

The statistics listed in this table were obtained after discarding a burn-in sample of 3000 iterations from the total of 96000 iterations in each chain, and subsequent thinning to 930 samples per chain. The number of burn-in samples was based on an initial Raftery-Lewis analysis. Having removed the effects attributable to the overdispersed starting values used for two of the three parallel chains, the Geweke *Z*-scores should satisfy the convergence criterion $-1.96<Z<+1.96$ with probability 0.95. Assuming convergence, shrink factor values and 97.5% quantiles (given in brackets) close to unity are expected (see Section 1.7.2). The Raftery-Lewis *N* provides an estimate of the number of iterations required to estimate the 0.025 and 0.975 quantiles to an accuracy of +/-0.0125 with probability 0.95. This gives a posterior interval coverage of between 0.925 and 0.975 with probability 0.95. Abbreviations: DC_1, fast DC-potential component; DC_2, slow DC-potential component; rate diff, difference between the ADC transition rates in the 2nd and 3rd animals; pden, posterior density.

Table 3.2. **ADC and DC-potential transition parameter population medians and 95% posterior intervals**

	Median	2.5% quantile	97.5% quantile
rate, DC_{fast} (min^{-1})	-2.09	-3.34	-1.35
rate, DC_{slow} (min^{-1})	-0.411	-0.51	-0.34
rate, ADC (min^{-1})	-1.18	-4.69	-0.45
Δrate(ADC - DC_{fast}) (min^{-1})	0.88	-2.52	2.31
Δrate(ADC - DC_{slow}) (min^{-1})	-0.77	-4.30	-0.027
$time_{1/2}$, DC_{fast} (min)	1.25	1.14	1.37
$time_{1/2}$, DC_{slow} (min)	2.06	1.69	2.43
$time_{1/2}$, ADC (min)	1.00	0.79	1.21
$\Delta time_{1/2}$(ADC - DC_{fast}) (min)	-0.24	-0.49	-0.02
$\Delta time_{1/2}$(ADC - DC_{slow}) (min)	-1.06	-1.49	-0.65

The rate parameters are normalised by scaling to a unitary transition as outlined in Section 3.2.2.3; $time_{1/2}$ indicates the transition half-response time. The 95% interval for the difference between the ADC and DC_{slow} transition rates [Δrate(ADC - DC_{slow})] excludes zero indicating a significant difference in median rates.

Fig. 3.20. **Temporal relationship between the ADC and DC-potential responses**

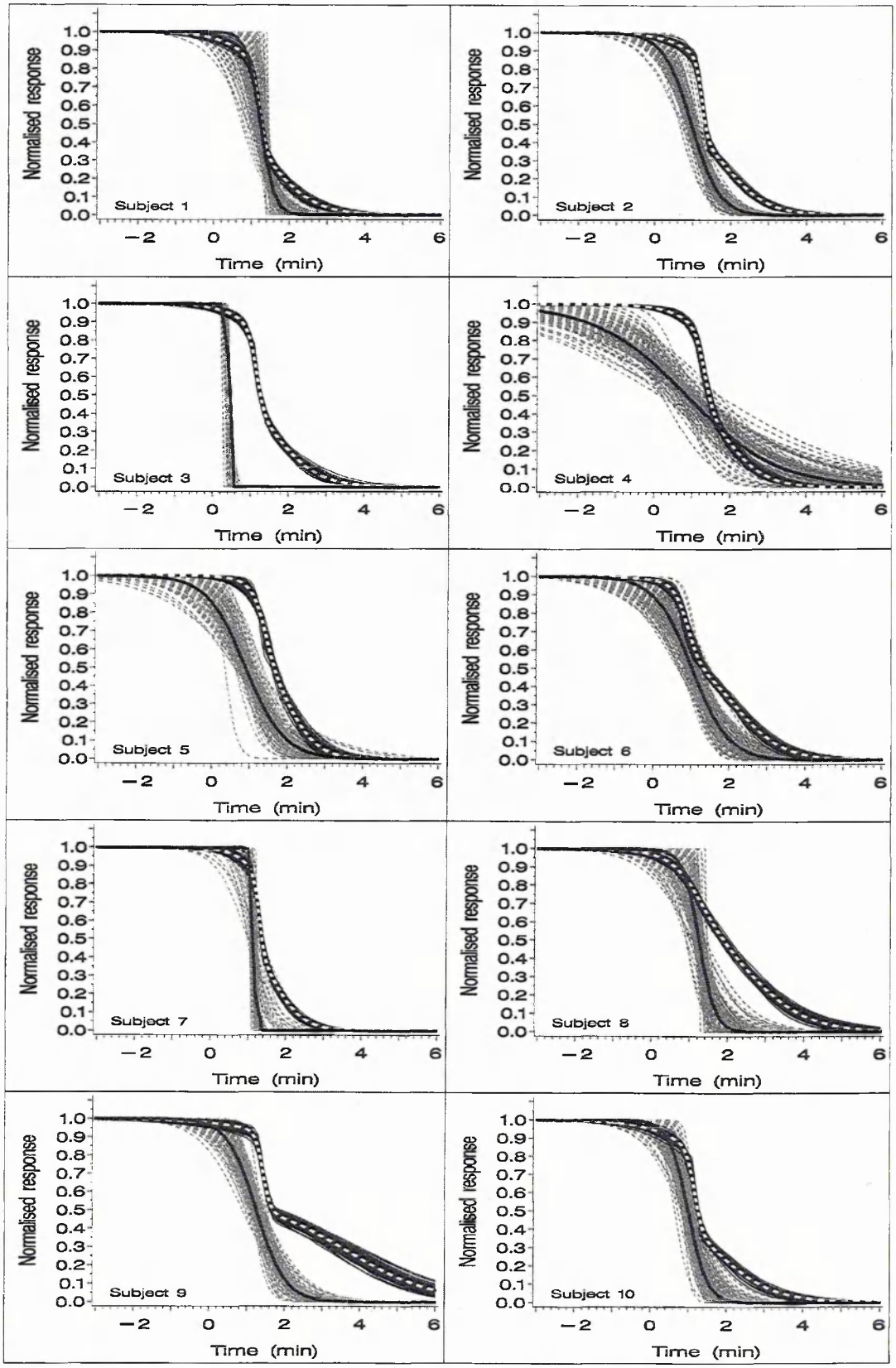


Figure legend on the following page

Fig. 3.20. continued

The normalised ADC (black continuous line) and DC-potential (white broken line) individual-specific posterior median response profiles obtained for each of the ten animals are shown, together with a set of 100 random time-course profiles generated by sampling from the posterior distribution (ADC profiles, grey broken lines; DC-potential profiles, black band surrounding the white broken line).

Fig. 3.21. The DC-potential data (+++) acquired from each of the ten animals are shown, together with the individual-specific posterior median (black line) and a set of 100 random time-course profiles generated by sampling from the posterior distribution (surrounding grey band).

Fig. 3.22. The ADC data (black irregular line) acquired from each of the ten animals are shown, together with the individual-specific posterior median (smooth black curve) and a set of 100 random time-course profiles generated by sampling from the posterior distribution (grey broken lines).

Fig. 3.21. **Individual specific DC-potential response profiles**

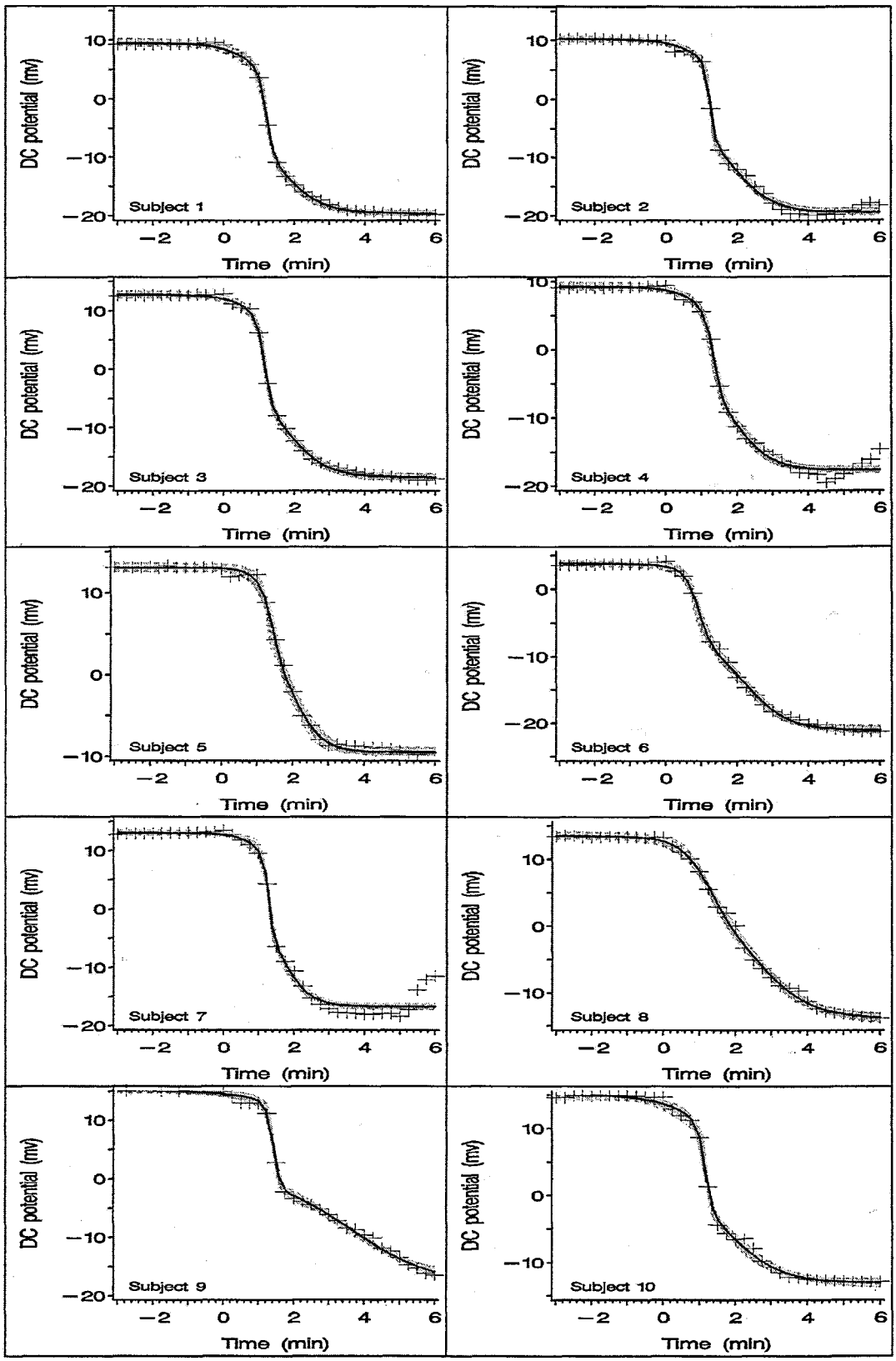


Figure legend on page 117.

Fig. 3.22. **Individual-specific ADC response profiles**

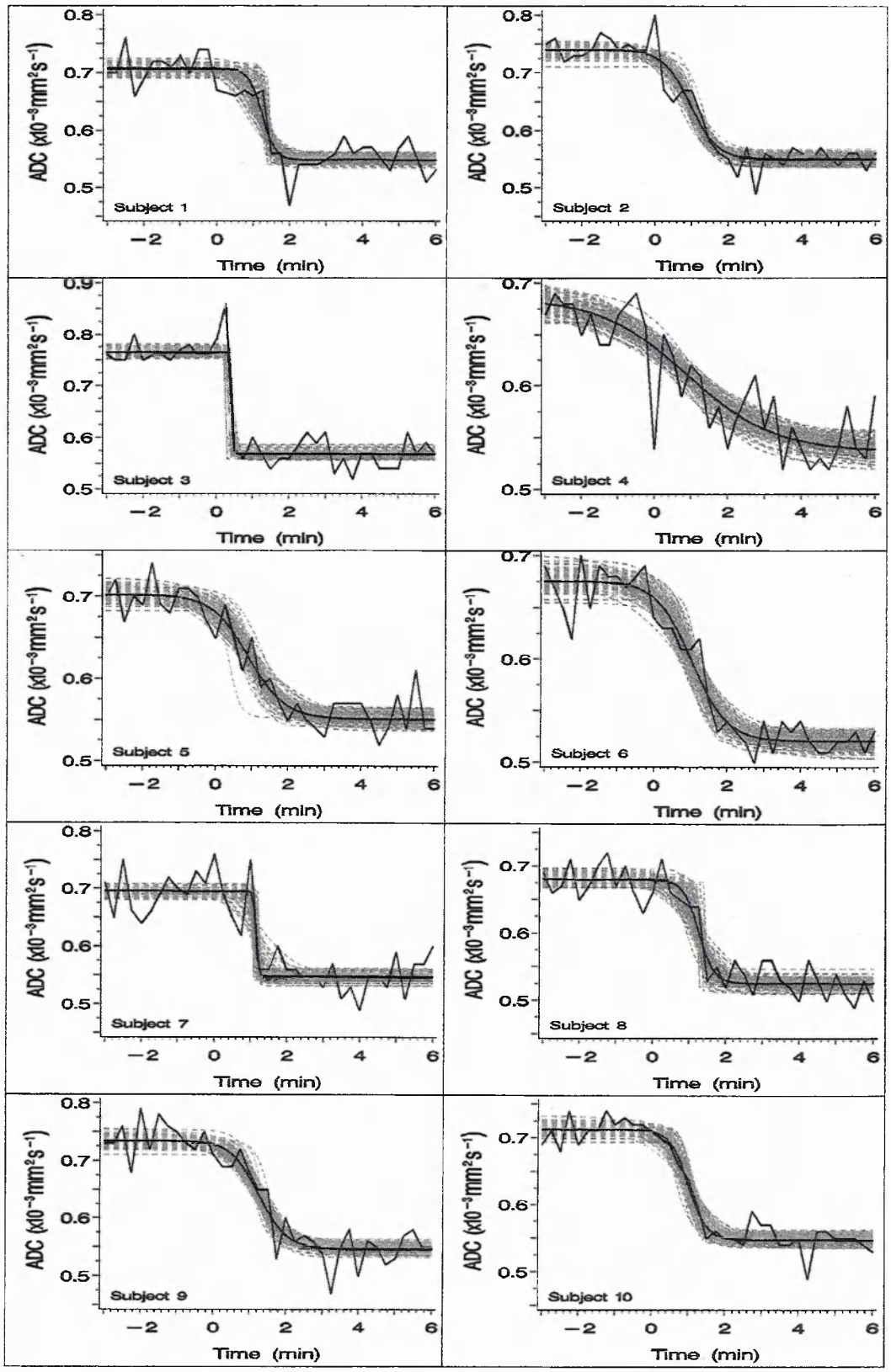


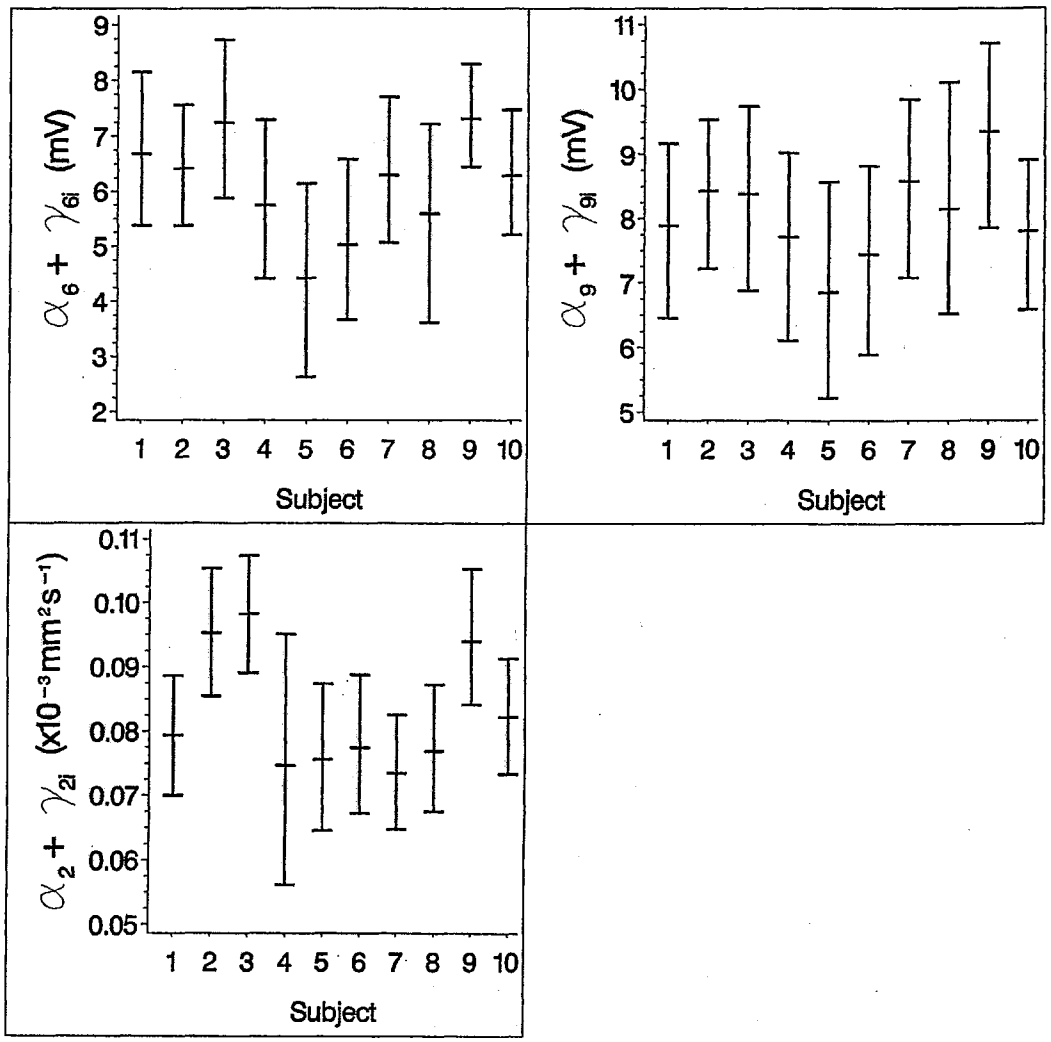
Figure legend on page 117.

profiles, together with a cluster of profiles obtained by sampling the posterior distribution. Fig. 3.21 shows the individual-specific median and sampled posterior DC-potential response profiles with the observed data superimposed, while Fig. 3.22 shows the median and sampled posterior ADC response profiles with the ADC observations superimposed.

Visual inspection of these figures indicates that the two-term logistic function provides a reasonable fit to the observed DC-potential responses (Fig. 3.21), while a single logistic term provides an adequate fit to the ADC data (Fig. 3.22). The DC-potential drift that occurred towards the end of the 6min post-occlusion period in a few cases is discussed in Section 3.4. In addition, the median and sampled response profiles obtained for animal 4 indicate a marked decline in ADC prior to the ischaemic insult (Figs. 3.20 and 3.22). This observation is discussed in detail in Section 4.3.3. The need to include two logistic terms in the model for the DC-potential response is demonstrated by the weighting coefficient posterior distributions. The population posterior medians (95% posterior interval) are: DC_{fast} (given by α_6), 6.12 (5.12, 7.20); DC_{slow} (given by α_9), 8.04 (7.04, 9.06). Furthermore, the individual-specific weighting-coefficient posterior intervals associated with each of the two DC-potential components ($\alpha_6+\gamma_{6i}$ and $\alpha_{9+}\gamma_{9i}$, $i = 1, 2,...,10$) all exclude zero (Fig. 3.23).

Among the striking features of these figures is the considerable between-animal variation in the rate of the ADC transition (Figs. 3.20 and 3.22), between-animal differences in the shape of the DC-potential response (Figs. 3.20 and 3.21), and an absence of any obvious and consistent relationship between the DC-potential and ADC transitions within each individual (Fig. 3.20). To expand on the latter observation it is informative to focus, for example, on the responses in animals 2 and 3. The immediate visual impression provided by Figs. 3.21 and 3.22 is a marked between-animal difference in ADC transition rate (Fig. 3.22, panels 2 and 3) despite the relative similarity of the DC-potential response profiles (Fig. 3.21, panels 2 and 3). In both animals the ADC transition tends to precede the DC-potential response (Figs. 3.2 and 3.20).

Fig. 3.23. **Individual-specific ADC and DC-potential scaling coefficients**



The plot shows the individual-specific scaling coefficient medians and 95% posterior intervals obtained for each of the two DC-potential components (DC_{fast}, $\alpha_6+\gamma_{6i}$ and DC_{slow}, $\alpha_{9+}\gamma_{9i}$, $i = 1, 2,...,10$). For completeness the individual-specific ADC transition scaling coefficients ($\alpha_{2+}\gamma_{2i}$, $i = 1, 2,...,10$) are included.

A visual impression of the relationship between the two components of the DC-potential response and the ADC response is given in Figs. 3.24 and 3.25. Fig. 3.24 focuses on the fast DC-potential component while Fig. 3.25 focuses on the slow DC-potential responses. Thus Fig. 3.24 shows each of the individual-specific fast DC-potential posterior median response profiles with the corresponding posterior median ADC response superimposed, while Fig. 3.25 shows the individual-specific posterior median slow DC-potential profiles, again with the ADC response superimposed. Neither of the two DC-potential components exhibits a consistent temporal relationship with the ADC response. Fig. 3.26 shows each of the ten individual pair of slow and

---

Fig. 3.24. Overlaid plots showing the normalised ADC (black line) and fast DC-potential component (white broken line within the black band) individual-specific posterior median responses obtained for each of the ten animals. Also shown is the set of 100 random time course profiles drawn from the posterior distribution (ADC profile set, grey broken lines; DC-potential profile set, black band).

Fig. 3.25. Overlaid plots showing the normalised ADC (black line) and slow DC-potential component (white broken line within the black band) individual-specific posterior median responses obtained for each of the ten animals. Also shown is the set of 100 random time course profiles drawn from the posterior distribution (ADC profile set, grey broken lines; DC-potential profile set, black band).

Fig. 3.26. Overlaid plots showing the fast and slow DC-potential component normalised posterior median profiles obtained for each of the ten animals (fast DC-potential posterior median, black continuous line; slow DC-potential posterior median, white broken line). Also shown is the set of 100 random time course profiles drawn from the posterior distribution (fast DC-potential sampled posterior profile set, broken grey lines; slow DC-potential sampled posterior profile set, continuous black lines).

Fig. 3.24. **Temporal relationship between the ADC and fast DC-potential responses**

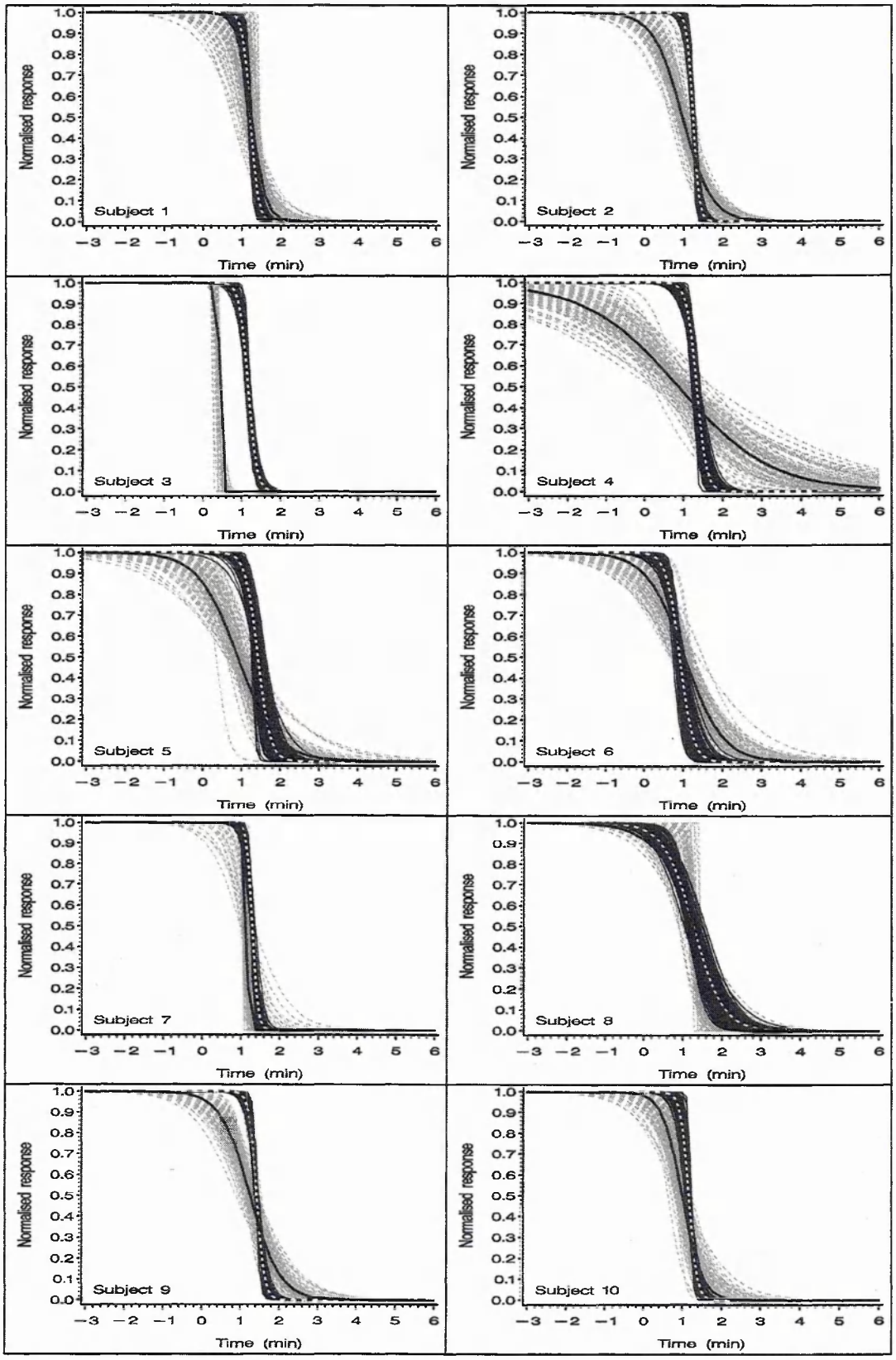


Figure legend on the preceding page.

Fig. 3.25. **Temporal relationship between the ADC and slow DC-potential responses**

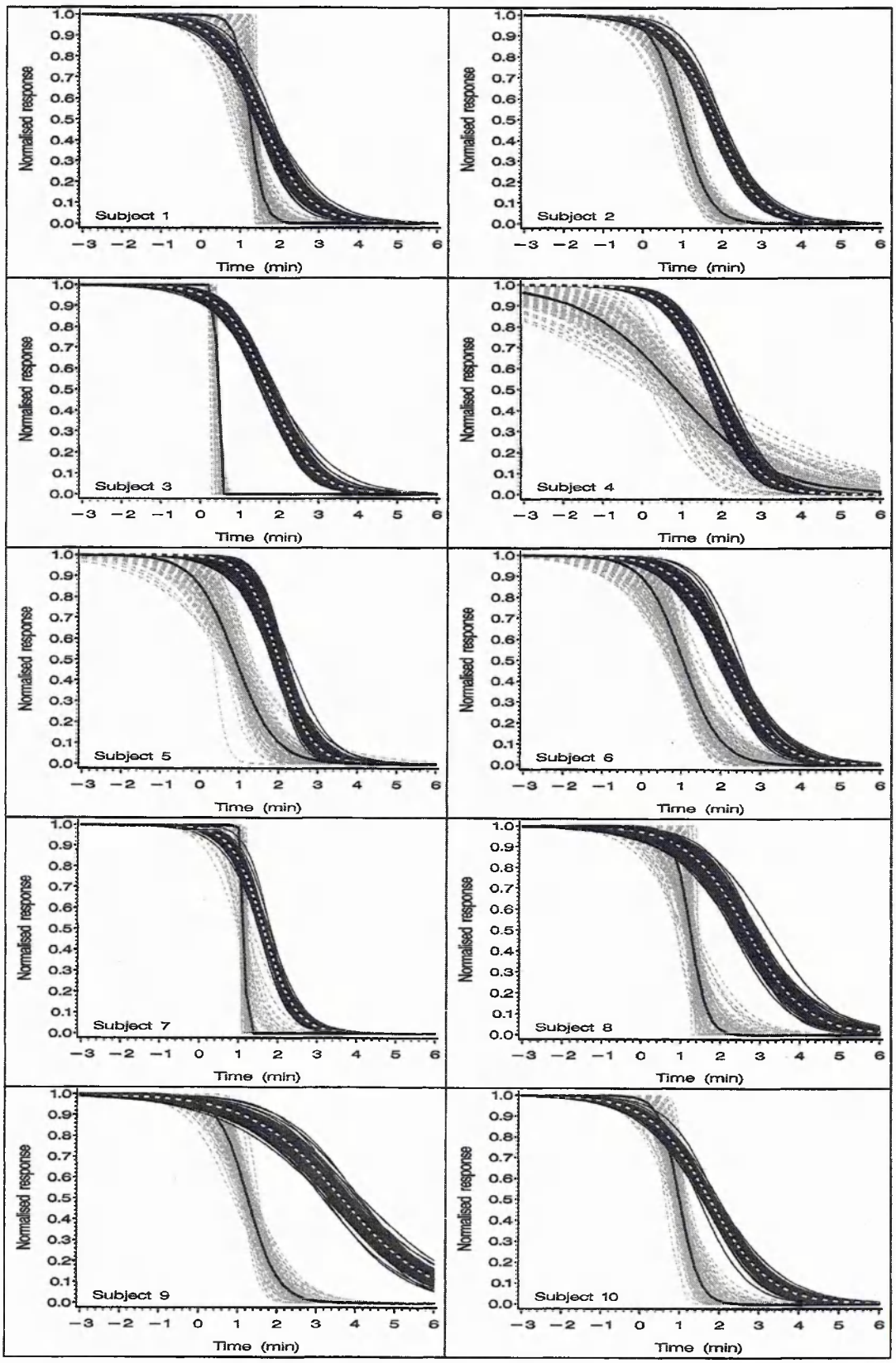


Figure legend on page 122.

Fig. 3.26. **Temporal relationship between the fast and slow DC-potential responses**

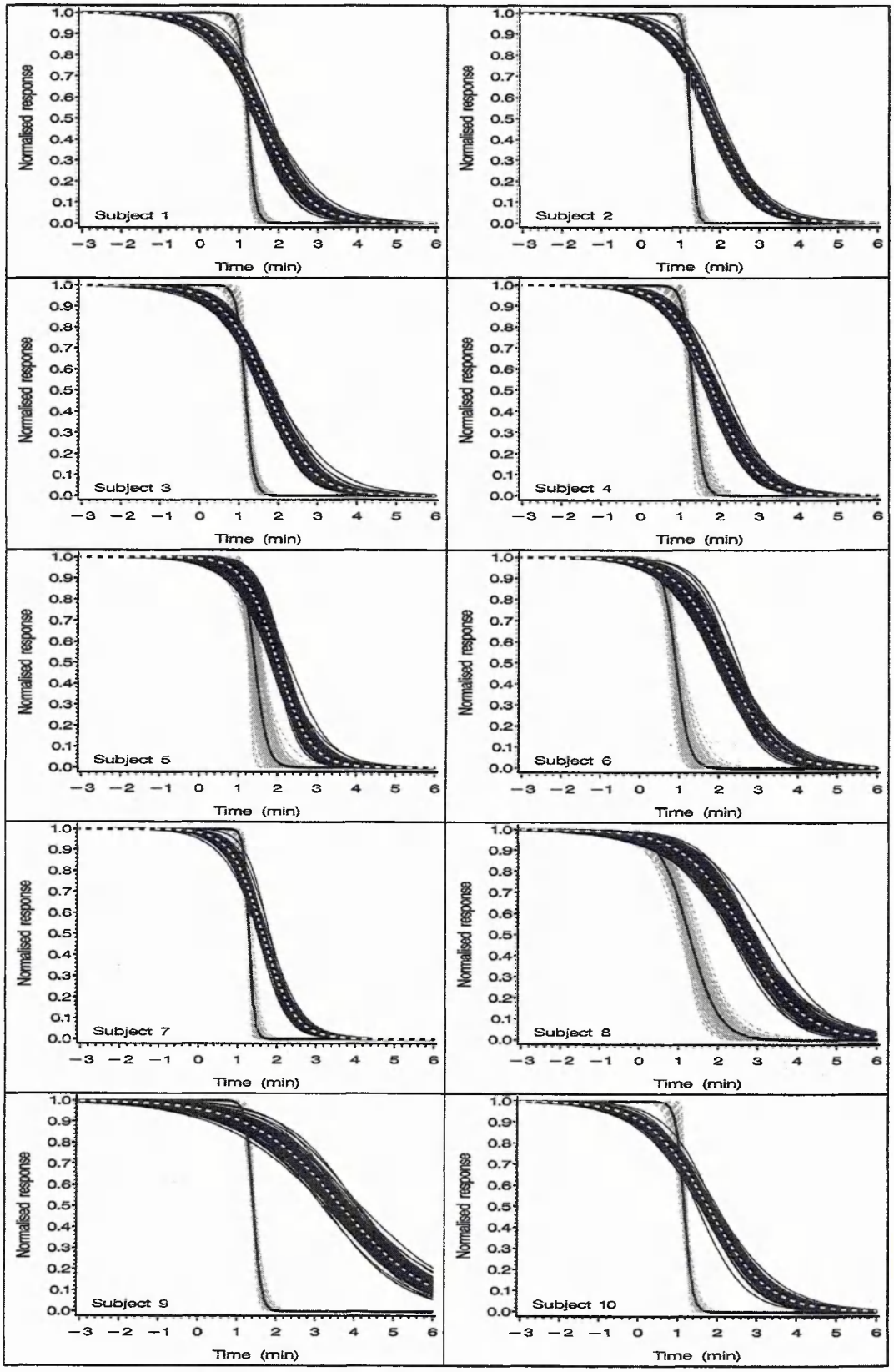


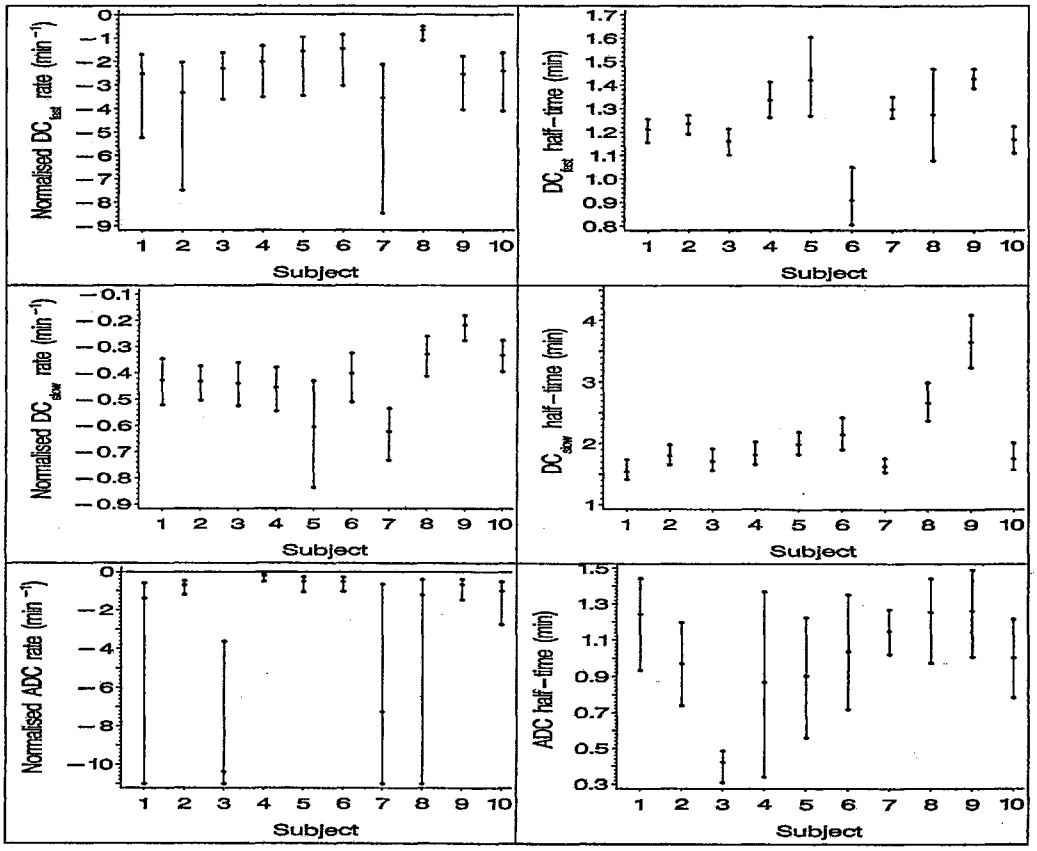
Figure legend on page 122.

fast DC-potential posterior median response profiles. Again, these plots reveal a substantial degree of between animal variation in response characteristics.

The ten individual DC_{fast}, DC_{slow} and ADC transition rate medians are displayed in Fig. 3.27 together with the 2.5% and 97.5% quantiles. Fig. 3.27 also displays the individual transition half-response time medians and associated quantiles. All six parameter medians (transition rate and half-response time for each of the three responses i.e., DC_{fast}, DC_{slow} and ADC responses) exhibit a considerable degree of between-subject variability. Four animals yield ADC transition-rate distributions that are heavily skewed, indicating that in these individuals the ADC transition sampling is insufficient to rule out a near step change in ADC. The ten individual ADC-DC_{fast} transition rate median differences and quantiles are shown in Fig. 3.28, together with the transition half-response time median differences and quantiles. Six of the ten individual ADC-DC_{fast} rate difference 95% posterior intervals exclude zero (Fig. 3.28, panel (a)), indicating a significant difference in rates in these subjects. Five of these six median differences are positive and one is negative. Four of the difference distributions are heavily skewed, this being a reflection of skewness in the corresponding ADC transition-rate distributions mentioned above. Four of the ten posterior intervals for the ADC-DC_{fast} difference in half-response time lie below zero (Fig. 3.28, panel (b)) indicating that the ADC half-response time occurs prior to the DC_{fast} half-response point in these individuals, the difference achieving statistical significance.

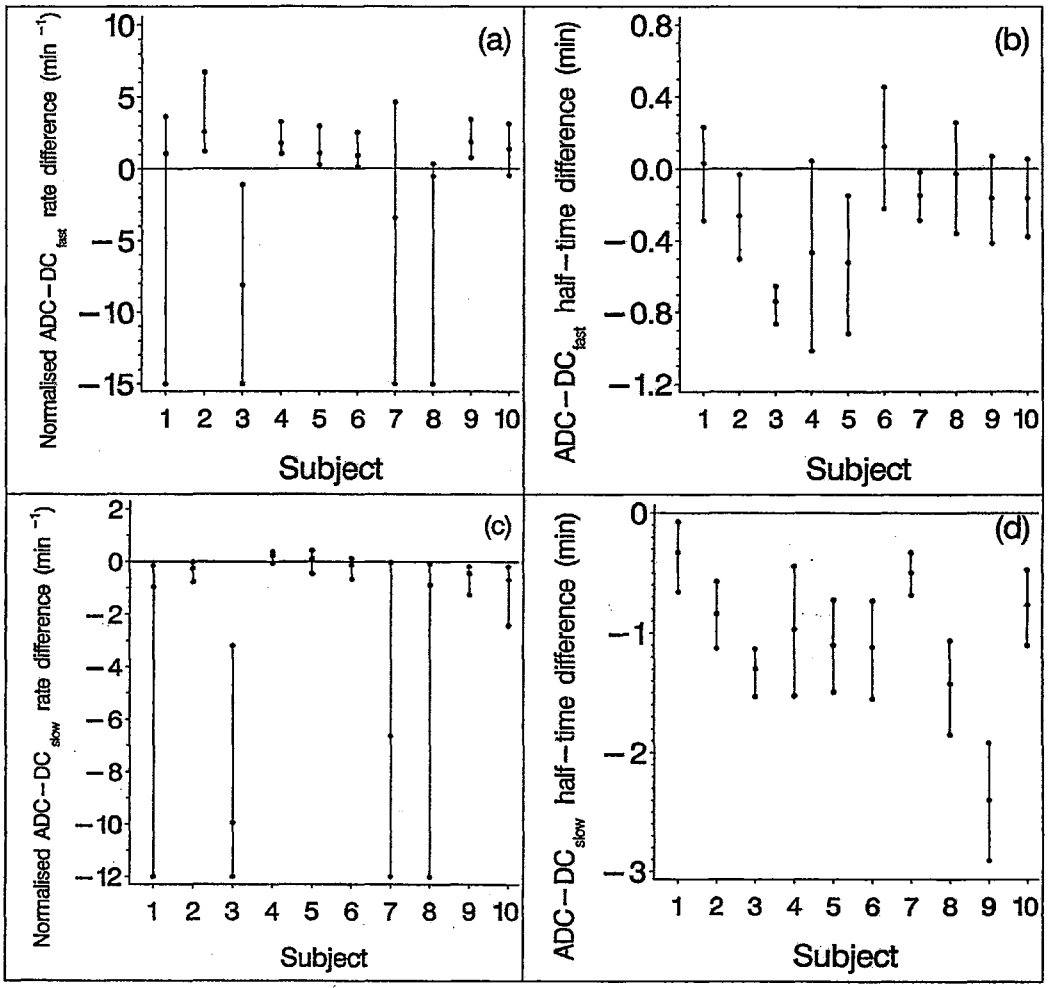
Figure 3.28 also shows the ADC-DC_{slow} transition rate and half-response time median differences and quantiles. Seven of the ten intervals for the difference in rate lie below zero (Fig. 3.28, panel (c)), indicating a significant difference in these individuals, while one interval is close to excluding zero with a positive median. All ten ADC-DC_{slow} half-response time difference intervals lie below zero, indicating that the ADC transition half-response time is consistently early relative to the DC_{slow} half-response time. Finally, animals 1 and 8 differ from the remaining animals; in these two animals the ADC and fast DC-potential component are similar in terms of both their transition rate and half-response time medians (Figs. 3.28, panels (a) and (b)). In addition to these

Fig. 3.27. **ADC and DC-potential normalised transition rates and half-response times**



Plots showing the ADC and DC-potential normalised transition rate and half-response time posterior medians and 95% intervals generated for each of the ten animals. The top two panels provide the fast DC-potential component results, while the middle two panels show the slow DC-potential component results. The lower limit of the ADC transition rate posterior interval for animals 1, 3, 7 and 8 is large and negative, indicating a near step change in ADC. For display purposes these have been truncated at –11.

Fig. 3.28. **Transition rate and half-response time difference between the ADC and DC-potential responses**



Plots showing the ADC-V_{DC} transition parameter difference medians and 95% posterior intervals generated for each of the ten individual animals. These data were obtained from the posterior sample by performing subtractions using the relevant normalised transition rate or half-response time parameters. The lower limit of the ADC-$V_{DCfast/slow}$ rate difference posterior intervals obtained for animals 1, 3, 7 and 8 is large and negative, reflecting a near step change in ADC (see Fig. 3.27). For display purposes these intervals have been truncated at −15 and −12 in (a) and (c), respectively.

two animals, animal 10 also yields transition rate and half-response time difference 95% intervals both of which span zero (Fig. 3.28, panels (a) and (b)). Animals 4 and 5 differ from the remaining animals with respect to their slow DC-potential transition rate medians, which are fast relative to the ADC transition. Furthermore, the 95% interval for animal 4 is close to excluding zero (Fig. 3.28, panel (c)).

3.4 Discussion

Despite the established role of diffusion imaging in stroke research and diagnosis, many aspects of the ADC response to an ischaemic insult are poorly understood. It is generally believed that anoxia-induced brain damage is the result of a cascade of events (Spetzler and Nehls, 1987; Siesjö, 1992) but the relationship between these events and the ADC response remains to be determined. This chapter focuses on the temporal relationship between the ADC and DC-potential responses to transient focal ischaemia. Until recently, a rigorous analysis of this type of longitudinal data was barely feasible. Although useful statistical models were available by the early 1980s (Laird and Ware, 1982), it is only with the more recent availability of substantial computing power that a proper modelling of longitudinal data has become routine. This study uses a Markov chain Monte Carlo simulation implementation of a nonlinear, full random coefficients model of the ADC and DC-potential responses to an ischaemic insult. Random coefficients are required to model adequately various features of the between-animal variation in the DC-potential and ADC responses. For example, the greater between-animal variance in baseline ADC, compared with the between-animal variance in the asymptotic post-occlusion ADC level is a striking feature of the ADC profiles (Fig. 3.1). This behaviour is captured by the full random coefficients model. Among the other advantages of the random coefficients modelling approach is its applicability to sparse data. This is particularly important in the present application because the ADC transition is undersampled and it is difficult to obtain precise estimates of the transition parameters by fitting individual data. Undersampling was unavoidable because signal-to-noise ratio considerations placed an upper limit of the order of $4min^{-1}$ on the image acquisition rate. Fortunately, a formal random coefficients analysis provides improved individual parameter

estimates (random coefficient estimates). This improvement is achievable because each individual contributes information to all the remaining individual-specific random coefficient estimates. This information borrowing behaviour is a central feature of Bayesian hierarchical models; it is examined in Appendix B. The MCMC simulation random coefficients modelling technique offers several additional advantages over the more traditional mixed-model/random-coefficients regression methods, including an elimination of the need for large-sample approximation. This applies to both linear and nonlinear applications. These issues are discussed in Chapter 6 in a little more detail.

The main purpose of the analysis outlined in this chapter was to determine whether the ADC and DC-potential changes might be driven by some common latent process. If some common process is directly responsible for the DC-potential and ADC responses (i.e., if the ADC and DC-potential are immediately and tightly coupled to some underlying driving process, with a linear dependence), the normalised individual-subject ADC and DC-potential transition parameters will be indistinguishable. It might be argued that any attempt to relate the temporal dependence of the ADC and DC-potential response is mistaken because the DC-potential recording electrode is sensitive to a highly localised region of tissue, while diffusion imaging is relatively low in resolution. (The present ADC data were acquired from a $1.5mm^2$ region within a 2.1mm slice, giving an acquisition volume of approximately $3mm^3$.) A high degree of spatial variation in DC-potential or ADC might invalidate the comparison. Nevertheless, several studies of this type have been performed. These have, however, led to disagreement regarding the temporal relationship between the two responses (as discussed by de Crespigny et al. (1999); see below for additional information). The present study attempts to resolve some aspects of the problem by adopting a formal statistical analysis. It is important to stress that a failure to demonstrate that any of the population mean ADC-V_{DC} transition parameter differences is non-zero (i.e., a failure to reject any of the null hypotheses) cannot be used as evidence for a common latent process since this might arise due to between-subject variability in transition characteristics. If a common latent process (as defined above) is involved, then the within-subject ADC and DC-potential transition

characteristics must be indistinguishable, irrespective of the between-subject variability in transition behaviour. Consistent with previous studies (see de Crespigny et al. (1999) Fig. 5, for example), the DC-potential response was found to consist of two components. Accordingly, it is necessary to examine the temporal relationship between the ADC response and each of the two DC-potential components.

Before entering into a detailed examination of the results obtained from this V_{DC}-ADC data re-analysis, a number of issues must be addressed. The first of these relates to the DC-potential drift that occurred, in a few cases, towards the end of the 6min post-occlusion period (Fig. 3.21, animal 7, for example). An inspection of Fig. 3.21 might lead to the conclusion that autoregressive (AR) error structure should have been incorporated into the model. An AR model was not adopted, however, for several reasons. Firstly, in the majority of cases an independent errors model was shown to be adequate, relative to an AR(1) model, when compared using AIC. Secondly, it was decided that the departure between the posterior median and observed DC-potential profiles that occurred in a few cases should not be modelled as a stochastic process because DC-potential drift is deterministic rather than stochastic. Any attempt to model DC-potential drift would be problematic, however, since it occurred in a few individuals and its occurrence was erratic. It is unfortunate that interference pickup is a common problem in this type of biological monitoring involving high impedance devices (Horowitz and Hill, 1980, p612). Stray capacitance is a well-known cause of interference coupling in this context. Only a few experiments were affected by DC-potential drift, however, and, when it did occur, it did not interfere with the transition. Time-course truncation could have been used as an approach to dealing with the problem, but this was not adopted in the interests of objectivity.

The second issue relates to the precision of the 2.5% and 97.5% quantile estimates given by the present simulation analysis. A case might be made for increasing the MCMC output chain length in order to achieve better accuracy. It was decided, however, that the limiting feature of this study is not the analysis, but the presence of noise spikes in several of the ADC transitions (Fig. 3.2).

The offending noise spikes were introduced when the radiofrequency shield was opened in order to perform the middle cerebral artery occlusion. An obvious solution to the noise spike problem is remote arterial occlusion, although a complete redesign of the radiofrequency probe and associated animal holder would be required to allow arterial occlusion without compromising the radiofrequency shield. It is this aspect of the study that requires attention, and improvements in this area are more pressing than refinements to the analysis.

A final matter that must be addressed is the marked decline in the posterior median ADC that occurs prior to the ischaemic insult in the fourth animal. The noise spike at zero time contributes to the problem, but there are other instances of similar, although less extreme behaviour. V_{DC}-ADC time-base synchronisation is discussed at length in Chapter 4, because it has important implications for the trigger threshold analysis that is the subject of that chapter. In contrast, the main conclusions of this chapter are based mainly on the transition rate data, and these are immune to time-base inaccuracy. Accordingly, no additional discussion of the issue is included here, except to mention the difficulty that arises because the ADC data were not available in real-time. In an ideal experiment a good quality baseline would be established in the interval immediately prior to occlusion. But ADC data generation requires a 2D Fourier transformation of the MR signal followed by ROI identification and post-processing. Given the 15s time resolution and the hardware used in this study, a real-time generation of the ADC data was not possible. (Simultaneous signal acquisition and processing was not feasible on a computer fitted with a single DSP (digital signal processing) board.)

Putting aside the occasional DC-potential drift that is discussed in the preceding paragraph, an adequate modelling of the DC-potential data is obtained using a sum of two logistic terms, while a single logistic term is sufficient with respect to the ADC response. It should be noted, however, that a poor signal-to-noise ratio causes a reduction in power to detect model inadequacy in the latter case. The observation that a sum of logistic terms provides a convincing description of the DC-potential and ADC responses is particularly interesting given the classical logistic form of the

relationship between binding-site saturation and log(concentration), as expressed by the Langmuir adsorption isotherm (i.e., the Michaelis-Menton equation; Colquhoun, 1971, Sections 12.8 and 14.6). An underlying binding process may be implicated, although this does assume an exponential ligand concentration dependence on time.

The results obtained from this MCMC analysis rule out the possibility that either of the two DC-potential components and the ADC response to an ischaemic insult are driven directly by some common latent process. This assertion is based on the following observations: six of the ten animals yielded distinguishable ADC and fast DC-potential transition rates; seven animals yield distinguishable ADC and slow DC-potential transition rates. The transition rate parameters have been made the focus of the analysis because it might be argued that differences in transition time could be an artefact of poor synchronisation between the MR and DC-potential recordings. Nevertheless the transition time statistics are consistent with the assertion that the ADC and DC-potential responses are driven independently. In four of the ten animals, the ADC half-response time preceded the fast DC-potential half-response time with a difference that achieves significance. All ten animals reached significance with respect to the difference between the ADC and slow DC-potential transition times. At this stage it is important to consider the possibility that these differences might arise as an artefact due to 'shrinkage' in individual parameter estimates towards the population estimate. Shrinkage towards the population mean is a characteristic of Bayesian hierarchical models (see, for example, Gelman et al., 1995, p133) and this is important in the present context due to the differing precision of the ADC and DC-potential measurements, and the differing transition rates. The ADC data were acquired with a lower sampling rate and lower precision. The ADC transition rate estimates are, therefore, more susceptible to shrinkage, especially when the transition is undersampled due to its rapidity. One can imagine a situation in which a within-animal pair of ADC/V_{DC} transitions with near identical rates appear to have differing rates due to shrinkage in the estimated ADC rate, while the DC-potential estimate remains close to its sample value. An inspection of the data indicates that estimation shrinkage does indeed take place (see Appendix B) but, with one exception, this occurs only when the ADC

transition is fast relative to both the DC-potential transition and the population median ADC rate. Thus, when it occurs, ADC parameter shrinkage reduces the estimated ADC-V_{DC} transition rate difference rather than creating a difference in rates as an artefact. The one exception, referred to above, is the ADC-DC_{fast} rate difference in Animal 1. In this instance the 95% posterior interval includes zero, and so shrinkage is not a concern.

To summarise, the main finding in this study is that the observed ADC and DC-potential changes are not driven by a common latent process. Furthermore, the statement (Harris et al., 2000) that the mean ADC half-response time is shorter than the mean DC-potential half response time is consistent with the population median half-response time statistics given in Table 3.2; thus this assertion is not altered when the biphasic nature of the DC-potential response is taken into consideration. The fundamental question is how the failure to find evidence for a common latent process relates to the known pathophysiology of the anoxic/ischaemic response. Although the electrical response to anoxia has been studied extensively, the biophysical basis of so-called anoxic depolarisation is not completely established; rather more is known about the related phenomenon of spreading depression. This literature is reviewed briefly in the following section.

3.4.1 The temporal relationship between the ADC and DC-potential responses to cerebral ischaemia/anoxia and spreading depression. The ADC and cell swelling

Many aspects of the biophysical responses to cerebral ischaemia/anoxia resemble those observed in spreading depression and, since there is a considerable literature on the latter, it is useful to draw on this information. Contradictory results appear to have been obtained concerning the temporal relationship between the ADC and DC-potential responses to spreading depression and cerebral ischaemia/anoxia. Thus, while it has been reported that the transient ADC response to spreading depression and ischaemia/anoxia appears simultaneously with the DC-potential shift (Gyngell et al., 1994; de Crespigny et al., 1998), others report that the ADC response to cerebral ischaemia occurs before anoxic depolarisation (Huang et al., 1997; de Crespigny et al., 1999;

Harris et al., 2000). de Crespigny et al. (1999) point out, however, that despite the earlier onset of the ADC response to ischaemia, a larger and more rapid ADC change occurs during depolarisation. It should be noted that technical limitations typically restricted the temporal resolution of early DW imaging work (scan times of 10min or more) with the result that any existing V_{DC}-ADC response mismatches may have been missed.

Reference has been made in Section 3.1.3 to early studies into the aetiology of stroke-induced ADC changes (Busza et al., 1992; Norris et al., 1994; Latour et al., 1994b; Szafer et al., 1995; Hoehn-Berlage et al., 1995; Anderson et al., 1996). Several researchers have suggested that cell swelling (cytotoxic oedema) is responsible for the observed reduction in ADC (Norris et al., 1994; Latour et al., 1994a, 1994b; Anderson et al., 1996). van der Toorn et al. (1996) state that there is a close correlation between the ADC time course and both the tortuosity and extracellular:intracellular volume fraction changes that occur in global ischaemia. They suggest, therefore, that these processes are mechanistically linked. This conclusion is, however, based on sample-average behaviour. A formal analysis of the within-subject temporal relationship between responses provides additional information, and this has been the focus of this chapter. Nevertheless, it is interesting to note that van der Toorn et al. (1996) report a difference between white and grey matter with respect to the latency of the ADC transition, a difference which is reflected in the tortuosity and volume fraction data. They suggest that cell swelling is the common link and that this is triggered by a depletion of phosphocreatine. In this context it is interesting to note that cell swelling has also been suggested as a direct cause of anoxic depolarisation (Balestrino, 1995).

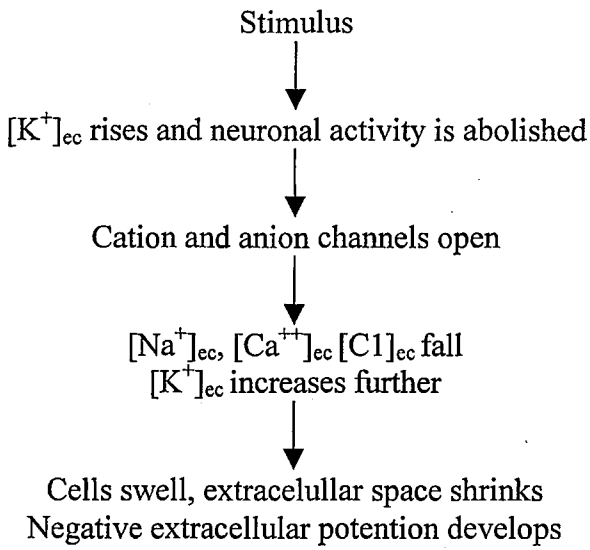
Several researchers have studied the electrophysiology of anoxia and reported a three-phase response (see Hansen (1985) and Lipton (1999) for reviews on the biophysical chemistry of ischaemia/hypoxia). A key feature of the second of these three phases is a rapid increase in extracellular potassium (K^+_{ec}) and an associated fast negative shift in extracellular DC-potential (anoxic depolarisation). These events are accompanied by a simultaneous movement of NaCl from

the extracellular space into the intracellular compartment and an increase in both extracellular impedance and extracellular marker concentration. The impedance and marker concentration changes are both attributed to a reduction in extracellular volume. It is well known that cell volume is maintained by osmotic pressure resulting from impermeable intracellular anions counterbalanced by extracellular Na^+. Thus, the redistribution of Na^+ that occurs during anoxia causes an osmotically-driven movement of water from the extracellular space into the intracellular compartment. Cell swelling and the observed reduction in extracellular volume are a manifestation of this process. Osmotically active products of anaerobic metabolism, including lactate, contribute to this process. Hansen (1978, 1985) states that passive diffusion in the absence of pump activity is insufficient to account for the observed increase in K^+_{ec} during anoxic depolarisation and that the ion movements that occur during this phase of the anoxic response are due to a sudden increase in overall cell membrane ion permeability. He suggests that this permeability change is caused by K^+_{ec}-induced cell membrane depolarisation, arising from the initial elevation in K^+_{ec} that occurs during the first phase of the anoxic response. Similar DC-potential, K^+_{ec}, impedance and extracellular marker responses are observed in spreading depression (see, for example, Collewijn and van Harreveld, 1966; Hansen and Olson, 1980; Nicholson and Kraig, 1981; Hansen, 1985). In fact, as stated previously, the biophysical chemistry of spreading depression is well established (the key events are summarised in Fig. 3.29[1]) and this has enabled Nicholson and Kraig (1981) to arrive at a working hypothesis for the sequence of events that underlie the process. A schematic of this hypothesis is outlined in Fig. 3.30[1]. It provides an explanation for the present results. A key observation is that Na^+_{ec} and Cl^-_{ec} decrease in spreading depression. Furthermore, both spreading depression and the associated $NaCl_{ec}$ changes persist in the presence of tetrodotoxin, showing that the increase in sodium permeability does not involve the conventional voltage-dependent channel. Thus action potentials are not an essential component, as had been assumed previously. (Tetrodotoxin is a puffer fish toxin that blocks the inward sodium current of the action potential.) Nicholson and Kraig suggested that a chemically-mediated channel is involved, and that the

---

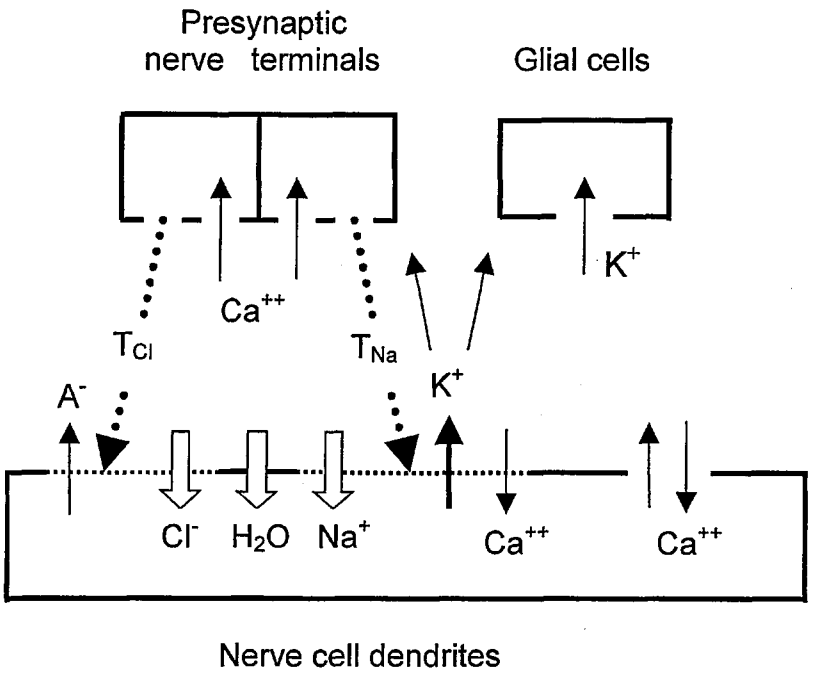
[1] Figs. 3.29 and 3.30 are based on Nicholson and Kraig (1981) Figs. 8 and 9, respectively.

Fig. 3.29. **Key events in spreading depression**



A schematic summary of the major events that occur in spreading depression (based on Nicholson and Kraig, 1981, Fig. 8). The square brackets indicate concentration of the species within the bracket and the subscript *ec* indicates extracellular.

Fig. 3.30. **The Nicholson-Kraig hypothesis**



A schematic summary of the Nicholson-Kraig (1981) hypothesis for the events that occur in spreading depression. An initial increase in extracellular K^+ causes presynaptic depolarisation and an associated Ca^{++} channel opening. The resulting influx of Ca^{++} leads to excitatory and inhibitory transmitter (T_{Na} and T_{Cl}) release. The excitatory transmitter causes dendritic channels to open allowing Na^+ and Ca^{++} to enter, which causes subsequent cell depolarisation. This, in turn, causes an additional release of K^+ into the extracellular space. Cl^- leaves the extracellular space, driven by its electrochemical gradient. Transmitter release continues which, in turn, causes NaCl to enter the dendrites as a neutral flux. This massive entry of NaCl is accompanied by the entry of water, either through separate water pores or via ion channels. The accumulation of K^+ in the extracellular space causes additional presynaptic depolarisation. It also causes glial cell depolarisation, with the result that glial cells act as a K^+ sink. Nicholson and Kraig suggest that this is the main cause of the observed negative shift in extracellular DC-potential. A^- is an unidentified anion. (A modification of Fig. 9 in Nicholson and Kraig, 1981.)

movement of water that accompanies the redistribution of NaCl occurs either via separate aqueous pores or via ion channels. This movement of water is assumed to be associated directly with the observed decrease in ADC. Nicholson-Kraig further suggest that an accumulation of K^+_{ec} produces glial cell depolarisation and that this is the cause of the negative extracellular DC-potential shift. The assertion that glial cell depolarisation and not neuronal depolarisation is the essential component of spreading depression had been suggested previously on the basis of experiments in which neuronal and glial cell intracellular DC-potential recordings were made in the presence and absence of tetradotoxin (Sugaya et al., 1975). A close temporal relationship between glial depolarisation and spreading depression was observed, while the neuronal intracellular DC-potential failed to show a consistent temporal relationship. Thus the Nicholson-Kraig interpretation of the sequence of events in spreading depression provides a plausible explanation for the temporal separation of the ADC and DC-potential responses observed in the present study. Furthermore, the notion that the volume and DC-potential components of spreading depression are separate processes has been demonstrated by experiments in which NaCl was replaced by sucrose. The negative potential associated with spreading depression was maintained, while the light-scattering and impedance changes associated with cell swelling were abolished (see Nicholson and Kraig, 1981, p221 for references). Although a simple picture of the sequence of events occurring during anoxic depolarisation appears to emerge, a number of experimental observations complicate the issue. Thus, although anoxic depolarisation might be attributed to $(Na^+\text{-}K^+)$-ATPase pump failure, tissue ATP concentrations are relatively unaffected in spreading depression[1] (Hansen, 1985, p126), despite the similarity of both the ion-concentration and DC-potential responses to spreading depression and ischaemia/anoxia. An additional complication emerges from the demonstration (Obrenovitch et al., 1990) that ischaemia-induced depolarisation occurs in isolated perfused brain, despite the maintenance of ATP at a significant level (one third of the control concentration). Product inhibition of the $(Na^+\text{-}K^+)$-ATPase by ADP (Apell et al., 1986) might underlie this finding, in which case a partial depletion of ATP suffices to cause

---

[1] A marked depletion of ATP occurs in ischaemia/anoxia (Hansen, 1985, p122 et seq.).

transmembrane ion-gradient disruption. To summarise, it is established that ischaemia/anoxia can lead to massive ion shifts and an associated decrease in extracellular DC-potential (anoxic depolarisation; Dijkhuizen et al., 1999, and references therein). Furthermore, it has been demonstrated that in spreading depression the changes in extracellular volume attributable to cell swelling occur in synchronisation with changes in extracellular DC-potential (Hansen and Olsen, 1980; Hansen, 1985, p119). Nevertheless the present diffusion study indicates that the ADC and DC-potential changes associated with anoxia are not driven by a common process.

3.4.2 The ADC and DC-potential responses to focal ischaemia. A summary of the latent process analysis

In this chapter the ADC and DC-potential data obtained in a previously published study has been re-examined, with a view to addressing the common latent process issue. The analysis focuses on the ADC and DC-potential transition characteristics. A simple comparison of the normalised transition rates and half-response times is valid provided the biological system in question behaves as an approximately linear system, i.e., provided the ADC and DC-potential responses are approximately linearly related to the driving process(es). Nonlinearity in the system will give rise to a difference in the ADC and DC-potential transition characteristics unless the dependencies in the two responses are identical. But identical dependencies are not necessarily expected. For example, identical DC-potential and ADC dependencies on ion concentration are not expected if ion movement is the common driving process. Nevertheless, any resulting difference in transition characteristics should be consistent across animals given a common latent process. No such regularity in transition behaviour is observed, however, thus ruling out both linear and nonlinear common latent processes. In particular, Fig. 3.28a shows that the median difference in the ADC and DC_{fast} transition rates is positive in some animals, while it is negative in others. Heterogeneity is also observed in the ADC-DC_{slow} median rate differences (Fig. 3.28c). It might be argued that it is wrong to suggest that the ADC-DC_{slow} rate differences shown in Fig. 3.28 provide evidence against a common latent process, since eight of the ten median differences are negative, the remaining two being close to zero. On the other hand, an inspection of the ten individual overlaid

component DC-potential and ADC posterior cluster plots (Figs. 3.24 and 3.25) lends credence to the conclusion that there is no consistent temporal relationship between the two transitions. A common latent process involving nonlinear responses therefore appears unlikely. A triggered process model provides an alternative to the common latent process hypothesis and might appear as an extreme form of nonlinearity in the temporal relationship between the ADC and DC-potential responses. In a triggered system the stimulus for change in one variable is threshold crossing in the value of the driver variable. The threshold in question might be the concentration of some trigger species, for example. A comprehensive analysis of the time-series data is required to explore fully the possible involvement of some triggered process. This work is the subject of Chapter 4.

To conclude, the present study suggests that the ADC and DC-potential responses to ischaemia are not driven by a common latent process, although some caution is required since this conclusion is based on the between-animal differences in temporal behaviour. The analysis shows that a restricted focus on the mean responses can be misleading. The temporal relationship between the ADC and the established biophysical/biochemical responses to ischaemia clearly requires further investigation in order to understand fully the mechanism underlying the ADC response. The possible involvement of some triggering mechanism is examined in Chapter 4.

4. IS ANOXIC DEPOLARISATION ASSOCIATED WITH AN ADC THRESHOLD?

4.1 Introduction

4.1.1 Diffusion imaging in cerebral ischaemia

Diffusion imaging is used extensively in stroke research and as a routine clinical tool in the clinical assessment of stroke patients due to its well-documented responsiveness to the tissue changes that occur following a period of cerebral ischaemia. Nevertheless, the mechanistic basis of the observed stroke-induced ADC changes remains to be established, despite the numerous publications on this subject. [Several review papers on the role of diffusion imaging in cerebral ischaemia have been published, including King et al. (1997b) and Thomas et al. (2000).] A key to understanding the cerebral ADC response to ischaemia is a detailed examination of the temporal relationship between the ADC response and the well documented biophysical and biochemical changes that occur following an ischaemic insult. In one such study Harris et al. (2000) examined the temporal relationship between the ADC and DC-potential responses to transient focal ischaemia, and concluded that these responses are distinguishable in their temporal characteristics, with the decline in ADC preceding the negative DC-potential shift, the latter of which is indicative of anoxic depolarisation. A subsequent and formal Bayesian analysis of these data was performed using a Markov chain Monte Carlo (MCMC) simulation approach, the main focus of which was a within-animal comparison of the ADC and DC-potential transition rates and transition half-response times. The results, which are presented in Chapter 3, reveal an absence of any consistent temporal relationship between the ADC and DC-potential transition responses to focal ischaemia among the sample of ten animals, an observation which appears incompatible with the notion that the ADC and DC-potential responses are driven by a common latent process. Despite this finding, one cannot necessarily conclude that the two responses are entirely uncoupled in time, since the possibility remains that one response is triggered by the other, albeit with uncoupled transition behaviour. Thus the DC-potential transition might be initiated at that point in time when the ADC

reaches some threshold. Once triggered, the DC-potential transition rate might be determined by any number of factors unrelated to the ADC. In this chapter the combined ADC and DC-potential data are examined from this perspective. This is not to suggest that a change in the diffusivity of water is likely to act as a direct trigger, merely that the ADC might be coupled to the cellular events that do trigger the DC-potential response. A role for ADC thresholds in ischaemia has been discussed previously (Hoehn-Berlage et al., 1995; de Crespigny et al., 1999), prompted, presumably, by the well-documented existence of ischaemic response thresholds (Astrup et al., 1981; Spetzler and Nehls, 1987; Heiss, 1992; Busza et al., 1992 ; Hoehn-Berlage et al., 1995). The purpose of the analysis outlined in this chapter is to examine the previously published data of Harris et al. (2000) for evidence indicating the involvement of an ADC threshold in central nervous system anoxic depolarisation.

4.1.2 ADC thresholds in cerebral ischaemia

The concept of ADC thresholds is not new. A cascade of events is known to occur following an ischaemic/anoxic insult, starting with functional and metabolic disturbances and culminating in the onset of energy failure and anoxic depolarisation. It has been suggested that a cerebral blood flow (CBF) threshold exists for the commencement of each event (Astrup et al., 1981; Spetzler and Nehls, 1987; Heiss, 1992). This has prompted a search for the existence of related ADC thresholds since this has implications for clinical diagnosis and should lead to an improved understanding of the biophysical chemistry of stroke-induced tissue damage. In one such study Hoehn-Berlage et al. (1995) matched the areas of ATP depletion and acidosis after 2h of ischaemia in the rat with co-registered ADC images to obtain a mean ADC threshold of 77±3% (±SD, threshold expressed as a percentage of the ADC in the hemisphere contralateral to the ischaemic lesion) for ATP depletion (energy failure) and 90±4% for acidosis. Using these data they distinguish the unsalvageable lesion core, which is destined to become necrotic, and the surrounding oligemic corona in which blood flow is sufficient to prevent energy failure, but at the cost of severe tissue acidosis. Thus, the latter corresponds to the area in which the ADC lies in the range 77% to 90%. It is emphasised that these thresholds apply only to the specific animal model

used in the study and a 2h-period of ischaemia, and that an extensive series of studies is required to determine whether these thresholds are time and species dependent. de Crespigny et al. (1999) suggest that the existence of different ADC thresholds for acidosis and energy failure is related to the observation that, in their studies, the ADC response is biphasic.

4.2 Methods

4.2.1 Experimental methods

A detailed description of the experimental and image processing procedures is given in Harris et al.(2000) and Roussel et al.(1994). A summary is provided in Section 3.2.

4.2.2 Longitudinal data analysis

A complete description of the MCMC simulation and convergence diagnostic test procedures is given in Sections 2.2 and 3.2.2. The trigger threshold analysis outlined in this chapter reuses the MCMC output of Chapter 3, but it is based mostly on an alternative, model-free approach to characterising the temporal relationship between the DC-potential and ADC responses, with an emphasis on response onset time. The majority of the analyses use the ADC and DC-potential normalised 0.9 level as a marker of the onset of the two transitions and hence to determine the relationship between them. In contrast, the ADC median threshold and 95% posterior interval data shown in Fig. 4.3 were obtained by equating the ADC threshold to the ADC at the maximum in the V_{DC}-ADC curve second derivative, as obtained using the logistic model for the ADC and DC-potential transitions. The V_{DC}-ADC curve second derivative (dV_{DC}/dD) is given by

$$\frac{d^2V_{DC}}{dD^2} = \frac{dV_{DC}}{dt}\frac{d^2t}{dD^2} + \left(\frac{dt}{dD}\right)^2 \frac{d^2V_{DC}}{dt^2}, \qquad [4.1]$$

where D indicates the ADC. Thus, given transitions of the form

$$y = \frac{\beta_a}{\left[1+\exp(\beta_b + \beta_c t)\right]}, \qquad [4.2]$$

with $y = V_{DCfast}$, V_{DCslow} or D, rearrangement yields

$$t = \frac{-\beta_b + \log\left(\frac{\beta_a - y}{y}\right)}{\beta_c}, \qquad [4.3]$$

from which the following derivatives are obtained:

$$\frac{dt}{dy} = \frac{-\beta_a}{\beta_c(\beta_a y - y^2)},$$

$$\frac{d^2t}{dy^2} = \frac{\beta_a(\beta_a - 2y)}{\beta_c(\beta_a - y)^2 y^2}, \qquad [4.4]$$

$$\frac{dy}{dt} = \frac{-\beta_a\beta_c \exp(\beta_b + \beta_c t)}{\left[1 + \exp(\beta_b + \beta_c t)\right]^2},$$

$$\frac{d^2y}{dt^2} = \frac{2\beta_a\beta_c^2 \exp(2\beta_b + 2\beta_c t)}{\left[1 + \exp(\beta_b + \beta_c t)\right]^3} - \frac{\beta_a\beta_c^2 \exp(\beta_b + \beta_c t)}{\left[1 + \exp(\beta_b + \beta_c t)\right]^2}.$$

These equations were used in conjunction with a simple search algorithm and the MCMC output of Chapter 3 to obtain the threshold medians and posterior intervals. Essentially the same results are obtained by adopting a simple finite difference approach.

4.3 Results

Fig. 4.1 shows the DC-potential and ADC responses acquired from each of the ten animals, plotted individually, together with a set of 100 random time-course profiles generated by sampling from the posterior distribution. The observed DC-potential and ADC response profiles are also shown in Chapter 3 (Fig. 3.2) together with the individual-specific ADC and DC-potential posterior median responses (Figs. 3.20 to 3.22). These plots indicate that there is a considerable uncertainty in the onset of the ADC response. Furthermore, a proportion of the randomly sampled

---

Fig. 4.1. Overlaid plots showing the normalised ADC (continuous irregular line) and DC-potential (white broken line) response profiles acquired from each of the 10 animals, together with a set of 100 animal-specific random profiles obtained by sampling the posterior distribution (ADC profiles, grey broken lines; DC-potential profiles, black band surrounding the white broken line).

Fig. 4.1. **Normalised individual-specific ADC and DC-potential response profiles**

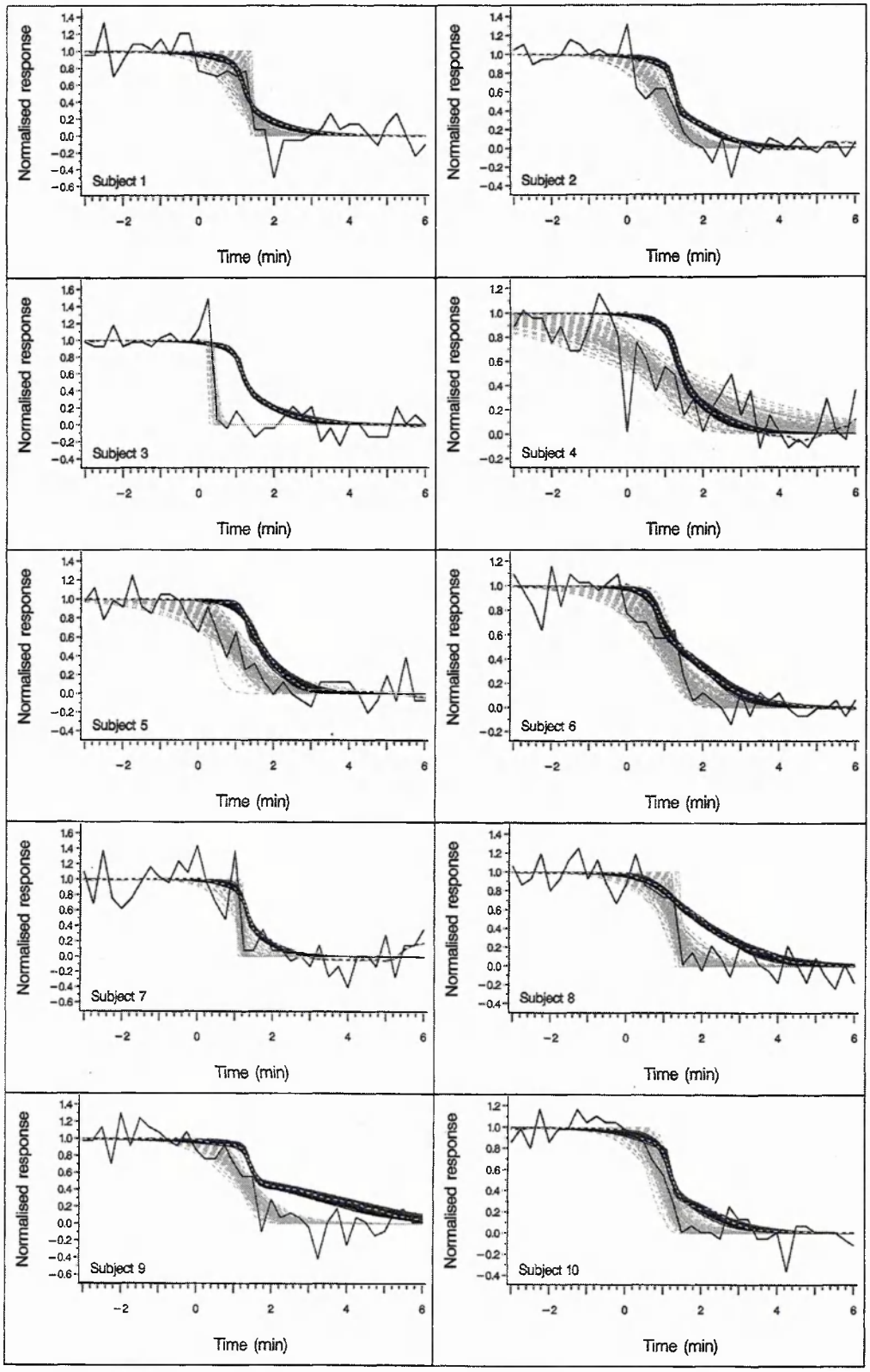


Figure legend on the preceding page.

ADC response profiles obtained for some animals (4, 5 and 6, in particular) show a non-negligible decline in ADC prior to zero time, and this appears to be reflected in the ADC observations. The reasons for this behaviour and the resulting implications are dealt with in Section 4.4.3. Suffice it to say that the unusual behaviour observed in these three animals must be taken into consideration in the following analyses.

The MCMC output was used to perform a number of calculations with a view to identifying behaviour consistent with the existence of an ADC trigger threshold. These analyses were exploratory in nature and relied heavily on various graphical displays. An initial search for the existence of an ADC threshold was performed by generating DC-potential versus ADC (V_{DC}-ADC) plots (Fig. 4.2), using the MCMC output. Separate plots were produced for each animal using the MCMC posterior median fixed- and random-effect coefficients to generate values for the ADC and DC-potential as a function of time, these serving as input to a plotting routine. In addition, a set of credible V_{DC}-ADC profiles was generated by resampling from the posterior distribution, i.e., by resampling the MCMC output. The median and sampled profile sets are both shown in Fig. 4.2. No consistent behaviour is observed across all ten animals. Thus, the plots obtained for animals 2, 4, 5, 9, 10 and possibly 6 might be taken as evidence for a threshold centred in the region of 0.60 to 0.62, say, but the remaining plots are inconsistent with a well-defined threshold in this range. Furthermore, the V_{DC}-ADC plots generated for animals 4, 5 and 6 should be viewed with caution due to possible baseline instability. The shape of the plot obtained for animal 1 indicates that the DC-potential and ADC essentially move in synchrony over much of

---

Fig. 4.2. V_{DC}-ADC curves were generated for each of the 10 animals using the posterior random coefficient estimates. The curve generated from the individual-specific posterior medians is shown for each animal (white), together with 100 random curves generated by sampling the posterior distribution (black). Each curve spans a time interval starting 3 minutes pre-occlusion (lower left region of the plot) and ending 6 minutes post-occlusion (upper right region).

Fig. 4.2. **Individual-specific DC-potential dependence on ADC**

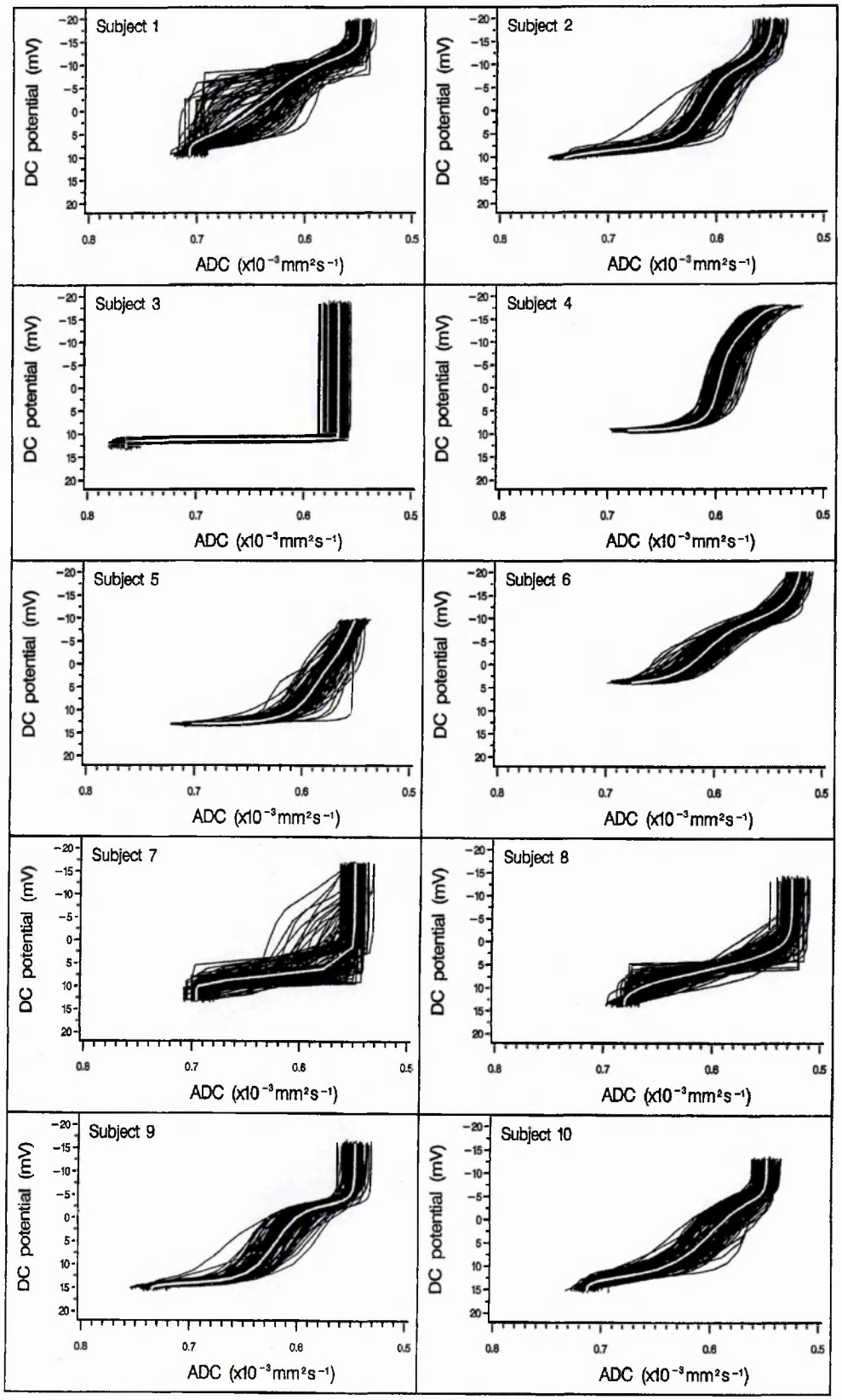
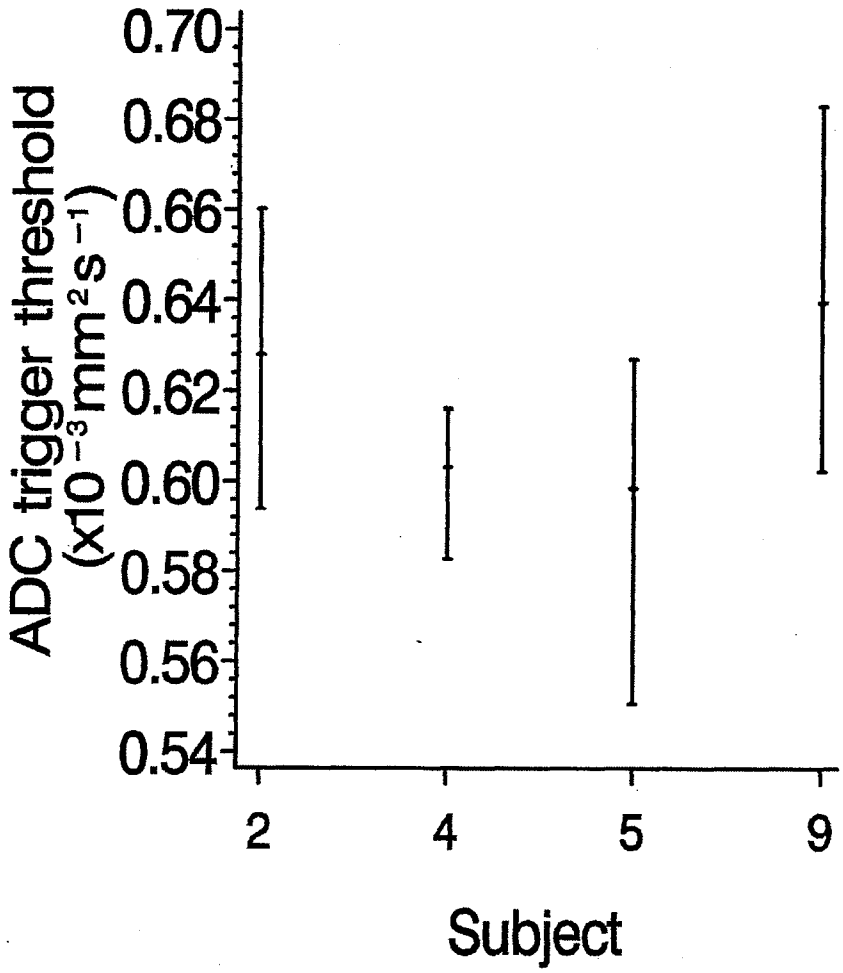


Figure legend on the preceding page.

their range. Similarly, proportional ADC and DC-potential responses were seen in animal 8 over a substantial part of the ADC transition. An additional feature of the response observed in this animal was the absence of the rapid DC-potential component that was observed in the remaining nine cases (see top left panel of Fig. 3.27). The unique plot obtained for the third animal is a reflection of the ADC transition having finished prior to the start of the DC-potential transition. The median ADC transition obtained for animal 7 is close to a step transition and so this plot contains little useful information about ADC threshold points. Each of the accompanying credible profile sets reveals a considerable degree of uncertainty, but the preceding remarks are unaffected by this imprecision. The V_{DC}-ADC credible profile set obtained in the subset of animals in which the median V_{DC}-ADC curve has a shape consistent with a triggered DC-response were used to generate a set of animal-specific ADC threshold posterior distributions, a subset of which is shown in Fig.4.3. This was achieved by using an ADC threshold defined as the V_{DC}-ADC curve maximum second derivative. (This definition is justified in Section 4.4.3.) The trigger threshold posterior distributions obtained for animals 6 and 10 cover a wide range of ADC values and are not, therefore, included in the figure since this contravenes the definition of a well-defined threshold. Although a cursory inspection of the V_{DC}-ADC plots (Fig. 4.2) obtained from the selected subset of animals might be taken to indicate the presence of an ADC threshold, Fig. 4.3 reveals a considerable degree of between-animal variation among the subset of posterior intervals. This is reflected in the joint posterior probability distribution. Calculation shows, for example, that the ADC range $0.600 \leq \text{threshold} \leq 0.615$ has a joint posterior probability of only 0.12% in the subset of four animals with a narrow individual-specific posterior distribution. This lack of consistent behaviour, together with the observation that several of the remaining V_{DC}-ADC plots (1, 3, 7 and 8) lack any threshold signature, leads to the conclusion that, taken as a whole, the V_{DC}-ADC plots provide no consistent evidence for a well-defined ADC threshold.

As an alternative approach to the trigger analysis the ADC values at the start of the DC-potential transition have been examined using a more direct indicator of the latter. Given the logistic function used to model the DC-potential response and the associated asymptotic behaviour, there

Fig. 4.3. **Individual-specific ADC thresholds**

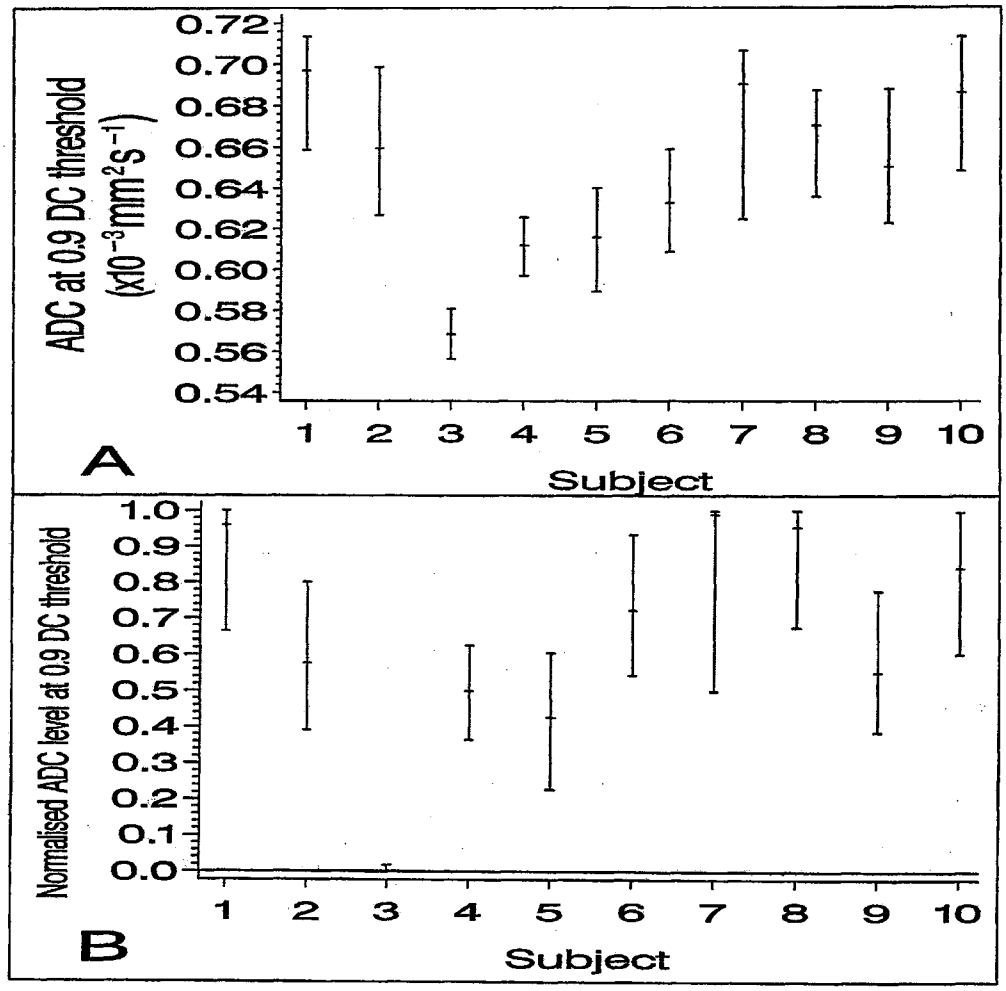


A plot showing the individual-specific ADC median threshold estimates and 95% posterior intervals generated for a subset of animals with a well defined turning point in the V_{DC}-ADC curve. The ADC threshold is defined as the maximum in the V_{DC}-ADC curve second derivative.

is no well-defined start to the DC-potential transition. A normalised DC-potential value of 0.9 was selected as marking the onset of the DC-potential response. Similar results were obtained using other normalised threshold crossing values, provided they were greater than 0.85. The argument for using the 0.9 normalised DC-potential level to define the start of the DC-potential transition is that it provides an estimate of the ADC at the point in time at which the DC-potential starts to show a marked excursion away from the pre-insult level and this should provide an estimate of the ADC trigger threshold if it exists. There is a considerable degree of between-animal variation in the resulting ADC trigger point (Fig. 4.4A) but this is attributable, at least in part, to the substantial between-animal variation in pre-insult ADC. The effects of between-animal variation in the pre-insult and asymptotic post-insult ADC levels can be removed by performing the analysis on a normalised ADC scale. Fig. 4.4B shows, however, that a substantial level of between-animal variation remains in the normalised trigger thresholds. No evidence for a consistent ADC threshold emerges from this analysis. As mentioned previously, animals 4, 5 and 6 might be disregarded as being unreliable, but substantial between-animal differences exist among the remaining threshold posterior intervals and so the conclusion remains unaltered. This is consistent with the impression obtained by inspecting the overlaid ADC and DC-potential response profiles (Fig. 4.1). In animal 3, for example, the ADC transition reached completion prior to the start of the DC-potential response while in animal 1 the two transitions are more synchronised.

To complete this part of the trigger-point analysis, the existence of latency in the DC-potential response must be considered. The above analysis was therefore repeated but using a modified model in which the DC-potential response is preceded by a latent period. Exploratory analyses were performed using a range of plausible latency values (Fig. 4.5). The ADC trigger threshold medians and 95% posterior intervals obtained without normalisation and assuming a latent period of 30s are shown in Fig. 4.6A. A considerable degree of between-animal variation is evident. A greater degree of similarity between animals is observed in the results obtained using the normalised ADC scale, and these results might be taken as consistent with a normalised ADC

Fig. 4.4. **ADC level at the start of the DC-potential transition**



Plots showing the estimated individual-specific median ADC level at the start of the DC-potential transition, together with the 95% posterior intervals. The start of the DC-potential transition is defined as the time at which the DC-potential fell to a level of 0.9 on a normalised scale. (A) Without ADC normalisation. (B) With ADC normalisation.

Fig. 4.5. **Relationship between the DC-potential and ADC. Delayed-response model results**

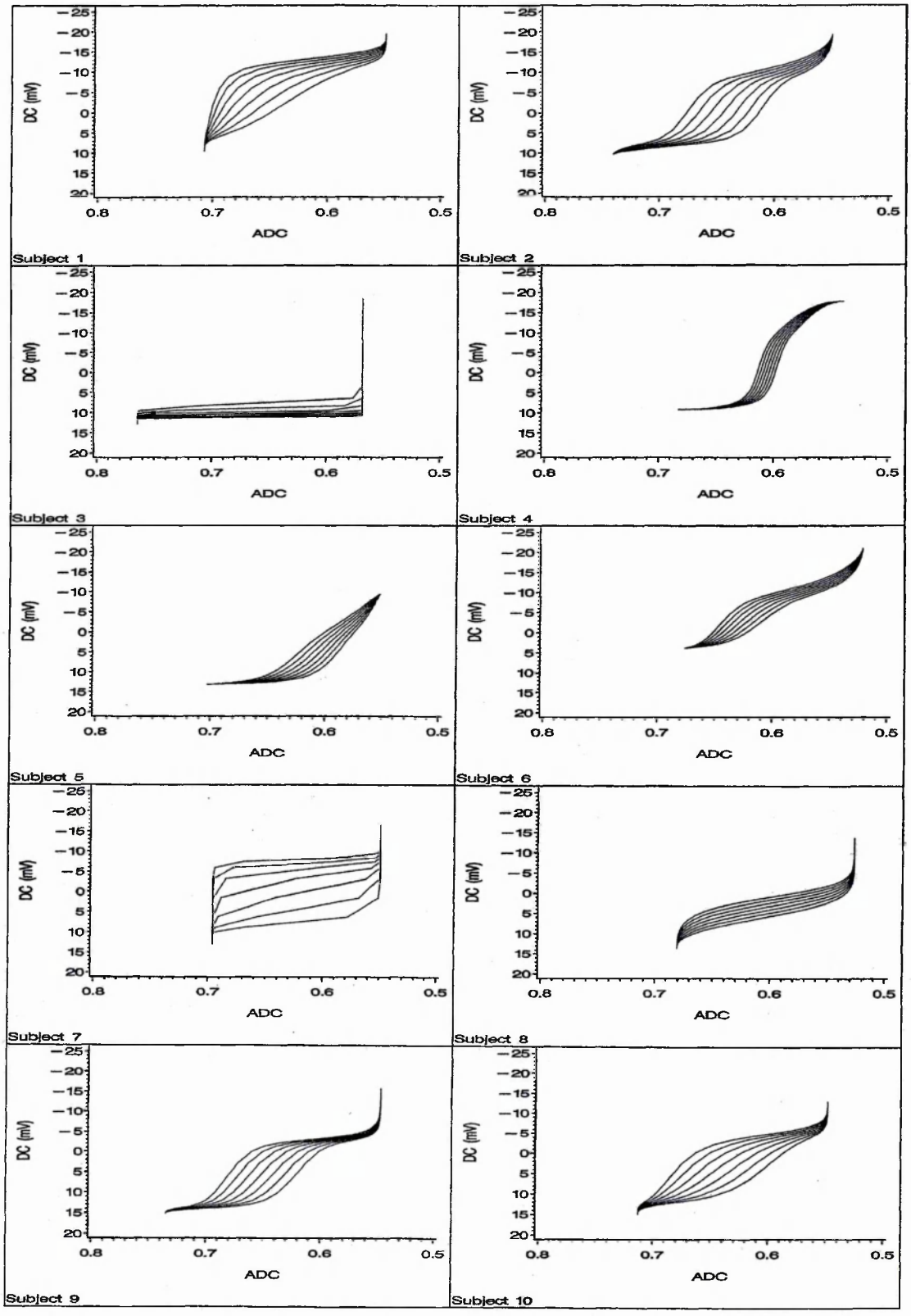
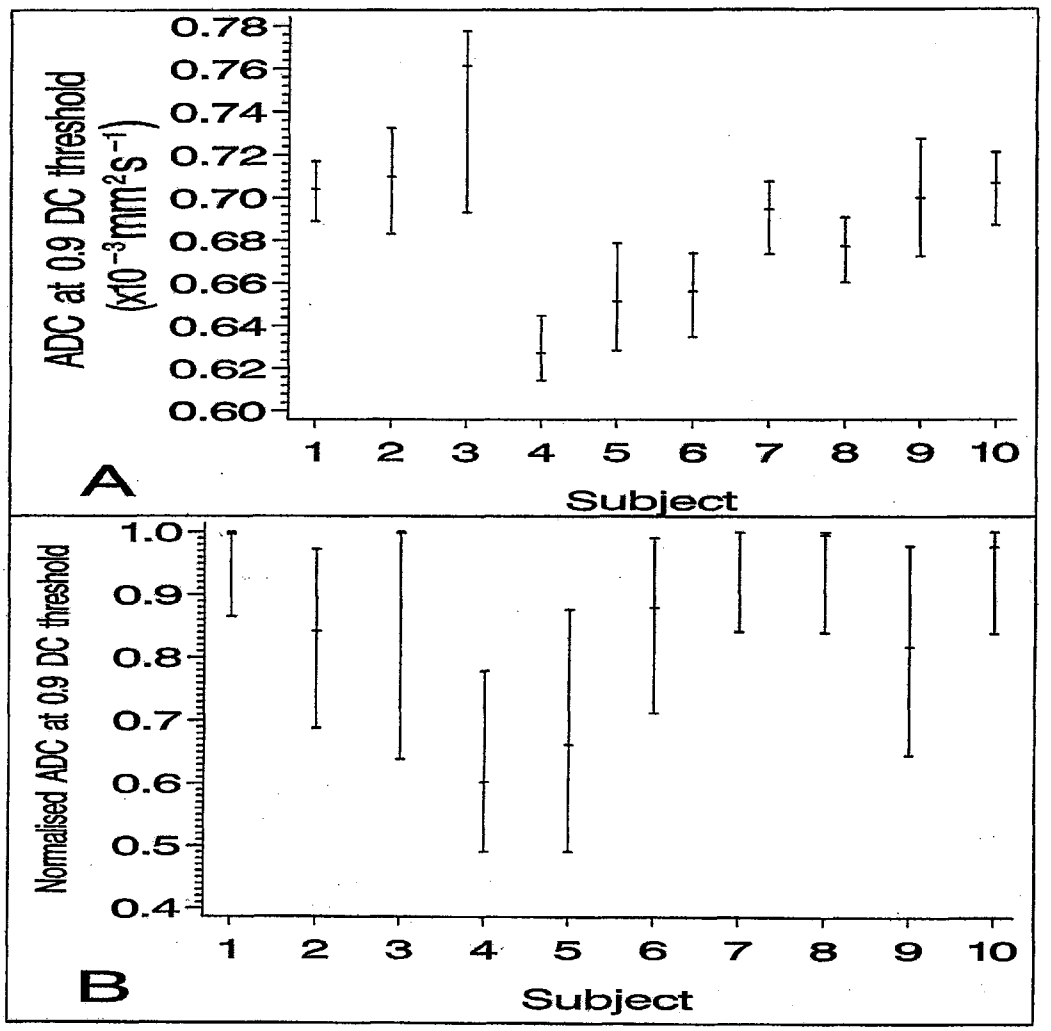


Figure legend on the following page.

Fig. 4.5. continued

A set of seven individual-specific V_{DC}-ADC curves were generated for each of the ten animals using the posterior random coefficient median estimates and assuming a latency in the DC-potential response ranging from zero (bottom curve) to 30 s (uppermost curve). The intermediate curves were generated at 5 s increments in latency. Each curve spans a time interval starting 3 min pre-occlusion (lower left region of the plot) and ending 6 min post-occlusion (upper right region).

Fig. 4.6. **ADC threshold estimates based on a delayed-response model**



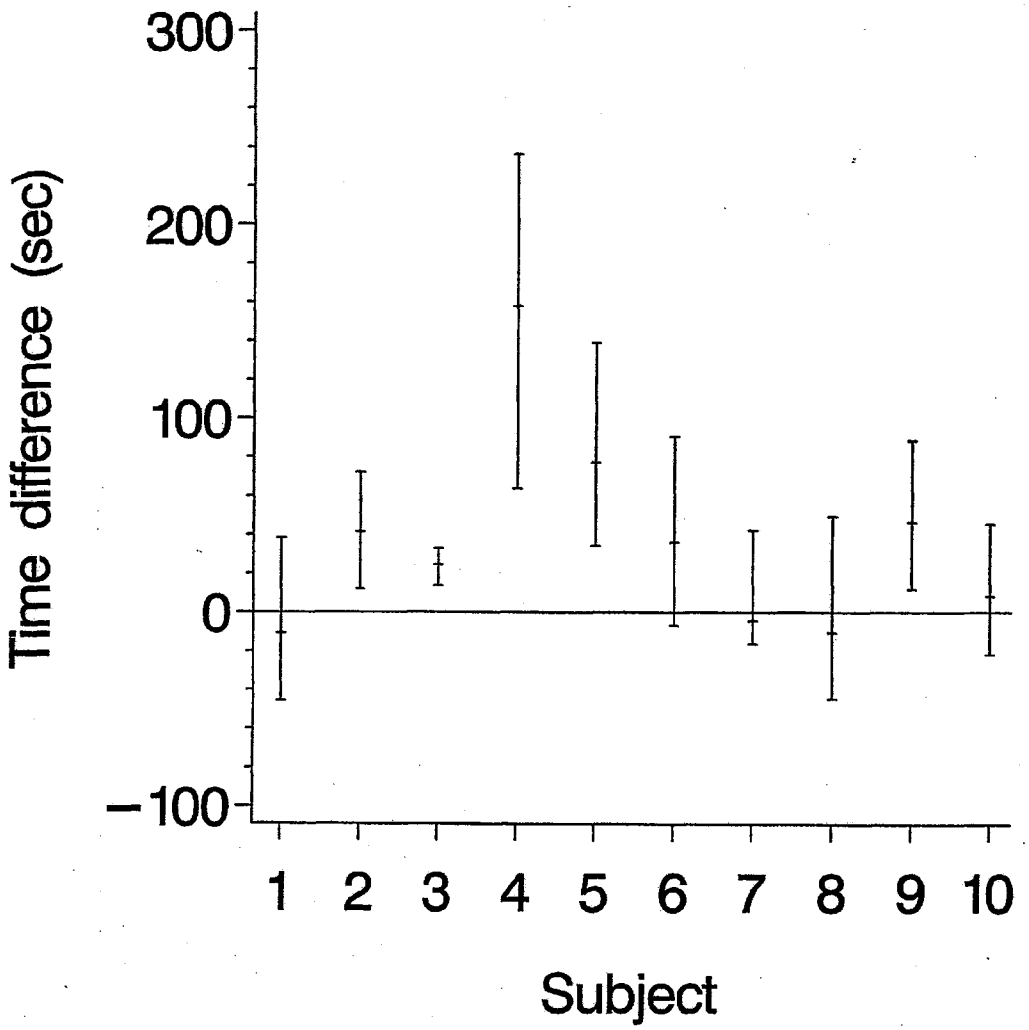
Plots showing the estimated individual-specific median ADC level 30 s before the start of the DC-potential transition, together with the 95% posterior intervals. The start of the DC-potential transition is defined as the time at which the DC-potential fell to a level of 0.9 on a normalised scale. (A) Without ADC normalisation. (B) With ADC normalisation. In subjects 1, 3 and 7 the median normalised threshold and 97.5% posterior quantile are both close to unity and appear superimposed in the figure.

threshold close to the baseline value of unity, given the need for caution concerning the data acquired from animals 4, 5 and 6 (Fig. 4.6B). Thus, if the notion of an ADC trigger is correct, the results suggest that it lies close to the pre-insult ADC level and that it is associated with a latency in the DC-potential response. Prompted by this finding an alternative presentation of the data was produced to show the individual-specific medians and 95% posterior intervals obtained for the difference in time between the normalised 0.9 level crossing points in the ADC and DC-potential responses, i.e., the individual-specific latencies, based on a normalised value of 0.9 for the ADC threshold and using a normalised value of 0.9 as an indicator of the start of the DC-potential response (Fig. 4.7). With the exception of animal 4, a latency of the order of 34s lies within the 95% posterior interval for all animals. The population median V_{DC}-ADC onset time difference is 22s (95% posterior interval: -6 to 70s).

Component Responses

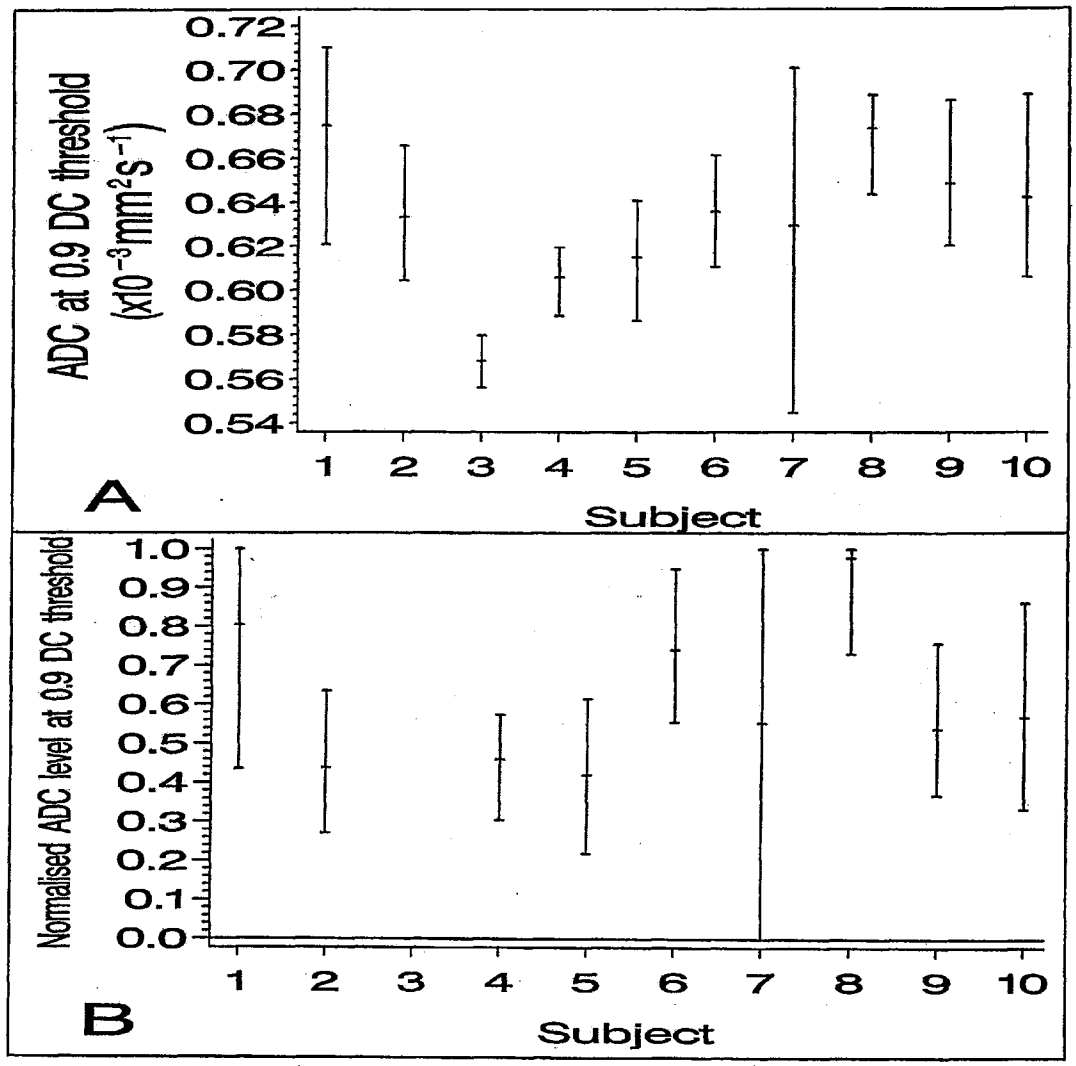
The analysis outlined in Chapter 3 shows that the DC-potential response is biphasic. Furthermore, an examination of Figs. 3.21 and 3.26 indicates that the two DC-potential components are simultaneous rather than sequential, both components making a significant contribution to the early stages of the response. Thus, to complete the analysis outlined in this chapter, it is necessary to examine each component separately in order to explore the possibility that one of the DC-potential components exhibits a consistent trigger-type relationship with the ADC. The failure to see this in the analysis of the composite DC-potential response might be due to its being obscured by the second, overlapping DC-potential component. Figs. 4.8A and 4.9A show the individual-specific median ADC threshold estimates and 95% posterior intervals obtained for the fast and slow DC-potential components, respectively. The former are not dissimilar to those obtained using the composite DC-potential response (Fig. 4.4A), and fail to provide evidence for a consistent ADC threshold in relation to the fast DC-potential component. The same applies to the normalised threshold data shown in Fig. 4.8B. Similarly, there is considerable between-animal variation in the slow DC-potential component ADC threshold posterior intervals, with or without normalisation (Fig. 4.9), although several of the posterior intervals extend towards unity on the normalised ADC

Fig. 4.7. **DC-potential to ADC 0.9 level crossing point time difference**



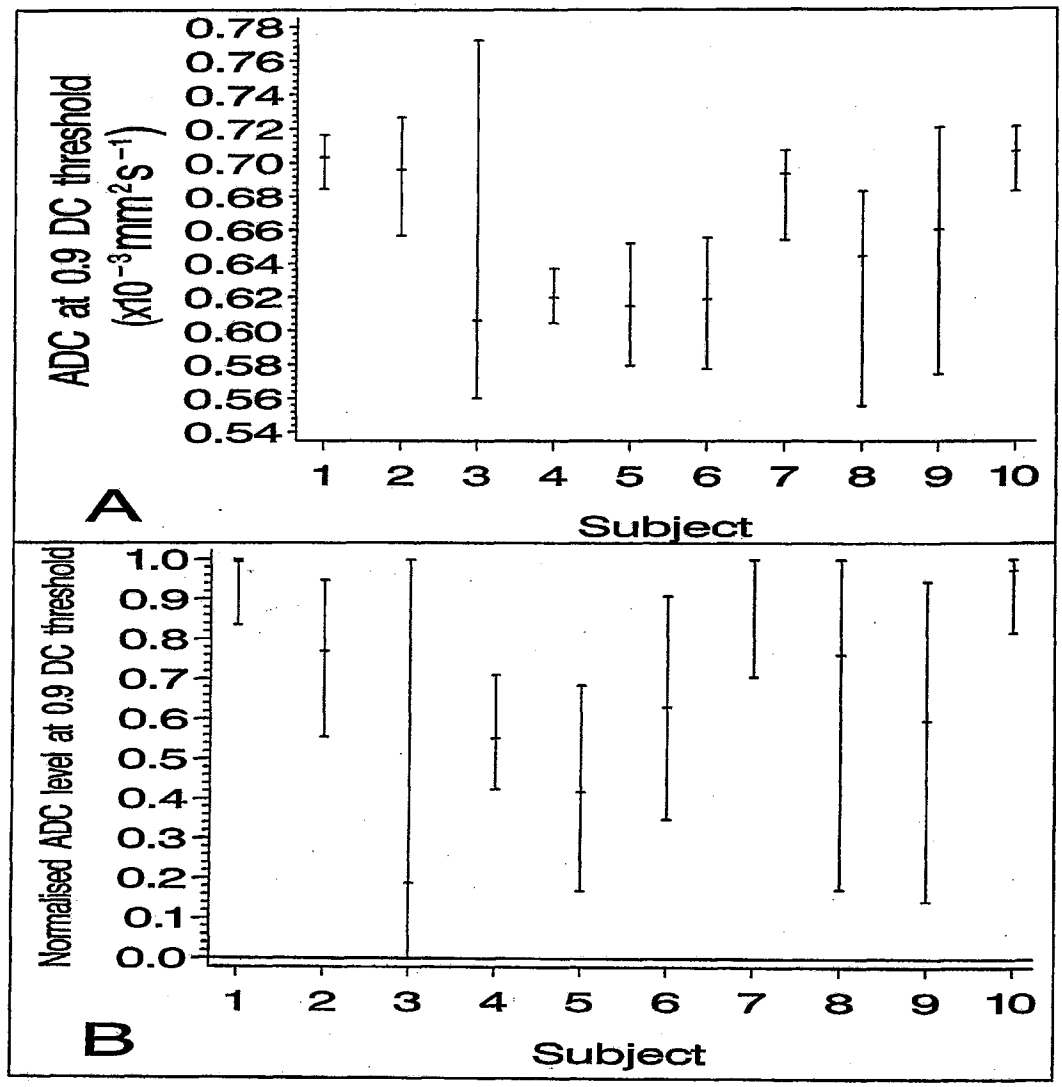
Individual-specific medians and 95% posterior intervals obtained for the difference in time between the normalised 0.9 level crossing points in the ADC and DC-potential responses.

Fig. 4.8. **ADC level at the start of the fast DC-potential transition**



Plots showing the estimated individual-specific median ADC level at the start of the fast DC-potential transition, together with the 95% posterior intervals. The start of the fast DC-potential transition is defined as the time at which the fast DC-potential component fell to a level of 0.9 on a normalised scale. (A) Without ADC normalisation. (B) With ADC normalisation.

Fig. 4.9. **ADC level at the start of the slow DC-potential transition**



Plots showing the estimated individual-specific ADC median level at the start of the slow DC-potential transition, together with the 95% posterior intervals. The start of the slow DC-potential transition is defined as the time at which the slow DC-potential component fell to a level of 0.9 on a normalised scale. (A) Without ADC normalisation. (B) With ADC normalisation.

scale. In addition, a comparison of Figs. 4.4B and 4.9B indicates that, with the exception of animals 5, 6 and 8, the individual-specific slow DC-potential median normalised thresholds are closer to unity than the corresponding composite DC-potential thresholds. This arises because, in the remaining cases, the onset of the slow DC-potential component precedes that of the fast-DC component (Fig. 3.26). Despite the resulting increase in similarity among the individual-specific medians, the slow DC-potential component ADC normalised threshold posterior intervals are too disparate to be taken as evidence for a consistent threshold across all animals (Fig. 4.9B).

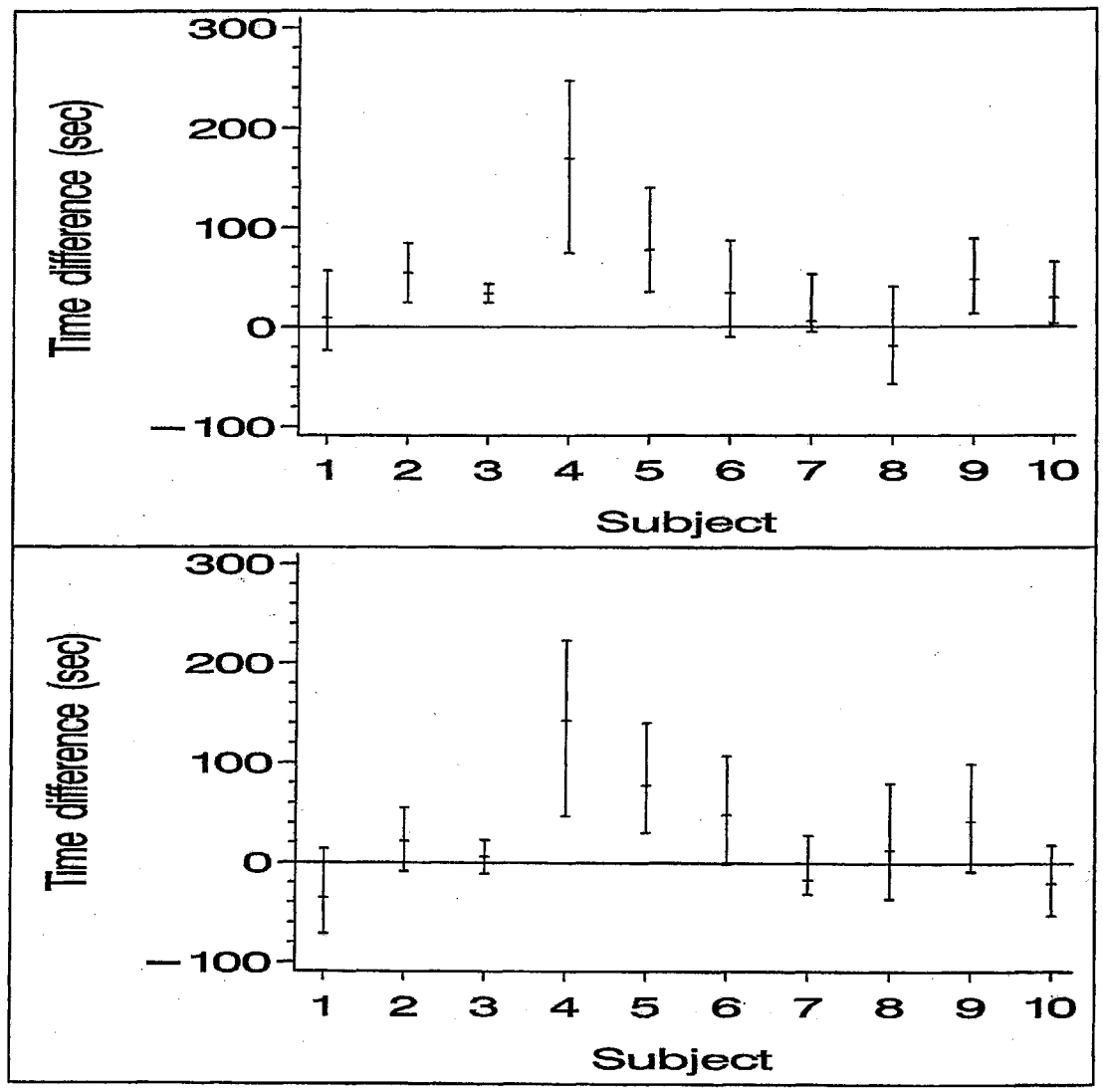
The preceding single DC-potential component analysis should be extended to consider the possibility of a well-defined threshold combined with a latency in the response. Inspection of Fig. 3.26 indicates, however, that a marked latency in the slow DC-potential response is not, in general, plausible since its onset is close to zero time in every case. In contrast, a latency in the fast DC-potential component is a possibility. A latent-period analysis was performed, but failed to uncover any consistent threshold-type relationship between the fast DC-potential component and the ADC response (data not shown). Finally, Fig. 4.10A shows the median difference between the individual-specific ADC and fast DC-potential normalised 0.9 level crossing times, together with their associated 95% posterior intervals. Fig. 4.10B provides this information in relation to the slow DC-potential component. The earlier onset of the slow DC-potential response, relative to the fast component, in seven of the ten animals (the exceptions are animals 5, 6 and 8; see Fig. 3.26) is reflected in the difference between the two sets of results. With the exception of two cases that were flagged as having an unstable baseline (animals 4 and 5), the posterior intervals for the slow DC-potential component all span zero, which is consistent with the preceding comment on the implausibility of a marked delay in the slow DC-potential response.

4.4 Discussion

4.4.1 ADC thresholds in cerebral ischaemia

Ischaemia causes a variety of well documented biochemical and electrophysiological changes in the affected tissue. It is established that, in general, each is associated with a well-defined CBF

Fig. 4.10. **Fast and slow DC-potential component to ADC 0.9 level crossing point time differences**



Plots showing the individual-specific medians and 95% posterior intervals generated for the difference in time between the normalised 0.9 level crossing points in the ADC and each of the two DC-potential components. (A) Fast DC-potential component. (B) Slow DC-potential component.

threshold, and that tissue perfusion must fall below some defined level to elicit each of the responses (Astrup et al., 1981; Spetzler and Nehls, 1987; Heiss, 1992). ADC measurements are used routinely in the clinical assessment of stroke patients, and it has been suggested that the ADC response might be similarly graded. For example, Hoehn-Berlage et al. (1995) have shown that, after two hours of ischaemia, rat brain tissue with an ADC lower than 77% of normal is destined to become necrotic, while poorly perfused tissue with an ADC in the range 77% to 90% of normal is salvageable, despite the accompanying acidosis. The authors emphasise that these ADC thresholds are expected to depend on the duration of the ischaemic period, and to be species specific. Prompted by the concept of a graded response to the decline in ADC that occurs in ischaemia, an analysis was undertaken to determine whether anoxic depolarisation is associated with a well-defined ADC threshold. This might be taken as evidence for a mechanism in which the ADC is tightly coupled with a process that triggers the DC-potential response. A reanalysis of previously published ADC and DC-potential data (Harris et al., 2000) was undertaken for this purpose, the results of which are presented in this chapter.

4.4.2 On the question of an ADC threshold for anoxic depolarisation

It is necessary at this point to clarify the use of the term *ADC threshold for depolarisation*. For the purpose of the present V_{DC}-ADC analysis, this is defined as a narrow and reproducible ADC level at which the DC-potential response is triggered, a level that is essentially independent of other prevailing conditions. In particular this definition requires the threshold to be independent of the duration of the ischaemic insult. A triggered response in which the ADC level depends on one or more additional factors would be interpreted as contravening the definition of a well-defined threshold. It might be argued that this definition is inappropriate and that the threshold associated with any triggered process might be modulated by one or more variables and fail, therefore, to meet the reproducibility criterion. This kind of modulation might be expected when the triggering and response processes are separated by a sequence of intermediate events. For example, the CBF threshold required to elicit an ADC response is reported to depend on the duration of the anoxic insult (Kohno et al., 1995, Table 3). Similarly, a dependence on insult duration is indicated in the

CBF threshold for other key components of the ischaemic response, including the functional/ electrical (Heiss, 1992), biophysical (i.e., volume/electrical impedance; Matsuoka and Hossmann, 1982b), biochemical (Kohno et al.,1995, Table 3) and pathological (Heiss, 1992; Lassen and Astrup, 1987, and references therein) components. In defence of the present definition it should be noted that the purpose of the V_{DC}-ADC analysis is to explore the possibility that the ADC and DC-potential responses are closely linked in the sense that both are intimately coupled to the transmembrane ion movements that occur subsequent to an ischaemic/anoxic insult. In this context the notion of a well-defined ADC threshold that is independent of the history of the system (occlusion rate and CBF level, for example), and other factors, is sensible.

4.4.3 V_{DC}-ADC MCMC simulation analysis

As a starting point for the analysis, V_{DC}-ADC plots were constructed using the MCMC output. If both the DC-potential and ADC measurements had been obtained with high precision and sufficient temporal resolution then the raw DC-potential and ADC data might have been used for this purpose. But the ADC measurements are relatively imprecise and in several cases the ADC transition is undersampled. For this reason the MCMC parameter medians were used to generate values for the ADC and DC-potential as a function of time, these serving as input to the plotting routine in preference to using the observed DC-potential and ADC values. In addition, a set of individual-specific credible V_{DC}-ADC profiles was generated by resampling the MCMC output. The V_{DC}-ADC curve maximum second derivative was used as a threshold parameter (Fig. 4.3) on the basis of its potential sensitivity to the onset of cascade-like processes. The V_{DC}-ADC curve analysis was supplemented by a threshold crossing-point analysis. It is appropriate, however, to acknowledge that a formal changepoint analysis (Carlin et al., 1992; Chapter 5 is devoted to MCMC changepoint analyses) modified to deal with smooth transitions (Bacon and Watts, 1971) might provide an alternative approach, but this was not pursued. Among the advantages of the method outlined in this chapter is the manner in which the logistic function allows a complete characterisation of the sigmoid transition. Furthermore, a logistic-type response has a plausible biological basis.

The main finding arising from this ADC threshold analysis is that, taken together, the data appear contradictory to the notion of a well-defined ADC threshold for anoxic depolarisation. Five of the ten animals yielded response profiles that appear consistent with the existence of an ADC trigger, but further analysis indicates that even among this subset of animals a considerable degree of variation is evident among the individual-specific threshold indices, suggesting the absence of a well-defined threshold. Disregarding those animals (4, 5 and 6) in which the baseline ADC is ill-defined, the data might be considered as consistent with a trigger-point close to the pre-insult ADC level (i.e., close to unity on the normalised ADC scale), together with a delay of the order of 30s in the onset of the DC-potential response. It should be noted, however, that this observation might be a reflection of the lower gradient in the early portion of the logistic ADC response curve. In the absence of any well-defined trigger point, the slower change in ADC close to zero time guarantees a better consistency between animals if a normalised trigger level close to unity is assumed, together with an appropriate value for the duration of the latent period. Expanding on this remark, each of the calculated subject-specific normalised trigger thresholds tends to unity with increasing latency, causing a reduction in the between-animal variation in the normalised trigger threshold estimate. The range of plausible values for the duration of the latent period is, however, subject to the time constraint that the trigger point must occur at a time greater than zero, at least to within experimental error. Thus, in the absence of a well-defined trigger threshold, a minimum between-subject variation in the normalised threshold estimate will be obtained by assuming a latency in the DC-potential response with a maximum plausible duration. Exploratory calculations show that even among the subset of animals with a well-defined maximum second derivative in the V_{DC}-ADC curve (animals 2, 4, 5, 9), no combination of ADC trigger level and plausible latency eliminates the considerable between-animal variation in ADC trigger threshold, with or without normalisation. This is seen in Fig. 4.6, which shows the individual-specific median thresholds, and 95% posterior intervals, derived by assuming a latent period of 30s and using the normalised 0.9 level crossing point as an approximate marker of the onset of the DC-potential response. Thirty seconds is close to the maximum possible latency under the condition of non-negative individual-specific trigger timing, and yet a considerable level of between-animal

difference remains among the posterior intervals, even if attention is restricted to the subset of animals with a well-defined V_{DC}-ADC curve turning point.

The latent variables analysis presented in Chapter 3 shows that the DC-potential response is biphasic, and consists of overlapping fast and slow components (Figs. 3.20, 3.23 and 3.26). Thus, to complete the analysis outlined in this chapter, it is necessary to examine each component separately in order to explore the possibility that one of the DC-potential components exhibits a consistent trigger-type relationship with the ADC. The failure to see this in the analysis of the composite DC-potential response might be due to its being obscured by the second, overlapping DC-potential component. No evidence for a well-defined threshold was found in either of the two DC-potential components, however.

Given the overall heterogeneity in the two ADC threshold indicators, i.e., the V_{DC}-ADC curve maximum second derivative and the normalised 0.9 level crossing point, the data as a whole (or, if preferred, the subset with animals 4 to 6 excluded) provide no strong evidence for the existence of a well-defined ADC trigger point. On the other hand, it is not possible to rule out a mechanism in which a small departure of the ADC from the pre-insult level is associated with an event, or series of events, that leads to delayed anoxic depolarisation, with a substantial degree of between-animal variation in the duration of the latent period. It should be noted that, in the present analysis, latency does not appear as a model parameter, per se, and is not subject to the same distributional constraint that operates on the random coefficients. Thus the model places no direct distributional constraint on the relationship between the ADC and DC-potential onset times. A different statistical model in which the ADC and DC-potential transitions are coupled through a latency random coefficient might provide supplementary information, but this approach was not pursued. Bearing this in mind, Fig. 4.7 shows the individual-specific median difference in time between the normalised 0.9 level crossing points in the ADC and DC-potential responses. These data provide evidence for a temporal separation in the initiation of the ADC and DC-potential responses in 5 animals, as indicated by those 95% confidence intervals that exclude zero. In contrast, however,

the population median difference 95% posterior interval spans zero (median: 22s; 95% posterior interval: -6 to 70s), which is consistent with the results obtained in a study of global ischaemia indicating that the ADC and DC-potential responses are indistinguishable in terms of their mean onset time (see Table 1 in de Crespigny et al., 1999). It has been suggested (de Crespigny et al., 1999) that this appears inconsistent with the report (Huang et al., 1997) that the ADC response precedes the negative DC-potential shift. The latter observation is based, however, on a comparison of the ADC and DC-potential data acquired in separate studies. [The DC-potential information was taken from Hansen (1978)]. Furthermore, the results presented in this chapter show that population mean/median onset-time difference estimates can be deceptive and that an analysis of within-subject differences provides additional information.

Some clarification is required regarding the very notion of an ADC threshold for anoxic depolarisation. Irrespective of the outcome of the analysis, changing water diffusivity is unlikely to act as a direct trigger. A more plausible hypothesis is that the ADC is coupled tightly to some underlying process that acts as a trigger for depolarisation. But the present analysis provides no firm evidence for a triggered DC-potential response and alternative mechanisms must be considered. Among the possibilities is a process in which the DC-potential and ADC respond independently to events that take place at a time close to the start of the ischaemic insult. This assumes that both transitions occur after a latent period, the duration of which tends to be longer for the DC-potential response. The distinction between this and the trigger hypothesis is that the former implies that both processes are passive, while the trigger hypothesis suggests that the ADC change is associated directly with the process that initiates anoxic depolarisation. Additional information is provided by the latent variables analysis of Chapter 3, the results of which suggest that the temporal relationship between the DC-potential and ADC responses is inconsistent with a latent variables model. The distinction between a latent process and a triggered process is important. According to a latent-process model the ADC and DC-potential are immediately and tightly coupled to some underlying driving process. The coupling is continuous over the duration of the response. In contrast the trigger model assumes that the two processes are linked at one

point in time only, i.e., the trigger point. Once initiated, the triggered response occurs at a rate that is independent of the triggering process. The distinction is between a driving process and a process that involves initiation only.

Baseline instability

Reference is made in Section 4.3 to the fact that in some animals the ADC transition appears to precede the ischaemic insult, as indicated by a decline in ADC prior to zero time. These data have not been excluded from the analysis because this was judged inappropriate. For the same reason a constrained model was not adopted, i.e., one in which the start of each of the two transitions is constrained to occur at a time greater than zero. Doubts regarding the validity of either of these approaches arise for the following reason. Baseline data were collected for a period of about 7min prior to occlusion, although it was considered appropriate to truncate the baseline period for modelling purposes, hence the 3min baseline span shown in Fig. 4.1. An inspection of the entire baseline shows, however, that the ADC declined slowly over the entire 7min period in the fourth animal, while an erratic, cyclic instability in ADC is evident in the fifth animal. Given the relative stability of the baseline ADC in the majority of the remaining animals, these time-dependent changes in ADC appear to reflect real changes in physiology as opposed to machine instability or measurement error. The resulting pre-insult decline in ADC appears to be genuine and it has been captured by the logistic function in these two animals. Thus the posterior parameter estimates and associated ADC time-course profiles obtained for animals 4 and 5 are a reflection of the superimposed declining baseline and the driven ADC transition. Given the preceding definition of an ADC threshold (Section 4.4.2) this summation is appropriate since the manner in which the proposed ADC threshold is reached is immaterial. It is fortunate, therefore, that any systematic decrease in baseline ADC is captured by the logistic function, although this necessarily introduces some degree of uncertainty in the normalised parameter estimates. The latter is inevitable because the pre-insult asymptote is not well-defined in the presence of an unstable baseline. Furthermore, a declining ADC during the baseline period implies some degree of model failure in the region of the post-insult asymptote due to the symmetry of the logistic function, although this is lost in the

noise. In an ideal experiment, the baseline ADC would be monitored in real-time and the experiment terminated if the baseline ADC is unstable. But the ADC time-course data were available only after a considerable amount of image processing and subsequent calculation, and an evaluation of the ADC data in real-time was not feasible (see Section 3.4 for further discussion). The data were, therefore, used in their entirety, but taking care to ensure that none of the conclusions is a consequence of baseline instability. To summarise, this study was undertaken to address important questions concerning the biophysical chemistry of the ADC response and the possibility of a threshold-type relationship between the ADC response and anoxic depolarisation. A formal analytical approach to the problem is outlined with the aim of making optimum use of the data while acknowledging the limitations of the study. The main conclusion is that the data provide no convincing evidence for the existence of a consistent ADC threshold. This conclusion is robust, despite the obvious baseline ADC instability that occurred in a few cases.

4.4.4 The sparse data problem, ill-conditioning and Bayesian random coefficient models

It is well known that random coefficients modelling can offer considerable advantages when dealing with problems that are under-determined or ill-conditioned in an ordinary regression context (see Section 1.4.3). This is especially true of Bayesian random coefficient (hierarchical) models. There were several instances in the present study in which attempts to fit an individual DC-potential and ADC profile pair led to failure, either because the ADC transition is corrupted with noise and/or is inadequately sampled, or because the biphasic nature of the DC-potential response is poorly identified. On the other hand, the Bayesian random coefficients model succeeded in providing the required subject-specific parameter estimates and associated posterior intervals. The Bayesian random coefficients model succeeds where a simple regression approach fails because it makes proper use of distributional information provided by the collection of individual response profiles. Thus, a formal modelling of the distribution of each of the random coefficients influences the subject-specific parameter estimates and associated statistics, and moderates the influence of spurious observations. This is related to the parameter shrinkage

behaviour discussed in Section 1.4.2.2, and is especially important if the response profile sampling within a given individual is sparse. It is, therefore, particularly relevant to the ADC data acquired in this study and, for this reason, shrinkage in the ADC transition-rate estimates is examined in Appendix B. Despite the advantages of a formal random-coefficients analysis, an improvement in the precision of the ADC random coefficient estimates would be helpful. This can, however, be achieved only by a reduction in the signal-to-noise ratio and/or an increase in sampling rate. Unfortunately, the signal-to-noise ratio is inversely related to the sampling rate, and an improvement in the quality of the data was not feasible given the available hardware.

4.4.5 Corroborative versus conflicting evidence for an ADC threshold in anoxic depolarisation

Several studies provide information that is relevant to the ADC threshold question. Hossmann (1971) reported that the onset of the extracellular volume response to ischaemia, as determined by cortical impedance measurements, coincides with the appearance of cortical potential changes (anoxic depolarisation), suggesting that these are related. On the other hand it has been noted (Kohno et al., 1995) that the CBF threshold for the electrical impedance response is far higher than the threshold for anoxic depolarisation [see, for example, the 2h CBF data given in Fig. 3 of Matsuoka and Hossmann, 1982b, which indicate a threshold of approximately 25-30 ml $100g^{-1}$ min^{-1} for the extracellular-space response, as measured by cortical impedance, compared with a threshold of the order of 8-10 ml $100g^{-1}$ min^{-1} for anoxic depolarisation (Lassen and Astrup, 1987, p462)]. This demonstrates that these processes can occur independently, despite the earlier and apparently contradictory report (Branston et al., 1978) that the CBF threshold for the electrical impedance response is 9.6 ml $100g^{-1}$ min^{-1}, which is close to the threshold for potassium release [of the order of 10 ml $100g^{-1}$ min^{-1} (Astrup et al., 1981; Lassen and Astrup, 1987, p462)]. Assuming a causative relationship between the extracellular volume change and the ADC response to ischaemia, the demonstration that the former can occur independently of anoxic depolarisation is consistent with the present V_{DC}-ADC data. In particular a latent variables relationship between the ADC and DC-potential responses has been ruled out (Chapter 3), and the

analysis outlined in this chapter suggests the absence of a threshold-type relationship between the two processes. In recognition of the observation that factors other than depolarisation must contribute to the volume changes that occur during an ischaemic insult, Hossman and co-workers (Matsuoka and Hossmann, 1982a, 1982b; Kohno et al., 1995) have discussed the well-documented osmotic disturbances that take place. Macromolecular breakdown products and lactic acid, the latter produced through anaerobic glycolysis, are both implicated. These components of the osmotic response are not expected, however, to make a significant contribution to the initial ADC-sensitive component of the extracellular volume change. Macromolecular degradation, for example, occurs at a later stage. It is suggested in Chapter 3 that components of the Nicholson-Kraig (1981) model for the electrophysiology of spreading depression provide an explanation for the present V_{DC}-ADC data. The threshold analysis outlined in this chapter does not alter this assertion.

4.4.6 Conclusion

The analysis reported in this chapter yields no consistent evidence for the existence of a well-defined ADC threshold for anoxic depolarisation. Despite the sampling-rate and signal-to-noise ratio limitations of the study, the precision of the subject-specific ADC threshold estimates is sufficient to warrant the conclusion that this is a reflection of true between-animal differences in ADC at the start of the DC-potential response.

5. MAGNETIC RESONANCE CHANGEPOINT STUDIES

5.1 Introduction

The motivation for the work outlined in this chapter arose from the realisation that in some pathological conditions [birth asphyxia (Hope et al., 1984; Hope and Reynolds, 1985; see Section 5.5.1 for details) and stroke[1] (Heiss, 1992), for example] clinical practice is driven, in part, by opinion based on results provided by time-of-onset studies. The concept of a *therapeutic window*, as applied to birth asphyxia and stroke, is an important example that arises in the paediatric environment in which the candidate works. Similarly, a *window of opportunity for intervention* has been discussed in relation to head trauma, again based on time-of-onset studies indicating that the primary insult leads to a series of events that culminates in delayed axonal damage (i.e., secondary axotomy; Povlishock, 1992; McKenzie et al., 1996; Teasdale and Graham, 1998; Laurer and McIntosh, 2001). Likewise, a progression of axonal changes have been identified in minor head injury, again leading to the suggestion that some form of therapeutic intervention might be possible (Povlishock et al., 1983). It is in the birth asphyxia context that a study is planned, and formal changepoint analyses of onset-time data are envisaged as an important component of the proposal. An evaluation of MCMC changepoint analytical methods was therefore undertaken. The changepoint problem arises in many other areas of MR research, including several biomedical disciplines, thus providing an additional incentive to undertake this work. A variety of Gibbs sampler and Metropolis MCMC analytical approaches to the changepoint problem were examined using two MR datasets. A selection of results are presented in this chapter. The following Sections provide an overview of several MR applications in which formal changepoint analyses are required. A brief discussion of the birth asphyxia therapeutic window issue is provided in Section

---

[1] In adult stroke a therapeutic window of up to 3-6 hours duration is consistent with clinical trial results (Fisher and Ratan, 2003) and neurological outcome studies in animals (Spetzler and Nehls, 1987).

5.5.1, while an appraisal of current practice in biomedical changepoint data analysis is given in Section 5.5.2.

5.1.1 The changepoint problem in MR research

Departure from an expected linear relationship between two variables is a phenomenon of considerable importance that arises in numerous disciplines, including a number of magnetic resonance applications. Among the well-known examples is the departure of a variety of systems from classical Arrhenius behaviour ($\log(rate) \propto 1/temp$ [1], temperature in °K), as observed in many studies into the effect of temperature on reaction rates and other processes. There are a number of applications in which Arrhenius behaviour is of considerable practical and/or economic importance. In other areas of research Arrhenius behaviour is of fundamental academic interest. For example, Arrhenius-plot changepoints have been associated with phase transitions in some systems, including lipid systems and biological membranes, electron paramagnetic resonance (EPR, also called electron spin resonance and abbreviated ESR) and NMR being prominent among the methods that have been used to study phase behaviour. Arrhenius-plot discontinuity studies are typical of an application in which formal changepoint analyses are required. A brief overview of selected industrial and scientific applications is provided in the following pages, mainly to make

---

[1]Absolute reaction rate theory provides the reaction rate expression (Silvius and McElhaney, 1981) $\text{rate} = (\kappa NkT/h)\exp(-\Delta H^{\#}/RT)\exp(\Delta S^{\#}/R)$ where κ is a constant, N is the number of reactant molecules, k and h are the Boltzmann and Planck constants, respectively, $\Delta H^{\#}$ and $\Delta S^{\#}$ are the enthalpy of activation and entropy of activation, respectively, R is the gas constant and T is absolute temperature. The pre-exponential term has a weak dependence on temperature relative to the first exponential term, thus giving rise to the Arrhenius equation which takes the form $k_{rate} = Ae^{-E/RT}$ where k_{rate} is the rate constant, A is treated as a constant, and E is the activation energy. The numerical difference between the activation energy and activation enthalpy is negligible in those systems under consideration.

the case for the importance of changepoint analyses, and hence to provide a justification for this chapter. It must be stressed that no attempt has been made to provide a comprehensive list of references. On the contrary, the majority of papers were selected using the criterion that each should include a figure illustrating the manner in which changepoints have been defined/detected. In addition, many of these studies were selected because they included an MR component. These references provide an entry point into the relevant literature.

5.1.1.1 Arrhenius discontinuity behaviour with industrial research applications

Among the economic and commercial contexts in which Arrhenius-plot discontinuities have been discussed are chill-injury/chill-sensitivity in plants (Bagnall and Wolfe, 1978, 1982), the shelf-life of both foods (Karmas et al., 1992) and pharmaceuticals (see Hancock and Zografi, 1997, for a review) and the longevity of seeds (Buitink et al., 1999, 2000). To elaborate on the shelf-life/longevity problem, a variety of pharmaceutical and food preparations exist as an amorphous solid. The chemical and physical properties of amorphous state materials are, therefore, hugely important in the food and drug industries because these have a direct impact on the chemical and physical stability and consequent shelf-life of numerous preparations. The major distinction between crystalline and amorphous solids is that the latter are characterised by short-range order and, in general, a greater degree of molecular mobility. Amorphous materials tend to be more reactive and relatively unstable. But an important characteristic of the amorphous state is the so-called glass transition (or glass to rubber transition). At temperatures lower than the glass transition temperature, T_g, an amorphous material exists in a glassy-state, while above T_g the material exists as a rubbery-solid. The former takes the form of a highly viscous glassy medium in which rotational and translational molecular mobility are restricted. In general, molecular mobility is a key factor in determining reaction rates, including rates of deterioration. Thus, deterioration tends to occur at a reduced rate in the glassy-state. For example, macromolecules, including protein and peptide therapeutic agents, tend to be more stable in a glassy-state medium (Chang et al., 1996). Consequently, the storage stability and shelf-life of amorphous pharmaceutical and food materials is critically dependent on storage temperature relative to the glass transition temperature. For these

reasons, numerous studies of the rotational correlation time and reaction-rate Arrhenius behaviour of amorphous materials have been carried out, with a focus on the changepoint associated with T_g (see, for example, Buitink et al., 1999, 2000; Karmas et al., 1992). One reason for the considerable interest in the amorphous state within the pharmaceutical industry is the improved solubility and bioavailability associated with amorphous materials, compared with crystalline preparations, despite the accompanying stability issues (Hancock et al., 1995; Hancock and Zografi, 1997). A similar interest in the amorphous state occurs in agriculture because the cytoplasm of seeds has been shown to form an intracellular glassy-state medium at low temperature and low water content, with a resulting increase in the longevity of stored seeds (Buitink et al., 1999, 2000).

The importance of the glass transition in amorphous materials is clear, and although differential scanning calorimetry (DSC) might be the method of choice for detecting phase changes, this is not always feasible (Buitink et al., 2000, p290). An alternative is to identify the transition temperature using ESR or NMR molecular mobility measurements [either directly via rotational correlation time or translational mobility measurement, or indirectly via T_2 measurement or solid-state magic-angle spinning spectroscopic studies (Kalichevsky et al., 1992)] in conjunction with Arrhenius-plot changepoint analysis (see, for example, Fig. 2 in Buitink et al., 2000). This approach has the advantage of providing additional information about each of the phases, including a variety of parameter estimates (for example, rotational correlation times, thermodynamic variables, etc; see Hancock and Zografi, 1997, pages 6-7 for references relating to the pharmaceutical sciences) which, it is suggested, can be used as predictors of shelf-life (Hancock et al., 1995, page 800). The relationship between the glass transition temperature and reaction-rate Arrhenius changepoint behaviour is the focus of attention in some studies (see, for example, Karmas et al., 1992, p877).

5.1.1.2 The Arrhenius changepoint problem in biomembrane research

The preceding section outlines the central role of Arrhenius plot discontinuity studies in some areas of pharmaceutical and food-science research. Similarly, Arrhenius discontinuity studies have provided important information relating to a number of basic science questions. In particular,

studies of the Arrhenius behaviour of membrane lipid mobility and the activity of various integral membrane enzymes has provided the basis for various theories relating to membrane structure and function. Arrhenius-plot changepoints have been the focus of much of this research (Kumamoto et al., 1971; Inesi et al., 1973; Hesketh et al., 1976; Wynn-Williams, 1976). A brief overview is provided in the following sections. ESR has played an important part in this work. [See Inesi et al. (1973), Raison and McMurchie (1974), Cannon et al. (1975) and Hesketh et al. (1976) for examples of early work on the muscle sarcoplasmic reticulum membrane ATPase and other membrane systems.] An example of the kind of information that can be derived from an ESR spectroscopic spin-labelled membrane study is provided by the first changepoint problem examined in this chapter. It is taken from an ESR study of the rabbit muscle sarcoplasmic reticulum (Ca^{++}-Mg^{++})-ATPase. A brief summary of the ESR method, as applied to biomembrane research, is provided below (Section 5.1.4) together with an overview of the literature on the Arrhenius behaviour of the sarcoplasmic reticulum membrane (Section 5.1.1.4). This is preceded by a brief description of one area in membrane biology in which temperature effects have an immediate functional relevance. The main objective is to provide support for the claim that changepoint analyses are fundamental to a number of important biological questions and to counter any impression that piecewise-linear models have no theoretical basis (Bagnall and Wolfe, 1982), thus undermining the suggested need for formal changepoint analytical methods. The material presented in this chapter indicates several research areas in which a rigorous changepoint treatment is required.

5.1.1.3 Low temperature membrane biology in plants and animals

Chill injury in plants

Many plants, including a variety of cereals, vegetables and fruits are susceptible to so-called chill injury. Below some critical temperature these plants fail to grow in the short term, and eventually die. Several studies have been performed to examine the relationship between the temperature-dependence of some growth-rate index and a variety of biochemical/biophysical observables, including mitochondrial function, chloroplast function and membrane lipid mobility (see, for

example, Raison and Chapman, 1976). The suggestion was that plant growth-rate and enzyme reaction-rate Arrhenius-plot discontinuities are both related directly to lipid mobility Arrhenius-plot changepoints and that changes in membrane lipid ordering/fluidity are the underlying determinant of the observed Arrhenius behaviour (Raison and Chapman, 1976). The relevance of these observations to the chill-sensitivity of certain plants is contentious, however, and it has been argued that membrane lipid mobility changes are an unlikely cause of abrupt temperature-induced changes in enzyme activation energies or plant growth rate characteristics (Bagnall and Wolfe, 1978). Nevertheless, interest in this kind of phenomenon continues in a variety of research areas, including low temperature animal biochemistry.

Low temperature animal biology

Hibernating mammals, reptiles and amphibians provide a number of fascinating examples of the manner in which cellular biochemistry has adapted to enable some species to deal with periods of low temperature and lack of food. For example, hibernators tolerate a degree of hypothermia that would be severely damaging in non-hibernating mammals (Mehrani and Storey, 1997). The transition between the hibernating and non-hibernating states is a temperature-driven process involving kinase-mediated protein phosphorylation. An important component of the arousal from hibernation is substrate mobilisation and subsequent energy production, as required for warming (MacDonald and Storey, 1998). The provision of free fatty acids as fuel for mitochondrial oxidation is a critical part of this process (MacDonald and Storey, 1998), brown adipose tissue being the primary source (MacDonald and Storey, 1998). Related to hibernation is the freeze-tolerance phenomenon exhibited by some hibernating amphibians and reptiles. This has been studied in a freeze-tolerant frog in which glucose is produced as a cryoprotective agent. Glucose serves to minimise the cell volume reduction that would otherwise occur in the presence of extracellular ice crystals (Holden and Storey, 2000). This has prompted several studies into the effects of temperature on the properties of the protein kinase involved in glucose production in this amphibian (Holden and Storey, 2000).

Returning to more general questions on low-temperature biochemical adaptation, various ATPases (an essential component of various ion pumps), kinases (enzymes involved in protein phosphorylation) and other integral membrane enzymes have been studied in the search for differences between hibernating and non-hibernating animals. In some cases the enzyme in question has been isolated and characterised (MacDonald and Storey, 1998; Holden and Storey, 2000), while other studies have been performed using membrane suspensions. The focus of these studies has been the sensitivity of the enzyme to changing temperature. Arrhenius discontinuities have been reported together with significant differences in kinetic/changepoint behaviour between hibernating and non-hibernating species (McMurchie et al., 1973; Mehrani and Storey, 1997; Holden and Storey, 1998). Some of these differences have been discussed in relation to the signalling processes involved in arousal from the hibernating state (MacDonald and Storey, 1998; Holden and Storey, 1998) and the adaptive biochemical regulation that is an essential part of mammalian hibernation (Mehrani and Storey, 1997). Low temperature biology studies of this type are among those in which rigorous changepoint analytical methods might be used to advantage. In particular, a random coefficients extension of the MCMC method outlined in this chapter might be used to perform a statistical assessment of the between-species differences in Arrhenius-plot changepoint behaviour. It should be noted that a particularly careful ESR spin-label study of a mitochondrial preparation obtained from cold-adapted hamsters (hibernators) and rats (non-hibernators) led to the conclusion that spin-label correlation time Arrhenius discontinuities can occur as an artefact if inappropriate approximations are used to obtain the rotational correlation time estimates (Cannon et al., 1975). In fact, Cannon and co-workers failed to find a relationship between lipid motional freedom within the mitochondrial membrane of hamster brown adipose tissue and the capacity to adapt to low temperature, casting doubt on the reliability of some earlier ESR studies.

It is established that, in general, many enzymes and other proteins respond to changing temperature by undergoing conformational transformation. Accordingly, some integral membrane proteins may exhibit a sensitivity to temperature that is mediated directly via an alteration in conformation

(Holden and Storey, 1998). On the other hand, there is a consensus of opinion that in many membrane systems the lipid matrix is involved intimately in the enzymatic response to changing temperature. The sarcoplasmic reticulum membrane (Ca^{++}-Mg^{++})-ATPase has been extensively studied in this context and the role of the membrane lipids has been debated for several decades, especially in relation to so-called protein-lipid interactions. (Marsh and Horváth, 1998, provide a review of the literature on the structure and dynamics of the lipid-protein interface.) Interest in the temperature dependent properties of this and other homothermous integral membrane systems stems from the premise that their thermodynamic behaviour is a key determinant of both structure and function and that protein-lipid interactions are a central factor. To give a tangible example, recent high resolution studies of membrane protein-lipid complexes [Pebay-Peyroula and Rosenbusch (2001) and Fyfe et al. (2001) provide an overview] have led to the intriguing hypothesis that hydrophobic surface matching provides a general mechanism for protein-lipid sorting, giving rise to a lateral partitioning/organisation of membrane lipid and protein components and consequent membrane compartmentation (Dumas et al., 1999). [The concept that hydrophobic interactions provide a passive driving force for membrane self-assembly is founded on well-established thermodynamic principles (Tanford, 1980; King and Marsh, 1987).] The question regarding the involvement of membrane lipids as a cause of sarcoplasmic reticulum (Ca^{++}-Mg^{++})-ATPase Arrhenius discontinuities remains unresolved, however (Godiksen and Jessen, 2002, and references therein). The first of the changepoint examples used in this chapter was taken from an ESR study of the rabbit hind leg muscle sarcoplasmic reticulum membrane. A brief summary of the relevant literature is given in the following section.

5.1.1.4 The Arrhenius behaviour of spin-labelled sarcoplasmic reticulum membrane

Muscle sarcoplasmic reticulum (SR) regulates muscle contraction through the release and subsequent re-uptake of Ca^{++}. [Murray et al. (2000) and Ganong (2001) provide a brief summary of the role of the sarcoplasmic reticulum in muscle contraction.] The sequestration of Ca^{++} is driven by an ATP-dependent Ca^{++} pump and brings about myofibril relaxation. The sarcoplasmic

reticulum membrane can be isolated in the form of closed vesicles (a so-called microsomal fraction, isolated using differential centrifugation methods) which retain the capacity for Ca^{++} translocation. Rabbit muscle SR microsomal preparations have been used extensively as a model system in biomembrane research, the Arrhenius behaviour of the (Ca^{++}-Mg^{++})-ATPase having received considerable attention. At first sight this focus on the temperature-dependent properties of the rabbit muscle SR might appear unjustified since the rabbit is a warm-blooded animal and thermally-induced membrane changes are not expected to occur in homothermous biological systems. Nevertheless, interest in using spin-labelling, NMR and other techniques to study temperature driven changes in this kind of membrane preparation has spanned several decades, motivated by the belief that the thermodynamic properties of these systems are key to understanding structure-function relationships. For example, it was suggested that univalent and divalent cations might interact with biological membrane lipid surfaces to bring about changes in lipid structure that are related to the crystalline to liquid-crystalline phase transition [see Träuble and Eibl (1974) and Träuble et al. (1976) for example]. Träuble and Eibl (1974) discussed the biological significance of this observation with specific reference to nerve excitation. Despite the rigour of the thermodynamic theory on which this proposition was based, the concept that lipid phase transitions are an important component of normal membrane function was speculative. Nevertheless, considerable interest in lipid phase changes, as detected by MR and other techniques, continues, with an emphasis on integral membrane enzyme regulation. [Zakim et al. (1992) have examined the relationship between the thermodynamic behaviour of the enzyme and lipid components of a variety of membrane systems and discussed their findings in relation to membrane enzyme regulation.] The central philosophy remains unchanged, namely that an understanding of the relationship between membrane structure and function is a prerequisite for the eventual understanding of the molecular basis of membrane-regulated biological processes. Controlled temperature change provides a convenient mechanism for producing experimental changes in membrane structure (crystalline to liquid-crystalline phase transitions in reconstituted membrane systems, for example), and for studying the accompanying effects on membrane function. The work of Inesi and co-workers (Inesi et al., 1973) provides an early example of this approach to

biomembrane research. They used ESR to monitor temperature-driven changes within the lipid matrix of the SR membrane, and related these to the discontinuity that occurs in the Arrhenius plot of the (Ca^{++}-Mg^{++})-ATPase hydrolytic activity. Their ESR data provided evidence of a lipid structural change at about 20°C (293°K), as indicated by a step discontinuity in the Arrhenius plot of the order parameter obtained with a spin-labelled fatty acid. This is close to the enzyme reaction-rate Arrhenius-plot changepoint temperature. They attributed this behaviour to an order-disorder lipid phase transition which, they suggested, causes a perturbation in the ATPase molecule. SR membrane lipids are, however, highly unsaturated, and the involvement of a crystalline to liquid-crystalline phase transition was consequently questioned by some researchers. A variety of alternative explanations were suggested to account for the nonlinear Arrhenius spin-label mobility behaviour, including a change in the anisotropy of the motion in parts of the lipid molecule (Davis et al., 1976), and the formation of lipid clusters (Lee et al., 1974; the term cluster was used in the sense of short-lived populations of densely packed molecules within a more freely dispersed collection of molecules). On the other hand, Hesketh et al. (1976) suggested that an interaction between the ATPase molecule and a tightly bound lipid annulus is a central determinant of the response to changing temperature and a consequent cause of the departure from classical Arrhenius behaviour. Other researchers raised the possibility that the Arrhenius discontinuity is a reflection of some structural change within the ATPase protein itself and not a property of the surrounding lipid (Dean and Tanford ,1978). Nevertheless, the original proposition (Inesi et al., 1973) that the observed behaviour is due to the presence of strong protein-lipid interactions, and that these interactions are fundamental to membrane function, persists as a general theory to the present day (Dumas et al., 1999; Fyfe et al., 2001).

Several observations cast doubt on some of the experimental data on SR membrane protein-lipid behaviour, including the demonstration that various properties of the SR membrane vesicular preparation depend on the medium used in the isolation procedure. In particular, the addition of the reducing reagent dithiothreitol to the isolation and assay media yields a greater stability in the calcium pump and a change in Arrhenius behaviour (references given in King and Quinn, 1983,

and King et al., 1987). The ESR data used in this chapter were acquired as part of a comparative study of the Arrhenius behaviour of ATPase activity and spin-label mobility in vesicles isolated in the presence and absence of dithiothreitol. A rigorous changepoint analysis is central to this kind of study.

5.1.2 Changepoint problems in pharmacology and medicine

The changepoint problem arises in a number of disciplines in addition to those discussed in the preceding sections, including pharmacology and medicine. For example, a fluorescent probe study of the Arrhenius changepoint behaviour of lipids in a plasma membrane preparation, combined with an examination of the associated Arrhenius behaviour of the plasma membrane (Na^+-K^+)-ATPase has led to the suggestion that changes in membrane lipid structure/fluidity may be an important component of the toxicity of the antitumour drug adriamycin (Deliconstantinos et al., 1987). Similarly, plasma membrane (Na^+-K^+)-ATPase and (Ca^{++}-Mg^{++})-ATPase Arrhenius behaviour has been examined in relation to the toxic side effects of some local anaesthetics [Kopeikina (1997); see Rosenberg and Alila (1982) for references to early studies of the effect of local anaesthetics on lipid membranes]. An intriguing application arises in transplant surgery. Following kidney transplantation, serum creatinine measurement is sometimes used to assess kidney function, since kidney failure is accompanied by a systematic, time-dependent decrease in the inverse of the serum creatinine concentration ($[\text{creatinine}]^{-1}$). Thus the onset of rejection is associated with a changepoint in the $[\text{creatinine}]^{-1}$ time-course data (Smith and Cook, 1980). A surgical setting provided the second of the two changepoint datasets used in this chapter. It relates to the serious complications that sometimes occur in association with major surgery due to ischaemia-reperfusion injury.

5.1.3 Intestinal ischaemia-reperfusion injury and hepatic energy failure

Intestinal ischaemia-reperfusion injury is a life-threatening state that is associated with a number of conditions, including mesenteric arterial occlusion and midgut volvulus [twisting of the midgut around its axis causing obstruction and vascular compromise (Spitz, 1995)]. The gut appears to be

particularly sensitive to ischaemia-reperfusion injury, with the consequence that this condition can also occur as a secondary response to a period of haemorrhagic shock or prolonged hypotension caused by a variety of primary insults, including burns, mechanical trauma and major surgery, including but not necessarily intestinal surgery (Deitch, 1992; Pastores et al., 1996; Stechmiller et al., 1997). In each of these conditions the restoration of arterial flow and tissue perfusion is a priority. This is achieved by surgical correction (midgut volvulus) , revascularisation (mesenteric arterial occlusion), or resuscitation (shock). Unfortunately, multiple organ failure is a complication that sometimes arises following reperfusion, hence the term intestinal ischaemia-reperfusion injury. It is, perhaps, paradoxical that reperfusion exacerbates the damage that occurs during the hypoxic phase with such devastating consequences, although progress has been made towards understanding the underlying pathophysiology and biochemistry (Welbourn et al., 1991; Deitch, 1992; Pastores et al., 1996; Stechmiller et al., 1997). Circulating bacteria and endotoxins[1] with subsequent macrophage[1] activation, neutrophil[1] release and cytokine[1] production all feature as essential components of the response. Remote organ endothelial cell damage, caused by oxygen free-radicals (via xanthine oxidase activation) and neutrophil accumulation, is a major cause of remote organ dysfunction and subsequent failure (Welbourn et al., 1991; Deitch, 1992; Pastores et al., 1996). Although it is established that mucosal cell damage and consequent impaired gut barrier function is a central component of multiple organ failure, the causal relationship between intestinal damage and remote organ failure remains speculative (Deitch, 1992; Pastores et al., 1996; Spechmiller et al., 1997).

---

[1] Endotoxins are toxins, contained within bacteria, that are released when bacterial cells die or become degraded. Macrophages are a component of the immune system which function by engulfing and destroying foreign cells, including bacteria. Neutrophils are a type of white blood cell. Cytokines are signalling proteins produced by activated immune cells in response to infection or injury.

The majority of midgut volvulus cases are encountered in newborn infants, this providing the motivation for several studies into the condition at The Institute of Child Health, UCL. The liver is among the organs that are commonly affected in intestinal ischaemia-reperfusion injury, and the data shown in Fig. 5.5 were acquired as part of a comprehensive experimental study into the effects of intestinal ischaemia and subsequent reperfusion on liver metabolism in the rat (Vejchapipat et al., 2001, 2002). Experimental intestinal ischaemia-reperfusion was produced surgically by clamping and subsequently unclamping the superior mesenteric artery. Liver energy failure was found to occur within a short period following reperfusion, as indicated by an increase in the inorganic phosphate (P_i) to adenosine triphosphate (ATP) ratio (PAR) and a decrease in pH (Fig. 5.5). The latter is caused by a switch to anaerobic metabolism and consequent lactic acid production. Intestinal ischaemia failed to produce hepatic energy failure in the absence of reperfusion (data not shown). Among the aims of the study was an examination of the temporal relationship between the pH and PAR changes. In particular, intracellular acidosis might be implicated as a causative factor in the liver failure associated with intestinal ischaemia-reperfusion injury, as opposed to a passive response, if the pH change precedes the PAR response. Consequently, a rigorous analysis is required to determine the pH and PAR changepoints and the difference between them. Among the other components of the study was an examination of the therapeutic effects of hypothermia and perfluorocarbons on the liver response to intestinal ischaemia-reperfusion injury (data not shown). Formal changepoint models are required to assess some aspects of these treatment effects.

5.1.4 Electron spin resonance spectroscopy in biomembrane research

Electron spin resonance (ESR) spectroscopy (also referred to as electron paramagnetic resonance spectroscopy and abbreviated EPR) is a magnetic resonance technique used to study paramagnetic centres arising from unpaired electrons. The first of the two changepoint problems presented in this chapter arose in an ESR spin-label study of the muscle sarcoplasmic reticulum membrane. The term spin label is used to describe any free radical that is used as a molecular environmental probe. An important class of spin labels is based on the nitroxide radical, which is a stable organic free-

radical species in which the free electron is localised within a nitrogen-oxygen bond (see insert to Fig. 5.1). Nitroxide spin labels are used extensively in biomembrane research due to their sensitivity to both motion and orientation in an external magnetic field, as reflected in the ESR spectrum (Berliner, 1976). The ESR data used in this chapter were acquired in a spin-label study of the temperature dependence of the molecular mobility (fluidity) of the sarcoplasmic reticulum membrane. Arrhenius plot changepoints were the focus of the study. These were based on order parameter measurements made using a nitroxide-labelled fatty acid incorporated into the membrane.

A brief overview of the motional sensitivity of nitroxide spin-labels is appropriate. The nitrogen nucleus possesses spin angular momentum with an associated quantum number (I) equal to one[1]. An interaction of the paramagnetic electron with the magnetic field associated with the magnetic moment of the neighbouring nitrogen nucleus produces hyperfine structure in the ESR spectrum consisting of three spectral lines (number of lines = $2I$+1). The positions of the lines and the magnitude of the hyperfine splittings depend on the relative orientation of the nitroxide molecular axes (see insert to Fig. 5.1) with respect to the external magnetic field. This gives rise to so-called spectral anisotropy (orientation dependence in the ESR spectrum) that is the basis of much spin-labelling work. Spectral anisotropy is usually specified in terms of the hyperfine constants, A_{xx}, A_{yy} and A_{zz}, associated with each of the three principal molecular axes ($A_{xx} = \sim A_{yy} < A_{zz}$ in the co-ordinate system of Fig. 5.1), together with the g-values that determine the absolute position of the resonances. Anisotropy in the hyperfine splitting is the main cause of the motional sensitivity of the spectrum, and rotational reorientation is the dominant form of motion within the lipid environment of a typical biological membrane. A nitroxide spin label undergoing rapid and isotropic rotational motion gives rise to a simple three-line spectrum due to rotational averaging, while increasing immobilisation leads to a loss of averaging and a resulting intermediate spectrum, i.e., intermediate

---

[1] The electron possesses spin angular momentum with an associated quantum number of 1/2, indicating that the component of spin angular momentum in a given direction takes on two values.

Fig. 5.1. **Spin-labelled sarcoplasmic reticulum membrane ESR spectrum**

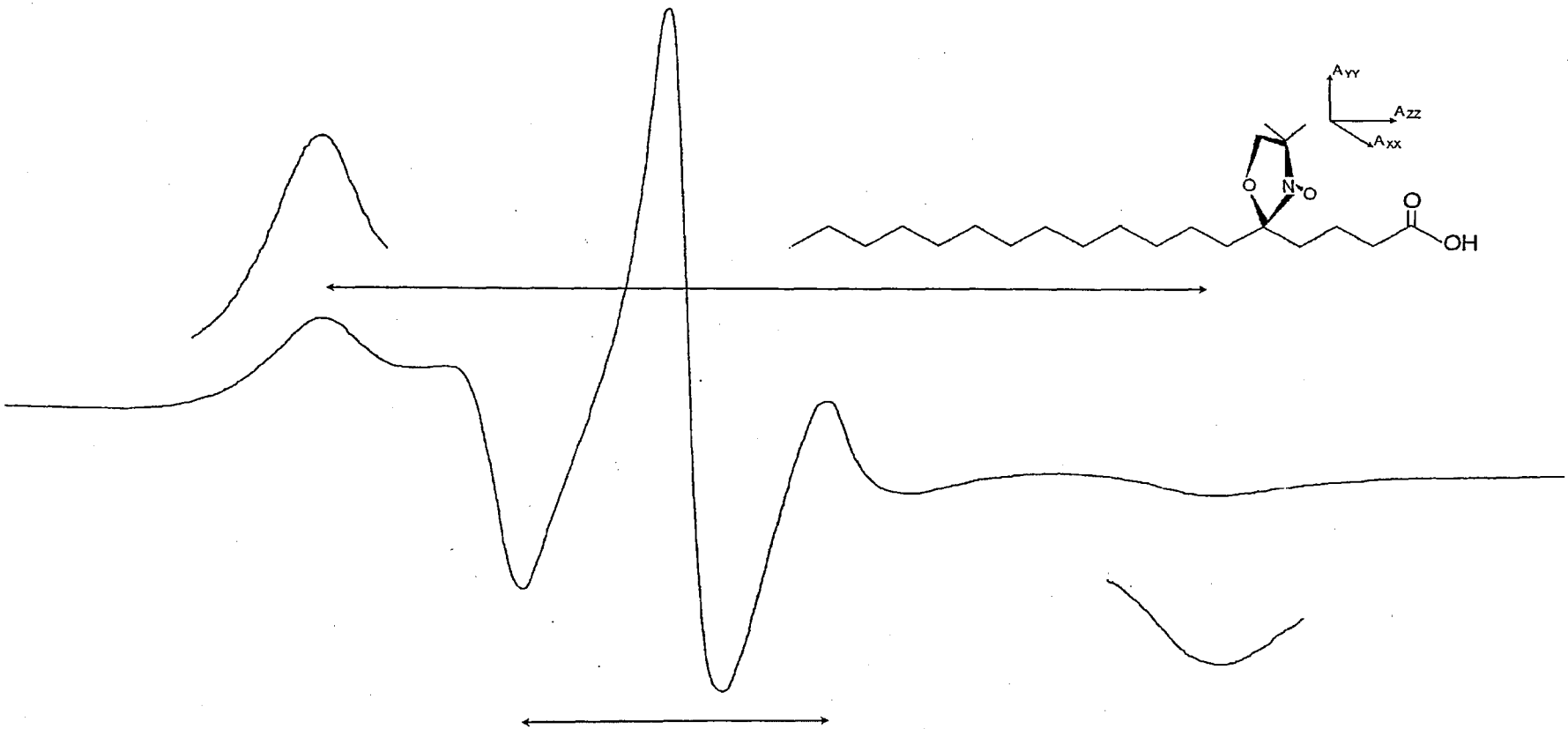


Figure legend on the following page

Fig. 5.1. continued

An ESR spectrum acquired at 24.8°C from a microsomal sarcoplasmic reticulum membrane preparation isolated in the absence of dithiothreitol and labelled with 3-doxylstearic acid. The insert provides a schematic showing the N-oxyl-4',4'-dimethyloxazolidine stearic acid spin label (n=3, m=12 in the nomenclature of Hidalgo et al., 1976; abbreviated 3-doxylstearic acid) together with the nitroxide principal molecular axes. The parallel and perpendicular hyperfine splittings are indicated by the two arrows (shorter arrow, $2A_{min}$, $A_{min} \simeq A_{\perp}$; longer arrow, $2A_{max}$, $A_{max} \simeq A_{\parallel}$; see Griffith and Jost (1976) pages 473 to 484 for additional information). For clarity, the outer hyperfine extrema are reproduced on an expanded scale above/below the spectrum.

to the fully averaged and the completely immobilised pseudo-powder spectrum. This intermediate condition arises when a fatty acid or lipid nitroxide spin label is incorporated into a lipid dispersion or membrane preparation. The nitroxide molecules become randomly oriented with respect to the external magnetic field, leading to a composite spectrum, an example of which is given in Fig. 5.1. Furthermore, although each of the nitroxide molecules undergoes rapid rotational motion, reorientation of the long molecular axis is restricted leading to a reduced averaging of the hyperfine splittings. Thus $A_{\parallel}$ (parallel component of the time-averaged hyperfine tensor) becomes smaller than the principal hyperfine value A_{zz}, while $A_{\perp}$ (perpendicular component of the time-averaged hyperfine tensor) becomes greater than the principal value A_{xx}. $A_{\parallel}$ and $A_{\perp}$ can be estimated from the ESR spectral turning points [see Fig. 5.1; $A_{\parallel} \simeq A_{max}$ and $A_{\perp} \simeq A_{min}$; Griffith and Jost (1976) pages 473 to 484 provide additional information] and the magnitude of the motional averaging determined using the order parameter (S), which is a function of A_{max} A_{min}, A_{zz}, A_{yy} and A_{xx} (Griffith and Jost, 1976). The order parameter is, therefore, related to the angular amplitude of the anisotropic molecular motion (Griffith and Jost, 1976; McConnell, 1976).

5.2 Experimental methods

5.2.1 Sarcoplasmic reticulum spin-labelling study

Sarcoplasmic reticulum vesicles were prepared from the white muscle of rabbit hind leg according to the method of Robinson et al. (1972) and Warren et al. (1974), excepting that N-2-hydroxyethylpiperazine-N'-2-ethane-sulphonic acid was substituted for the histidine buffer. An N-oxyl-4',4'-dimethyloxazolidine (doxyl) stearic acid derivative [n=3, m=12, in the nomenclature of Hidalgo et al. (1976, page 4226); structure given in the insert to Fig. 5.1] was incorporated into the membrane as described by Hidalgo et al. (1976). ESR spectra (conventional first harmonic absorption spectra) were acquired using a Varian E-104 spectrometer. Order parameter estimates were calculated with polarity correction from the first-derivative spectral turning points, as outlined by Griffith and Jost (1976, pages 481 to 484, Method II), using the spin-labelled fatty acid hyperfine tensor data given in Appendix II of Berliner (1976).

5.2.2 Intestinal ischaemia-reperfusion study

The pH-PAR data used in this chapter were acquired within the RCS Unit of Biophysics at the Institute of Child Health, UCL, by Paisarn Vejchapipat, while working for the degree of Doctor of Philosophy. Full experimental details, including a description of the surgical procedures and MR spectroscopic data acquisition method, are provided by Vejchapipat (2001) and Vejchapipat et al. (2001, 2002).

5.3 Statistical methods

The following models were used in the changepoint analyses outlined in this chapter.

Model 1. Order parameter (S) Arrhenius plot analysis with continuous changepoint (k_S):

$$\begin{aligned} &\ln(S)_i \sim N(\mu_i, \sigma^2) \\ &\mu_i = \beta_0 + \beta_{11}(1/T - k_S)_i,\ (1/T)_i < k_S;\ \mu_i = \beta_0 + \beta_{12}(1/T - k_S)_i,\ (1/T)_i \geq k_S. \end{aligned} \qquad [5.1]$$

Model 2. pH-PAR data analysis. Full model with discrete changepoints (k_l, l=pH, PAR):

$$\begin{aligned} &y_{i(pH)} \sim N(\hat{y}_{i(pH)}, \tau_{i(pH)}) \\ &\hat{y}_{i(pH)} = \beta_{0(pH)}, i = 1,2,...,k'_{pH};\ \ \hat{y}_{i(pH)} = \beta_{0(pH)} + \beta_{12(pH)}(t_i - k'_{pH}), i = k'_{pH}+1, k'_{pH}+2,...,n \\ &\tau_{i(pH)} = \tau_{1(pH)}, i = 1,2,..,k'_{pH};\ \tau_{i(pH)} = \tau_{2(pH)}, i = k'_{pH}+1, k'_{pH}+2,...,n \\ &k'_{pH} \in \{1,2,...,n\}, k_{pH} = t(k'_{pH}) \end{aligned} \qquad [5.2]$$

$$\begin{aligned} &y_{i(par)} \sim N(\hat{y}_{i(par)}, \tau_{i(par)}) \\ &\hat{y}_{i(par)} = \beta_{0(par)} + \beta_{11(par)}(t_i - k'_{par}), i = 1,2,...,k'_{par}, \\ &\hat{y}_{i(par)} = \beta_{0(par)} + \beta_{12(par)}(t_i - k'_{par}), i = k'_{par}+1, k'_{par}+2,...,n \\ &\tau_{i(par)} = \tau_{1(par)}, i = 1,2,...,k'_{par}, \tau_{i(par)} = \tau_{2(par)}, i = k'_{par}+1, k'_{par}+2,...,n \\ &k'_{par} \in \{1,2,...,n\}, k_{par} = t(k'_{par}) \end{aligned}$$

where k'_l is the changepoint occasion, l=pH, PAR, and $t(k'_l)$ is time at the changepoint.

Model 3. pH-PAR data analysis. Full model with continuous changepoints (k_l, l=pH, PAR):

$$
\begin{aligned}
&y_{i(pH)} \sim N(\hat{y}_{i(pH)}, \sigma^2_{i(pH)}), \quad y_{i(par)} \sim N(\hat{y}_{i(par)}, \sigma^2_{i(par)}) \\
&\hat{y}_{i(pH)} = \beta_{0(pH)}, t_i < k_{pH} \\
&\hat{y}_{i(pH)} = \beta_{0(pH)} + \beta_{12(pH)}(t_i - k_{pH}), t_i \geq k_{pH} \\
&\hat{y}_{i(par)} = \beta_{0(par)} + \beta_{11(par)}(t_i - k_{par}), t_i < k_{par} \\
&\hat{y}_{i(par)} = \beta_{0(par)} + \beta_{12(par)}(t_i - k_{par}), t_i \geq k_{par} \\
&\sigma^2_{i(pH)} = \sigma^2_{1(pH)}, t_i < k_{pH}, \sigma^2_{i(pH)} = \sigma^2_{2(pH)}, t_i \geq k_{pH} \\
&\sigma^2_{i(par)} = \sigma^2_{1(par)}, t_i < k_{par}, \sigma^2_{i(par)} = \sigma^2_{2(par)}, t_i \geq k_{par}.
\end{aligned}
\qquad [5.3]
$$

Reduced models were derived from each of the two pH-PAR full models by imposing the constraint $\beta_{11(par)} = 0$. The continuous changepoint Metropolis simulation analyses were performed using a transformation of the form

$$
\begin{aligned}
&\varepsilon'_l = \frac{e^{\varepsilon_l}}{1+e^{\varepsilon_l}}, \quad -\infty < \varepsilon_l < +\infty \\
&k_l = a_l + (b_l - a_l)\varepsilon'_l
\end{aligned}
\qquad [5.4]
$$

where ε_l is the changepoint parameter in Metropolis space for the lth variable, $l = pH, PAR, S$. k_l, $l = pH, PAR$, is the changepoint in absolute time (pH-PAR analysis), and k_S is the changepoint in reciprocal absolute temperature (order parameter analysis). a_l and b_l are the lower and upper limit of the changepoint prior distribution, which were set equal to the minimum and maximum of the observation range, respectively (l = pH, PAR: $a_l = t_{\min} = 54$s, $b_l = t_{\max} = 186$s; $l = S$: $a_l = 1000/T_{\max} = 3.2373$ K^{-1}, $b_l = 1000/T_{\min} = 3.6630$ K^{-1}).

The Metropolis simulations were performed using a normal likelihood, as described in Chapter 3. Uniform prior distributions were used for all parameters except the variance and changepoint parameters. The prior $p(\sigma) \propto \sigma^{-1}$ was used for each of the variances, while a bounded uniform prior distribution, $p(k_l) \sim \text{Unif}(a_l, b_l)$, was used for each changepoint parameter. The change of variables rule yields a prior of the form $p(\varepsilon_l) \sim (b_l - a_l)e^{\varepsilon_l} / (1+e^{\varepsilon_l})^2$ for each changepoint parameter in Metropolis space. A total of 50,000 samples was stored using a runtime thinning ratio of 1:100 in the ESR analysis and 1:200 in the pH-PAR analysis. Additional details regarding the Metropolis

simulation analyses are given in Chapter 2 (Section 2.2). Each analysis was performed using a transition kernel obtained by performing a preliminary training run with transition kernel updating (Section 2.2.1). Three parallel chains were generated in each case. Convergence diagnostic tests were carried out using CODA (Sections 2.2.2).

Gibbs sampling was performed using WinBUGS (downloaded from http://www/mrc-bsu.cam.ac.uk/bugs. The Gibbs sampler results shown in this chapter were all generated using discrete changepoint models. Priors of the form $N(0.0, 10^{-6})$ were used for all parameters, except the precision and changepoint parameters. A gamma(0.001, 0.001) prior was adopted for each of the precision parameters, while a discrete uniform prior distribution on K_n, $K_n = \{1,2,..,n\}$, was used for each of the changepoint parameters, where n is the number of design points. Three parallel chains (100,000 samples per chain) were generated and convergence diagnostic tests carried out using CODA (Section 2.2.2).

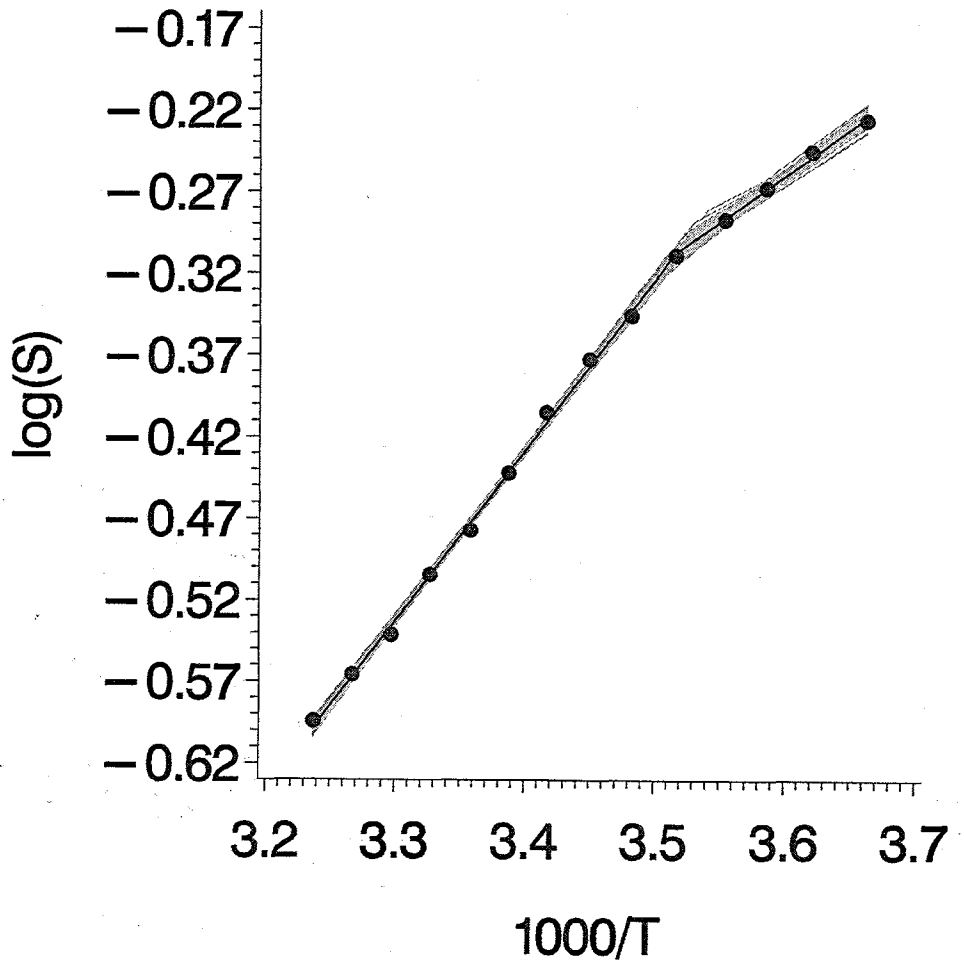
5.4 Results

5.4.1 ESR spin-labelled sarcoplasmic reticulum membrane Arrhenius study

ESR spectra were acquired from a fatty-acid spin-labelled sarcoplasmic reticulum membrane microsomal preparation at intervals of temperature in the range 0.1 to 38.4°C. The spectrum acquired at 24.8°C (298.0°K) is shown in Fig. 5.1 for illustrative purposes; it is typical of the kind of spectrum obtained from a partially immobilised nitroxide. Order parameter estimates were derived from the spectra as described in Section 5.2.1 and used in a subsequent Metropolis MCMC simulation analysis based on the model given in [5.1]. The simulation was performed as described in Sections 5.3. The order parameter data are displayed in the form of an Arrhenius plot in Fig 5.2.

Formal convergence testing was preceded by a visual inspection of the MCMC output, as displayed in the form of overlaid chain trace plots. These plots, which were generated for all model parameters, are shown in Fig. 5.3, together with the log(posterior density) overlaid trace plot. Various 2D trajectory plots were also examined. The latter are shown in Fig. 5.4 as primary model

Fig. 5.2. **Spin-labelled sarcoplasmic reticulum membrane order parameter Arrhenius plot**



3-doxylstearic acid-labelled sarcoplasmic reticulum membrane order parameter (S) data displayed in the form of an Arrhenius plot. The observed order parameter data are displayed together with the posterior median piecewise-linear Arrhenius profile (black lines). Also shown is a set of 100 random profiles drawn from the posterior distribution (grey lines). The parameter medians are (95% posterior intervals in parentheses): k_S = 3.515 (3.493, 3.536), β_0 = -0.3090 (-0.3290, -0.2896), β_{11} = 1.042 (1.010, 1.077), β_{12} = 0.5600 (0.4633, 0.6391).

Fig. 5.3. **Order parameter Arrhenius-plot changepoint analysis. Overlaid chain plots**

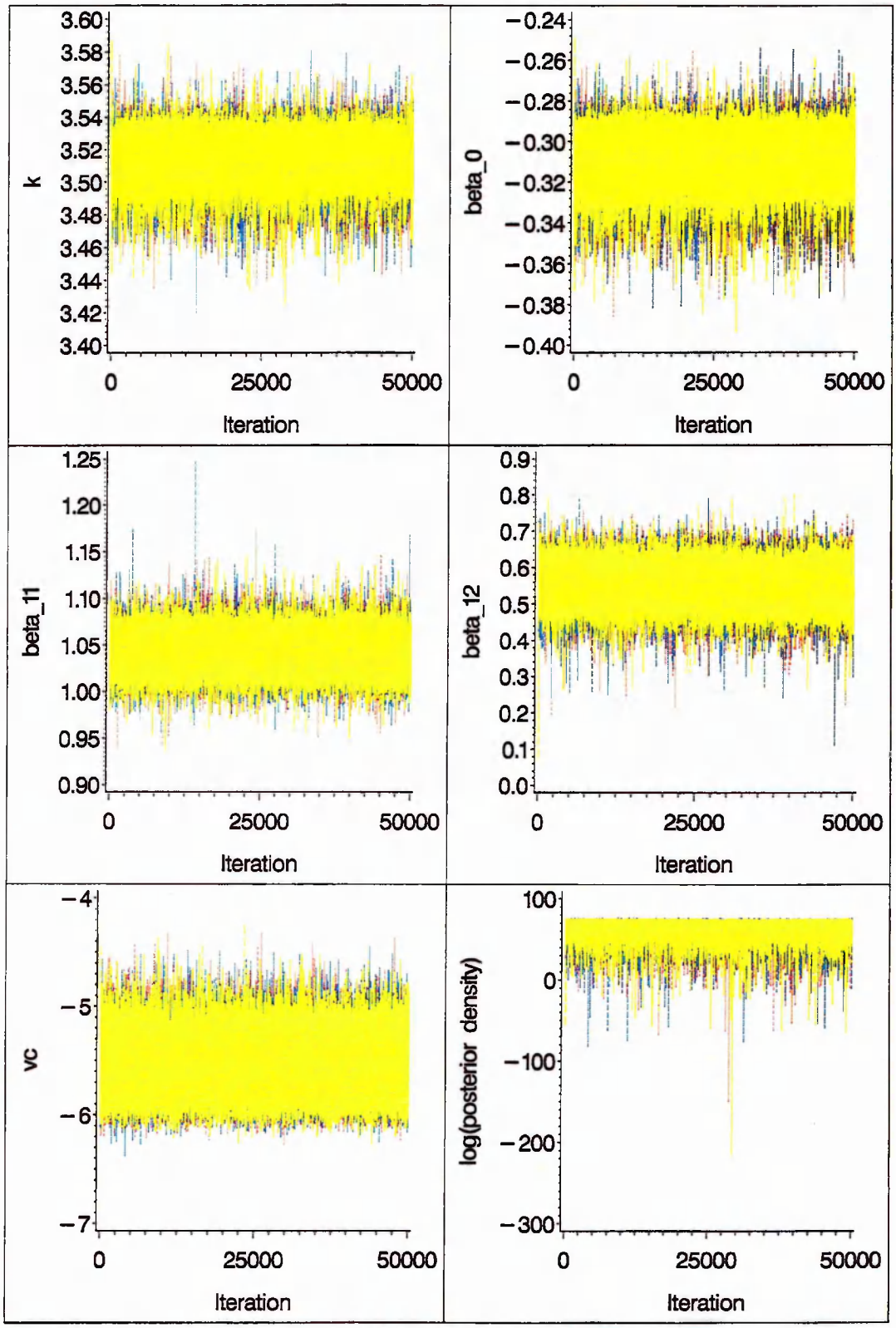


Figure legend on page 194.

Fig. 5.4. **Order parameter Arrhenius-plot changepoint analysis. Overlaid chain trajectory plots and histogram**

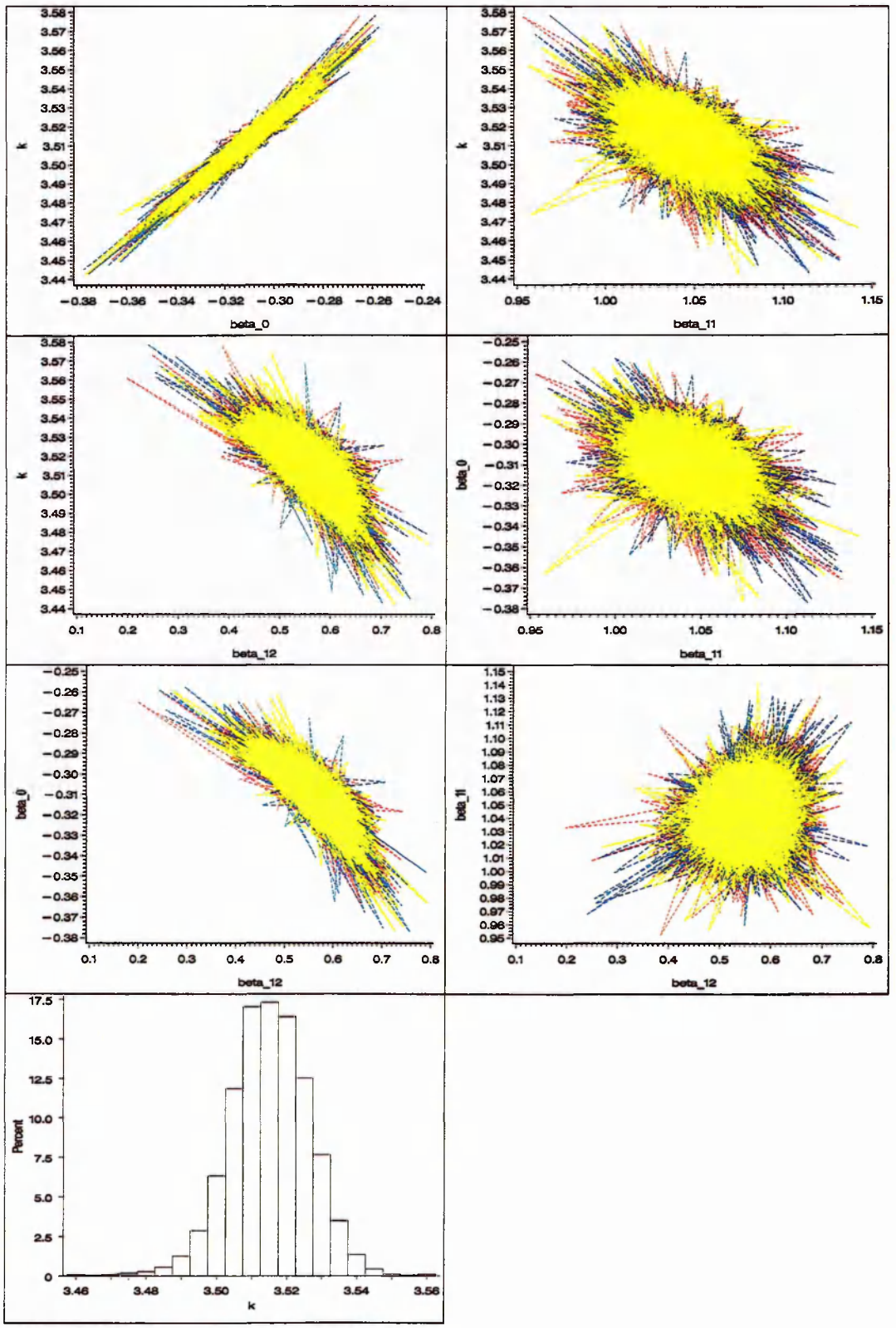


Figure legend on the following page.

parameter trajectories, as opposed to the Metropolis space trajectories, to allow an immediate assessment in relation to the underlying Arrhenius behaviour displayed in Fig. 5.2. The overlaid trace plots and trajectory plots both indicate a good coverage of parameter space. Furthermore, a satisfactory overlap of the three parallel-chain trace plots is obtained for every parameter. Similarly, the trajectory plots indicate no major problems in the form of substantial chain separation, although they do show a few instances of individual chain migration into isolated regions of parameter space. Nevertheless, there is no indication of prolonged residence within these regions, suggesting that these occur due to occasional excursions into regions of low probability. These low probability excursions are not expected to be reproducible across the set of parallel chains. A substantial correlation among the parameters is evident (Fig. 5.4). This is not unexpected, given the form of the model. The inverse-temperature changepoint parameter (k_S) and intercept (β_0) are, for example, highly correlated, but this does not cause significant convergence problems due to the low dimensionality of the Metropolis space. Finally, histograms were produced for each parameter, including the changepoint parameter, the latter of which is included in Fig. 5.4. All have an acceptable shape and there are no instances in which the tails reveal a sampling problem. In summary, the overlaid trace plots, trajectory plots and histograms all appear satisfactory; the trace

---

Fig 5.3. Overlaid chain plots for each of the model parameters, including the residual error variance (displayed in the Metropolis-space metric) and the log(posterior density). Two chains (yellow and blue) were started using overdispersed values for all model parameters while the third chain (red) was started at a position close to the expected posterior median. A run-time thinning ratio of 1:100 was applied giving a post thinning chain length of 50,000 iterations.

Fig 5.4. Overlaid chain trajectory plots for each pair of regression parameters. Two chains (yellow and blue) were started using overdispersed values for all model parameters while the third chain (red) was started at a position close to the expected posterior median. The trajectories are displayed after additional thinning to show the separate paths with greater clarity. Also shown is the changepoint parameter histogram.

and trajectory plots show no signs of convergence failure.

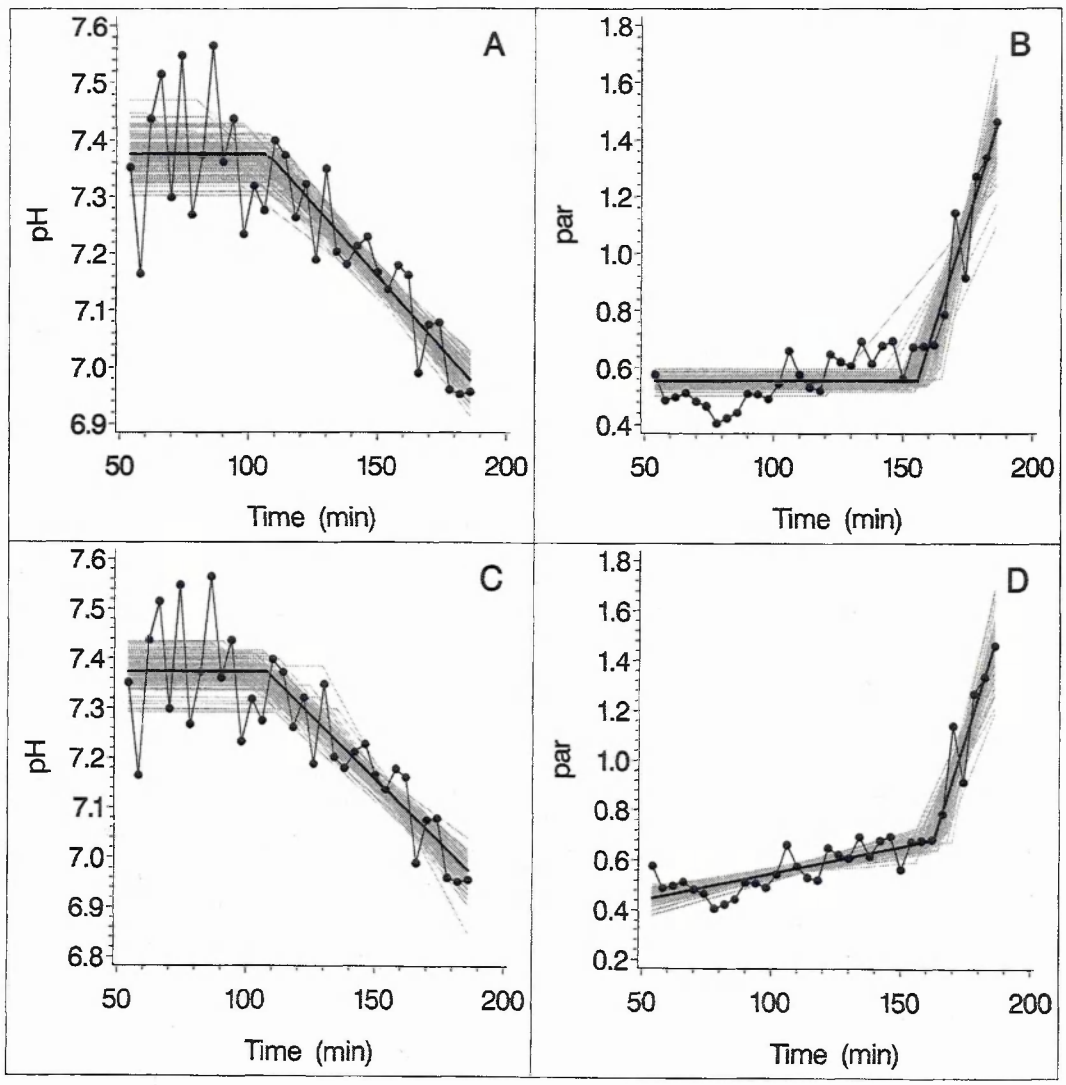
A more formal assessment of convergence was performed using the Gelman-Rubin shrink factor, the Geweke time-series *Z*-score and the Raftery-Lewis procedure (see Sections 1.7 and 2.2.2 for details). Satisfactory results were obtained for all parameters. The majority of the Geweke *Z*-score points lie within the range $-2<Z<2$, and each of the Gelman-Rubin shrink-factor plots indicates a value close to unity towards the end of each set of chains. The Raftery-Lewis results indicate that each chain is of sufficient length to yield all 0.025 and 0.975 quantile estimates with an accuracy of at least +/-0.0015 with 0.95 probability.

Fig. 5.2 shows the Arrhenius plot derived from the parameter median estimates, together with a cluster of profiles obtained by sampling the posterior distribution, and superimposed observed order parameter data. The Arrhenius plot changepoint 95% posterior interval has a temperature range of 9.7 to 13.2 °C (median 11.4 °C) indicating a well-defined Arrhenius-plot discontinuity. The complete set of posterior medians is listed in the legend to Fig. 5.2 together with the 95% posterior intervals.

5.4.2 Liver response to intestinal ischaemia-reperfusion injury

Rat liver pH and PAR (inorganic phosphate to adenosine triphosphate ratio) data were acquired, starting 50min after the production of intestinal ischaemia, and continued during a subsequent reperfusion phase, which was initiated at 90 minutes, i.e., 40min into the data acquisition period. The data collected from one of six animals are shown in Fig. 5.5. This pair of pH/PAR profiles was used in an assessment of various changepoint models, each of which was evaluated in relation to its behaviour within the MCMC setting. In each case a simultaneous pH/PAR changepoint MCMC simulation analysis was performed using the Metropolis algorithm. In some cases the analysis was repeated using the Gibbs sampler. Among the models investigated were two intersecting biphasic-linear models, both of which assume the absence of a discontinuity in the observed variable at the changepoint. Details are provided in Section 5.3. One of the two models, which is referred to as the

Fig. 5.5. **Liver metabolite response to intestinal ischaemia-reperfusion injury**



Plots showing the liver pH and PAR (inorganic phosphate to ATP ratio) responses to intestinal ischaemia-reperfusion. Data were acquired starting 50 minutes after the initiation of intestinal ischaemia, and continued during a subsequent reperfusion phase, which was started at 90 minutes, i.e., 40 minutes into the data acquisition period. The reduced model$^{(A \text{ and } B)}$ and full model$^{(C \text{and } D)}$ posterior median time-dependencies are shown (black lines) superimposed on the experimental observations. Also shown is a set of 100 random profiles drawn from the posterior distribution (grey lines).

reduced model, constrains PAR to be constant prior to its changepoint ([5.2] and [5.3] with the constraint $\beta_{11(par)} = 0$), while this constraint is lifted in the second model ([5.2] and [5.3], referred to as the full model). Both models constrain pH to be constant prior to its changepoint. Gibbs sampling was performed using a discrete changepoint variant of the two models ([5.2] and its reduced form), while the Metropolis-based simulations were performed using a continuous changepoint implementation ([5.3] and its reduced form).

Metropolis chain overlaid trace plots were generated for all model parameters in each of the simulation analyses (shown for the full model in Fig. 5.6 and for the reduced model in Fig. 5.7). The log(posterior density) overlaid trace plot is included in each of these figures. Various 2D trajectory plots were also produced (Figs. 5.8 and 5.9). These trajectory plots are shown in the primary model parameter space (i.e., pre-transformation) as opposed to the Metropolis parameter space. Thus the trajectories displayed in Figs. 5.8 and 5.9 relate directly to the underlying time-dependent pH and PAR behaviour shown in Fig. 5.5. A good overlap of the three parallel chains is achieved for every parameter with no sign of convergence failure. The truncated appearance of several of the PAR full model trajectory plots shown in Fig.5.8 [(k versus $\beta_0)_{par}$, (k versus $\beta_{11})_{par}$ and (k versus $\beta_{12})_{par}$] is among the striking features of this graphical output. This is attributable to an abrupt change in likelihood in the region of k_{par}=170, which is entirely consistent with the observed PAR time-course data (Fig. 5.5). The trajectory plots indicate no major problems in the form of substantial chain separation. An occasional migration into an isolated region of parameter space occurs and is commented on in Section 5.4.1. The pH and PAR changepoints and PAR-pH changepoint difference are the main focus of the analysis. Histograms for these three parameters are shown in Fig. 5.10. Histograms were also generated for the remaining parameters. All have an acceptable shape and show no sign of sampling problems. In summary, the histograms, overlaid chain trace plots and trajectory plots all appear satisfactory. In particular, the trace and trajectory plots show no signs of convergence failure. Gelman-Rubin shrink factor and Geweke time-series Z-score diagnostic tests were performed on a subset of parameters, and yielded satisfactory results.

Fig. 5.6. **Full model pH-PAR changepoint analysis. Overlaid chain plots**

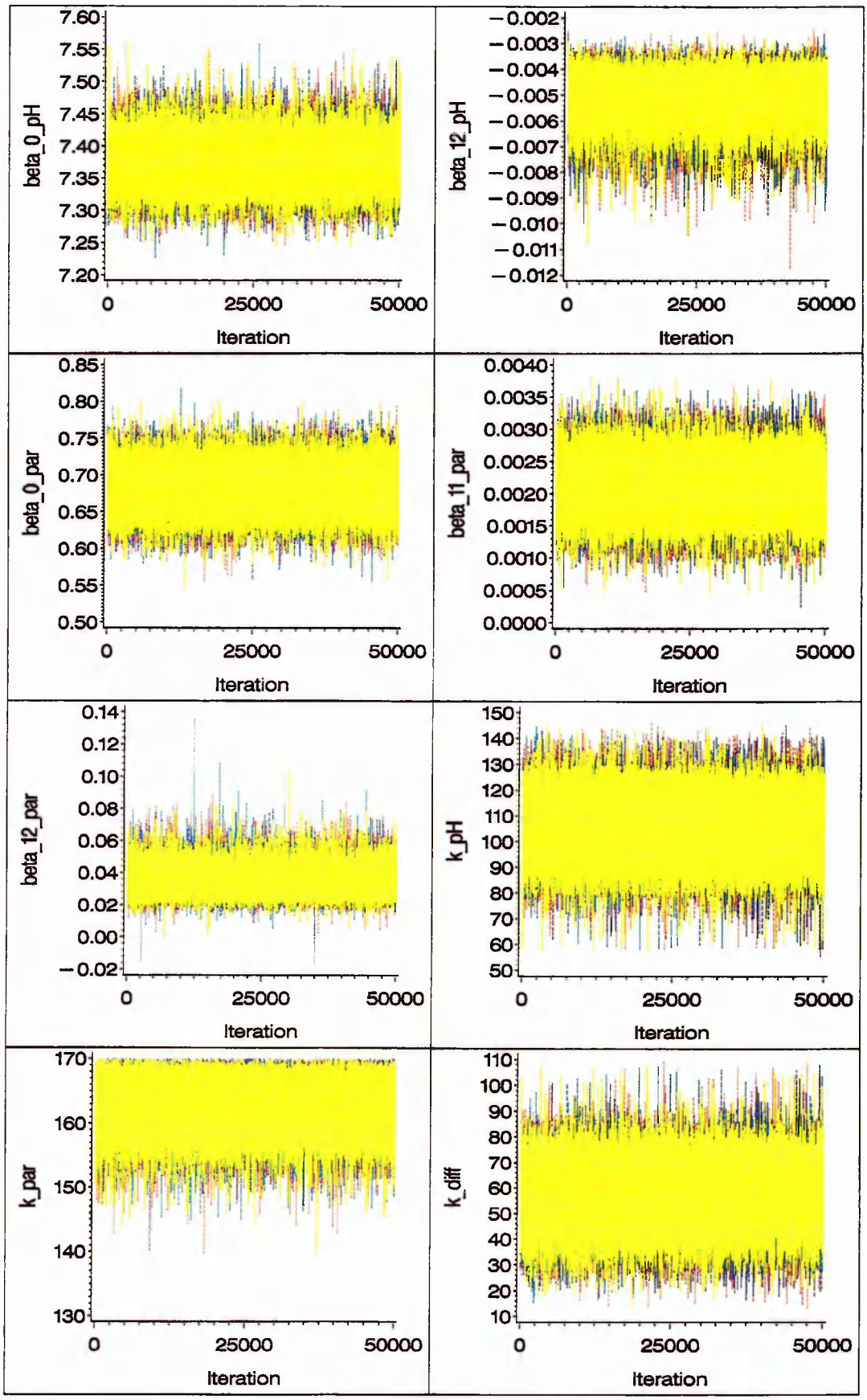
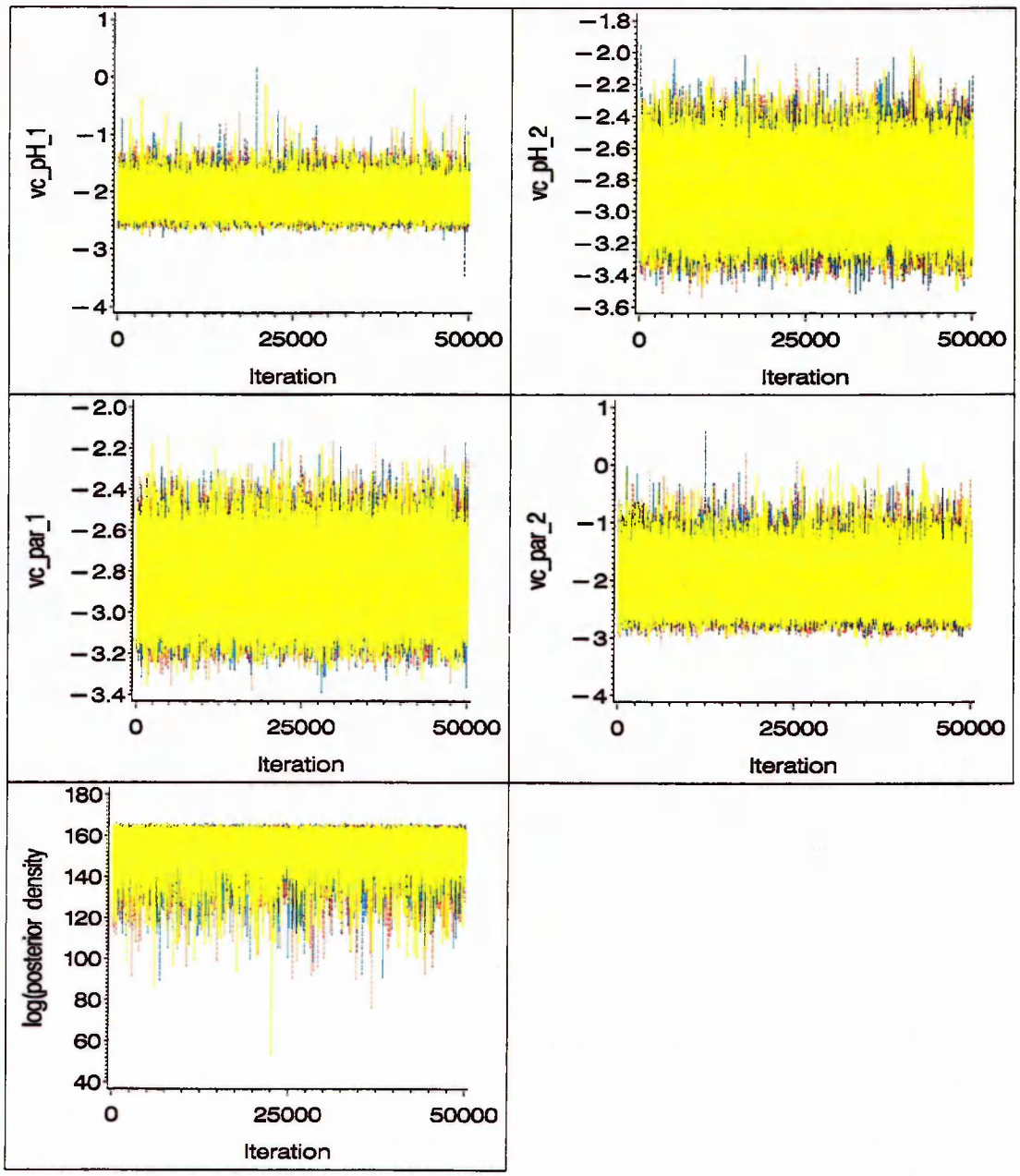


Fig. 5.6. Continued.



Overlaid chain plots are shown for each of the model parameters, including the two residual error variances, together with plots for the change-point time difference ($k_{diff} = k_{par} - k_{pH}$) and the log(posterior density). Two chains (yellow and blue) were started using overdispersed values for all model parameters while the third chain (red) was started at a position close to the expected posterior median. A run-time thinning ratio of 1:200 was applied giving a post thinning chain length of 50,000 iterations.

Fig. 5.7. **Reduced model pH-PAR changepoint analysis. Overlaid chain plots**

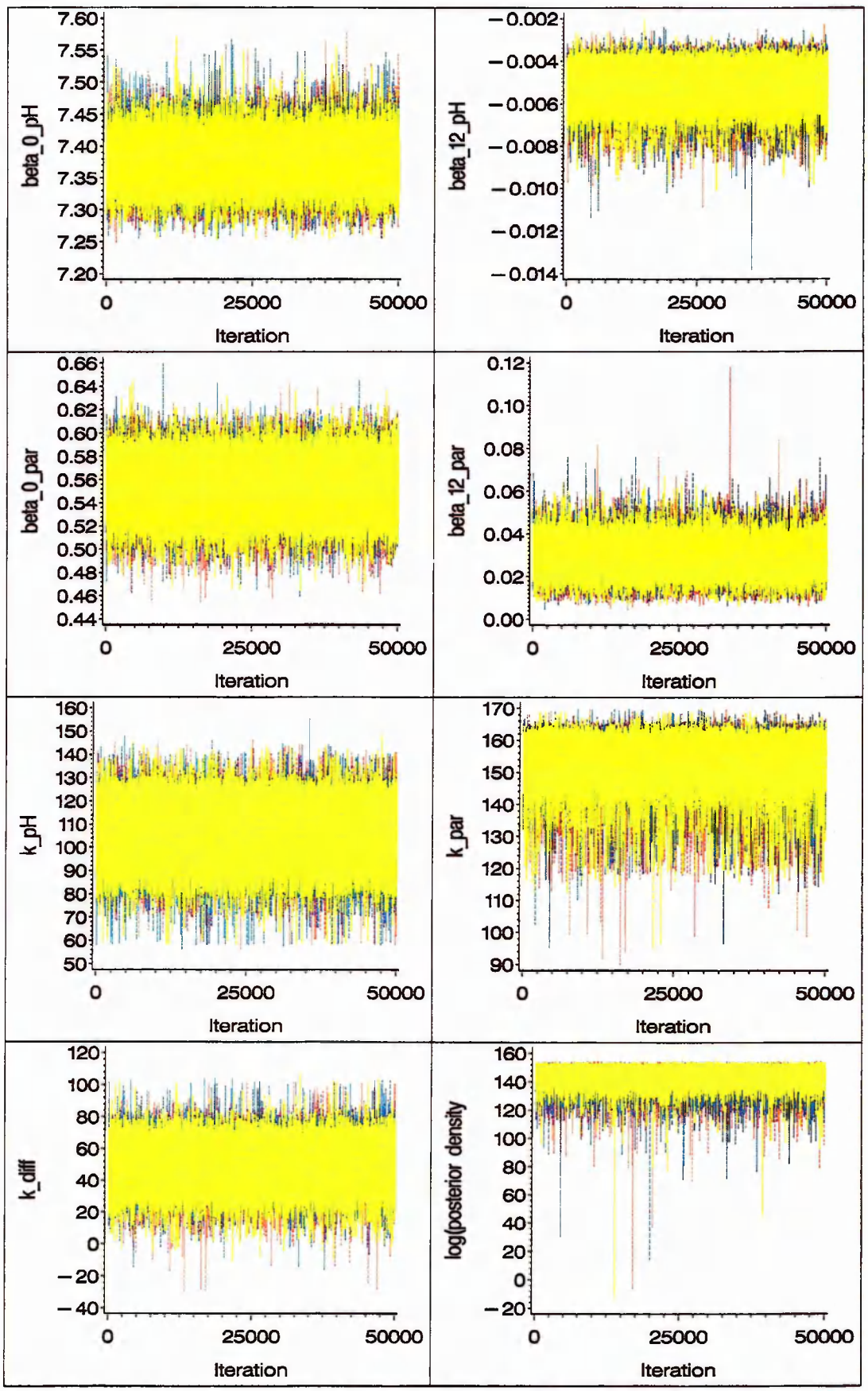
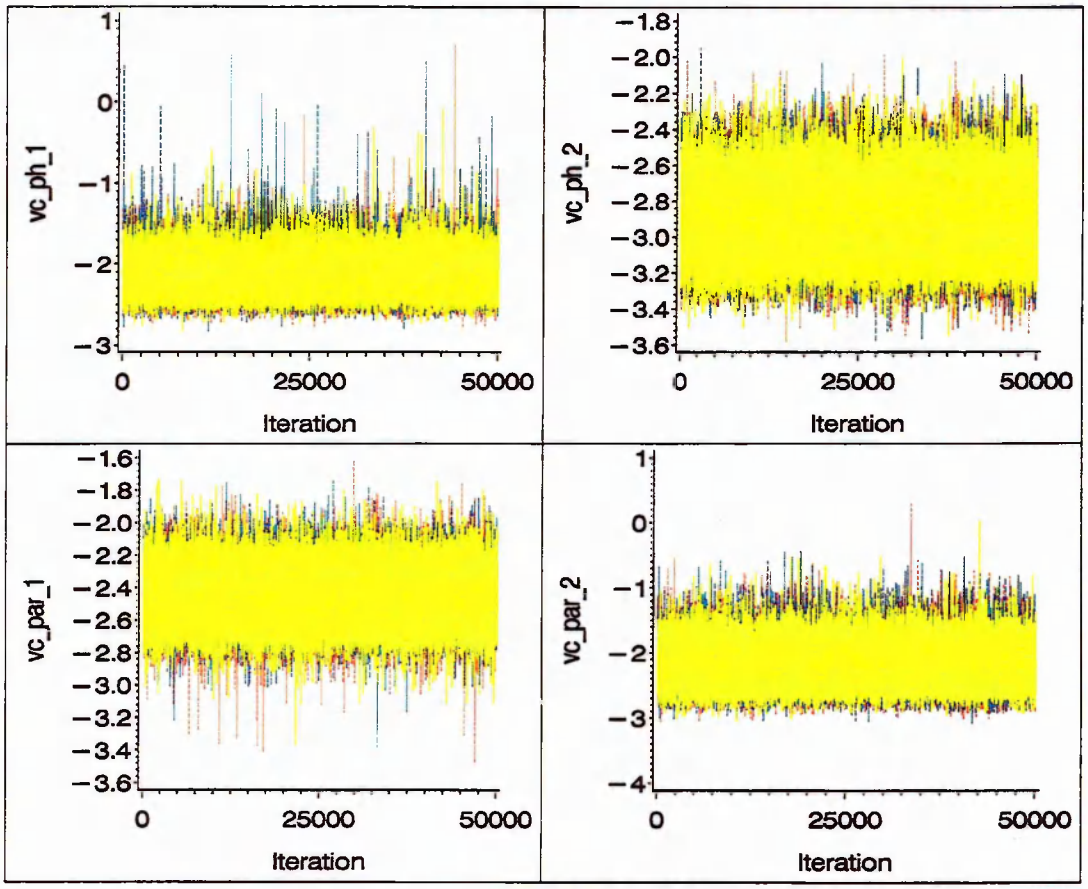


Fig. 5.7. Continued.



Overlaid chain plots are shown for each of the model parameters, including the two residual error variances, together with plots for the change-point time difference ($k_{diff} = k_{par} - k_{pH}$) and the log(posterior density). Two chains (yellow and blue) were started using overdispersed values for all model parameters while the third chain (red) was started at a position close to the expected posterior median. A run-time thinning ratio of 1:200 was applied giving a post thinning chain length of 50,000 iterations.

Fig. 5.8. **Full model pH-PAR changepoint analysis. Overlaid chain trajectory plots**

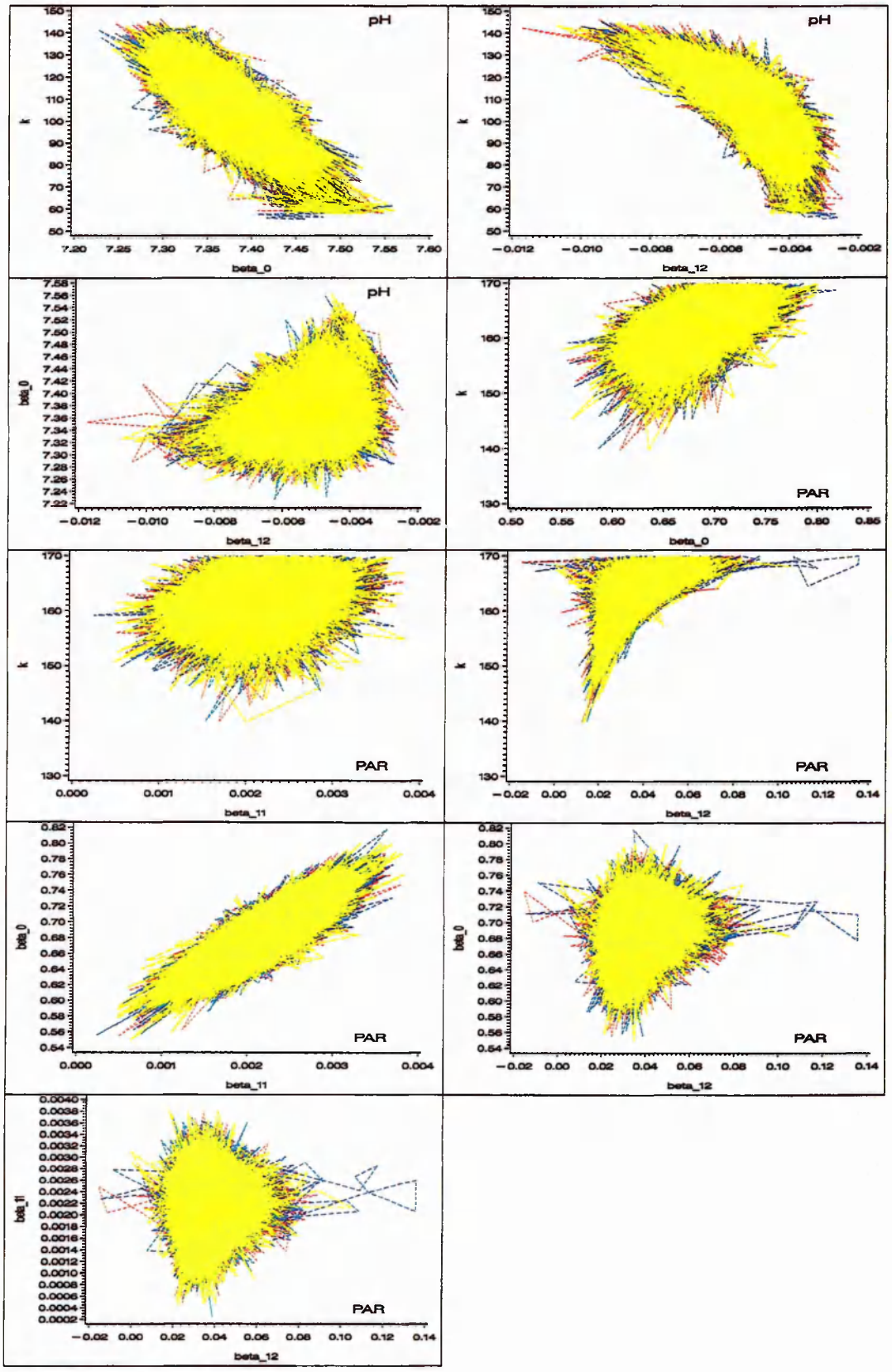


Figure legend on page 205.

Fig. 5.9. **Reduced model pH-PAR changepoint analysis. Overlaid chain trajectory plots**

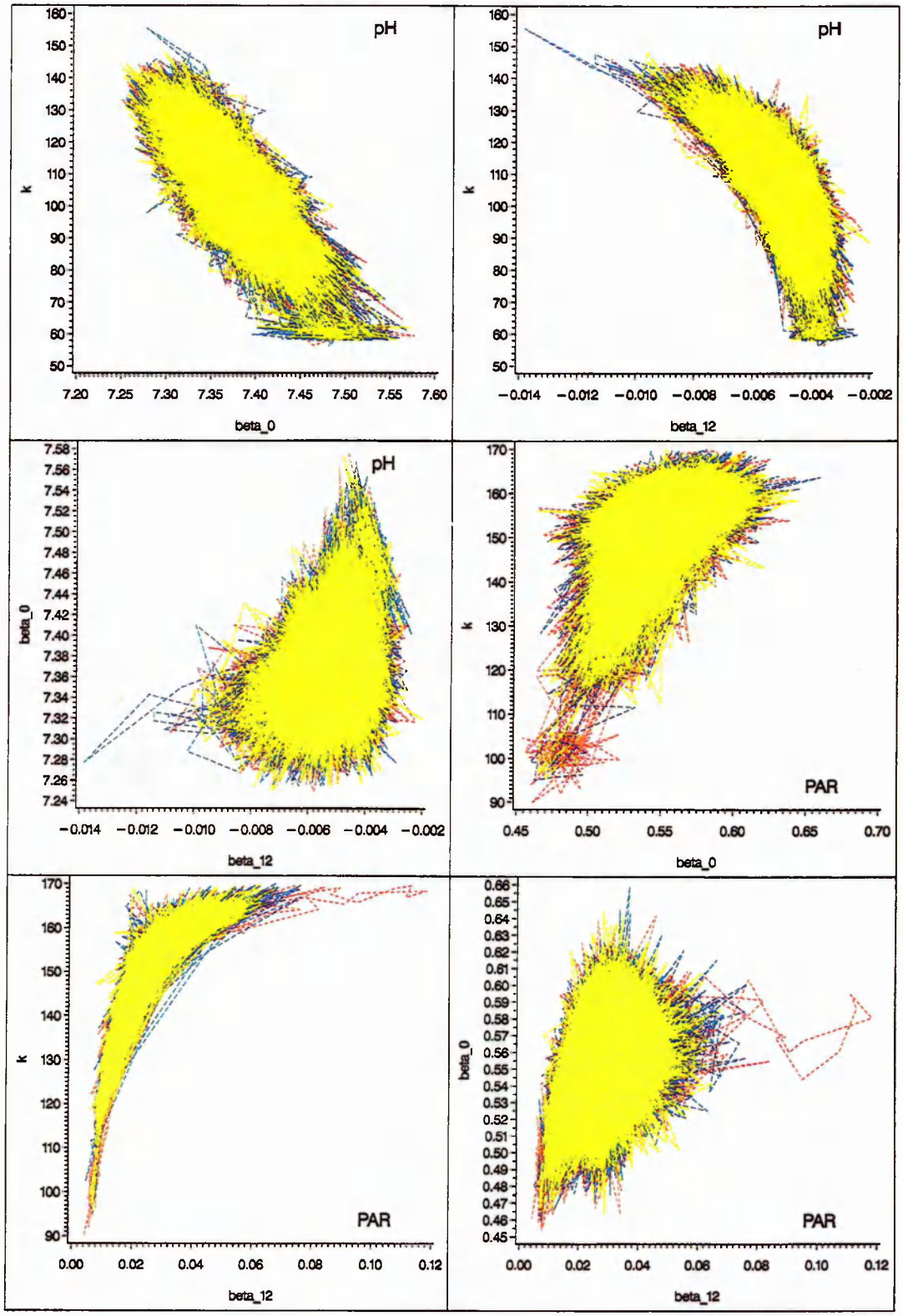


Figure legend on page 205.

Fig. 5.10. **Liver metabolite changepoint analysis. Changepoint parameter histograms**

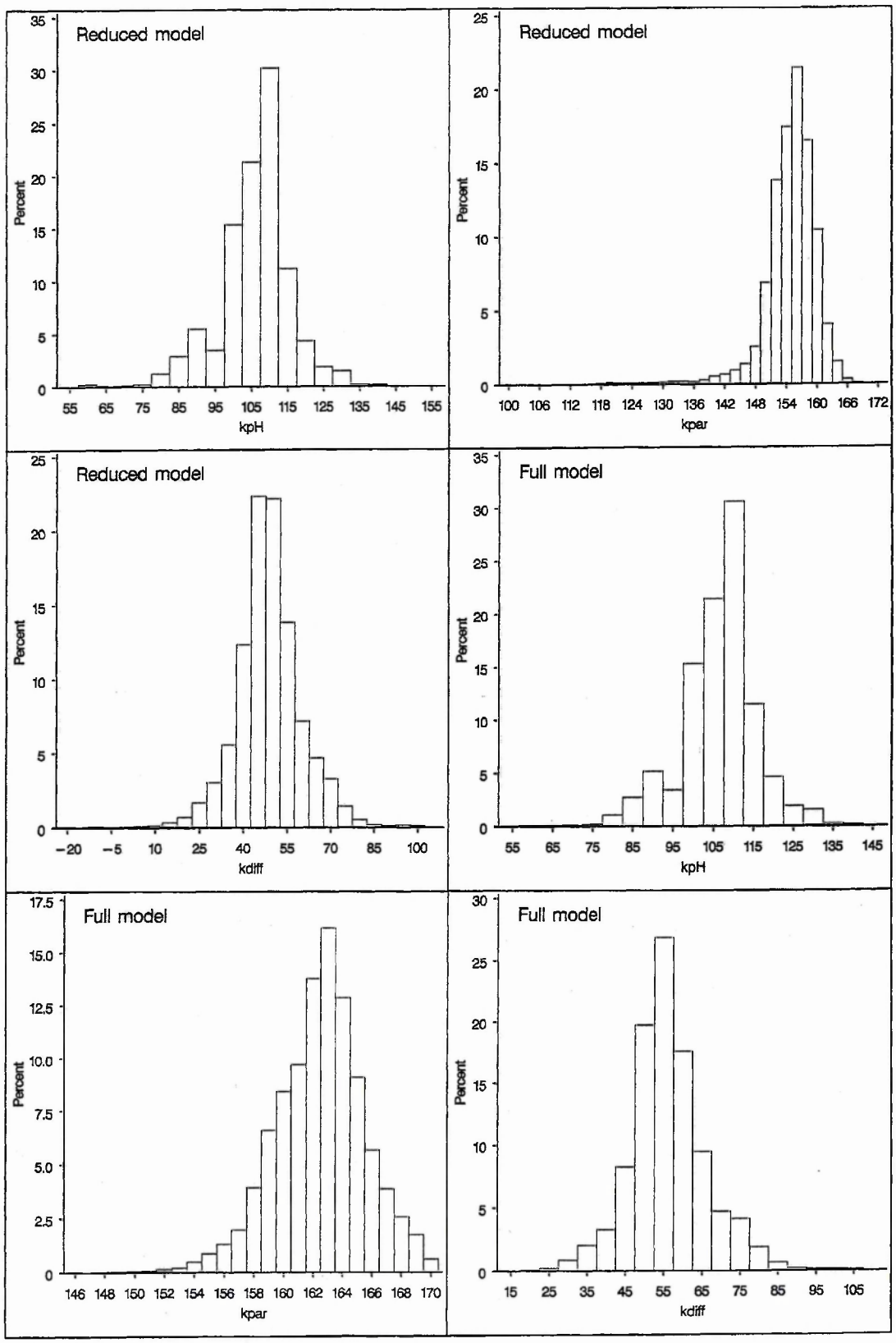


Figure legend on the following page.

The Raftery-Lewis procedure indicates that each individual chain provides the required full and reduced model changepoint parameter 0.025 and 0.975 quantile estimates to an accuracy of at least 0.004, with probability 0.95.

The posterior medians and 95% posterior intervals derived from the Metropolis output are listed in Table 5.1. The results obtained using Gibbs sampling are included for comparison. Gibbs sampling was performed using a discrete changepoint model in contrast to the continuous changepoint model used in the Metropolis-based simulations. The two MCMC simulation methods are not, therefore, expected to produce identical posterior distributions. Nevertheless, no substantial discrepancy between the two sets of results is evident. Posterior median pH and PAR time-course data were generated using the Metropolis MCMC output. These are shown in Fig. 5.5 together with the observed data and a cluster of profiles obtained by sampling the posterior distribution. The 95% posterior interval for the difference between the pH and PAR changepoints ($k_{diff} = k_{par} - k_{pH}$) is the main focus of the analyses. Both models yield a 95% posterior interval for k_{diff} that lies on the positive side of zero, leading to the conclusion that the pH response precedes the increase in PAR.

Fig. 5.8. Overlaid chain trajectory plots are shown for selected pairs of regression parameters. Two chains (yellow and blue) were started using overdispersed values for all model parameters while the third chain (red) was started at a position close to the expected posterior median. The trajectories are displayed after additional thinning to show the separate paths with greater clarity.

Fig. 5.9. Overlaid chain trajectory plots are shown for selected pairs of regression parameters. Two chains (yellow and blue) were started using overdispersed values for all model parameters while the third chain (red) was started at a position close to the expected posterior median. The trajectories are displayed after additional thinning to show the separate paths with greater clarity.

Fig. 5.10. A histogram is shown for each of the pH and PAR changepoint parameters and for the changepoint time difference ($k_{diff} = k_{par} - k_{pH}$), as generated using the full and reduced models.

Table 5.1. **Intestinal ischaemia-reperfusion changepoint analysis. Metropolis and Gibbs sampler posterior medians and 95% intervals**

FULL MODEL

Parameter	Method	Median	Lower	Upper
$\beta_{0(pH)}$	M	7.37	7.32	7.43
	B	7.38	7.32	7.45
$\beta_{0(par)}$	M	0.682	0.632	0.733
	B	0.683	0.633	0.734
$\beta_{12(pH)}$	M	-0.00508	-0.00656	-0.00382
	B	-0.00498	-0.00670	-0.00369
$\beta_{11(par)}$	M	0.00216	0.00144	0.00285
	B	0.00217	0.00144	0.00288
$\beta_{12(par)}$	M	0.0338	0.0232	0.0505
	B	0.0332	0.0219	0.0491
k_{pH}	M	107.6	85.8	126.7
	B	106	78	126
k_{par}	M	162.6	156.0	168.4
	B	162	158	166
k_{diff} (k_{par}- k_{pH})	M	55.3	36.0	78.2
	B	56	36	84
$\sigma^2_{1(pH)}$	M	0.0136	0.00695	0.0345
	B	0.0150	0.00742	0.0433
$\sigma^2_{2(pH)}$	M	0.00302	0.00160	0.00650
	B	0.00337	0.00174	0.00736
$\sigma^2_{1(par)}$	M	0.00326	0.00197	0.00593
	B	0.00347	0.00209	0.00636
$\sigma^2_{2(par)}$	M	0.0143	0.00500	0.0733
	B	0.0187	0.00594	0.1183

REDUCED MODEL

Parameter	Method	Median	Lower	Upper
$\beta_{0(pH)}$	M	7.37	7.32	7.44
	B	7.38	7.32	7.45
$\beta_{0(par)}$	M	0.554	0.516	0.590
	B	0.555	0.516	0.592
$\beta_{12(pH)}$	M	-0.00506	-0.00657	-0.00379
	B	-0.00498	-0.00664	-0.00370
$\beta_{12(par)}$	M	0.0290	0.0172	0.0422
	B	0.0286	0.0163	0.0432
k_{pH}	M	107.4	84.4	126.6
	B	106	82	126
k_{par}	M	155.4	142.0	162.5
	B	154	142	162
k_{diff} (k_{par}- k_{pH})	M	48.2	26.1	72.4
	B	52	24	76
$\sigma^2_{1(pH)}$	M	0.0136	0.00694	0.0351
	B	0.0150	0.00740	0.0421
$\sigma^2_{2(pH)}$	M	0.00304	0.00160	0.00653
	B	0.00337	0.00175	0.00737

Table 5.1. continued

$\sigma^2_{1(par)}$	M	0.00728	0.00427	0.0134
	B	0.00779	0.00456	0.0145
$\sigma^2_{2(par)}$	M	0.0118	0.00472	0.0433
	B	0.0147	0.00533	0.0636

Changepoint parameter median estimates and 95% posterior intervals generated from the Metropolis (M) and Gibbs sampler (WinBUGS, B) output.

5.5 Discussion

5.5.1 Time-of-onset (changepoint) phenomena in medicine

The changepoint problem arises in numerous disciplines, including many MR applications. The motivation for the work presented in this chapter comes from the need for a rigorous statistical analysis of MR data obtained in *time-of-onset* studies in which delayed biochemical and/or physiological responses are believed to occur subsequent to some event or stimulus. The pH-PAR MR spectroscopic study presented in this chapter is typical of this type of investigation; a changepoint analysis is required to address the key research question. Among the several areas in which the changepoint question arises is birth asphyxia and it is in this context that the candidate was prompted to undertake a study of changepoint analytical methods.

Hypoxic-ischaemic injury during delivery causes major neurodevelopmental abnormality in a significant proportion of birth asphyxia survivors. Phosphorus NMR spectroscopic studies have shown that the ratio of phosphocreatine to inorganic phosphate (PCr/P_i ratio; an index of cellular energy status) declines over a period of several days in those infants destined to die or to suffer severe neurodevelopmental abnormalities (Hope et al., 1984; Hope and Reynolds, 1985). The PCr/P_i ratio is, however, reported to be normal in these individuals on the day of delivery, and this has led to the notion of a latent period during which therapeutic intervention might be used to bring about a reduction in central nervous system damage following birth asphyxia. (The latter process is sometimes referred to as *programmed cell death* although, strictly speaking, this term refers to a programmed cell death mechanism that is specific to neurons.) Mild hypothermia is among the possible interventions (Thoresen et al., 1995). The expression secondary (delayed) energy failure is used in relation to the period of slow decline in PCr/P_i because it is believed that this decline is initiated by a primary acute hypoxic-ischaemic episode of energy failure that occurs before birth and which resolves on resuscitation (Lorek et al., 1994; Thoresen et al., 1995). The concept of a latent phase followed by a period of secondary energy failure is, in essence, the *therapeutic window* that has been discussed in many papers on cerebral ischaemia (Heiss, 1992, for example). A number of clinicians have asked questions regarding the exact duration of the therapeutic window

in birth asphyxia, the widely held view being that it lasts only a few hours (2-6 hours) after birth (Toet et al., 1999, and references therein). Head trauma provides another example in which time-of-onset studies have demonstrated the involvement of delayed/secondary responses, and led to the suggested existence of *a window of opportunity for therapeutic intervention* (Povlishock, 1992; McKenzie et al., 1996; Teasdale and Graham, 1998; Laurer and McIntosh, 2001).

In the birth asphyxia context, considerable clinical significance has been attached to studies suggesting that the PCr/P_i ratio is apparently normal on the first day after delivery, as indicated by a failure to demonstrate a statistically significant PCr/P_i ratio difference between birth asphyxiated and normal infants (Hope et al., 1984; Hope and Reynolds, 1985). Only after several days does the difference reach significance, as judged by consecutive two-sample statistical tests. The resulting concept of a therapeutic window in birth asphyxia is hugely important, but it may be a statistical artefact. Larger studies with a greater power to detect group mean differences may lead to a different conclusion. Experiments have been performed using newborn piglets in order to investigate various aspects of the birth asphyxia problem (Lorek et al., 1994; Thornton et al., 1998). As expected, the PCR/P_i ratio is severely depressed during an episode of experimentally induced ischaemia, and it recovers rapidly after the restoration of blood flow. This is followed by a subsequent reduction (secondary energy failure). In summary, the PCr/P_i response in the newborn piglet is reported to match that observed in birth-asphyxiated human infants (Lorek et al., 1994; Thornton et al., 1998), reinforcing the belief that a secondary response occurs with a delay of many hours. But this conclusion is based on an identical statistical approach, namely, a comparison of separate consecutive significance tests obtained at each of several measurement occasions. The latency/therapeutic window concept appears to imply a delay in the event that triggers the secondary response, but this might be a misconception. The term *secondary* appears to have become confused with the concept of response latency/delay. The ischaemia-reperfusion insult appears to cause a response in the PCr/P_i ratio having the form of an exponentially-damped polynomial of the type described by Crowder and Tredger (1981) and, in this sense, the final slow decline is part of a continuum, initiated when reperfusion commences (see Thornton et al., 1998,

Fig. 2a; Hope et al., 1984, Fig. 3, noting that the pre-reperfusion PCr/P_i ratio in inaccessible in birth asphyxiated infants). In this case the notion of a latent period is misleading. The observation that in human infants the PCr/P_i ratio is close to normal on the day of delivery might be of little therapeutic significance. Given an initial restoration in PCr/P_i followed by a gradual decline, any estimate of the duration of the supposed latent period, as obtained using the longitudinal data in a series of significance tests, might be determined solely by the power of the study. Estimates obtained in this manner may have little or no clinical relevance. The hypothesis that the secondary response is preceded by a well-defined latent period suggests some kind of changepoint behaviour, and the data should be examined in this context. Hence the need for a rigorous analytical approach and the decision to explore MCMC changepoint methods.

Two MR changepoint datasets unrelated to the birth asphyxia problem were used in this development work, based on the premise that the methods are expected to apply to changepoint problems in general. The first of these is an Arrhenius discontinuity problem. Among the reasons for selecting this example is the observation that it is in this context that the changepoint problem is encountered in much of the biochemical/biomedical MR literature. Furthermore, Arrhenius discontinuity studies occur in many areas of research, both academic and industrial (Section 5.1.1). Biomembrane research is prominent among those areas of biophysical chemistry in which MR techniques have made an important contribution. ESR spectroscopy, in particular, has been used extensively in the study of the Arrhenius properties of biomembranes (Sections 5.1.1 and 5.1.4). NMR methods, including T_2 relaxation and solid-state magic-angle spinning techniques (Kalichevsky et al., 1992) have also been used, but to a lesser extent. Much of this work focuses on phase transition behaviour. Thus, although differential scanning calorimetry is, in general, the definitive method for studying thermotropic phase transitions, magnetic resonance methods play an important part and it is in this context that formal changepoint analytical methods are required. The order parameter Arrhenius data presented in this chapter are typical; they are taken from an ESR study of the rabbit muscle sarcoplasmic reticulum membrane.

The second dataset was acquired in an NMR study of the hepatic response to intestinal ischaemia-reperfusion. The analysis is more challenging than the typical Arrhenius discontinuity problem, because the response is bivariate. An important aspect of the analysis is the need for a rigorous statistical assessment of the apparent difference between two changepoints, in addition to evaluating the evidence for a changepoint in each variable. Furthermore, the changepoint is somewhat ill-defined compared with a typical Arrhenius-plot changepoint, and in this sense it is more representative of the kind of data that might be encountered in a majority of clinical or biomedical MR research studies. This chapter presents some results obtained during an investigation of MCMC simulation approaches to these two changepoint problems.

5.5.2 Statistical approaches to the changepoint problem

A point that emerges from an appraisal of the literature on *Arrhenius models* (Section 5.5.3) is the need for a formal comparison of competing models (see, for example, Londesborough, 1980; Silvius and McElhaney, 1981; Bagnall and Wolfe, 1982; Godiksen and Jessen, 2002). In many applications this takes the form of a comparison between a biphasic linear model and some curvilinear model, as required to distinguish between a discontinuity versus some other form of nonlinearity. The standard textbook methods for comparing models based on the likelihood ratio test assumes nested models. This approach is, therefore, inadmissible since the models under consideration do not satisfy the nested criterion. This is not an issue if the Bayes factor is adopted, but was a considerable problem when, in the past, computational/numerical intractability precluded such an approach. Methods were devised with biological problems in mind. Among these was a simulation approach (Williams, 1970) in which the distribution of a likelihood ratio discriminating function is generated under each of the two competing models. In the case discussed by Williams (1970) the discriminating function reduced to the ratio of the residual sum of squares obtained for the competing piecewise-linear and curvilinear models. The distributions obtained under the two competing models, together with the observed ratio of residual sum of squares, form the basis for discriminating between the models. This kind of approach might retain its appeal among those reluctant to adopt a more formal method for model discrimination.

In those applications in which a changepoint/discontinuity is demonstrated, as opposed to some curvilinear form, changepoint estimation is usually required. In some cases the changepoint may be sufficiently well defined that a visual assessment of the piecewise linear regression plot and superimposed data, together with a summary parameter statistical analysis (Section 1.4.1.1), might be adequate for the task in hand. In contrast, a formal analysis is required in many applications, including studies in which between-species or between-sample differences in changepoint are the focus. For example, the need exists for a rigorous analysis of the nonlinear Arrhenius-plot behaviour that arises in a variety of biological membrane systems, as was noted several decades ago (Londesborough, 1980; Silvius and McElhaney, 1981). In those studies in which Arrhenius behaviour is determined by a continuous temperature sweep within individual samples (possibly with temperature-sweep replication), as opposed to an independent samples/observations approach, the analysis takes the form of a random coefficients changepoint problem.

A number of methods have been suggested in the biological literature for defining/estimating changepoints, including approaches based on the residual sum of squares (Bagnall and Wolfe, 1978, page 1237) and informal residual distributional considerations (Wolfe and Bagnall, 1979). Wolfe and Bagnall (1980) provide a brief summary of the methods reported in the earlier biological literature. Although largely ignored in many areas of biomedical research, the statistical literature on the changepoint problem is substantial. Carlin et al. (1992) provide a useful entry-point to this literature and, additionally, outline an Hierarchical Bayesian approach to the changepoint problem, a summary of which follows.

Given

$$\begin{aligned} Y_i &\sim f(.), \mathrm{i} \le k \\ Y_i &\sim g(.)\ \mathrm{i} > k \end{aligned} \qquad [5.5]$$

where k is the changepoint, and likelihood

$$L(\mathbf{Y}|k,\theta,\eta) = \prod_{i=1}^{k} f(Y_i|\theta) \prod_{i=k+1}^{n} g(Y_i|\eta), \qquad [5.6]$$

then, assuming a prior distribution $\pi(k,\theta,\eta)$, the joint posterior distribution is

$$p(k,\theta,\eta|\mathbf{Y}) \propto L(\mathbf{Y}|k,\theta,\eta)\pi(k,\theta,\eta). \qquad [5.7]$$

The marginal distribution of k is

$$p(k|\mathbf{Y}) \propto \int p(k,\psi|Y)d\psi \qquad [5.8]$$

with $\psi = (\theta,\eta)$. The posterior odds for *no change* is given by $p(k=n|\mathbf{Y}) / \{1-p(k=n|\mathbf{Y})\}$.

Carlin et al. (1992) give a number of changepoint examples including one, referred to as a *changing linear regression* model, which is relevant to the sarcoplasmic reticulum membrane order parameter and intestinal ischaemia-reperfusion analyses. They use this piecewise linear model to re-examine the surfactant changepoint data presented previously by Bacon and Watts (1971). A three-stage hierarchical implementation of the regression model is outlined, in which inverse gamma distributions are assigned to the variance components at the second stage, together with independent multivariate normal distributions for the piecewise regression coefficients and a discrete uniform prior distribution on K_n, $K_n = \{1,2,..,n\}$, for the changepoint parameter, where n is the number of design points. Thus the Carlin-Gelfand-Smith three-stage hierarchical model takes the form:

STAGE 1

$$\begin{aligned} &Y_i \sim N(\alpha_1 + \beta_1 x_i, \sigma_1^2),\ i = 1,\ldots,k \\ &Y_i \sim N(\alpha_2 + \beta_2 x_i, \sigma_2^2),\ i = k+1,\ldots,n \\ &k \in \{1,2,\ldots,n\} \end{aligned} \qquad [5.9]$$

STAGE 2

$$\begin{aligned} &\theta_1 = (\alpha_1,\beta_1)^T \sim N(\theta_0,\Sigma) \\ &\theta_2 = (\alpha_2,\beta_2)^T \sim N(\theta_0,\Sigma) \\ &\sigma_1^2 \sim IG(a_0,b_0) \\ &\sigma_2^2 \sim IG(a_0,b_0) \\ &k \sim dunif(1,n) \end{aligned} \qquad [5.10]$$

STAGE 3

$$\theta_0 \sim N(\mu, C)$$
$$\Sigma^{-1} \sim W[(\rho V)^{-1}, \rho]. \qquad [5.11]$$

The inverse gamma distribution parameters are chosen to produce a vague prior for each of the variances and a vague third-stage prior is adopted for θ_0. Using this model they compare the posterior distribution obtained using a *somewhat informative* prior for Σ^{-1} [Wishart prior with ρ=4, V=diag(0.001, 0.3)] with the posterior distribution generated under a less informative prior [Wishart, with ρ=2, V=diag(0.0, 0.0)]. Finally, these results are compared with the posterior distribution given in Bacon and Watts (1971), which is based on the continuous changepoint Bayesian model described below ([5.12] with the transition function $\tanh\{(x-x_0)/\gamma\}$) and the single-stage noninformative prior $p(\beta,\sigma) \propto 1/\sigma$, where $\beta = (\alpha_0, \alpha_1, \alpha_2, x_0, \gamma)$.

A useful feature of the Bacon and Watts treatment of the biphasic linear problem is the introduction of a transition function, $\mathrm{trn}\{(x-x_0)/\gamma\}$, which they use in a model of the form

$$y = \alpha_0 + \alpha_1(x-x_0) + \alpha_2(x-x_0)\mathrm{trn}\{(x-x_0)/\gamma\} \qquad [5.12]$$

to obtain a smooth transition between the two linear dependencies. Among the transition functions discussed is $\tanh\{(x-x_0)/\gamma\}$. This function has the property that $\lim_{\gamma\to\infty} \mathrm{trn}\{(x-x_0)/\gamma\}$ gives simple linear dependence, while $\lim_{\gamma\to 0} \mathrm{trn}\{(x-x_0)/\gamma\} = \mathrm{sgn}(x-x_0)$ gives a discontinuity in the first derivative at x-x_0. This aspect of their treatment is relevant to the Arrhenius-plot modelling problem, especially in relation to the curvilinear issue, mentioned above, which arises in the context of the *finite-transition range* models referred to below (Section 5.5.3). Furthermore, models of this kind might be examined in connection with the intestinal ischaemia-reperfusion study, and any other context in which a discontinuous first derivative is not stipulated, a priori, on the grounds of well-established

theory. An alternative approach to dealing with this kind of behaviour is segmented polynomial regression (Gallant and Fuller, 1973) which is based on piecewise models of the form

$$\begin{aligned} y &= g(x) + \varepsilon \\ g(x) &= g_1(x,\beta_1), a \le x \le \alpha_1 \\ &= g_2(x,\beta_2), \alpha_1 < x \le \alpha_2 \\ &\cdot \\ &\cdot \\ &\cdot \\ &= g_r(x,\beta_r), a_{r-1} < x \le b, \end{aligned} \qquad [5.13]$$

where $\alpha = (\alpha_1, \alpha_2, ..., \alpha_r)$ is an unknown vector of knots and $g(x,\beta_j)$ is continuous with a continuous first derivative in x over the interval $[a,b]$. Gallant and Fuller (1973) outline an approach to finding the least-squares estimates of the unknown parameters. A piecewise linear-quadratic-linear model might apply to some of the Arrhenius dependencies discussed by Silvius and McElhaney (1981) in which the transition enthalpy is insufficient to cause a near discontinuity in the first derivative at the transition, assuming an empirical approach is sought in preference to thermodynamic modelling. Arrhenius plot shape and its dependence on transition enthalpy is among the topics discussed in the following section.

5.5.3 Arrhenius plot models

The Arrhenius-plot changepoint model used in this chapter attributes the departure from classical Arrhenius behaviour to a discontinuity in $d\ln(S)/d(1/T)$, while assuming continuity in $\ln(S)$. This kind of intersecting biphasic linear model has been adopted in numerous enzyme and biomembrane research studies. In general, a discontinuity in $d\ln(\text{rate})/d(1/T)$ is attributed to an abrupt change in activation energy. (The numerical difference between the activation energy and activation enthalpy is negligible in the systems under consideration. The two terms are used synonymously in much of the biological literature, although *activation enthalpy* is a concept that originates from transition state theory, while the term *activation energy* arises from the empirical Arrhenius equation treatment of reaction rates.) In any serious application the validity of the model under consideration must be assessed in comparison with plausible competing models. In particular, a generalisation to

include a discontinuity in rate should be considered. At the other extreme, a smooth dependence of rate on inverse temperature should be examined. The competing Arrhenius models question has been the subject of several papers, many of which have addressed this issue in relation to the kinetics of biomembrane enzyme systems.

Two simple competing reactions cannot produce a true discontinuity in an Arrhenius plot, although a biphasic linear model might provide a good approximation to the curve that is obtained given a sufficient difference in activation enthalpy (Kumamoto et al., 1971). A commonly held view is that the apparent Arrhenius plot discontinuities that have been reported to occur in many integral membrane enzyme systems are real and that membrane lipid phase transitions are the cause. Lipid-protein interactions are believed to be an important component of the coupling between integral membrane enzymes and the lipid matrix in which they are embedded. Accordingly, a phase transition within the lipid matrix might be expected to affect the kinetic properties of the embedded enzymes. [See Table 1 in Zakim et al. (1992) for information on the relationship between lipid phase transitions and integral membrane enzyme properties. Griffith and Jost (1976) and McConnell (1976) provide a review of early ESR biomembrane and lipid-protein interaction studies. The more recent literature is reviewed by Marsh and Horváth (1998).] In some cases, however, the observed departure from classical Arrhenius behaviour might be an artefact of the measurement. For example, Silvius et al. (1978) discuss temperature-dependent substrate binding in relation to reaction rate Arrhenius plot discontinuity artefacts, while Cannon et al. (1975) and Schreier et al. (1978) refer to artefacts arising from the inappropriate approximation methods that are sometimes used to calculate correlation-time and order parameter estimates based on ESR spectral measurements.

Returning to the competing models issue, a discontinuity in $d\ln(\text{rate})/d(1/T)$ in the absence of a discontinuity in rate gives rise to intersecting biphasic linear behaviour. Several researchers have discussed this model (Kumamoto et al., 1971; Wynn-Williams, 1976; Londesborough, 1980; Silvius and McElhaney, 1981). Among the explanations that have been suggested for the continuity

in rate is enthalpy-entropy compensation. According to this model the phase transition is associated with a change in activation entropy equal to the enthalpy change (Kumamoto et al., 1971). Subsequent to the early publications on Arrhenius plot nonlinearity Silvius and McElhaney (1981) provided a particularly detailed treatment of the problem, and made a number of important comments. They pointed out that most biological thermotropic changes occur over a finite temperature range which, depending on the magnitude of the transition enthalpy, can be broad (a range of several decades, for example). Biological membranes, in particular, are not expected to undergo sharp phase transitions due to heterogeneity in their lipid composition. Phase diagrams are especially useful in this context, since they provide a schematic representation of the membrane-lipid response to changing temperature, at least in principle (Bagnall and Wolfe, 1978). For example, the phase diagram for a simple 2-component lipid mixture shows the presence of two critical temperatures, between which separate lipid domains of differing composition coexist. Only a few models predict true biphasic linear Arrhenius behaviour (Silvius and McElhaney, 1981). Silvius and McElhaney (1981) and Wynn-Williams (1976) both discuss the special conditions under which an abrupt change in activation energy can occur in the absence of a discontinuity in rate. An important conclusion is that the absence of a discontinuity in rate does not depend on an exact enthalpy-entropy compensation. Among the models outlined by Silvius and McElhaney (1981) is a kinetic model that applies to lipid motion, as monitored using spin label and fluorescent probes. They show that although a sharp lipid phase transition will yield an Arrhenius plot discontinuity, the former is not a necessary condition for the latter, and that a discontinuity can arise under other conditions. In particular, a transition enthalpy of sufficient magnitude can produce an apparent Arrhenius plot break, although the transition itself will span a wide temperature range. It is, in fact, difficult to determine the starting and endpoints of a thermotropic transition from an Arrhenius plot. Thus, although membrane probes can provide the transition midpoint, they do not provide an estimate of the transition temperature range. An independent calorimetric assessment of phase behaviour is clearly desirable. (Fig. 4 in Buitink et al., 1999, provides an example in which rotational correlation time Arrhenius plot discontinuities are shown in relation to transition onset temperatures determined by differential scanning calorimetry.) The Silvius-McElhaney treatment

explains why studies reporting sharp Arrhenius changepoints are so common in the biomembrane literature. They also discuss the conditions under which a discontinuity in rate (referred to as jump discontinuities) might be observed. Although an infinite transition enthalpy is required to produce a true discontinuity in rate, enthalpies of several hundred kcal mol^{-1} might produce an apparent discontinuity, because the 1-2°C interval over which the transition then occurs is smaller than the temperature sampling interval used in most studies. To summarise, a number of important points emerge. (1) Activation energies cannot be determined with reliability from an Arrhenius-plot slope; (2) lipid mixtures do not have a sharp phase transition, as indicated by their phase diagrams (Bagnall and Wolfe, 1978); (3) an apparent activation enthalpy discontinuity can occur, without a discontinuity in rate, at the phase separation/transition starting and endpoint temperatures. This behaviour does not depend on enthalpy-entropy compensation.

An important conclusion is that biphasic linear Arrhenius plots can arise in the absence of a phase transition, and that the former should be interpreted with caution. The need for a rigorous analysis of the Arrhenius plot shape is clear. Thus, Londesborough (1980) points out that although the application of some rigorous statistical analytical method might improve the accuracy with which an Arrhenius-plot changepoint is estimated, it does not necessarily establish the validity of the underlying model. This can be achieved only through a rigorous comparison of competing models. The latter places stringent demands on the quality of the data.

5.5.4 A comparison of the Gibbs sampling and Metropolis changepoint analyses

Various approaches were adopted in an attempt to implement the two changepoint analyses using the Gibbs sampler but, on the whole, these were unsuccessful. In particular, Gibbs sampler implementation of the continuous changepoint variant of the models was problematic, and output was invariably generated in which the changepoint parameter moved rapidly to one end of the range allowed under the prior distribution (uniform over the entire measurement range), after which the parameter became trapped. The same Gibbs sampler behaviour was encountered with the

discrete changepoint treatment of the order parameter data, which is based on a uniform prior density over the set of design points. Carlin et al. (1992, p390) interpret the result $k = n$ as indicating no change, stating that a posterior probability for $k = n$ close to zero is required in order to claim strong evidence for a changepoint. The Gibbs sampler result suggesting the absence of a discontinuity in the order parameter Arrhenius-plot gradient is inconsistent with the changepoint posterior distribution generated with the Metropolis algorithm, which indicates a well-defined changepoint. In contrast to the difficulty experienced with the Gibbs sampler analysis of the order parameter Arrhenius data, the pH-PAR changepoint posterior distributions generated by the Metropolis and Gibbs sampler algorithms are in good agreement (Table 1).

5.5.5 Spin-labelled sarcoplasmic reticulum Arrhenius-plot changepoint analysis

The spin-label order parameter data used in this chapter were acquired as part of a study of the effect of dithiothreitol on the Arrhenius behaviour of the sarcoplasmic reticulum ATPase reaction rate and membrane lipid motion (King and Quinn, 1983; King et al., 1987). Several studies had shown that the sarcoplasmic reticulum calcium pump is stabilised in vesicles isolated in the presence of dithiothreitol (see King et al., 1987, for references) and that the Arrhenius properties of the ATPase are altered if dithiothreitol is added to the assay medium. Various properties of the dithiothreitol stabilised preparation were therefore examined and compared with the behaviour of the preparation isolated in the absence of this reagent. The data shown in Fig. 5.2 were acquired from a sarcoplasmic reticulum microsomal preparation isolated in the absence of dithiothreitol. A departure from classical linear Arrhenius dependence is demonstrated, consistent with other studies (for example, Inesi et al., 1973; Hesketh et al., 1976). A changepoint in the region of 9.7 to 13.2°C is low, however, compared with the values reported in the literature, which lie in the range 20 to 30°C. When these data were acquired in 1979, questions regarding the precision of the changepoint estimates were raised but not addressed formally. The Arrhenius plot appeared to indicate a well-defined changepoint. This is consistent with the 95% posterior interval of 9.7 to 13.2°C, provided by the present MCMC simulation analysis. This interval is well removed from 20°C. A visual

inspection of data not included in this chapter indicates possible differences in changepoint behaviour during the two phases of a heating-cooling cycle, raising the possibility of hysteresis. The order parameter data used in this chapter were acquired during the cooling phase. Hysteresis might, therefore, account for the discrepancy. It is noted in passing that the ESR spectrum obtained at 0° is close to the limit at which the position of the high-field resonance can be estimated. Furthermore, the method used to extract parameter estimates from the ESR spectra has been questioned. ESR rotational correlation time data are frequently presented as Arrhenius plots, and it has been suggested that spin-label Arrhenius changepoint anomalies can occur due to a violation of the motional narrowing assumption that underlies the standard approximations used to obtain the correlation time estimates. Rigorous slow-motion mathematical-modelling/simulation (Freed, 1976) has been recommended (Schreier et al., 1978). Similarly, Cannon et al. (1975) suggest that order parameter Arrhenius plot discontinuities can occur as an artefact if approximation is used to extract the order parameter estimates. In fact the physical significance of the measured order parameter depends on correlation time (Griffith and Jost, 1976, p474; McConnell, 1976, p542). Again, slow-motion modelling has been recommended (Cannon et al., 1975; Schreier et al., 1978) which, at the time, was a considerable computational undertaking. These issues are beyond the scope of this chapter. The present changepoint MCMC simulation study was carried out for the sole purpose of investigating the performance of various MCMC approaches to problem. This ESR study predates the upsurge of papers on MCMC in the statistical literature. Given additional experimental data with adequate replication (the funding period expired prior to completion of the study) an extension of the Metropolis-based MCMC analysis outlined in this section should provide definitive answers to a number of questions, including the Arrhenius-behaviour differences between various microsomal sarcoplasmic reticulum preparations and questions relating to hysteresis.

The order parameter analysis outlined in this chapter deals with only one of several outstanding issues regarding spin-label Arrhenius studies, namely the changepoint estimation problem. In addition a rigorous comparison of competing models is required, and the need for robust order

parameter (or rotational correlation time) estimates is obvious. Finally, the uncertainty that surrounds the thermodynamic and/or structural significance of uncorroborated spin-label Arrhenius-plot discontinuities must be addressed. Differential scanning calorimetry provides definitive information on the phase behaviour of membrane lipids, but is not always feasible (Buitink et al., 2000, p290).

5.5.6 Liver response to intestinal ischaemia-reperfusion injury

Intestinal ischaemia-reperfusion injury is a serious clinical condition, which is associated with a number of complications, including the common occurrence of multiple organ failure. The latter is the most common cause of morbidity and mortality in patients with intestinal ischaemia-reperfusion injury (see Section 5.1.3). An essential feature of the condition is the damaging effect of reperfusion, as opposed to ischaemia, per se. Multiple organ failure involving gut injury is a major cause of intensive care unit (ICU) deaths, including surgical ICU mortality.

The pH-PAR data used in this chapter were acquired as part of an extensive study of several aspects of multiple organ failure in intestinal ischaemia-reperfusion, including a detailed assessment of the hepatic metabolic response and an examination of the protective potential of hypothermia and perfluorocarbon administration (Vejchapipat et al., 2001, 2002). Among the principal aims of the study was an assessment of the effect of hypothermia on the onset of liver failure, as indicated by the PAR index. Furthermore, the time difference between the onset of the pH and PAR changes is of considerable interest from a mechanistic viewpoint. A formal changepoint analysis is required, therefore, to address properly several of the key research questions. A complete analysis of the entire set of pH-PAR responses was not performed because the main objective of the work outlined in this chapter was to explore alternative MCMC approaches to the changepoint problem. It is interesting to note, however, that the posterior distribution obtained for the PAR-pH onset-time difference indicates that the pH response preceded the onset of the increase in PAR in the one case examined. A previous informal treatment of the data had led to the conclusion that the decline in pH was relatively late and that intracellular

acidosis was an unlikely cause of hepatic energy failure. The present result indicates that a re-evaluation of the data is required; a modification of the conclusion would follow if the pH response was found to precede energy failure in every animal.

The model assessment component of the pH-PAR data analysis has not been addressed in this chapter. The intersecting biphasic linear model is empirical and, as far as the candidate is aware, it is not possible to rule out alternative models using arguments based on established biochemistry and physiology. Insufficient is known about the mechanism that underlies hepatic energy failure in intestinal ischaemia-reperfusion injury, and any of the empirical models that were discussed in the Arrhenius-plot context might apply. Thus a curvilinear response, at one extreme, and a biphasic response with a discontinuity in each of the response variables and their respective derivatives might be considered. The latter model was examined but, unfortunately, insufficient pre-reperfusion data were collected to allow a useful assessment of this type of model. In contrast, sensible results were obtained given the constraint imposed by assuming continuity in pH and PAR at the changepoint. Subject to the need for model validation, confidence in the Metropolis and Gibbs sampling results presented in this chapter is justified given the good agreement between the posterior intervals obtained using these alternative methods. Agreement at the level of simulation error is not expected because a continuous changepoint model was adopted in the Metropolis simulations, while a discrete changepoint model was used in the Gibbs sampler analysis.

The pH-PAR modelling work outlined in this chapter was not extended to include a full random-coefficients analysis of the complete dataset (the normothermic group consisted of six animals) mainly because the study was somewhat compromised by the absence of spectroscopic data for the first 54 minutes of the ischaemic period. As stated previously, the remaining 36 minutes of pre-reperfusion data are not sufficient to allow a proper assessment of competing models. This MR study was performed using a purpose-built animal-holder/RF-coil assembly designed specifically to permit remote control of the superior mesenteric arterial blood flow (Vejchapipat et al., 2001). Accordingly, it might have been possible to commence data acquisition soon after the start of the

ischaemic phase. The question of experimental design and its optimisation should, of course, always be considered. In the present context simulations might be used to improve the study design as required to enable an assessment of competing models and to obtain changepoints estimates with sufficient precision.

5.5.7 Final comments

In conclusion, the purpose of the work outlined in this chapter was to explore the MCMC approach to changepoint modelling using ESR spin-label and MR spectroscopic data to achieve this. The model assessment part of the problem has not been addressed. For this reason no attempt is made to offer an interpretation of the Arrhenius changepoint estimate and other parameters extracted from the ESR data. Furthermore, the observed spectral changes and resulting Arrhenius behaviour provide an indirect indicator of membrane lipid phase behaviour, at best. The MR intestinal ischaemia-reperfusion study is less problematic in the sense that a direct observation of the processes under consideration (i.e., acidosis and energy failure) is possible. Consequently, the posterior probability distributions provide a direct estimate of the parameter that is sought, i.e., the difference in time between the onset of the pH and PAR responses. Although the application of an inadequate model is expected to yield a biased estimate, this has no impact on the validity of the interpretation regarding the temporal/causal relationship between acidosis and liver energy failure.

6. DISCUSSION

6.1 The role of Bayesian hierarchical modelling in longitudinal MR data analysis and paediatric neurology

MR imaging and spectroscopy can be used in a variety of modes to provide surrogate markers of pathology, effectively allowing the non-invasive acquisition of simultaneous biochemical, physiological and anatomical data. For this reason multivariate longitudinal studies are extremely common in biomedical MR research. Invariably, huge amounts of data are generated within a single study, and frequently the typical MR researcher is required to undertake analytical problems of considerable complexity. Although a number of textbooks (including Crowder and Hand, 1990; Jones, 1993; Goldstein, 1995; Hand and Crowder, 1996) on longitudinal data analysis serve to supplement the extensive statistical literature on the subject, this work has been somewhat ignored by the MR research community. Thus, despite the major methodological advances that have been made since the early 1980s, it is unfortunate that inferior statistical methods continue to be used by the majority MR researchers, with the result that many studies are severely compromised.

The seminal paper by Laird and Ware (1982) on random coefficients modelling provides the basic framework for both Bayesian and non-Bayesian longitudinal data analysis. Among the attractive features of the approach is the considerable degree of model flexibility that is permitted, including a wide range of within-subject error structures (equicorrelated errors, autoregressive error structure, etc.). This is coupled with a rigorous modelling of the between-subject variability in behaviour. For several reasons, random coefficient models are especially relevant to many of the clinical MR studies carried out within the tertiary referral centre in which the candidate works (Great Ormond Street Hospital for Children). The first is the fundamental need for a formal modelling of subject variability. The V_{DC}-ADC data presented in Chapters 3 and 4 illustrate the point; it is typical of the MR data acquired in longitudinal MR studies. In many cases mean behaviour might be of secondary interest, and the main focus is the manner in which some

measure evolves within individuals. This applies to both clinical and non-clinical investigations. The V_{DC}-ADC investigation falls into this category because an improved understanding of the processes that drive the ADC response to focal ischaemia is provided by the study of the within-individual temporal relationship between the ADC and DC-potential responses. Thus, although a formal treatment of the between-individual differences in response behaviour might be required solely for the purpose of obtaining reliable mean statistics in some studies, the V_{DC}-ADC investigation provides an example in which a modelling of individual behaviour is required to address the primary research question. The distinction between analyses in which the population mean behaviour is the focus and those in which individual-specific inferences are required is discussed in Section 1.4.1.5.

A related context in which an explicit random coefficients modelling is required is in paediatric prognostic modelling. Prognostic models are relevant to both paediatric research and clinical practice. A number of circumstances arise in which prognostic models, based on MR data, might be used to advantage, including premature birth. Chapter 5 gives a brief account of the manner in which brain phosphorus NMR spectroscopic PCr/P_i ratio data (an index of cell energy status) acquired during the first few days after delivery from asphyxiated infants, born at term, provides an indicator of those destined to die or to suffer severe neurodevelopmental abnormalities (Hope et al., 1984; Hope and Reynolds, 1985). In individuals with a reduced PCr/P_i ratio the insult is so severe that it is incompatible with survival and the maintenance of normal neurodevelopmental function. The birth prematurity scenario is not dissimilar, and a proportion of children that survive prematurity at birth exhibit major neurological impairment, subsequently. Of those that survive without major impairment (85 to 90%), up to 30% develop cognitive and behavioural problems. Furthermore, MR volumetric analysis based on images acquired within a few weeks of delivery indicates a relationship between early right sensorimotor white matter volume and neurodevelopmental outcome assessed at between 18 and 20 months (Peterson et al., 2003). Early right midtemporal white matter volume is similarly related to neurodevelopmental outcome at 18-20 months (Peterson et al., 2003). Recent studies have shown, however, that among those

individuals appearing to survive unimpaired initially, a proportion go on to develop subtle functional abnormalities that become apparent only later in childhood (Bhutta et al., 2002). It is suggested that, in some cases, these latent cognitive deficits are caused by elusive but identifiable brain anatomical/structural abnormalities (see Gadian, 2002, for a minireview). The latter are not necessarily identified by visual inspection of the MR images, and it is only through the development of new image processing techniques that the underlying abnormalities have been detected. It is in this context that a study of the relationship between structural pathology and impaired function in preterm children has been undertaken as a major component of the MR research programme at the Great Ormond Street Hospital for Children. Among the aims is the development of prognostic models. It is anticipated that MR imaging information acquired shortly after birth, together with a subsequent longitudinal MR characterisation of the manner in which pathology evolves in the early stages, will provide a basis for prognosis. Growth-curve prognostic models of the type outlined by Berzuini (1996) will receive particular attention. The study will be exploratory in nature, various structural pathologies and cognitive functions requiring investigation. It is envisaged that a characterisation of the manner in which pathology evolves within individual subjects will lead to an improved understanding of the causes of some types of mild cognitive impairment. A characterisation of the population mean progressive behaviour is, in contrast, relatively unimportant.

A number of special but unavoidable circumstances contrive to compromise the studies outlined above. Fortunately, however, Bayesian hierarchical modelling, as implemented using MCMC, provides a credible solution to the resulting analytical challenges. Foremost among the difficulties that arise in paediatric MR monitoring is the need to sedate young children during scanning and the resulting sparse data problem. Sedation is required for two reasons. Firstly, MR images are, in general, degraded if the subject moves during data acquisition. Diffusion images are especially prone to degradation, caused by motion artefacts, because a sensitivity to displacement is fundamental to diffusion imaging. Without sedation few young children can be expected to co-operate for the length of time required to perform an extensive MR investigation, and the quality

of the resulting images will be compromised in most cases. Secondly, anxiety will cause some young children to become generally unco-operative, a few simply refusing to enter the scanner. Thus, although MR imaging can be undertaken in newborn infants while they sleep, sedation is required from about 2 months onwards. MR data can, therefore, be obtained only if an MR investigation is required on clinical grounds, since sedation for the collection of research data is ethically unacceptable and strictly forbidden. Longitudinal MR data in children under the age of about 7 years are, therefore, necessarily sparse. In older children imaging is usually possible without sedation. The sparse data problem is compounded by the implausibility of acquiring MR data on every individual at pre-specified design points. This rules out the development of simple prognostic models based on a defined set of early observations. The growth-curve approach provides the solution, and Bayesian hierarchical modelling is the method of choice for dealing with undersampled longitudinal data. In some studies very little imaging data might be collected on a few individuals, but it is anticipated that the improved individual-specific parameter estimates afforded by Bayesian random coefficient (hierarchical) models will provide useful information from studies that would otherwise be of limited value due to the undersampling that inevitably occurs given the need to employ sedation. Thus the growth-curve prognostic modelling approach should allow forecasting for every individual. It remains to be determined whether useful prognostic models can be constructed using MR growth curve data.

The motivation for these studies is the notion that corrective or ameliorative action becomes a possibility given the identification of individuals at risk of a poor outcome. The concept of a therapeutic window is discussed in relation to birth asphyxia in Section 5.5.1. It has to be said that, despite the passage of several decades since therapeutic approaches to hypoxic-ischaemic brain damage were first discussed (Astrup et al., 1981; Hope et al., 1984; Lassen and Astrup, 1987; Heiss, 1992), little progress has been made[1]. It is suggested, however, that a *special needs* approach might be used to advantage in the mild cases that are under discussion.

---

[1]Among recent developments is an assessment of potential therapeutic interventions through formal clinical trials. For example, several multicentre trials are in progress to determine whether

A similar need for prognostic modelling arises in temporal lobe epilepsy (TLE). This is a debilitating condition commonly associated with mesial temporal sclerosis, which is a pathology characterised by changes in the hippocampus, including a reduction in volume and an increase in T_2 relaxation time. (A review of magnetic resonance imaging as it relates to mesial temporal sclerosis and TLE is provided by Kuzniecky and Jackson, 1995.) This form of epilepsy tends to be unresponsive to medication, and temporal lobe resection is performed in a proportion of cases. In fact, mesial temporal sclerosis is the most common structural abnormality seen in patients requiring surgery for intractable epilepsy. An important clinical observation is the association between mesial temporal sclerosis and prolonged febrile convulsions (PFC), between 4 to 20% of PFC cases subsequently developing TLE with associated mesial temporal sclerosis. Furthermore, approximately 50% of patients undergoing surgery for TLE have a history of PFC. Newly emerging MR data has led to the proposition that in these TLE cases an episode of prolonged febrile convulsions has initiated a process that has caused hippocampal damage, culminating in mesial temporal sclerosis and TLE. Components of the process appear to be MR visible due to the associated T_2 relaxation time, volume and, possibly, water diffusivity changes. Accordingly, it is reasonable to suppose that the evolution of pathology might be monitored using MR imaging, and that growth curve prognostic modelling is a possibility. The implications for clinical practice are important. At present the development of TLE is not always recognised in the initial stages with

---

cooling reduces the risk of death or neurodevelopmental injury following birth asphyxia. These include the *total body hypothermia for perinatal asphyxia* (TOBY) trial, which is a multicentre UK investigation into the effects of reducing body temperature to 33-34°C for 72h, compared with the maintenance of body temperature at 36.8-37.2°C in the control group. Randomisation takes place within 6 hours of delivery, and the primary outcome is the combined rate of mortality and severe neurodevelopmental impairment in survivors at 18 months of age (www.npeu.ox.ac.uk/TOBY).

the result that there can be a considerable delay in the prescription of suitable medication. The identification of individuals at risk of developing TLE would allow close monitoring, with the result that the onset of seizure activity should be recognised without delay and suitable medication started. Among the advantages of prompt intervention is a minimisation of secondary tissue damage, circumstantial evidence indicating that the electrical activity associated with TLE causes additional injury (Scott et al., 2003). Looking to the future, a number of clinical trials have been conducted to investigate the therapeutic effects of potential antiepileptogenic drugs. [Epileptogenesis is the pathological process leading to epilepsy. Schachter (2002) and Walker et al. (2002) discuss the case for antiepileptogenic therapy.] Although no effective treatment has emerged so far (Temkin, 2001), success in this pursuit would provide a treatment that might be used to prevent the development of TLE following a PFC. Given the relatively low proportion of PFC cases destined for eventual TLE, long term drug-administration following every occurrence of a PFC may be unacceptable. Given the serious nature of TLE, however, drug administration to a subset of high risk individuals might be appropriate. Growth curve data acquired following a PFC might provide the required prognostic information.

In some paediatric conditions there may be no effective treatment. Nevertheless prognostic models still have a place in clinical practice, since parents frequently express the need to know what the future holds. Invariably, clinicians are required to deal with questions regarding eventual outcome after a medical emergency, even if the outcome is inescapable due to the absence of any effective treatment. Sometimes this information is sought in order to plan for the future. When dealing with hereditary conditions, information of this type is fundamental to decisions regarding future pregnancies. Non-clinicians might find the emphasis on the provision of prognostic information surprising, but it is among the duties that a paediatrician is required to fulfil on a regular basis.

To summarise, random coefficients modelling provides an approach to achieving two objectives. The first of these is a characterisation of the pattern with which pathology evolves within individuals; the second is the development of growth-curve prognostic models. A characterisation

of the time-dependent behaviour of the pathologies under consideration is of interest in its own right from a clinical viewpoint and, in some cases, from a basic neuroscience perspective. The latter applies to those pathologies that lead to the abstruse neurodevelopmental/cognitive deficits referred to earlier. Prognostic models are required primarily as an aid to clinical decision making, but it is anticipated that, in some cases, these models will provide an insight into the processes that underlie some types of cognitive impairment.

Returning to the special circumstances that impact on paediatric studies, small-sample problems are sometimes an inevitable consequence of the rarity of the condition under investigation. Several cognitive abnormalities are rare, for example, and a number of ongoing paediatric studies at the Great Ormond Street Hospital for Children are concerned with conditions in which scarcity places an absolute upper limit on the number of patients that are available for recruitment. This adds to the analytical challenge. The V_{DC}-ADC data presented in Chapters 3 and 4 provide a non-clinical example in which the small sample and sparse data problems are combined, since in some animals the ADC transition was undersampled due to its rapidity. An additional complication arose due to the need to adopt some type of nonlinear model. Thus a full random-coefficients treatment is required to achieve a proper modelling of the between-animal variation in transition behaviour that is central to the analysis, while the form of the DC-potential and ADC responses demanded the use of a nonlinear model. A number of future investigations are envisaged with the same combination of features that were encountered in the V_{DC}-ADC study, namely multivariate, longitudinal and small-sample studies in which the analytical challenge is compounded by a degree of unavoidable undersampling and the need to adopt some form of nonlinear model. These challenges are met in a number of other areas of research, including pharmacokinetics-pharmacodynamics. In particular sparse data are a major problem in many pharmacokinetic studies. This provided the stimulus for the development of improved methods for dealing with these issues. Recently, much of this effort has been directed at the MCMC simulation approach to statistical inference. Relevant aspects of the pharmacokinetic/pharmacodynamic modelling problem are discussed briefly in the following pages.

Referring specifically to the undersampled data problem, a commonly encountered difficulty in pharmacokinetics/pharmacodynamics is the impracticality of acquiring sufficient data on an individual to obtain the required subject-specific kinetic parameter estimates with the necessary precision. In fact it is not uncommon to face the situation in which the data obtained for some subjects are so sparse that the resulting analytical problem is either underdetermined or too ill-conditioned to allow any useful estimates to be obtained (Gelman et al., 1996a). An advantage of the Bayesian analytical approach is that it can be used to turn an ill-conditioned problem into one that is well-conditioned. This is achieved through a combination of two features of the Bayesian formalism. Firstly, the explicit incorporation of prior information on one or more of the model parameters is a mechanism through which constraints can be imposed, together with a consequent improvement in conditioning. In many applications reliable and independent data are available, and this information can be used to generate the required informative prior distributions (Gelman et al., 1996a; Wakefield, 1996). This approach is not always adopted, however, since the pursuit of objectivity will sometimes prohibit the use of the informative priors that are required to provide a sufficient constraint on the poorly identified parameters. On the other hand, prognostic modelling is an application in which informative priors might be used to advantage. A second mechanism through which the Bayesian approach provides an alleviation of the effects of ill-conditioning is through the formal combination of information afforded by the individuals that make up the sample, but taken as a whole in the form of distributional information. Thus, although the data obtained for some individuals might be insufficient to provide parameter estimates, if used in isolation, the distributional information provided by the remaining subjects has a constraining effect that can yield useful estimates for otherwise poorly identified subject-specific parameters. The V_{DC}-ADC study of Chapters 3 and 4 provides an example of this behaviour, which arose because signal-to-noise ratio considerations placed an upper limit on the effective ADC acquisition rate. In those animals in which the ADC transition was fast, the ischaemia-induced transition was undersampled. Consequently, reliable estimates of the ADC transition parameters could not be obtained by fitting the individual ADC time-series data in these cases. A

formal Bayesian analysis provides best estimates of the individual-specific parameters in question, best in the sense of being an optimum weighted average of the estimate that would be obtained from the individual time series, taken in isolation, and the group average. The weighting is determined by the adequacy of the sampling rate and the relative magnitude of the within-subject and between-subject variances. This behaviour is shared by both non-Bayesian and Bayesian random coefficient models. In the former context Laird and Ware (1982, p966-7) described this behaviour as Empirical Bayes. The resulting parameter shrinkage is discussed in Sections 1.4.1.3 and 1.4.2.2, and it is examined in Appendix B in the V_{DC}-ADC study context, where it is shown that the level of shrinkage in the individual-specific ADC transition rate estimates is related to the adequacy with which each transition is sampled. When the data available for a particular transition are sparse the estimated rate is more heavily weighted by the population mean, thus using the group behaviour to moderate the individual-specific value, giving an *improved* (safe) estimate. In contrast, in those individuals in which the response is well defined, the estimate is dominated by the data acquired for that individual.

Another characteristic of the paediatric prognostic and longitudinal data modelling studies that are envisaged is the expectation that nonlinear models will be a necessity. As stated previously, nonlinear random effect models play a major role in pharmacokinetic and pharmacotoxicology research, and given the importance of these research areas to the drug industry, this provided the impetus for a considerable effort, extending over several decades, in a search for useful statistical methods (see, for example, Vonesh, 1992; Chapter 7 in Jones (1993); Wakefield, 1996; Gelman et al., 1996a; Roe, 1997). Much of the earlier work was based on Taylor series expansion (Racine-Poon and Smith, 1990; Vonesh, 1992; Davidian and Giltinan, 1993; Chapter 7 in Jones, 1993). A number of algorithms became established, but each is based on a low-order expansion. The accuracy of the resulting parameter estimates is, therefore, uncertain. Furthermore, the traditional approach to both linear and nonlinear random-effect models is to use the Hessian matrix to approximate the parameter covariance matrix. Large samples are required to ensure the validity of this approximation (see, for example, Sections 4.1 and 9.4 in Gelman et al., 1995; Section 5.2 in

Carlin and Louis, 1996; Vonesh and Carter, 1992; Vonesh, 1992). In summary large-sample theory is central to the standard approach to both linear and nonlinear random effects inferential statistics. Clearly, the application of the resulting asymptotic statistics to small sample studies is dubious. In nonlinear modelling applications the accuracy of the approximation and resulting statistics is difficult to determine in practice (Mager and Göller, 1997). Furthermore, it is not possible, in general, to determine the sample size required for the validity of the asymptotic statistics that are obtained. For this reason the MCMC simulation approach is an attractive alternative.

MCMC offers several advantages over the more traditional random-coefficients regression methods. Firstly statistics can be generated without resorting to large-sample approximation. As indicated previously, a number of ongoing paediatric studies at the Great Ormond Street Hospital for Children involve extremely rare conditions. Small samples are inevitable and asymptotic statistics inappropriate. By adopting an MCMC implementation, full random-coefficient nonlinear models can be adopted, without resorting to approximation. All of the advantages relating to the optimum use of sparse data apply. [Wakefield (1996) and Gelman et al. (1996a) both discuss the advantages of the simulation-based Bayesian approach to sparse data analyses, with specific reference to nonlinear hierarchical models.] An additional and powerful feature of the MCMC approach is that it allows considerable flexibility in the manner in which the primary model parameters are treated. In particular these primary variables can be incorporated into complicated functions, including derivatives, as required to generate derived variables of interest. The V_{DC}-ADC curve analysis of Chapters 4 provides an example. The required posterior intervals are readily obtained from the MCMC output, again avoiding the kind of approximation required in a traditional analysis involving complicated nonlinear functions of the primary variables. It is no exaggeration to suggest that MCMC methods have revolutionised the approach to statistical inference in many areas of research. These methods are expected to become an essential component of some of the paediatric MR research that is in progress at the Great Ormond Street Hospital for Children.

Appendix A. Longitudinal data analysis. A comparison of mixed model regression methods

A.1 Introduction

A number of approaches to longitudinal data analysis are outlined in Chapter 1, with an emphasis on MCMC, as implemented using the Metropolis/Metropolis-Hastings and Gibbs sampler algorithms (Section 1.5.5). In addition, brief accounts of the GEE and traditional mixed model regression methods are included (GEE, Section 1.4.1.5; mixed-model regression, Section 1.4.1.3). During the developmental stages of the Metropolis simulation work outlined in this thesis, several well-known longitudinal datasets were used to assess the performance of the Metropolis algorithm and to compare the resulting MCMC statistics with those obtained using well-established programs based on standard frequentist statistical methods (mixed-model regression and GEE). In addition, a comparison was made with Gibbs sampler results. Both the Potthoff-Roy growth curve data (Potthoff and Roy, 1964; Jennrich and Schluchter, 1986) and the RATS data of Gelfand et al. (1990) were used for this purpose. A selection of results obtained with the latter are shown in Table A.1. Frequentist mixed model regression, GEE and Gibbs sampling results are included in the table, together with the Metropolis simulation results. Gibbs sampling was performed using several types of noninformative prior.

A.1.1 Noninformative prior distributions in random coefficients modelling

Traditionally, Gibbs sampling growth curve hierarchical modelling analyses are performed using normal priors for the fixed-effect and random coefficients, and inverse gamma priors for the variance components (or, equivalently, gamma priors if the model is specified in terms of precision parameters; Gelfand et al., 1990; Spiegelhalter et al., 1995a; Carlin and Louis, 1996, p166 et seq.). In many applications, each of the normal and inverse gamma prior distributions is specified using parameter values chosen to render the resulting distribution noninformative. This

specification is common among the BUGS/WinBUGS example programs provided by the Cambridge MRC Biostatistics Unit (Spiegelhalter et al., 1995b) and this is, no doubt, a major reason their widespread use. A different approach is required, of course, in those applications in which informative priors are a prerequisite. Informative priors are sometimes used, for example, to express current medical opinion in clinical trials (Spiegelhalter et al., 1994). The following applies to those applications in which objectivity demands the use of noninformative priors.

A number of research papers have addressed the issue of variance component prior distributions (Morris and Normand, 1992; Kass and Wasserman, 1996, and references therein; Daniels, 1999). The classical noninformative prior for the normal standard deviation is its inverse, i.e., $p(\sigma) \propto 1/\sigma$ (Box and Tiao, 1973, p29). Alternatively, prior probability might be specified in terms of σ^2, in which case the classical noninformative prior is $1/\sigma^2$ (Lee, 1989, p55). These follow from the requirement that each noninformative prior should satisfy the criterion that the distribution is uniform in a metric in which the likelihood is data translatable (Box and Tiao, 1973, p26; Lee, 1989, p48). The latter requires that the dispersion character of the distribution is unaltered under a location shift. A number of statisticians have pointed out, however, that a prior of the form $1/\sigma^2$ (or $1/\sigma$) can be problematic in some random coefficient modelling applications because $\sigma^2 = 0$ might be supported by a non-negligible likelihood, giving rise to an improper posterior distribution (Dumouchel and Waternaux, 1992; Spiegelhalter, 1995a, Section 9.2; Daniels, 1999; Kass and Wasserman, 1996, pages 1359 and 1361). A similar problem can arise if an inverse gamma prior is used with an extreme scale-parameter value. The latter approach is common in applications requiring vague prior distributions. The resulting minimally informative priors are described as *just proper* (Spiegelhalter et al., 1995a, Section 9.2; Congdon, 2001, p20), but problems can nevertheless occur with very large inverse-gamma scale parameter values. A number of solutions have been suggested including the use of the Pareto prior distribution for the variance components in random coefficient modelling applications (Spiegelhalter et al. 1995a, Section 9.2.2).

A.1.1.1 The Pareto distribution

One solution to the impropriety problem discussed in the preceding paragraph is to adopt the bounded uniform prior $p(\sigma^k) \sim Uni(0,r)$ (Spiegelhalter et al., 1995a, Section 9.2.2). The Pareto distribution

$$\tau \sim Pareto(\alpha,c) \rightarrow p(\tau) = \alpha c^{\alpha} \tau^{-(\alpha+1)}, \tau > c \qquad [A.1]$$

with parameters (k/2, $r^{-2/k}$) generates the required uniform prior. τ ($\tau = 1/\sigma^2$) is the precision parameter. Thus the prior distribution σ^2~Uni(0, r) can be obtained using the specification τ~Pareto(1, r^{-1}), while the prior σ~Uni(0, r) can be generated using τ~Pareto(0.5, r^{-2}). These relationships follow from the *change of variables* rule. Thus

$$\begin{aligned} p(\sigma^2) &= p(\tau)\left(\frac{d\tau}{d\sigma^2}\right) \\ &\propto (\sigma^2)^{\alpha+1}(\sigma^2)^{-2} \\ &= (\sigma^2)^{\alpha-1} \end{aligned} \qquad [A.2]$$

which is a constant given $a = 1$, while

$$\begin{aligned} p(\sigma) &= p(\tau)\left(\frac{d\tau}{d\sigma}\right) \\ &\propto \sigma^{2\alpha+2}\sigma^{-3} \\ &= \sigma^{2\alpha-1} \end{aligned} \qquad [A.3]$$

which is a constant given $a = 0.5$. These two variance component prior distributions are among several that were examined in the RATS Gibbs-sampling trial analyses.

A.2 Methods

The Metropolis-based MCMC simulation method is described in detail in Section 2.2. The Metropolis results are derived from three parallel chains each consisting of 5000 samples obtained using a run-time thinning ratio of 1:400. GEE, classical mixed model regression and Gibbs sampler analyses were performed using SAS PROC GENMOD (Version 6.12), SAS PROC MIXED (Version 6.12) and WinBUGS (Version 1.3), respectively. Each of the Gibbs sampler results is derived from a single chain of length 10000, not including a burn-in run of 2000 samples. The Gibbs sampler results listed in Table A.1 are naïve in the sense that the single chain

MCMC output was used in the absence of formal convergence testing. Visual inspection of the single-chain trace plots suggested satisfactory behaviour, however, as indicated by a rapid movement over parameter space. The following BUGS code was used to perform the Gibbs sampling analysis with τ~Pareto(1, r^{-1})

```
model
{
for( i in 1 : N )
  {
  for( j in 1 : T )
    {
    Y[i , j] ~ dnorm(mu[i , j],tau.c)
    mu[i , j] <- alpha[i] + beta[i] * x[j]
    }
  alpha[i] ~ dnorm(alpha.c,alpha.tau)
  beta[i] ~ dnorm(beta.c,beta.tau)
  }
tau.c ~ dpar(1.0,1.0E-6)
sigma <- 1 / sqrt(tau.c)
alpha.c ~ dnorm(0.0,1.0E-6)
alpha.tau ~ dpar(1.0,1.0E-6)
beta.c ~ dnorm(0.0,1.0E-6)
beta.tau ~ dpar(1.0,1.0E-6)
}
```

while the analysis based on τ~Pareto(0.5, r^{-2}) was performed using the code

```
model
{
for( i in 1 : N )
  {
  for( j in 1 : T )
    {
    Y[i,j] ~ dnorm(mu[i,j],tau.c)
    mu[i,j] <- alpha[i] + beta[i] * x[j]
    }
  alpha[i] ~ dnorm(alpha.c,alpha.tau)
  beta[i] ~ dnorm(beta.c,beta.tau)
  }
tau.c ~ dpar(0.5,1.0E-6)
sigma <- 1 / sqrt(tau.c)
alpha.c ~ dnorm(0.0,1.0E-6)
alpha.tau ~ dpar(0.5,1.0E-6)
beta.c ~ dnorm(0.0,1.0E-6)
beta.tau ~ dpar(0.5,1.0E-6)
}
```

A.3 Results and Discussion

A.3.1 RATS growth curve parameter estimates and standard errors

The population intercept and slope median/mean estimates given by each of the methods are close, relative to their respective standard errors, with the exception of the GEE[AR(1)] intercept (Table A.1). The standard error estimates are similar, with the result that the 95% posterior intervals/confidence intervals are close, with the exception of the interval given for the intercept by the GEE[AR(1)] calculation, which is shifted to a lower value due to the shift in the mean estimate.

A.3.2 Sensitivity to changes in the form of the precision-parameter priors

The RATS posterior mean, median and 95% posterior interval estimates obtained with Gibbs sampling are insensitive to changes in the form of the noninformative precision/variance parameter priors (Table A.1). This insensitivity might be expected because the two random coefficient precision parameter estimates are well removed form zero. For example, using priors of the form $\tau \sim \text{Pareto}(1.0, 10^{-6})$ for each of the precision parameter priors, the Gibbs sampler 95% posterior intervals for τ_α and τ_β are 0.0041 to 0.018 (median, 0.0083) and 1.84 to 7.09 (median, 3.66), respectively. The remaining four precision-parameter priors led to 95% posterior intervals further removed from zero, although the differences between the five sets of simulation results were negligible in this respect. (The remaining four Gibbs sampler simulations generated 95% posterior interval lower limits in the range of 0.0043 to 0.0049 for τ_α and 1.96 to 2.12 for τ_β.)

TABLE A.1. **A comparison of methods for longitudinal data analysis. RAT growth curve parameter estimates, standard errors, 95% posterior intervals and confidence intervals**

Parameter	Method	Estimate		Std. Error	2.5% limit	97.5% limit
		Mean	Median			
Intercept	MM-reg (diag)	106.6		2.24	102.0	111.1
	MM-reg(un)	106.6		2.30	101.9	111.3
	GEE[ind]	106.6		2.26	102.1	111.0
	GEE[ar(1)]	103.7		2.15	99.5	108.0
	Gibbs [U(logσ)]	106.6	106.7	2.27	102.2	111.0
	Gibbs [U(logσ^2)]	106.6	106.6	2.29	102.1	111.2
	Gibbs τ~Par(1,10^{-6})	106.6	106.6	2.38	101.9	111.2
	Gibbs τ~Par(0.5,10^{-6})	106.5	106.6	2.33	102.0	111.1
	Gibbs τ~Gam(10^{-3},10^{-3})	106.6	106.6	2.28	102.0	111.0
	Metropolis	106.5	106.7	2.28	101.4	110.5
Slope	MM-Reg (diag)	6.19		0.10	5.98	6.40
	MM-Reg(un)	6.19		0.11	5.97	6.40
	GEE[ind]	6.19		0.10	5.98	6.39
	GEE[ar(1)]	6.17		0.10	5.97	6.36
	Gibbs [U(logσ)]	6.18	6.18	0.11	5.97	6.39
	Gibbs [U(logσ^2)]	6.18	6.18	0.11	5.98	6.39
	Gibbs τ~Par(1,10^{-6})	6.19	6.19	0.11	5.97	6.40
	Gibbs τ~Par(0.5,10^{-6})	6.19	6.19	0.11	5.97	6.40
	Gibbs τ~ Gam(10^{-3},10^{-3})	6.19	6.19	0.11	5.98	6.39
	Metropolis	6.18	6.18	0.11	5.97	6.40

Two REML mixed-model regression (MM-reg) analyses were performed, one using a diagonal random-coefficients covariance matrix (diag), the other assuming an unstructured covariance matrix structure (un). GEE calculations were performed using two alternative working correlation matrix structures, namely an independent structure [ind] and a first-order autoregressive structure [ar(1)]. (A GEE calculation based on an unstructured working correlation matrix failed to converge.) Among the Gibbs sampler simulation results are those obtained with variance component priors $p(\log\sigma) \propto$ constant [abbreviated U(logσ)], which is equivalent to $p(\sigma) \propto 1/\sigma$, and $p(\log\sigma^2) \propto$ constant [abbreviated U(logσ^2)], which is equivalent to $p(\sigma^2) \propto 1/\sigma^2$.

Appendix B. Parameter estimator shrinkage. A comparison of nonlinear regression and MCMC results

Fig. B.1 shows six individual normalised ADC time courses taken from Chapter 3 (see Figs. 3.1 and 3.2), together with a set of 100 random time-course profiles generated by sampling the posterior distribution. Superimposed on each figure are two individual-specific fitted ADC profiles, one based on the nonlinear regression parameter estimates obtained by fitting each ADC profile individually, the other derived from the MCMC random-coefficient medians, as obtained by modelling all ten V_{DC}-ADC response profiles simultaneously (see Chapter 3). The corresponding transition rate estimates are included in the figure. These individual-specific median transition rates are listed again in Table B.1, together with their 2.5% and 97.5% posterior quantiles. The population median normalised ADC transition rate is -1.18 min^{-1}, with 2.5% and 97.5% posterior quantiles at -4.69 and -0.45, respectively. The subset of results presented in Fig. B.1 were selected to demonstrate a fundamental property of the individual-specific parameter estimates provided by a Bayesian random coefficients analysis, namely parameter shrinkage. This behaviour is discussed in Section 1.4.1.3 (in relation to the empirical Bayes property of mixed model regression estimators), in Section 1.4.2.2 (in the context of Bayes estimators), and again in Section 4.4.4 with reference to the sparse data problem. Referring to the sparse data issue, a

---

Fig. B.1. Overlaid plots showing the normalised ADC data acquired from each of six animals (continuous irregular line), together with a pair of individual-specific calculated ADC response profiles, one based on the nonlinear regression parameter estimates obtained by fitting each ADC profile individually (broken black curve), the other derived from the MCMC random-coefficient medians, as obtained by modelling all ten V_{DC}-ADC response profiles simultaneously (continuous black curve; the MCMC median curve generated for animal 10 is visually indistinguishable from the nonlinear regression profile). The MCMC and nonlinear-regression transition-rate estimates are included in each panel. Also shown is a set of 100 animal-specific random profiles obtained by sampling the posterior distribution (grey broken lines).

Fig. B.1. **ADC transition-rate parameter shrinkage**

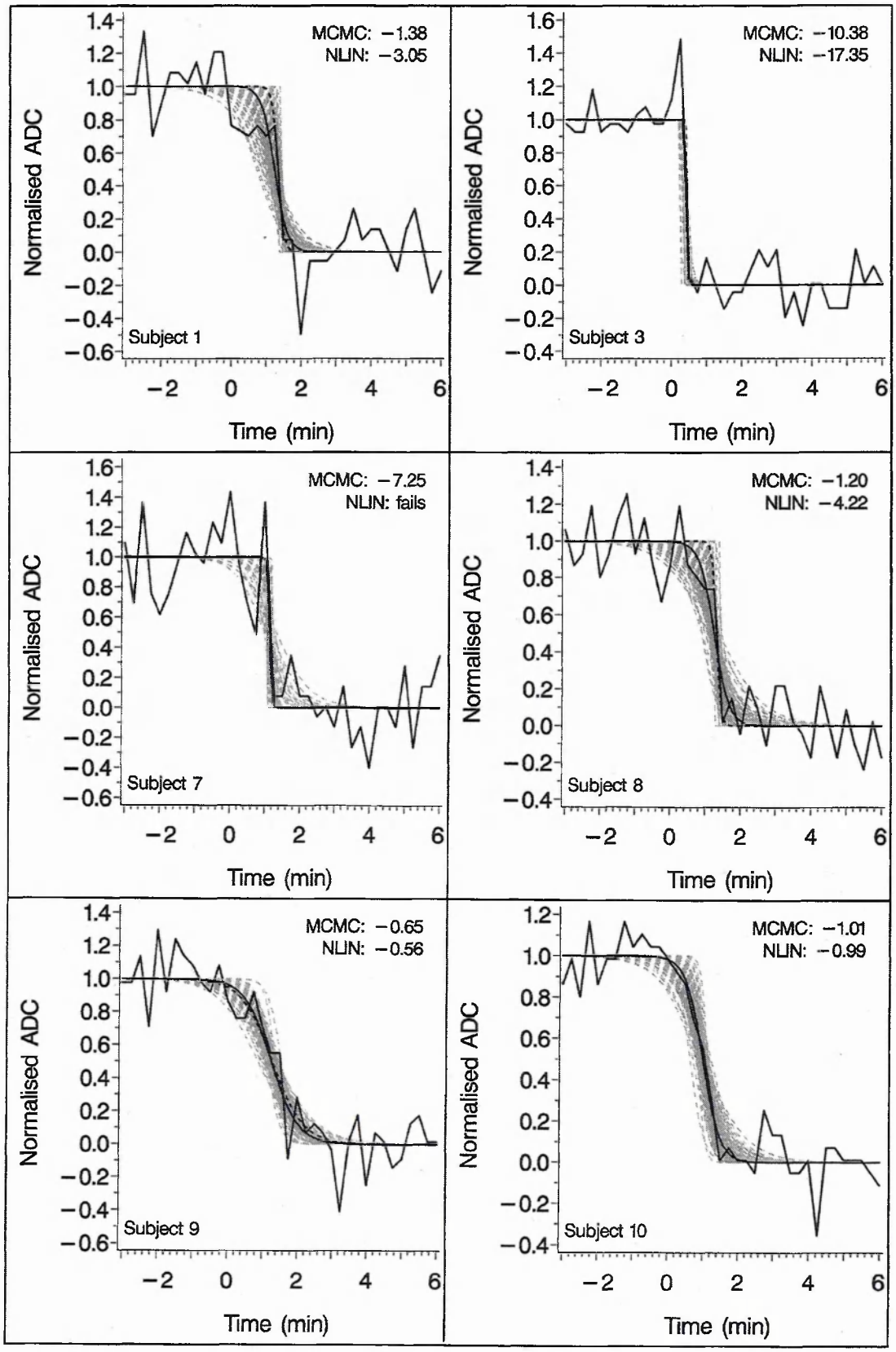


Figure legend on the preceding page.

Table B.1. **Comparison of the ADC median transition rates obtained by MCMC simulation and the values given by a nonlinear regression analysis of the individual profiles**

Animal ID and Method	2.5% quantile (min^{-1})	Median (estimate) (min^{-1})	97.5% quantile (min^{-1})
S1 mcmc	-37.53	-1.382*	-0.585
S1 nlin	-11.740	-3.013 (-3.053)	-0.790
S3 mcmc	-229.9	-10.38*	-3.629
S3 nlin	No estimate	-17.35 (-17.35)	No estimate
S8 mcmc	-29.663	-1.1955*	-0.416
S8 nlin	-31.033	-4.133 (-4.217)	-0.568
S9 mcmc	-1.459	-0.653	-0.395
S9 nlin	-0.999	-0.556 (-0.559)	-0.313
S10 mcmc	-2.743	-1.009	-0.504
S10 nlin	-1.711	-0.987 (-0.991)	-0.574

The MCMC individual-specific normalised ADC transition rate estimates were calculated using $\beta_{4i} / 4 = \exp(\alpha_3 + \gamma_{3i}) / 4(\alpha_4 + \gamma_{4i})$. Section 3.2.2.3 provides notation and regression model details. The normalised nonlinear regression median rate estimates and quantiles (nlin) were generated by first sampling from a MVN distribution with mean vector equal to the nonlinear regression estimate, and covariance matrix calculated using the asymptotic standard errors and the asymptotic correlation matrix provided by the nonlinear regression program. The resulting sample was subsequently used with the preceding expression for rate to generate a second stage random transition-rate sample, and thus to obtain the required statistics. The estimates in brackets were calculated using the nonlinear regression parameter estimates. The population median rate is -1.18 min^{-1}.

*The ADC transition is inadequately sampled and the MCMC estimate has shrunk towards the population median.

dominant feature of the ADC data acquired in the focal ischaemia study presented in Chapters 3 and 4 is an undersampling of the ADC transition in a subset of animals, due to the rapidity of the transition relative to the sampling rate. Fig. B.1 shows that in those cases in which an adequate sampling of the ADC transition is obtained (animals 9 and 10), the discrepancy between the nonlinear regression and MCMC transition rate estimates is not marked. In contrast, when the transition is poorly sampled (animals 1, 3 and 8) a greater discrepancy is observed with the MCMC estimates exhibiting shrinkage towards the population median.

These results serve to illustrate the point that a formal random coefficients modelling analysis yields parameter estimates and associated statistics based on an optimum use of the information provided by the experiment. Each subject-specific coefficient is a weighted average of the value obtained by fitting the response profiles individually and the mean value, with a weighting determined by the sampling adequacy and the relative magnitude of the within-subject and between-subject variances. In those individuals in which the response is well defined, the estimate is dominated by the data acquired for that individual, while in those cases in which relatively little information is available, the mean behaviour is used to modify the individual-specific value to give an improved estimate. It should be noted in this context that the V_{DC}-ADC analysis was performed using uniform priors for the fixed- and random-effect coefficients, thus ensuring that all parameter estimates are dominated by the data. Parameter shrinkage behaviour is shared by both non-Bayesian and Bayesian random coefficient models. In the former context Laird and Ware (1982) describe the behaviour as Empirical Bayes (see pages 966-7).

Random coefficients modelling is especially advantageous when dealing with models that are underdetermined or ill-conditioned in an ordinary regression sense. For example, the V_{DC}-ADC dataset includes several cases that are not amenable to individual analysis, and attempts to fit these individual DC-potential and ADC profile pairs ended in failure, either because the ADC transition is corrupted with noise and/or is sampled inadequately or because the biphasic nature of the DC-potential response is poorly identified. For example, in one case (animal 7) the nonlinear

regression routine failed to converge due to an inadequate sampling of the ADC transition, caused by its rapidity coupled with noise corruption. The DC-potential data acquired from animal 8 are problematic because the response lacks the marked biphasic character of the remaining DC-potential profiles (see Fig. 3.21 and Fig. 3.27, top left panel), with the result that the two-term logistic model is extremely ill-conditioned, as applied to these data in isolation. On the other hand, the Bayesian random coefficients model succeeded in providing the required subject-specific posterior median estimates and associated intervals. This is attributable to the proper use of distributional information derived from the sample as a whole. Thus, a formal modelling of the distribution of each of the random coefficients influences the individual-specific parameter estimates and moderates the effect of spurious observations. This property is referred to as information borrowing.

Despite information borrowing, some random coefficient models remain ill-conditioned or underdetermined. Bayesian random coefficient models (hierarchical models) offer an additional mechanism for dealing with the problem, namely the use of informative priors. Ill-conditioned random coefficient models are, for example, a common occurrence in pharmacokinetics. In some cases this is due to undersampling; sometimes it arises because the physiological model under investigation is fundamentally under-identified. It has been suggested that informative priors derived from well established physiological data provide a useful solution to this problem (Gelman et al., 1996a). Thus, the constraints imposed by adopting carefully constructed informative priors provides a mechanism for turning an ill-conditioned problem into one that is well-conditioned (see Section 1.4.3).

Appendix C. Hierarchical centering

C.1 Introduction

Sections 1.6.3 and 1.6.4 of the Introduction deal with the convergence and poor performance problems that are a common occurrence in MCMC simulation work. A variety of solutions have been proposed, a summary of which is included with an emphasis on hierarchical centering in random coefficients modelling applications. This appendix demonstrates the substantial improvement in performance that can be achieved by adopting a hierarchical centered formulism. To this end a well-known multilevel dataset is used to compare the Gibbs sampler output obtained with a naïve, non-centered random-coefficients model with the output generated using an equivalent hierarchical centered model specification. The example is a *variance components* problem described by Littell et al. (1996, p155) and is based on a study, performed in a semiconductor plant, to establish the cause of variability in semiconductor oxide thickness. The experimental design stipulated that 3 silicon wafers were drawn from each of 8 randomly selected *lots* (each lot consisting of 25 wafers). Oxide thickness was measured at 3 randomly selected *sites* on each wafer. Following Littell et al. (1996) the model is

$$y_{ijk} = \mu + \alpha_i + w_{j(i)} + s_{k(ij)}, \quad i = 1,2,...,8, \quad j = 1,2,3, \quad k = 1,2,3, \qquad [C.1]$$

where $a_i \sim N(0, \sigma_l^2)$ is the random *lot* effect, $w_{j(i)} \sim N(0, \sigma_w^2)$ is the random *wafer* variation and $s_{k(ij)} \sim N(0, \sigma_s^2)$ the random site (residual within wafer) variation.

C.2 Gibbs sampling (WinBUGS) simulation

The non-centered simulation analysis was performed using the WinBUGS code

```
model
{
for(i in 1: lots)
  {
  for( j in 1 : wafers)
    {
    for( k in 1 : sites)
      {
      y[i , j, k] ~ dnorm(p[i,j], tau.site)
```

```
      }
    p[i,j]<-mu + l[i] + w[i,j]
    w[i,j] ~ dnorm(0, tau.wafer)
    }
  l[i]~dnorm(0,tau.lot)
  }
sigma2.lot <- 1 / tau.lot
sigma2.wafer <- 1 / tau.wafer
sigma2.site <- 1 / tau.site
tau.lot ~ dgamma(0.001, 0.001)
tau.wafer ~ dgamma(0.001, 0.001)
tau.site ~ dgamma(0.001, 0.001)
mu ~ dnorm(0.0, 1.0E-10)
}
```

while the hierarchical centered model was implemented using the WinBUGS code

```
model
{
for(i in 1: lots)
  {
  l[i]~dnorm(mu,tau.lot)
  for( j in 1 : wafers)
    {
    w[i,j] ~ dnorm(l[i], tau.wafer)
    for( k in 1 : sites )
      {
      y[i , j, k] ~ dnorm(w[i,j], tau.site)
      }
    }
  }
sigma2.lot <- 1 / tau.lot
sigma2.wafer <- 1 / tau.wafer
sigma2.site <- 1 / tau.site
tau.lot ~ dgamma(0.001, 0.001)
tau.wafer ~ dgamma(0.001, 0.001)
tau.site ~ dgamma(0.001, 0.001)
mu ~ dnorm(0.0, 1.0E-10)
}
```

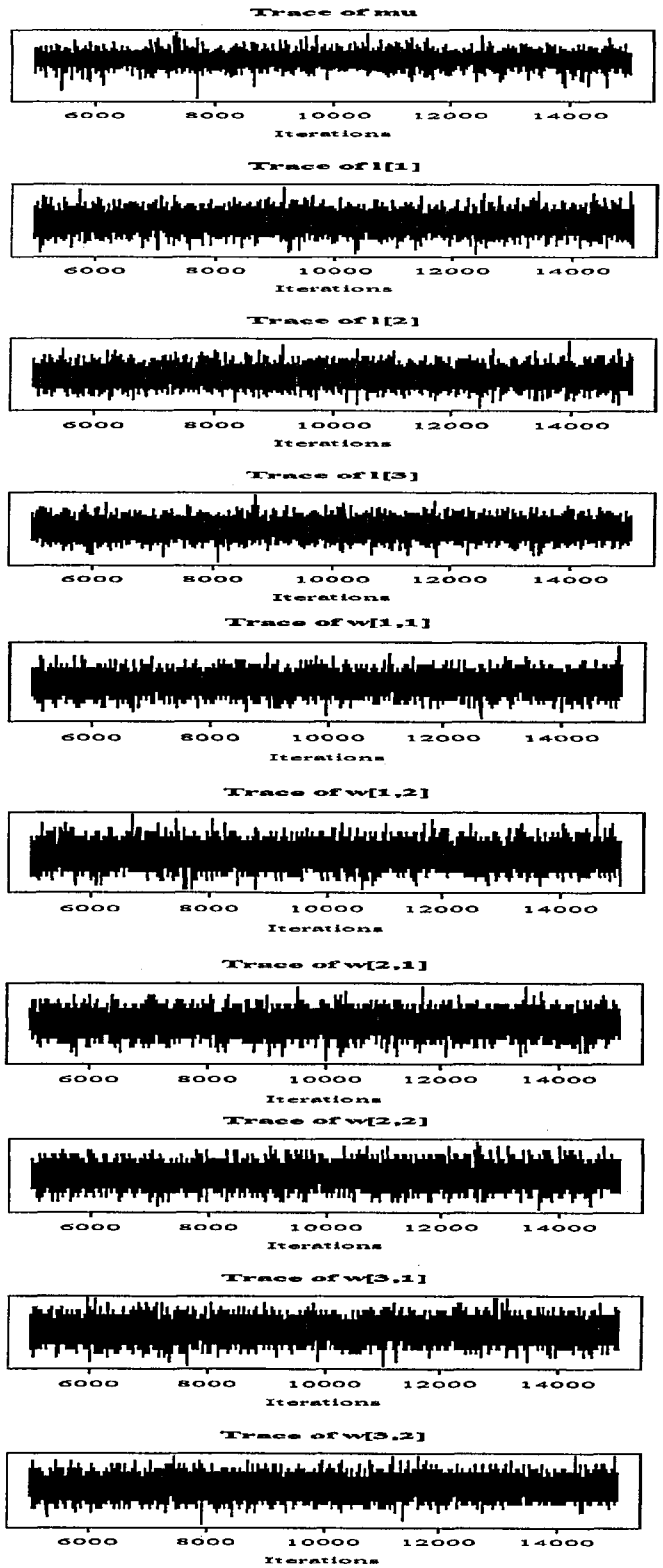
A single chain of 10000 samples was generated for each model subsequent to a 5000 sample burn-in run.

C.3 Results and Discussion

Figure C.1a shows the MCMC chain trace plots obtained for ten selected parameters, as generated using the hierarchical centered model. The corresponding trace plots generated without

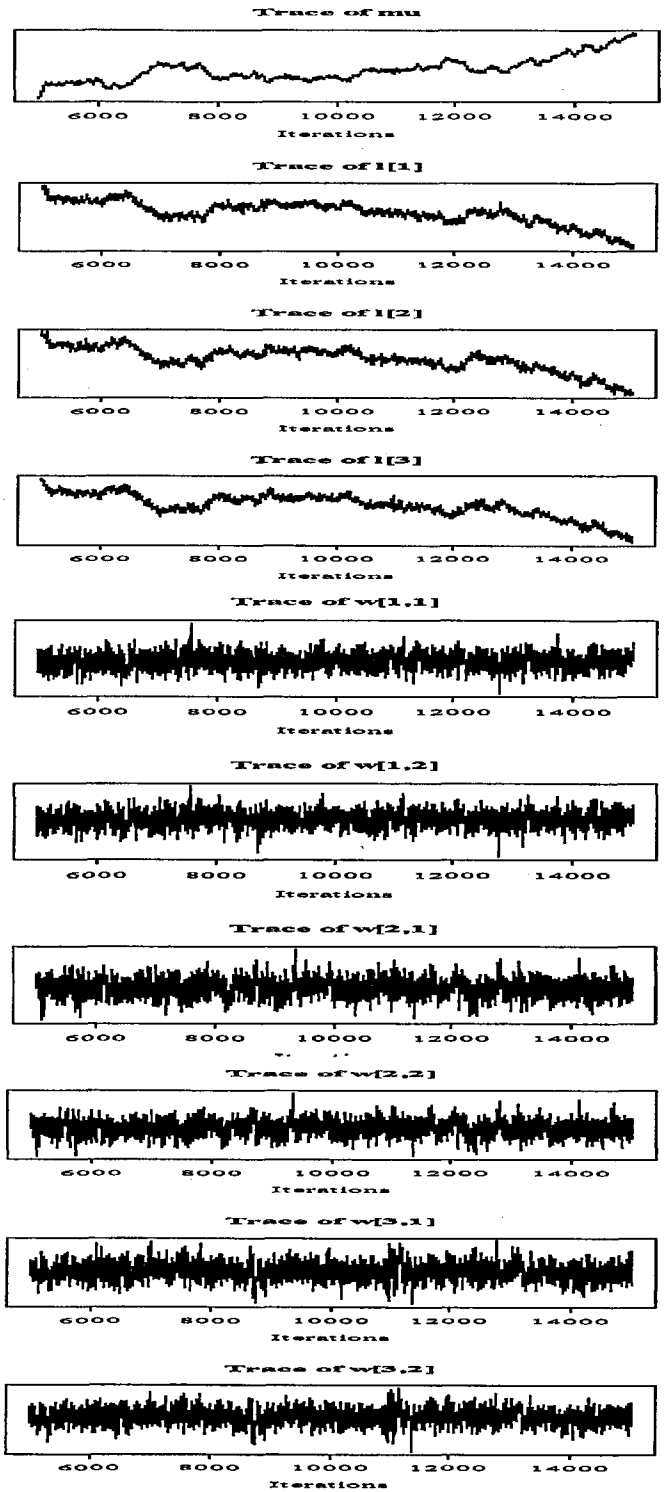
hierarchical centering are shown in Fig. C.1b. The latter indicate an extremely slow movement over parameter space, and are typical of the badly behaved chains encountered with non-centered random-coefficient models. The hierarchical centered simulation yields a dramatic improvement in parameter space coverage. This visual impression of the enhanced performance afforded by hierarchical centering is reflected in the within-chain autocorrelation data listed in Table C.1. The naïve, non-centered model within-chain lag(5) autocorrelations are substantial with some close to unity while, in contrast, the hierarchical centered model yields negligible lag(5) autocorrelation coefficients. The centered-model single-chain variance component estimates are consistent with those provided by Littell et al. (1996). [MCMC median estimates (95% posterior intervals in parenthesis): $\sigma_l^2 = 138.0$ (44.7, 562.6); $\sigma_w^2 = 37.1$ (17.4, 88.6); $\sigma_s^2 = 12.8$ (8.7, 20.1). SAS PROC MIXED (Littell et al., 1996, p156): $\sigma_l^2 = 129.9$; $\sigma_w^2 = 35.9$; $\sigma_s^2 = 12.6$.]

Fig. C.1a. **A comparison of centered and non-centered multilevel-model Gibbs sampler performance. Centered model output.**



Trace plots are shown for ten selected parameters (μ, l_i, i=1, 2, 3; w_{ij}, i=1, 2, 3 and j=1, 2), as obtained using the hierarchical centered model.

Fig. C.1b. **A comparison of centered and non-centered multilevel-model Gibbs sampler performance. Non-centered model output.**



Trace plots are shown for ten selected parameters (μ, l_i, i=1, 2, 3; w_{ij}, i=1, 2, 3 and j=1, 2), as obtained using the non-centered model.

Table C.1. **Multilevel model MCMC simulation performance. A comparison of the Gibbs sampler within-chain autocorrelation coefficients obtained with and without hierarchical centering**

		$l[1]$	$l[2]$	$l[3]$	$w[1,1]$	$w[1,2]$	$w[2,1]$	$w[2,2]$	$w[3,1]$	$w[3,2]$
C	lag(1)	0.123	0.153	0.112	0.080	0.128	0.050	0.059	0.054	0.047
	lag(5)	0.002	0.012	0.003	-0.019	0.024	0.013	-0.010	3.0e-4	-0.006
NC	lag(1)	0.997	0.997	0.997	0.764	0.771	0.768	0.772	0.775	0.766
	lag(5)	0.987	0.987	0.986	0.506	0.515	0.537	0.532	0.531	0.520

The lag 1 and lag 5 within-chain autocorrelation coefficients are listed for nine selected parameters (l_i, i=1,2,3; w_{ij}, i=1,2,3 and j=1,2), as obtained using the centered (C) and non-centered (NC) hierarchical models.

Appendix D. p-Values as evidence against the null hypothesis. The inverse probability problem

D.1 Introduction

The motivation for this appraisal of p-values and their use as evidence against the null hypothesis is the realisation that their misuse might be widespread among clinicians, with a resulting potential to mislead medical practitioners involved in making important clinical decisions. On occasions these decisions are life determining. In each of the four tertiary referral centres in which the candidate has worked, clinicians undertake a dual research and clinical role. In this setting clinicians are faced with unusual and challenging problems, and current research has a potential to impact on patient management. Consequently, a misunderstanding of some aspects of statistical inference is potentially harmful. An abundance of statistical texts have been written aimed specifically at biomedical and clinical researchers, but many simply add to the problem. An impression of the nature of the *p-value problem* is provided by the following quotes. Goodman (1999a) published a paper on *the p-value fallacy* in which he asserts that the common statistical procedure, based on p-values *is widely perceived as a mathematically coherent approach to inference.* But it *is an amalgam of incompatible elements, whose utility for scientific inference has been the subject of intense debate among statisticians for almost 70 years*. Browne (1995) discusses the concept of *probability* as a measure of belief and concludes that the *problem is that neither classical nor Bayesian methods are able to provide the kind of answers clinicians want. That classical methods are flawed is undeniable - I wish I had an alternative*. Edwards (1992, p180) stated that w*hat used to be called judgement is now called prejudice, and what used to be called prejudice is now called a null hypothesis ... it is dangerous nonsense (dressed up as the 'scientific method'), and will cause much trouble before it is widely appreciated as such.* Jeffreys (1980, p453) states that *I have always considered the arguments for the use of P absurd. They amount to saying that a hypothesis that may or may not be true is rejected because a greater departure from the trial value was improbable; that is, that it has not predicted something that has not happened.* Several solutions

have been suggested, aimed at clinical and epidemiological audiences. The purpose of this appendix is to provide a brief outline of the relevant literature.

D.2 Mathematical and subjective probability. Deductive and inductive reasoning

A misunderstanding of the distinction between mathematical and subjective probability underlies the common misuse of the p-value. The former refers to the objective or physical probability that applies to sequences of events. It has little meaning in the context of single events. Mathematical probability is underpinned by a set of consistent axioms and a probability calculus. In contrast, subjective probability refers to belief and relative degrees of belief. Subjective probability applies to any declarative statement, including hypotheses, and it applies to single events.

Fundamental to understanding the role of the p-value in scientific inference is the distinction between *deductive* and *inductive* reasoning. The distinction is important because Fisher's significance test forms a basis for an inductive scientific philosophy while the Neyman-Pearson hypothesis test is deductive. These two approaches are distinct and, possibly, irreconcilable. Yet the common approach to statistical inference in biomedical research is based on some improper combination of the two. Deductive reasoning involves a flow of logic from the hypothesis (the null hypothesis, for example) to the observations. In contrast, the inductive process involves a reasoning in which the flow of logic is from a particular observation to some general theory, hypothesis or fundamental truth.

D.3 Experiments and statistical inference. Bayes, Fisher, Neyman & Pearson

The majority of experimental biomedical studies, including some phase I and phase II clinical trials, involving statistical inference are of a type in which the researcher seeks to determine the "truth" in relation to some hypothesis, i.e., the probability of the hypothesis, conditional on the data

provided by the experiment. It appears that the majority of clinical researchers work under the assumption that an absolute probability can be provided. Unfortunately, this goal is unattainable. Given prob($D|H$), where D are the data provided by the experiment and H is the hypothesis, the researcher seeks to make inductive statements of inverse probability (i.e., prob($H|D$). The resulting probability is subjective and necessarily uncertain.

Although Bayes theorem provides a formal method for performing probability inversion, several alternative approaches to statistical inference have been suggested. Popper (1974) dismissed formal induction and argued that scientific inference should be entirely deductive. He worked within a refutation/falsification framework in which the researcher makes predictive statements, constructed by deduction and based on the theory or hypothesis under investigation. Testable predictions are central to the process; these are compared with experimental observations. If the hypothesis passes every test it is temporarily retained. On the other hand it is discarded if the comparison results in falsification.

Fisher's significance test

Fisher advocated the significance test [as distinct from the NP *acceptance* procedure (see Fisher, 1973, p79 et seq.)] as a basis for statistical inference in the natural sciences , based on ideas that were formulated and developed over several decades, starting in earnest when he moved to the Rothamsted Experimental Research Station in 1919 (Green, 2003). Inductive reasoning was central to his view of the scientific method. He believed that reasoned argument should be used, together with relevant experimental evidence, to make the case for some general theory. Although Fisher used the p-value as a measure of evidence, Fisher's p-value is subjective; it has no meaning as an absolute probability and their is no obligation to compare the p-value with some predefined critical value. A sufficiently small p-value is presented as evidence against some (null) hypothesis, and it is used in conjunction with other information to present a reasoned argument for some proposal. The procedure is necessarily subjective and distinct from the deductive approach outlined by Neyman-Pearson. The latter is based on a formal hypothesis testing (acceptance test) philosophy.

Neyman-Pearson hypothesis testing

Neyman and Pearson (1933) were motivated to provide a rule, based on a statistical test, for deciding whether a hypothesis should be rejected. Their procedure requires that the null (H_0) and alternative hypotheses (H_1) are specified, together with the Type I error rate (α, the probability that the null hypothesis is rejected when true, under repetition of the process), and Type II error rate (β, the probability that the alternative hypothesis is rejected when true, under replication). Finally, a critical region is defined for the statistic on which the test is based. The Neyman-Pearson (NP) deductive approach replaces Fisher's subjective measure of evidence with a strict decision rule, based on the frequencies of the possible outcomes under repetition of the experiment. NP clarify the point by stating (1933, p291) that *as far as a particular hypothesis is concerned, no test based upon the theory of probability can by itself provide any valuable evidence of the truth or falsehood of that hypothesis*. NP further state that (p291) *without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviour with regard to them, in following which we insure that, in the long run of experience, we shall not be too often wrong*. The NP decision rule *tells us nothing as to whether in a particular case H is true... But... if we behave according to such a rule, then in the long run we shall reject H when it is true not more, say, than once in a hundred times, and in addition we may have evidence that we shall reject H sufficiently often when it is false.*

The Fisher Neyman-Pearson debate

Fisher and NP debated fiercely the relative merits of the philosophies underlying their competing methods, a debate that continues to receive attention (see Lehmann, 1993). It is extraordinary, therefore, that these two apparently irreconcilable methods appear to have become amalgamated. In a paper aimed at psychologists, Macdonald (1997, p333) states that *what has become institutionalized in psychology is not Fisherian statistics. It is an incoherent mishmash of some of Fisher's ideas on the one hand and some of the ideas of Neyman and Pearson on the other.* The reason for this adulteration is, of course, that the p-value underlies both methods.

Fisher was strongly opposed to the notion of a decision-making rule as part of the scientific process. In a comment on the difference between the NP philosophy and his own position, Fisher (1973, p80) stated that *the differences between these two situations seem to the author many and wide, and I do not think it would have been possible to overlook them had the authors* (NP) *of this reinterpretation had any real familiarity with work in the natural sciences.* In addition, he suggested (Fisher, 1935, p40) that *although some uncertain inferences can be rigorously expressed in terms of mathematical probability, it does not follow that mathematical probability is an adequate concept for the rigorous expression of uncertain inferences of every kind.* He clarified his position with the statements (Fisher, 1935, p40) that *inferences of the classical theory of probability are all deductive in character. They are statements about the behaviour of... samples, ... drawn from populations which are fully known. ...A mathematical quantity of a different kind .. appears to take its place .. as a measure of rational belief when we are reasoning from the sample to the population.* Thus it has been suggested that (Macdonald, 1997, p337) *the extent to which a statistical inference applies to a new situation goes beyond statistical theory. Those wishing to generalize from one situation to another are forced into subjective judgements which depend on their knowledge, experience and intuition.* Fisher (1973, p50) stated that *in choosing the grounds upon which a general hypothesis should be rejected, personal judgement may and should properly be exercised.* In addition, he made the assertion that *mathematical probability is inadequate to express the nature and extent of our uncertainty in the face of certain types of observational material, while in all cases the concept of mathematical likelihood will supply very helpful guidance, if we are prepared to give up our irrational urge to express ourselves only in terms of mathematical probability* (Fisher's correspondence with D.J. Finney, 1954; taken from Bennett, 1990, p92). Others have expressed a similar view, including Macdonald (1997, p343), who states that *the Fisherian approach ... views statistical testing as an aid to scientific inference and not as a characterization of it. By purporting to account for the entire process of belief revision in terms of probability models Bayesian approaches attempt the impossible and necessarily fail.* Fisher's hostility to the entire NP method is epitomised in his rejection of the *concept that the scientific worker can regard himself as an inert item in a vast co-operative concern working according to*

accepted rules, and the notion of the worker's *supposed duty mechanically to make a succession of automatic decisions, deriving spurious authority from the ... mathematics of the theory of decision functions* (Fisher, 1973, p104).

Fisher and NP were clear that central to their disagreement was the distinction between inductive behaviour (based on the accept/reject decision) and inductive inference (see Fisher, 1973, p105; Goodman, 1993, p488; Lehmann, 1993, p1243, and references therein for additional information). Fisher was confident regarding the validity of his approach to statistical inference as it applies to the natural sciences. Nevertheless, he expressed some reservation regarding the role of extreme values (i.e., the tails of the distribution). He stated that *this feature is indeed not very defensible save as an approximation* (Fisher, 1973, p71). Furthermore, it might appear a strange contradiction that Fisher should argue forcibly for the inductive approach in which the evidence provided by an experiment is used in conjunction with other relevant information to reach a conclusion, and yet he rejected the formal Bayesian treatment of prior information (Fisher, 1970, pages 9 and 20-1).

Among the ramifications arising from the present-day amalgamation of the Fisher and NP methods is a misconception regarding the p-value as a measure of evidence. Some researchers appear, for example, to suppose a legitimate dual role for the p-value, based on an interpretation of the observed (exact) p-value as a post-study error rate (i.e., an observed error rate) and, simultaneously, as an inductive measure of evidence. Among the important issues is the validity of this interpretation. Goodman (1993, p489-490) argues that the link between the p-value and the Type I error rate is broken by quoting exact p-values, and that the (exact) p-value is not an error rate. But statisticians disagree over this point (see, for example, the discussion involving Hinkley, 1987; Casella and Berger, 1987b, p134; Berger and Sellke, 1987b, p136; Berger and Delampady, 1987a, p 329, 1987b, p 348; Cox, 1987). Related to this issue is the connection between the p-value and power in hypothesis testing, and specificity/sensitivity in diagnostic testing. Several researchers have discussed the relationship between diagnostic and statistical tests (Diamond and Forrester, 1983; Browner and Newman, 1987; Goodman, 1993, p489, 1999a; Poole, 2001, p292). For

example, it has been stated that the p-value is the false-positive rate (1-specificity). But this statement assumes the validity of the error rate interpretation of an exact p-value (see Table 2 in Diamond et al., 1983).

A related issue that has been addressed by a number of statisticians concerns the possibility, or otherwise, of reconciling the Fisher and NP methods, and thus to achieve some kind of unification, despite the apparent contradiction that lay at the centre of the NP-Fisher controversy. Among the texts that discuss this issue is the paper by Lehmann (1993), who concludes that components of the two approaches might be adopted to produce a unified method.

D.4 Do p-values overstate the evidence against the null hypothesis?

Recognising that a majority of clinical/biomedical researchers appear to use p-values as a measure of evidence, the question arises regarding the strength of evidence afforded by a given p-value. Among numerous papers on *using the p-value as a measure of evidence against the null hypothesis*, Berger and Sellke (1987a[1]) have examined the problem in considerable detail. The following is a summary of some of their key results. (Additional information is provided by Berger and Delampady, 1987a[1].) Given the competing hypotheses $H_0{:}\theta = \theta_0$ and $H_1{:}\theta \neq \theta_0$, together with their prior probabilities (π_0, $0 < \pi_0 < 1$ for H_0 and $\pi_1 = (1\text{-}\pi_0)$ for H_1), BS_1 proceed by supposing that the mass on H_1 is distributed according to some density $g(\theta)$. In many applications the point null hypothesis $H_0{:}\theta = \theta_0$ is used as an approximation to the narrow interval hypothesis $H_0{:}|\theta - \theta_0| \leq \varepsilon$ for some small value of ε. Given these definitions, the marginal density of X is

$$m(x) = f(x|\theta_0)\pi_0 + (1-\pi_0)m_g(x) \qquad [D.1]$$

where $f(x|\theta_0)$ is the density of the observed data, given θ_0, and

---

[1] Berger and Sellke (1987a) is abbreviated BS_1 in the following pages, while their rejoinder (Berger and Sellke, 1987b) is abbreviated BS_2. Similarly Berger and Delampady (1987a) is abbreviated BD_1, while the accompanying rejoinder (Berger and Delampady, 1987b) is abbreviated BD_2.

$$m_g(x) = \int f(x|\theta)g(\theta)d\theta . \qquad [D.2]$$

It follows that the posterior probability takes the form

$$\begin{aligned} \Pr(H_0|x) &= f(x|\theta_0)\,\pi_0 / m(x) \\ &= \left[1 + \frac{(1-\pi_0)}{\pi_0}\frac{m_g(x)}{f(x|\theta_0)}\right]^{-1}. \end{aligned} \qquad [D.3]$$

Thus, the posterior odds and Bayes factor[1] are given by

$$\frac{\Pr(H_0|x)}{1-\Pr(H_0|x)} = \frac{\pi_0}{(1-\pi_0)}\frac{f(x|\theta_0)}{m_g(x)} \qquad [D.4]$$

and

$$B_g(x) = \frac{f(x|\theta_0)}{m_g(x)}, \qquad [D.5]$$

respectively. In general, $\Pr(H_0|x)$ is given by

$$\Pr(H_0|x) = \frac{1}{[1+(\pi_1/\pi_0)B(x)^{-1}]}. \qquad [D.6]$$

To summarise, the BS_1 treatment is based on the creation of a dichotomy, introduced via the prior (point prior mass on H_0 and diffuse mass on H_1, with prior weights π_0 and π_1). Thus the posterior can be viewed in terms of a dichotomy, as obtained by applying Bayes theorem to the dichotomous prior. This approach is adopted solely for the purpose of comparing the p-values provided by the NP acceptance procedure with a Bayesian measure of evidence. $P(H_0|x) > 0.5$ indicates that the evidence is in favour of H_0, although subsidiary cost-benefit considerations might cause the analyst to adopt a different threshold. It might be noted that several statisticians have criticised this approach (for example, Casella and Berger, 1987a, p110-111, 1987b, p133, 1987c, p345; Vardeman, 1987), as indicated later in this appendix.

---

[1] The term *weight of evidence* is given to the logarithm of the Bayes factor, which offers the advantage of being additive (Lee, 1989, p127)

BS_1 (p115) proceed to examine an example involving the normal densities $g \sim N(\theta_0, \sigma^2)$ and $\overline{X} \sim N(\theta, \sigma^2 / n)$), i.e., the case in which equality is assumed in the sample variance and the prior variance of the distribution under the alternative hypothesis. It is readily shown that in this case

$$\Pr(H_0|x) = \left[1 + \frac{(1-\pi_0)}{\pi_0}(1+n)^{-1/2} \exp\{t^2 / [2(1+n^{-1})]\}\right]^{-1}. \qquad [D7]$$

BS_1 provide a table of $\Pr(H_0|x)$ values for several combinations of t and n with π_0 fixed at 0.5 (Table 1; additional information is provided in BD_1, Table 1). The table shows, for example, that given fixed $t = 1.96$, $\Pr(H_0|x)$ increases from 0.52 with $n = 50$ to 0.82 with $n = 1000$. This result demonstrates the point that the traditional frequentist approach can lead to a rejection of H_0 at the 0.05 level although $\Pr(H_0|x)$ favours H_0. In addition, the example illustrates the point that, given fixed t (fixed p-value), $\Pr(H_0|x)$ tends to unity as n tends to infinity (the Jeffreys-Lindley paradox).

The main focus of the BS_1 paper is the provision of lower bounds on the Bayes factor and $\Pr(H_0|x)$ for various classes of prior distributions, including the class of all normal prior distributions. They show that, in the latter case, the Bayes factor lower bound is given by $\sqrt{e}\, t e^{-t^2/2}$, $t > 1$, which provides a posterior probability lower bound of

$$\Pr(H_0|x, G_{NOR}) = \left[1 + \frac{(1-\pi_0)}{\pi_0}\frac{\exp(t^2/2)}{\sqrt{e}\, t}\right]^{-1}, t > 1. \qquad [D.8]$$

Thus, given $t = 1.96$ and $\pi_0 = \pi_1 = 0.5$, the lower bound on $\Pr(H_0|x)$ is 0.321, again indicating that the p-value can be misleading as a measure of evidence against H_0. BS_1 also show that a lower bound on $\Pr(H_0|x)$ equal to 0.227 is obtained by taking the prior density on H_1 to be symmetric and distributed in a manner that is biased most favourably towards H_1 (BS_1 Table 2). Although the p-value and posterior probability are distinct, and equality is not expected, the difference between the two is alarming given the manner in which many researchers use the p-value as a measure of evidence against the null hypothesis. The reason for this disagreement is the manner in which the p-value is based on the observed statistic together with more extreme values in the tails of the distribution while, in contrast, $\Pr(H_0|x)$ depends on the observed data and prior. Including unobserved and more extreme data exaggerates the strength of evidence against the null (see, for example, BS_1,

p114; BD_1, p329; Berger and Berry, 1988). In a rejoinder to comments on their paper, BD_2 (p348) question whether it is *really fair to H_0 to hurl against it not just the mild evidence x_0* (i.e., the observations), *but also all the much stronger extreme values, when these extreme values did not occur.*

The central problem with the p-value is that the observed value might be a rare occurrence under both H_0 and H_1 (Morris, 1987; Sellke et al., 2001), in which case it is unreasonable to conclude that the data provide evidence against H_0. The fact is that the p-value can be misleading in several circumstances, including those in which H_1 is unlikely, a priori. It is also misleading when the sample size is small (inadequate power), and when the sample size is large. In the latter context several statisticians have pointed out that a critical value of 0.05 is far too lenient when n is large, since any arbitrarily small difference becomes statistically significant with very large n. Accordingly, it is suggested that the critical value should be adjusted downwards as n increases, possibly balancing the Type I and Type II error rates (Zellner, 1987, p340).

The results obtained for the two-sided precise null hypothesis and one-sided diffuse null appear to differ markedly since, in the latter case, it is possible to reconcile the posterior probability and p-value (Casella, and Berger, 1987a[1]; Lee, 1989, p128). Given this discrepancy, no consensus exists regarding the relevance of the BS_1 two-sided, precise null results, as applied to statistical inference in experimental research. Several statisticians and epidemiologists have recommended that p-values should be abandoned as the standard vehicle for presenting evidence in research publications [see, for example, Rothman (1978, 1998), Diamond and Forrester (1983), Gardner and Altman (1988), Goodman (1999a, 2001) and Poole, 2001] while others take the position that the p-value provides a reasonable measure of evidence against the null hypothesis, as it applies to

---

[1] Casella and Berger, 1987a, is abbreviated CB_1 in the following pages, while their rejoinder (Casella and Berger, 1987b) is abbreviated CB_2. Their comment (Casella and Berger, 1987c) on Berger and Delampady (1987a) is abbreviated CB_3.

the majority of experiments (see, for example, CB_1, CB_3). The reasons for this apparent disagreement are examined in the following paragraphs.

The inescapable conclusion appears to be that p-values overstate the evidence against the point null hypothesis. But no consensus exists on the relevance of this observation to applied statistics, the argument being that point null hypotheses are contrived and of little practical relevance. Thus it is suggested that many important problems are one-sided with no strong prior belief in the point null H_0:$\theta = \theta_0$ (CB_1, p106; CB_3, p345). CB_1 (p106) comment on the typical experiment and state that *saddling an experimenter with a two-sided test would not be appropriate*. This point has been discussed at length (see, for example, BS_2, p136 and BD_1, p326-7, which refers to CB_1). BS_2 suggest that the precise null applies to what they call testing problems, i.e., problems in which there can be real belief in the null hypothesis, with the implication that some precise theory is being tested. They suggest that, in contrast, most one-sided problems are concerned with estimation studies in which the investigator wishes to determine the magnitude of an effect (also referred to as decision problems (BS_2, p136; BD_1, p327), although the latter might be confused with aspects of the NP method). Thus, it appears that there is some agreement regarding the assertion that diffuse one-sided tests are appropriate in some settings. Having argued that the one-sided test applies to the majority of practical problems, CB_1 state that, for many classes of prior distributions, the infimum of the Bayesian posterior probability of H_0 is equal to the p-value, and that in other cases the infimum is less than the p-value. This result might be interpreted as a vindication of the use of p-values as a measure of evidence. On the other hand, Berger and Mortera (1999, p543) question the validity of using the lower bound in isolation since, given a posterior probability interval covering a considerable range with 0.05 as the lower limit, this amounts to providing a summary measure that is the least favourable to the null. Putting this argument aside, the cause of the apparent discrepancy between the BS_1/BD_1 results on point null hypotheses and the CB_1 results on the one-sided problem requires investigation. Among the important differences is the manner in which the point null treatment uses a prior in which a proportion of the prior density is concentrated at the point null, with the remaining density distributed over H_1. In contrast the CB_1 one-sided treatment is based on

diffuse prior distributions. Not surprisingly the two-sided precise null results are dominated by the point prior mass on H_0, and the apparent discrepancy between the results obtained by BS_1 and BD_1 on the one hand and CB_1 on the other is largely attributable to the difference in priors (see, for example, CB_1, p110-111; CB_3, p344; Morris, 1987, p132). Only if H_0 is a wide interval with prior mass distributed uniformly over it does the p-value agree with $Pr(H_0|x)$ (BS_2, p137). The question therefore shifts from one regarding the relevance of point null tests, and on to the relative merits of diffuse priors versus those that include a concentration of mass at a single point. Convincing arguments have been made for both types of prior. To summarise, in addition to the distinction between precise null tests and diffuse tests it is necessary to distinguish between concentrated mass priors and diffuse priors. The validity of the point mass prior appears to depend on the context.

Commenting on the point null result, Vardeman (1987) suggests that there should be little surprise in finding that essentially anything can be found for $Pr(H_0|x)$, *depending on how one is allowed to move prior mass around on H_0 and H_1.* The essential question is whether precise null tests with concentrated priors are meaningful in any practical context, as opposed to a mathematical convenience. CB_3, (p344) argue that *a great many practitioners should not be testing point nulls* (see CB_1, p106, for additional discussion). Similarly, Vardeman (1987) suggests that the notion of a concentration of prior mass at a single point is *completely unappealing* and incompatible with the usual scientific scenario. He states that, although a unimodal distribution about a best value is reasonable, extra mass at H_0 is wrong. These objections to the point null test with concentrated prior density might appear convincing, but several rejoinders deserve consideration. The first is the assertion that, regardless of the validity of point hypotheses, researchers use point nulls routinely and so precise null tests should be done properly (BS_1, p114). Secondly, the precise null is a good approximation to the narrow interval null H_0:$|\theta - \theta_0| \leq \varepsilon$, with ε *sufficiently small* (BS_1, p114 and 119). This approximation is adequate provided the likelihood is reasonably constant over the interval of H_0 (Lee, 1989, p131-133). Sellke et al. (2001) state that the point null hypothesis is a good approximation to the small interval null H_0:$|\theta - \theta_0| < \varepsilon$ provided $\varepsilon < \sigma/(4\sqrt{n})$. It is clear, therefore, that a range of cases exists, with point null hypotheses at one extreme and diffuse nulls at

the other. Hence we may conclude that the p-value is not useful over the entire range to which it is applied. It is, therefore, appropriate to examine in greater depth the validity of the point null with concentrated mass.

It has been argued that the point null applies in several kinds of study, including bioequivalence studies, data-dredging (data fishing) investigations and screening studies. In each case the prior expectation might be of a tiny effect or no effect. In this situation the point null test with prior mass concentrated on the null is appropriate, and the p-value overstates the evidence against H_0. Several questions arise. Should the analyst always use a prior that gives equal weight to H_0 and H_1, assuming objectivity is required as opposed to the expression of informative prior opinion? What kind of prior should be used if H_1 is implausible? To address these issues it is useful to note the assertion that at least three types of point null can be defined (CB_3, p345), namely, (1) point null tests of convenience, (2) precise null tests in which the null does correspond to a concentration of prior belief and (3) precise null hypotheses that describe an interesting and unique feature, but with no special prior belief. The BS_1/BD_1 results apply to the second type. With this in mind it is useful to list several points of disagreement among statisticians. Agreement between the p-value and Bayesian posterior probability (lower bound) occurs in the one-sided case when a diffuse prior is adopted because the prior probability on H_0 is low (i.e., a relatively small proportion of the area under the prior probability density function lies close to H_0; see BD_2 reply to CB_3, p350-1, including Table 1). In this case, the data contribute relatively little evidence and do little to update prior probability; the p-value agrees with the Bayesian posterior probability because the latter is dominated by the prior. The p-value hides the fact that much of the evidence against H_0 is due to its prior probability being small. The counter argument is that p-values appear to fail when prior mass is concentrated on H_0, but this concentration of prior mass is not reasonable (CB_3; CB_1, p110). One might conclude that situations do arise in which a point null test is reasonable, and that a concentration of prior density on the null is valid in these cases (BS_2, p136-7). For example, it has been argued that precise hypothesis testing (as opposed to studies involving estimation, using the terminology of BS_2, p136, and BD_1, p327, referred to above) has a place because there can be real

belief that the precise null is approximately true (BS_2, p 136-7). On the other hand, in those cases in which a concentration of prior density on the null is unreasonable, then the point null formalism is unreasonable at the outset. Hence the need to distinguish between studies in which existing experience is suggestive of a non-trivial effect, on the one hand, and data fishing, observational and screening studies on the other. The form of prior that applies in these two extremes is distinct, but the p-value based hypothesis test treats the two identically.

D.5 The Bayes factor and its lower bound

When dealing with simple hypotheses the Bayes factor offers the advantage that it reduces to a likelihood ratio and thus avoids the need to assign prior probability. Unfortunately, many hypotheses of interest are composite and a complete separation of prior information and the evidence provided by the data is not then possible (see, for example, Lee, 1989, p126). The composite case dependence on prior probability can, however, be avoided by using the lower bound on the Bayes factor. Among several lower bounds that have been suggested is the minimum Bayes factor (B_{min}) which applies to the normal model. It is the likelihood of H_0 relative to the best alternative, i.e., the maximum likelihood alternative. Thus it is the minimum likelihood ratio, and is calculated by taking the null and alternative hypotheses, each with a density concentrated entirely at the point most favoured by the data (Edwards et al., 1963, p228; Pratt, 1987, p123). As such it provides the worst case against the null hypothesis. For normal models $B_{min} = \exp(-Z^2/2)$ (Lee, 1989 p139; Goodman, 1993, 1999b). Substitution into [D.6] yields the corresponding minimum posterior probability (P_{min}). If B_{min} (or, alternatively, P_{min}) does not lead to a rejection of H_0, then the weaker evidence derived from a full Bayesian analysis cannot do so.

An alternative approach, based on a class of flexible p-value distributions under H_1, yields the Bayes factor lower bound $-e\,p\,ln(p)$, $p < 1/e$ (Sellke et al., 2001, p66). [Goodman (2001, Table 1) compares this lower bound with B_{min} and the p-value]. Substitution into [D.6] yields the corresponding lower bound on $\Pr(H_0|x)$ (P_{lb}). The following results are obtained by taking prior odds equal to unity and $t = 1.96$. Using $\exp(-Z^2/2)$: $B_{min} = 0.147$, $P_{min} = 0.128$; using $-e\,p\,ln(p)$: $B_{lb} =$

0.407, $P_{lb} = 0.289$; using $\sqrt{e}\, t\, e^{-t^2/2}$ (see [D.8] and related text): B_{lb}=0.473, $P_{lb} = 0.321$. The following miscellaneous points might be noted. Firstly, although P_{lb} is superior to the p-value as a measure of evidence in many cases, it is, nevertheless, a lower bound and might, therefore, fail to provide an accurate indication of the evidence against H_0. Thus, although the lower bound might be suggestive of evidence against H_0, the researcher remains uncertain regarding the validity of H_0 rejection (BS_1, p121). The upper bound is required, but this depends on the prior probability distribution adopted for the alternative hypothesis. It is, therefore, subjective. Secondly, it is interesting to observe that the p-value obtained by replacing t with t-1is close to the Bayes factor lower bound obtained for one class of reasonable H_1 prior distribution (BS_1, p120, Fig. 4). Thus it is suggested that t_{obs}= 2 indicates no more than mild evidence against H_0, while t_{obs}= 3 indicates significant evidence (BS_1, p120).

To conclude, it has been noted (BS_1, p114) that although a p-value of 0.05 was never intended as an absolute indicator of evidence against H_0, many researchers do use hypothesis testing in this manner. Nelder (1999, p257) has discussed the problem, referring to *the non-scientific ... obsession with significance tests*. He suggests *that many editors and referees will not accept papers unless they contain these non-scientific modes of inference, and that authors know this and act accordingly*. He asserts (p261) that *the most important task before us ... is to demolish the p-value culture*.

REFERENCES

Abe, K., Aoki, M., Kawagoe, J., Yoshida, T., Hattori, A., Kogure, K. and Itoyama, Y. (1995) Ischemic delayed neuronal death. A mitochondrial hypothesis. *Stroke*, **26**, 1478-1489.

Apell, H.-J., Nelson, M.T., Marcus, M.M. and Läuger, P. (1986) Effects of the ATP, ADP and inorganic phosphate on the transport rate of the Na^+, K^+-pump. *Biochim. Biophys. Acta*, **857**, 105-115.

Armitage, P. and Colton, T. (Eds.) (1998) *Encyclopedia of biostatistics*. Wiley, Chichester.

Anderson, A.W., Zhong, J., Petroff, O.A.C., Szafer, A., Ransom, B.R., Prichard, J.W. and Gore, J.C. (1996) Effects of osmotically driven cell volume changes on diffusion-weighted imaging of the rat optic nerve. *Magn. Reson. Med.*, **35**, 162-167.

Arnold, D.L., Matthews, P.M. and de Stefano, N. (1997) MRI and proton MRS in the evaluation of multiple sclerosis. In *Magnetic resonance spectroscopy and imaging in neurochemistry* (Ed. Bachelard, H.) pp. 267-288. Plenum Press, New York.

Astrup, J., Siesjö, B.K. and Symon, L. (1981) Thresholds in cerebral ischemia - The ischemic penumbra. *Stroke*, **12**, 723-725.

Bacon, D.W. and Watts, D.G. (1971) Estimating the transition between two intersecting straight lines. *Biometrika*, **58**, 525-534.

Bagnall, D.J. and Wolfe, J. (1982) Arrhenius plots: information or noise. *Cryo-Letters*, **3**, 7-16.

Bagnall, D.J. and Wolfe, J.A. (1978) Chilling sensitivity in plants: do the activation energies of growth processes show an abrupt change at a critical temperature? *J. Exp. Bot.*, **29**, 1231-1242.

Balestrino, M. (1995) Pathophysiology of anoxic depolarization: new findings and a working hypothesis. *J. Neurosci. Methods*, **59**, 99-103.

Barnard, J. and Meng, X.-L. (1999) Applications of multiple imputation in medical studies: from AIDS to NHANES. *Stat. Methods Med. Res.*, **8**, 17-36.

Bell, B.A. (1984) A history of the study of the cerebral circulation and the measurement of cerebral blood flow. *Neurosurgery*, **14**, 238-246.

Bennett, J.H. (Ed.) (1990) *Statistical inference and analysis. Selected correspondence of R.A. Fisher*. Clarendon Press, Oxford.

Berger, J.O. and Berry, D.A. (1988) Statistical analysis and the illusion of objectivity. *American Scientist*, **76**, 159-165.

Berger, J.O. and Delampady, M. (1987a) Testing precise hypotheses. *Statistical Sci.*, **2**, 317-335.

Berger, J.O. and Delampady, M. (1987b) Testing precise hypotheses. Rejoinder. *Statistical Sci.*, **2**, 348-352.

Berger, J.O. and Mortera, J. (1999) Default Bayes factors for nonnested hypothesis testing. *J. Am. Statist. Assoc.*, **94**, 542-554.

Berger, J.O. and Sellke, T. (1987a) Testing a point null hypothesis: the irreconcilability of *P* values and evidence. *J. Am. Statist. Assoc.*, **82**, 112-122.

Berger, J.O. and Sellke, T. (1987b) Testing a point null hypothesis: the irreconcilability of *P* values and evidence: rejoinder. *J. Am. Statist. Assoc.*, **82**, 135-139.

Berk, K. (1987) Computing for incomplete repeated measures. *Biometrics*, **43**, 385-398.

Berliner, L.J. (Ed.) (1976) *Spin labeling. Theory and applications*. Academic Press, New York.

Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. (Eds.) (1992) *Bayesian statistics 4*. Oxford University Press, Oxford.

Berzuini, C. (1996) Medical monitoring. In *Markov chain Monte Carlo in Practice* (Ed. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J.) pp. 321-337. Chapman and Hall, London.

Best, N., Cowles, M.K. and Vines, K. (1995) *CODA. Convergence diagnosis and output analysis software for Gibbs sampling output. Version 0.30.* Technical report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.

Bhutta, A.T., Cleves, M.A., Casey, P.H., Cradock, M.M. and Anand, K.J.S. (2002) Cognitive and behavioural outcomes of school-age children who were born preterm. *J. Am. Med. Assoc.*, **288**, 728-737.

Bloch, F., Hansen, W.W. and Packard, M. (1946) The nuclear induction experiment. *Phys. Rev.*, **70**, 474-485.

Box, G.E.P. and Tiao, G.C. (1973) *Bayesian inference in statistical analysis.* Addison-Wesley, Reading, Massachusetts.

Branston, N.M., Strong, A.J. and Symon, L. (1978) Impedance related to local blood flow in cerebral cortex. *J. Physiol.*, **275**, 81P-82P.

Brooks, S.P. and Roberts, G.O. (1998) Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing*, **8**, 319-335.

Browne, R.H. (1995) Bayesian analysis and the GUSTO trial. Letter to the Editor. *J. Am. Med. Assoc.*, **274**, 873.

Browner, W.S. and Newman, T.B. (1987) Are all significant *P* values created equal? *J. Am. Med. Soc.*, **257**, 2459-2463.

Buitink, J., Hemminga, M.A. and Hoekstra, F.A. (1999) Characterization of molecular mobility in seed tissues: an electron paramagnetic resonance spin probe study. *Biophys. J.*, **76**, 3315-3322.

Buitink, J., Hoekstra, F.A. and Hemminga, M.A. (2000) Molecular mobility in the cytoplasm of lettuce radicles correlates with longevity. *Seed Science Research*, **10**, 285-292.

Burton, P., Gurrin, L. and Sly, P. (1998) Tutorial in biostatistics. Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modelling. *Stat. Med.*, **17**, 1261-1291.

Busza, A.L., Allen, K.L., King, M.D., van Bruggen, N., Williams, S.R. and Gadian, D.G. (1992) Diffusion-weighted imaging studies of cerebral ischemia in gerbils. Potential relevance to energy failure. *Stroke*, **23**, 1602-1612.

Cady, E.B. (1997) Phosphorus and proton magnetic resonance spectroscopy of the brain of the newborn infant. In *Magnetic resonance spectroscopy and imaging in neurochemistry* (Ed. Bachelard, H.) pp. 289-327. Plenum Press, New York.

Calamante, F., Thomas, D.L., Pell, G.S., Wiersma, J. and Turner, R. (1999) Measuring cerebral blood flow using magnetic resonance imaging techniques. *J. Cerebr. Blood Flow Metab.*, **19**, 701-735.

Callaghan, P.T. (1991) *Principles of nuclear magnetic resonance microscopy*. Oxford University Press, Oxford.

Cannon, B., Polnaszek, C.F., Butler, K.W., Eriksson, L.E.G. and Smith, I.C.P. (1975) The fluidity and organization of mitochondrial membrane lipids of the brown adipose tissue of cold-adapted rats and hamsters as determined by nitroxide spin probes. *Arch. Biochem. Biophys.*, **167**, 505-518.

Carlin, B.P., Gelfand, A.E. and Smith, A.F.M. (1992) Hierarchical Bayesian analysis of changepoint problems. *Appl. Statist.*, **41**, 389-405.

Carlin, B.P. and Louis, T.A. (1996) *Bayes and empirical Bayes methods for data analysis.* Chapman and Hall, London.

Carlin, B.P. and Louis, T.A. (2000) *Bayes and empirical Bayes methods for data analysis*, 2nd edn. Chapman and Hall/CRC, Boca Raton.

Carr, H.Y. and Purcell, E.M. (1954) Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.*, **94**, 630-638.

Carrington, A. and McLachlan, A.D. (1967) *Introduction to magnetic resonance.* Harper and Row, New York.

Casella, G. and Berger, R.L. (1987a) Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *J. Am. Statist. Assoc.*, **82**, 106-111.

Casella, G. and Berger, R.L. (1987b) Testing a point null hypothesis: the irreconcilability of *P* values and evidence: rejoinder. *J. Am. Statist. Assoc.*, **82**, 133-135.

Casella, G. and Berger, R.L. (1987c) Testing precise hypotheses. Comment. *Statistical Sci.*, **2**, 344-347.

Casella, G. and George, E.I. (1992) Explaining the Gibbs sampler. *The American Statistician*, **46**, 167-174.

Chang, B.S., Beauvais, R.M., Dong, A. and Carpenter, J.F. (1996) Physical factors affecting the storage stability of freeze-dried interleukin-1 receptor antagonist: glass transition and protein conformation. *Arch. Biochem. Biophys.*, **331**, 249-258.

Chatfield, C. (1989) *The analysis of time series. An introduction*, 4th edn. Chapman and Hall, London

Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49**, 327-335.

Cnaan, A., Laird, N.M. and Slasor, P. (1997) Tutorial in biostatistics. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat. Med.*, **16**, 2349-2380.

Cochran, W.G. and Cox, G.M. (1957) *Experimental designs*, 2nd edn. Wiley, New York.

Collewijn, H. and van Harreveld, A. (1966) Membrane potential of cerebral cortical cells during spreading depression and asphyxia. *Exp. Neurol.*, **15**, 425-436.

Colquhoun, D. (1971) *Lectures on biostatistics. An introduction to statistics with applications in biology and medicine*. Oxford University Press, Oxford.

Congdon, P. (2001) *Bayesian statistical modelling*. Wiley, Chichester.

Conte, S.D. and de Boor, C. (1981) *Elementary numerical analysis. An algorithmic approach.* McGraw-Hill, Tokyo.

Cory, D.G. and Garroway, A.N. (1990) Measurement of translational displacement probabilities by NMR: an indicator of compartmentation. *Magn. Reson. Med.*, **14**, 435-444.

Cowles, M.K. and Carlin, B.P. (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. *J. Am. Statist. Assoc.*, **91**, 883-904.

Cox, D.R. (1987) Testing precise hypotheses. Comment. *Statistical Sci.*, **2**, 335-336.

Cox, D.R. (1992) *Planning of experiments*. Wiley, New York.

Crowder, M.J. and Hand, D.J. (1990) *Analysis of repeated measures*. Chapman and Hall, London.

Crowder, M.J. and Tredger, J.A. (1981) The use of exponentially damped polynomials for biological recovery data. *Appl. Statist.*, **30**, 147-152.

Daniels, M.J. (1999) A prior for the variance in hierarchical models. *Canadian J. Statistics*, **27**, 567-578.

Davidian, M. and Giltinan, D.M. (1993) Some general estimation methods for nonlinear mixed-effects models. *J. Biopharm. Statist.*, **3**, 23-55.

Davidian, M. and Giltinan, D.M. (1995) *Nonlinear models for repeated measurement data.* Chapman and Hall, London.

Davis, D.G., Inesi, G. and Gulik-Krzywicki, T. (1976) Lipid molecular motion and enzyme activity in sarcoplasmic reticulum membrane. *Biochemistry*, **15**, 1271-1276.

de Crespigny, A., Röther, J., van Bruggen, N., Beaulieu, C. and Moseley, M.E. (1998) Magnetic resonance imaging assessment of cerebral hemodynamics during spreading depression in rats. *J. Cerebr. Blood Flow Metab.*, **18**, 1008-1017.

de Crespigny, A.J., Röther, J., Beaulieu, C. and Moseley, M.E. (1999) Rapid monitoring of diffusion, DC potential, and blood oxygenation changes during global ischemia. Effects of hypoglycemia, hyperglycemia and TTX. *Stroke*, **30**, 2212-2222.

Dean, W.L. and Tanford, C. (1978) Properties of a delipidated, detergent-activated Ca^{2+}-ATPase. *Biochemistry*, **17**, 1683-1690.

Deitch, E.A. (1992) Multiple organ failure. *Ann. Surg.*, **216**, 117-134.

Deliconstantinos, G., Kopeikina-Tsiboukidou, L. and Villiotou, V. (1987) Evaluation of membrane fluidity effects and enzyme activities alterations in adriamycin neurotoxicity. *Biochem. Pharmacol.*, **36**, 1153-1161.

Diamond, G.A. and Forrester, J.S. (1983) Clinical trials and statistical verdicts: probable grounds for appeal. *Ann. Intern. Med.*, **98**, 385-394.

Diggle, P.J., Heagerty, P.J., Liang, K.-Y. and Zeger, S.L. (2002) *Analysis of longitudinal data*, 2nd edn. Oxford University Press, Oxford.

Dijkhuizen, R.M., Beekwilder, J.P., van der Worp, H.B., van der Sprenkel, J.W.B., Tulleken, K.A.F. and Nicolay, K. (1999) Correlation between tissue depolarizations and damage in focal ischemic rat brain. *Brain Res.*, **840**, 194-205.

Dixon, W.J. and Merdian, K. (1992) *ANOVA and regression with BMDP 5V.* Dixon Statistical Associates, Los Angeles.

Dobson, A.J. (1990) *An introduction to generalized linear models.* Chapman and Hall, London.

Doran, M., Hajnal, J.V., van Bruggen, N., King, M.D., Young, I.R. and Bydder, G.M. (1990) Normal and abnormal white matter tracts shown by MR imaging using directional diffusion weighted sequences. *J. Comput. Assist. Tomogr.*, **14**, 865-873.

Dumas, F., Lebrun, M.C. and Tocanne, J.-F. (1999) Is the protein/lipid hydrophobic matching principle relevant to membrane organization and functions? *FEBS lett.*, **458**, 271-277.

Dumouchel, W. and Waternaux, C. (1992) Hierarchical models for combining information for meta-analyses. In *Bayesian statistics 4* (Ed. Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M.) pp. 338-341. Oxford University Press, Oxford.

Edwards, A.W.F. (1992) *Likelihood. Expanded edition.* The Johns Hopkins University Press, Baltimore.

Edwards, W., Lindman, H. and Savage, L.J. (1963) Bayesian statistical inference for psychological research. *Psychol. Rev.*, **70**, 193-242.

Everitt, B.S. (1995) The analysis of repeated measures: a practical review with examples. *The Statistician*, **44**, 113-135.

Fisher, R.A. (1935) The logic of inductive inference. *J. R. Statist. Soc.*, **98**, 39-82.

Fisher, R.A. (1970) *Statistical methods for research workers*, 14th edn. Oliver and Boyd, Edinburgh.

Fisher, R.A. (1973) *Statistical methods and scientific inference*, 3rd edn. Collier Macmillan, London.

Fisher, M. and Ratan, R. (2003) New perspectives on developing acute stroke therapy. *Ann. Neurol.*, **53**, 10-20.

Fletcher, R. (1980) *Practical methods of optimization. Volume 1. Unconstrained optimization.* Wiley, Chichester.

Freed, J.H. (1976) Theory of slow tumbling ESR spectra for nitroxides. In *Spin labeling. Theory and applications* (Ed. Berliner, L.J.) pp. 53-132. Academic Press, New York.

Fyfe, P.K., McAuley, K.E., Roszak, A.W., Isaacs, N.W., Cogdell, R.J. and Jones, M.R. (2001) Probing the interface between membrane proteins and membrane lipids by X-ray crystallography. *Trends in Biochemical Sciences*, **26**, 106-112.

Gadian, D.G. (1995) *NMR and its applications to living systems*, 2nd edn. Oxford University Press, Oxford.

Gadian, D.G. (2002) Magnetic resonance approaches to the identification of focal pathophysiology in children with brain disease. *Dev. Sci.*, **5**, 279-285.

Gallant, A.R. and Fuller, W.A. (1973) Fitting segmented polynomial regression models whose join points have to be estimated. *J. Am. Statist. Assoc.*, **68**, 144-147.

Ganong, W.F. (2001) *Review of medical physiology*. Lange Medical Books/McGraw-Hill, New York.

Gardner, M.J. and Altman, D.G. (1988) Estimating with confidence. *Br. Med. J.*, **296**, 1210-1211.

Garthwaite, P.H., Jolliffe, I.T. and Jones, B. (1995) *Statistical inference*. Prentice Hall, London.

Gelfand, A.E., Hills, S.E., Racine-Poon, A. and Smith, A.F.M. (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. *J. Am. Statist. Assoc.*, **85**, 972-985.

Gelfand, A.E., Sahu, S.K. and Carlin, B.P. (1995) Efficient parametrizations for normal linear mixed models. *Biometrika*, **82**, 479-488.

Gelfand, A.E., Sahu, S.K. and Carlin, B.P. (1996) Efficient parameterizations for generalized linear mixed models. In *Bayesian statistics 5* (Ed. Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M.) pp. 165-180. Oxford University Press, Oxford.

Gelman, A. (1996) Inference and monitoring convergence. In *Markov chain Monte Carlo in practice* (Ed. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J.) pp. 131-143. Chapman and Hall, London.

Gelman, A., Bois, F. and Jiang, J. (1996a) Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *J. Am. Statist. Assoc.*, **91**, 1400-1412.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995) *Bayesian data analysis*. Chapman and Hall, London.

Gelman, A., Roberts, G.O. and Gilks, W.R. (1996b) Efficient Metropolis jumping rules. In *Bayesian statistics 5* (Ed. Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M.) pp. 599-607. Oxford University Press, Oxford.

Gelman, A. and Rubin, D.B. (1992a) A single series from the Gibbs sampler provides a false sense of security. In *Bayesian statistics 4* (Ed. Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M.) pp. 625-631. Oxford University Press, Oxford.

Gelman, A. and Rubin, D.B. (1992b) Inference from iterative simulation using multiple sequences. *Statistical Sci.*, **7**, 457-472.

Gerald, C.F. (1978) *Applied numerical analysis*, 2nd edn. Addison-Wesley, Massachusetts.

Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian statistics 4* (Ed. Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M.) pp. 169-193. Oxford University Press, Oxford.

Geyer, C.J. (1992) Practical Markov chain Monte Carlo. *Statistical Sci.*, **7**, 473-483.

Gillespie, D.T. (1992) *Markov processes: an introduction for physical scientists*, Academic Press, Boston.

Gilks, W.R. (1992) Derivative-free adaptive rejection sampling for Gibbs sampling. In *Bayesian statistics 4* (Ed. Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M.) pp. 641-649. Oxford University Press, Oxford.

Gilks, W.R., Best, N.G. and Tan, K.K.C. (1995) Adaptive rejection Metropolis sampling within Gibbs sampling. *Appl. Statist.*, **44**, 455-472.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (Eds.) (1996a) *Markov chain Monte Carlo in practice*. Chapman and Hall, London.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996b) Introducing Markov chain Monte Carlo. In *Markov chain Monte Carlo in practice* (Ed. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J.) pp. 1-19. Chapman and Hall, London.

Gilks, W.R. and Roberts, G.O. (1996) Strategies for improving MCMC. In *Markov chain Monte Carlo in practice* (Ed. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J.) pp. 89-114. Chapman and Hall, London.

Gilks, W.R., Thomas, A. and Spiegelhalter, D.J. (1994) A language and program for complex Bayesian modelling. *The Statistician*, **43**, 169-177.

Godiksen, H. and Jessen, F. (2002) Temperature and Ca^{2+}-dependence of the sarcoplasmic reticulum Ca^{2+}-ATPase in haddock, salmon, rainbow trout and zebra cichlid. *Comp. Biochem. Physiol. B*, **133**, 35-44.

Goldstein, H. (1987) *Multilevel models in educational and social research*, Charles Griffin, London.

Goldstein, H. (1995) *Multilevel Statistical Models*, 2nd edn. Arnold, London.

Goldstein, H., Healy, M.J.R. and Rasbash, J. (1994) Multilevel time series models with applications to repeated measures data. *Stat. Med.*, **13**, 1643-1655.

Goodman, S.N. (1993) p Values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate. *Am. J. Epidemiol.*, **137**, 485-496.

Goodman, S.N. (1999a) Towards evidence-based medical statistics. 1: the *P* value fallacy. *Ann. Intern. Med.*, **130**, 995-1004.

Goodman, S.N. (1999b) Towards evidence-based medical statistics. 2: the Bayes factor. *Ann. Intern. Med.*, **130**, 1005-1013.

Goodman, S.N. (2001) Of *P*-values and Bayes: a modest proposal. *Epidemiology*, **12**, 295-297.

Green, P.J. (2003) Notes on the life and work of R.A. Fisher. *The Statistician*, **52**, 299-301.

Griesemer, D., Zawar, C. and Neumcke, B. (2002) Cell-type specific depression of neuronal excitability in rat hippocampus by activation of ATP-sensitive potassium channels. *Eur. Biophys. J.*, **31**, 467-477.

Griffith, O.H. and Jost, P.C. (1976) Lipid spin labels in biological membranes. In *Spin Labeling. Theory and applications* (Ed. Berliner, L.J.) pp. 453-523. Academic Press, New York.

Gyngell, M.L., Back, T., Hoehn-Berlage, M., Kohno, K. and Hossmann, K.-A. (1994) Transient cell depolarization after permanent middle cerebral artery occlusion: an observation by diffusion-weighted MRI and localized ^{1}H-MRS. *Magn. Reson. Med.*, **31**, 337-341.

Hahn, E.L. (1950) Spin echoes. *Phys. Rev.*, **80**, 580-594.

Hammersley, J.M. and Handscomb, D.C. (1964) *Monte Carlo methods*. Methuen, London.

Hancock, B.C., Shamblin, S.L. and Zografi, G. (1995) Molecular mobility of amorphous pharmaceutical solids below their glass transition temperatures. *Pharm. Res.*, **12**, 799-806.

Hancock, B.C. and Zografi, G. (1997) Characteristics and significance of the amorphous state in pharmaceutical systems. *J. Pharm. Sci.*, **86**, 1-12.

Hand, D. and Crowder, M. (1996) *Practical longitudinal data analysis*. Chapman and Hall, London.

Hansen, A.J. (1978) The extracellular potassium concentration in brain cortex following ischemia in hypo- and hyperglycemic rats. *Acta Physiol. Scand.*, **102**, 324-329.

Hansen, A.J. (1985) Effect of anoxia on ion distribution in the brain. *Physiol. Rev.*, **65**, 101-148.

Hansen, A.J. and Olsen, C.E. (1980) Brain extracellular space during spreading depression and ischemia. *Acta Physiol. Scand.*, **108**, 355-365.

Harris, N.G., Zilkha, E., Houseman, J., Symms, M.R., Obrenovitch, T.P. and Williams, S.R. (2000) The relationship between the apparent diffusion coefficient measured by magnetic resonance imaging, anoxic depolarization, and glutamate efflux during experimental cerebral ischemia. *J. Cerebr. Blood Flow Metab.*, **20**, 28-36.

Heiss, W.-D. (1992) Experimental evidence of ischemic thresholds and functional recovery. *Stroke*, **23**, 1668-1672.

Helmer, K.G., Dardzinski, B.J. and Sotak, C.H. (1995) The application of porous-media theory to the investigation of time-dependent diffusion in *in vivo* systems. *NMR Biomed.*, **8**, 297-306.

Helpern, J.A., Ordidge, R.J. and Knight, R.A. (1992) The effect of cell membrane water permeability on the apparent diffusion coefficient of water. *Proc. Soc. Magn. Reson. Med. 11th Annual Meeting, Berlin,* p. 1201. Society of Magnetic Resonance in Medicine, Berkeley, CA.

Hesketh, T.R., Smith, G.A., Houslay, M.D., McGill, K.A., Birdsall, N.J.M., Metcalfe, J.C. and Warren, G.B. (1976) Annular lipids determine the ATPase activity of a calcium transport protein complexed with dipalmitoyllecithin. *Biochemistry*, **15**, 4145-4151.

Hidalgo, C., Ikemoto, N. and Gergely, J. (1976) Role of phospholipids in the calcium-dependent ATPase of the sarcoplasmic reticulum. *J. Biol. Chem.*, **251**, 4224-4232.

Hills, S.E. and Smith, A.F.M. (1992) Parameterization issues in Bayesian inference. In *Bayesian statistics 4* (Ed. Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M.) pp. 227-246. Oxford University Press, Oxford.

Hinkley, D.V. (1987) Testing a point null hypothesis: the irreconcilability of *P* values and evidence (comment). *J. Am. Statist. Assoc.*, **82**, 128-129.

Hoehn-Berlage, M., Norris, D.G., Kohno, K., Mies, G., Leibfritz, D. and Hossmann, K.-A. (1995) Evolution of regional changes in apparent diffusion coefficient during focal ischemia of rat brain: the relationship of quantitative diffusion NMR imaging to reduction in cerebral blood flow and metabolic disturbances. *J. Cerebr. Blood Flow Metab.*, **15**, 1002-1011.

Holden, C.P. and Storey, K.B. (1998) Protein kinase A from bat skeletal muscle: a kinetic study of the enzyme from a hibernating mammal. *Arch. Biochem. Biophys.*, **358**, 243-250.

Holden, C.P. and Storey, K.B. (2000) Purification and characterization of protein kinase A from liver of the freeze-tolerant wood frog: role in glycogenolysis during freezing. *Cryobiology*, **40**, 323-331.

Hope, P.L., Costello, A.M.de L., Cady, E.B., Delpy, D.T., Tofts, P.S., Chu, A., Hamilton, P.A., Reynolds, E.O.R. and Wilkie, D.R. (1984) Cerebral energy metabolism studied with phosphorus NMR spectroscopy in normal and birth-asphyxiated infants. *Lancet*, **2**, 366-370.

Hope, P.L. and Reynolds, E.O.R. (1985) Investigation of cerebral energy metabolism in newborn infants by phosphorus nuclear magnetic resonance spectroscopy. *Clinics in Perinatology*, **12**, 261-275.

Horowitz, P. and Hill, W. (1980) *The art of electronics*. Cambridge University Press, Cambridge.

Hossmann, K.-A. (1971) Cortical steady potential, impedance and excitability changes during and after total ischemia of cat brain. *Exp. Neurol.*, **32**, 163-175.

Huang, N.C., Yongbi, M.N. and Helpern, J.A. (1997) The influence of preischemic hyperglycemia on acute changes in the apparent diffusion coefficient of brain water following global ischemia in rats. *Brain Res.*, **757**, 139-145.

Inesi, G., Millman, M. and Eletr, S. (1973) Temperature-induced transitions of function and structure in sarcoplasmic reticulum membranes. *J. Mol. Biol.*, **81**, 483-504.

Jackman, S. (2000) Estimation and inference via Bayesian simulation: an introduction to Markov chain Monte Carlo. *Am. J. Political Sci.*, **44**, 375-404.

Jacobs, B.C. and Brant-Zawadzki, M. (1992) Ischemia. In *Magnetic resonance imaging*, 2nd edn. (Ed. Stark, D.D. and Bradley, W.G.) pp. 636-669. Mosby Year Book, St. Louis.

Jeffreys, H. (1980) Some general points in probability theory. In *Bayesian analysis in econometrics and statistics* (Ed. Zellner, A.) pp. 451-453. North-Holland Publishing Company, Amsterdam.

Jennrich, R.I. and Schluchter, M.D. (1986) Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, **42**, 805-820.

Jones, B. and Kenward, M.G. (1989) *Design and analysis of cross-over trials*. Chapman and Hall, London.

Jones, R.H. (1993) *Longitudinal data with serial correlation: a state-space approach.* Chapman and Hall, London.

Kalichevsky, M.T., Jaroszkiewicz, E.M., Ablett, S., Blanshard, J.M.V. and Lillford, P.J. (1992) The glass transition of amylopectin measured by DSC, DMTA and NMR. *Carbohydrate Polymers*, **18**, 77-88.

Karmas, R., Buera, M.P. and Karel, M. (1992) Effect of glass transition on rates of nonenzymatic browning in food systems. *J. Agric. Food. Chem.*, **40**, 873-879.

Kass, R.E., Carlin, B.P., Gelman, A. and Neal, R.M. (1998) Markov chain Monte Carlo in practice: a roundtable discussion. *The American Statistician*, **52**, 93-100.

Kass, R.E. and Wasserman, L. (1996) The selection of prior distributions by formal rules. *J. Am. Statist. Assoc.*, **91**, 1343-1370.

King, M.D., Houseman, J., Gadian, D.G. and Connelly, A. (1997a) Localized q-space imaging of the mouse brain. *Magn. Reson. Med.*, **38**, 930-937.

King, M.D., Houseman, J., Roussel, S.A., van Bruggen, N., Williams, S.R. and Gadian, D.G. (1994) q-Space imaging of the brain. *Magn. Reson. Med.*, **32**, 707-713.

King, M.D. and Marsh, D. (1987) Head group and chain length dependence of phospholipid self-assembly studied by spin-label electron spin resonance. *Biochemistry*, **26**, 1224-1231.

King, M.D. and Quinn, P.J. (1983) A passage saturation transfer paramagnetic resonance study of the rotational diffusion of the sarcoplasmic reticulum calcium-ATPase. *J. Bioenerg. Biomembr.*, **15**, 135-150.

King, M.D., Quinn, P.J., Munkonge, F.M. and Madden, T.D. (1987) The influence of calcium pump coupling on the Arrhenius behavior of sarcoplasmic reticulum Ca^{2+}-ATPase. *J. Bioenerg. Biomembr.*, **19**, 45-52.

King, M.D., van Bruggen, N., Busza, A.L., Houseman, J., Williams, S.R. and Gadian, D.G. (1992) Perfusion and diffusion imaging. *Magn. Reson. Med.*, **24**, 288-301.

King, M., van Bruggen, N., Busza, A. and Turner, R. (1997b) Diffusion-weighted magnetic resonance imaging. In *Magnetic resonance spectroscopy and imaging in neurochemistry* (Ed. Bachelard, H.) pp. 179-211. Plenum Press, New York.

Kohno, K., Hoehn-Berlage, M., Mies, G., Back, T. and Hossmann, K.-A. (1995) Relationship between diffusion-weighted MR images, cerebral blood flow, and energy state in experimental brain infarction. *Magn. Reson. Imaging*, **13**, 73-80.

Kopeikina, L.T., Kamper, E.F., Siafaka, I. and Stavridis, J. (1997) Modulation of synaptosomal plasma membrane-bound enzyme activity through the perturbation of plasma membrane lipid structure by bupivacaine. *Anesthesia and Analgesia*, **85**, 1337-1343.

Krizaj, D., Rice, M.E., Wardle, R.A. and Nicholson, C. (1996) Water compartmentalization and extracellular tortuosity after osmotic changes in cerebellum of *Trachemys scripta. J. Physiol.*, **492**, 887-896.

Kumamoto, J., Raison, J.K. and Lyons, J.M. (1971) Temperature "breaks" in Arrhenius plots: a thermodynamic consequence of a phase change. *J. Theor. Biol.*, **31**, 47-51.

Kuzniecky, R.I. and Jackson, G.D. (1995) *Magnetic resonance in epilepsy*. Raven Press, New York.

Laird, N.M. and Ware, J.H. (1982) Random-effects models for longitudinal data. *Biometrics*, **38**, 963-974.

Lassen, N.A. and Astrup, J. (1987) Ischemic penumbra. In *Cerebral blood flow. Physiologic and clinical aspects* (Ed. Wood, J.H.) pp. 458-466. McGraw-Hill, New York.

Latour, L.L., Hasegawa, Y., Formato, J.E., Fisher, M. and Sotak, C.H. (1994a) Spreading waves of decreased diffusion coefficient after cortical stimulation in the rat brain. *Magn. Reson. Med.*, **32**, 189-198.

Latour, L.L., Svoboda, K., Mitra, P.P. and Sotak, C.H. (1994b) Time-dependent diffusion of water in a biological model system. *Proc. Natl. Acad. Sci. USA*, **91**, 1229-1233.

Laurer, H.L. and McIntosh, T.K. (2001) Pharmacologic therapy in traumatic brain injury: update on experimental treatment strategies. *Current Pharmaceutical Design*, 7, 1505-1516.

Lauterbur, P.C. (1973) Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, **242**, 190-191.

Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E. and Laval-Jeantet, M. (1986) MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, **161**, 401-407.

Lee, A.G., Birdsall, N.J.M., Metcalfe, J.C., Toon, P.A. and Warren, G.B. (1974) Clusters in lipid bilayers and the interpretation of thermal effects in biological membranes. *Biochemistry*, **13**, 3699-3705.

Lee, P.M. (1989) *Bayesian statistics: an introduction*, 1st edn. Arnold, New York.

Lee, P.M. (1997) *Bayesian statistics: an introduction*, 2nd edn. Arnold, New York.

Lehmann, E.L. (1993) The Fisher, Neyman-Pearson theories of testing hypotheses: one theory or two? *J. Am. Statist. Assoc.*, **88**, 1242-1249.

Lindley, D.V. (1965) *Introduction to probability and statistics from a Bayesian viewpoint. Part 2. Inference*. 1970 Reprint. Cambridge University Press, Cambridge.

Lindley, D.V. and Smith, A.F.M. (1972) Bayes estimates for the linear model. *J. R. Statist. Soc. B*, **34**, 1-41.

Lipton, P. (1999) Ischemic cell death in brain neurons. *Physiol. Rev.*, **79**, 1431-1568.

Littell, R.C., Freund, R.J. and Spector, P.C. (1991) *SAS system for linear models*, 3rd edn. SAS Institute, Cary.

Littell, R.C., Milliken, G.A., Stroup, W.W. and Wolfinger, R.D. (1996) *SAS system for mixed models*. SAS Institute, Cary.

Londesborough, J. (1980) The causes of sharply bent or discontinuous Arrhenius plots for enzyme-catalysed reactions. *Eur. J. Biochem.*, **105**, 211-215.

Lorek, A., Takei, Y., Cady, E.B., Wyatt, J.S., Penrice, J., Edwards, A.D., Peebles, D., Wylezinska, M., Owen-Reece, H., Kirkbride, V., Cooper, C.E., Aldridge, R.F., Roth, S.C., Brown, G., Delpy, D.T. and Reynolds, E.O.R. (1994) Delayed ("secondary") cerebral energy failure after acute hypoxia-ischemia in the newborn piglet: continuous 48-hour studies by phosphorus magnetic resonance spectroscopy. *Pediatr. Res.*, **36**, 699-706.

MacDonald, J.A. and Storey, K.B. (1998) cAMP-dependent protein kinase from brown adipose tissue: temperature effects on kinetic properties and enzyme role in hibernating ground squirrels. *J. Comp. Physiol. B*, **168**, 513-525.

Macdonald, R.R. (1997) On statistical testing in psychology. *Br. J. Psychol.*, **88**, 333-347.

Mager, H. and Göller, G. (1997) Comments on the statistical analysis of toxicokinetic studies. *Drug Information J.*, **31**, 551-562.

Mansfield, P. and Grannell, P.K. (1975) Diffraction and microscopy in solids and liquids by NMR. *Phys. Rev. B*, **12**, 3618-3634.

Marsh, D. and Horváth, L.I. (1998) Structure, dynamics and composition of the lipid-protein interface. Perspectives from spin-labelling. *Biochim. Biophys. Acta*, **1376**, 267-296.

Matsuoka, Y. and Hossmann, K.-A. (1982a) Brain tissue osmolality after middle cerebral artery occlusion in cats. *Exp. Neurol.*, **77**, 599-611.

Matsuoka, Y. and Hossmann, K.-A. (1982b) Cortical impedance and extracellular volume changes following middle cerebral artery occlusion in cats. *J. Cerebr. Blood Flow Metab.*, **2**, 466-474.

Matthews, J.N.S. (1993) A refinement to the analysis of serial data using summary measures. *Stat. Med.*, **12**, 27-37.

Matthews, J.N.S., Altman, D.G., Campbell, M.J. and Royston, P. (1990) Analysis of serial measurements in medical research. *Br. Med. J.*, **300**, 230-235.

McCall, D.W., Douglass, D.C. and Anderson, E.W. (1959) Diffusion in liquids. *J. Chem. Phys.*, **31**, 1555-1557.

McCall, D.W., Douglass, D.C. and Anderson, E.W. (1963) Self-diffusion studies by means of nuclear magnetic resonance spin-echo techniques. *Ber. Bunsenges. Physik. Chem.*, **67**, 336-340.

McConnell, H.M. (1976) Molecular motion in biological membranes. In *Spin labeling. Theory and applications* (Ed. Berliner, L.J.) pp. 525-560. Academic Press, New York.

McKenzie, K.J., McLellan, D.R., Gentleman, S.M., Maxwell, W.L., Gennarelli, T.A. and Graham, D.I. (1996) Is β-APP a marker of axonal damage in short-surviving head injury? *Acta Neuropathol.*, **92**, 608-613.

McLean, R.A. and Sanders, W.L. (1988) Approximating degrees of freedom for standard errors in mixed linear models. *Proceedings of the Statistical Computing Section,* New Orleans. pp. 50-59. American Statistical Association.

McMurchie, E.J., Raison, J.K. and Cairncross, K.D. (1973) Temperature-induced phase changes in membranes of heart: a contrast between the thermal response of poikilotherms and homeotherms. *Comp. Biochem. Physiol. B*, **44**, 1017-1026.

Mead, R. (1988) *The design of experiments. Statistical principles for practical applications*. Cambridge University Press, Cambridge.

Mehrani, H. and Storey, K.B. (1997) Protein kinase C from bat brain: the enzyme from a hibernating mammal. *Neurochem. Int.*, **31**, 139-150.

Miller, D.H., Grossman, R.I., Reingold, S.C. and McFarland, H.F. (1998) The role of magnetic resonance techniques in understanding and managing multiple sclerosis. *Brain*, **121**, 3-24.

Mira, A. and Sargent, D.J. (2000) Strategies for speeding Markov chain Monte Carlo algorithms. Technical Report. Faculty of Economics, University of Insubria, Varese, Italy.

Morgan, B.J.T. (1984) *Elements of simulation*. Chapman and Hall, London.

Morris, C.N. (1987) Testing a point null hypothesis: the irreconcilability of *P* values and evidence (comment). *J. Am. Statist. Assoc.*, **82**, 131-133.

Morris, C.N. and Normand, S.L. (1992) Hierarchical models for combining information and for meta-analyses. In *Bayesian statistics 4* (Ed. Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M.) pp. 321-344. Oxford University Press, Oxford.

Moseley, M.E., Cohen, Y., Kucharczyk, J., Mintorovitch, J., Asgari, H.S., Wendland, M.F., Tsuruda, J. and Norman, D. (1990a) Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system. *Radiology*, **176**, 439-445.

Moseley, M.E., Cohen, Y., Mintorovitch, J., Chileuitt, L., Shimizu, H., Kucharczyk, J., Wendland, M.F. and Weinstein, P.R. (1990b) Early detection of regional cerebral ischemia in cats: comparison of diffusion- and T2-weighted MRI and spectroscopy. *Magn. Reson. Med.*, **14**, 330-346.

Müller, P. (1991) A generic approach to posterior integration and Gibbs sampling. Technical report. Department of Statistics. Purdue University.

Murray, R.K., Granner, D.K., Mayes, P.A. and Rodwell, V.W. (2000) *Harper's biochemistry.* Appleton and Lange, Stamford.

Neal, R.M. (1995) Suppressing random walks in Markov chain Monte Carlo using ordered overrelaxation. Technical report. Department of Statistics, University of Toronto.

Neil, J.J., Duong, T.Q. and Ackerman, J.J.H. (1996) Evaluation of intracellular diffusion in normal and globally-ischemic rat brain via ^{133}Cs NMR. *Magn. Reson. Med.*, **35**, 329-335.

Nelder, J.A. (1999) Statistics for the Millennium. From statistics to statistical science. *The Statistician*, **48**, 257-269.

Neyman, J. and Pearson, E.S. (1933) On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc. A*, **231**, 289-337.

Nicholson, C. and Kraig, R.P. (1981) The behavior of extracellular ions during spreading depression. In *The application of ion-selective microelectrodes* (Ed. Zeuthen, T.) pp. 217-238. Elsevier/North-Holland Biomedical Press, Amsterdam.

Norris, D.G., Niendorf, T. and Leibfritz, D. (1994) Healthy and infarcted brain tissues studied at short diffusion times: the origins of apparent restriction and the reduction in apparent diffusion coefficient. *NMR Biomed.*, **7**, 304-310.

Obrenovitch, T.P., Scheller, D., Matsumoto, T., Tegtmeier, F., Höller, M. and Symon, L. (1990) A rapid redistribution of hydrogen ions is associated with depolarization and repolarization subsequent to cerebral ischemia reperfusion. *J. Neurophysiol.*, **64**, 1125-1133.

Omar, R.Z., Wright, E.M., Turner, R.M. and Thompson, S.G. (1999) Analysing repeated measurements data: a practical comparison of methods. *Stat. Med.*, **18**, 1587-1603.

Pastores, S.M., Katz, D.P. and Kvetan, V. (1996) Splanchnic ischemia and gut mucosal injury in sepsis and the multiple organ dysfunction syndrome. *Am. J. Gastroenterol.*, **91**, 1697-1710.

Pebay-Peyroula, E. and Rosenbusch, J.P. (2001) High-resolution structures and dynamics of membrane protein-lipid complexes: a critique. *Current Opinion in Structural Biology*, **11**, 427-432.

Peterson, B.S., Anderson, A.W., Ehrenkranz, R., Staib, L.H., Tageldin, M., Colson, E., Gore, J.C., Duncan, C.C., Makuch, R. and Ment, L.R. (2003) Regional brain volumes and their later neurodevelopmental correlates in term and preterm infants. *Pediatrics*, **111**, 939-948.

Poole, C. (2001) Low P-values or narrow confidence intervals: which are more durable? *Epidemiology*, **12**, 291-294.

Polson, N.G. (1996) Convergence of Markov chain Monte Carlo algorithms. In *Bayesian statistics 5* (Ed. Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M.) pp. 297-321. Oxford University Press, Oxford.

Popper, K.R. (1974) *The logic of scientific discovery*, 7th impression. Hutchinson, London.

Potthoff, R.F. and Roy, S.N. (1964) A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313-326.

Povlishock, J.T. (1992) Traumatically induced axonal injury: pathogenesis and pathobiological implications. *Brain Pathol.*, **2**, 1-12.

Povlishock, J.T., Becker, D.P., Cheng, C.L.Y. and Vaughan, G.W. (1983) Axonal change in minor head injury. *J. Neuropathol. Exp. Neurol.*, **42**, 225-242.

Pratt, J.W. (1987) Testing a point null hypothesis: the irreconcilability of *P* values and evidence (comment). *J. Am. Statist. Assoc.*, **82**, 123-125.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1988) *Numerical recipes in C.* Cambridge University Press, Cambridge.

Priestley, M.B. (1981) *Spectral analysis and time series.* Academic Press, London.

Purcell, E.M., Torrey, H.C. and Pound, R.V. (1946) Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.*, **69**, 37-38.

Racine-Poon, A. and Smith, A.F.M. (1990) Population models. In *Statistical methodology in the pharmaceutical sciences* (Ed. Berry, D.A.S.) pp. 139-162. Marcel Dekker, New York.

Racine, A., Grieve, A.P., Flühler, H. and Smith, A.F.M. (1986) Bayesian methods in practice: experiences in the pharmaceutical industry. *Appl. Statist.*, **35**, 93-150.

Raftery, A.E. (1996) Hypothesis testing and model selection. In *Markov chain Monte Carlo in practice* (Ed. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J.) pp. 163-187. Chapman and Hall, London.

Raftery, A.E. and Lewis, S.M. (1992) How many iterations in the Gibbs sampler? In *Bayesian statistics 4* (Ed. Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M.) pp. 763-773. Oxford University Press, Oxford.

Raftery, A.E. and Lewis, S.M. (1996) Implementing MCMC. In *Markov chain Monte Carlo in practice* (Ed. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J.) pp. 115-130. Chapman and Hall, London.

Raison, J.K. and Chapman, E.A. (1976) Membrane phase changes in chilling-sensitive *vigna radiata* and their significance to growth. *Aust. J. Plant Physiol.*, **3**, 291-299.

Raison, J.K. and McMurchie, E.J. (1974) Two temperature-induced changes in mitochondrial membranes detected by spin labelling and enzyme kinetics. *Biochim. Biophys. Acta*, **363**, 135-140.

Rawlings, J.O. (1988) *Applied regression analysis: a research tool.* Wadsworth and Brooks, Belmont.

Roberts, G.O. (1996) Markov chain concepts related to sampling algorithms. In *Markov chain Monte Carlo in practice* (Ed. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J.) pp. 45-57. Chapman and Hall, London.

Robinson, J.D., Birdsall, N.J.M., Lee, A.G. and Metcalfe, J.C. (1972) ^{13}C and ^{1}H nuclear magnetic resonance relaxation measurements of the lipids of sarcoplasmic reticulum membranes. *Biochemistry*, **11**, 2903-2909.

Roe, D.J. (1997) Comparison of population pharmacokinetic modeling methods using simulated data: results from the population modeling workgroup. *Stat. Med.*, **16**, 1241-1262.

Rosenberg, P.H. and Alila, A. (1982) Hydrophobic membrane interaction of etidocaine, bupivacaine and 2-chloroprocaine. A spin and fluorescent probe study. *Naunyn-Schmiedeberg's Arch. Pharmacol.*, **319**, 95-100.

Ross, B.D., Ben-Yoseph, O. and Chenevert, T.L. (1997) In vivo magnetic resonance imaging and spectroscopy: application to brain tumors. In *Magnetic resonance spectroscopy and imaging in neurochemistry* (Ed. Bachelard, H.) pp. 145-178. Plenum Press, New York.

Rothman, K.J. (1978) A show of confidence. *N. Engl. J. Med.*, **299**, 1362-1363.

Rothman, K.J. (1998) Writing for epidemiology. *Epidemiology*, **9**, 333-337.

Roussel, S.A., van Bruggen, N., King, M.D., Houseman, J., Williams, S.R. and Gadian, D.G. (1994) Monitoring the initial expansion of focal ischaemic changes by diffusion-weighted MRI using a remote controlled method of occlusion. *NMR Biomed.*, **7**, 21-28.

SAS Institute (1990) *SAS/IML Software: usage and reference. Version 6.* SAS Institute, Cary, North Carolina.

SAS Institute (1999) *SAS/STAT user's guide, Version 8.* SAS Institute, Cary, North Carolina.

SAS Institute (2001) *SAS/STAT software: changes and enhancements, Release 8.2*. SAS Institute, Cary, North Carolina.

Schachter, S.C. (2002) Drug-mediated antiepileptogenesis in humans. *Neurology*, **59**, S34-S35.

Schafer, J.L. (1999) Multiple imputation: a primer. *Stat. Methods Med. Res.*, **8**, 3-15.

Schluchter, M.D. and Elashoff, J.D. (1990) Small-sample adjustments to tests with unbalanced repeated measures assuming several covariance structures. *J. Statist. Comput. Simul.*, **37**, 69-87.

Schreier, S., Polnaszek, C.F. and Smith, I.C.P. (1978) Spin labels in membranes. Problems in practice. *Biochim. Biophys. Acta*, **515**, 395-436.

Scott, R.C., Cross, J.H., Gadian, D.G., Jackson, G.D., Neville, B.G.R. and Connelly, A. (2003) Abnormalities in hippocampi remote from the seizure focus: a T_2 relaxometry study. *Brain*, **126**, 1968-1974.

Sellke, T., Bayarri, M.J. and Berger, J.O. (2001) Calibration of *p* values for testing precise null hypotheses. *The American Statistician*, **55**, 62-71.

Siesjö, B.K. (1981) Cell damage in the brain: a speculative synthesis. *J. Cerebr. Blood Flow Metab.*, **1**, 155-185.

Siesjö, B.K. (1992) Pathophysiology and treatment of focal cerebral ischemia. Part II: mechanisms of damage and treatment. *J. Neurosurg.*, **77**, 337-354.

Silvius, J.R. and McElhaney, R.N. (1981) Non-linear Arrhenius plots and the analysis of reaction and motional rates in biological membranes. *J. Theor. Biol.*, **88**, 135-152.

Silvius, J.R., Read, B.D. and McElhaney, R.N. (1978) Membrane enzymes: artifacts in Arrhenius plots due to temperature dependence of substrate-binding affinity. *Science*, **199**, 902-904.

Smith, A.F.M. and Cook, D.G. (1980) Straight lines with a change-point: a Bayesian analysis of some renal transplant data. *Appl. Statist.*, **29**, 180-189.

Smith, A.F.M. and Gelfand, A.E. (1992) Bayesian statistics without tears: a sampling-resampling perspective. *The American Statistician*, **46**, 84-88.

Smith, T.C., Spiegelhalter, D.J. and Thomas, A. (1995) Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat. Med.*, **14**, 2685-2699.

Sokoloff, L. (1976) Circulation and energy metabolism of the brain. In *Basic neurochemistry* (Ed. Siegel, G.J., Albers, R.W., Katzman, R. and Agranoff, B.W.) pp. 388-413. Little, Brown and Company, Boston.

Somjen, G.G., Aitken, P.G., Balestrino, M., Herreras, O. and Kawasaki, K. (1990) Spreading depression-like depolarization and selective vulnerability of neurons. A brief review. *Stroke*, **21 Suppl. III**, 179-183.

Spetzler, R.F. and Nehls, D.G. (1987) Cerebral protection against ischemia. In *Cerebral blood flow* (Ed. Wood, J.H.) pp. 651-676. McGraw Hill, New York.

Spiegelhalter, D.J., Freedman, L.S. and Parmar, M.K.B. (1994) Bayesian approaches to randomized trials. *J. R. Statist. Soc. A*, **157**, 357-416.

Spiegelhalter, D., Thomas, A., Best, N. and Gilks, W. (1995a) BUGS 0.5. Bayesian inference using Gibbs sampling. Technical report, MRC Biostatistics Unit, Institute of Public Health, Cambridge.

Spiegelhalter, D., Thomas, A., Best, N. and Gilks, W. (1995b) BUGS 0.5 examples. Technical report, MRC Biostatistics Unit, Institute of Public Health, Cambridge.

Spitz, L. (1995) Malrotation. In *Pediatric surgery*, 5th edn. (Ed. Spitz, L. and Coran, A.G.) pp. 341-347. Chapman and Hall Medical, London.

Stechmiller, J.K., Treloar, D. and Allen, N. (1997) Gut dysfunction in critically ill patients: a review of the literature. *Am. J. Critical Care*, **6**, 204-209.

Stejskal, E.O. and Tanner, J.E. (1965) Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *J. Chem. Phys.*, **42**, 288-292.

Strang, G. (1980) *Linear algebra and its applications*, Academic Press, New York.

Sugaya, E., Takato, M. and Noda, Y. (1975) Neuronal and glial activity during spreading depression in cerebral cortex of cat. *J. Neurophysiol.*, **38**, 822-841.

Sullivan, L.M., Dukes, K.A. and Losina, E. (1999) Tutorial in biostatistics. An introduction to hierarchical linear modelling. *Stat. Med.*, **18**, 855-888.

Szafer, A., Zhong, J. and Gore, J.C. (1995) Theoretical model for water diffusion in tissues. *Magn. Reson. Med.*, **33**, 697-712.

Tanford, C. (1980) *The hydrophobic effect: formation of micelles and biological membranes*, 2nd edn. Wiley, New York.

Tanner, M.A. (1996) *Tools for statistical inference. Methods for the exploration of posterior distributions and likelihood functions*, 3rd edn. Springer, New York.

Tanner, J.E. and Stejskal, E.O. (1968) Restricted self-diffusion of protons in colloidal systems by the pulsed-gradient, spin-echo method. *J. Chem. Phys.*, **49**, 1768-1777.

Teasdale, G.M. and Graham, D.I. (1998) Craniocerebral trauma: protection and retrieval of the neuronal population after injury. *Neurosurgery*, **43**, 723-737.

Temkin, N.R. (2001) Antiepileptogenesis and seizure prevention trials with antiepileptic drugs: meta-analysis of controlled trials. *Epilepsia*, **42**, 515-524.

Theil, H. (1971) *Principles of econometrics*, North-Holland Publishing Company, Amsterdam.

Thomas, D.L., Lythgoe, M.F., Pell, G.S., Calamante, F. and Ordidge, R.J. (2000) The measurement of diffusion and perfusion in biological systems using magnetic resonance imaging. *Phys. Med. Biol.*, **45**, R97-R138.

Thoresen, M., Penrice, J., Lorek, A., Cady, E.B., Wylezinska, M., Kirkbride, V., Cooper, C.E., Brown, G.C., Edwards, A.D., Wyatt, J.S. and Reynolds, E.O.R. (1995) Mild hypothermia after severe transient hypoxia-ischemia ameliorates delayed cerebral energy failure in the newborn piglet. *Pediatr. Res.*, **37**, 667-670.

Thornton, J.S., Ordidge, R.J., Penrice, J., Cady, E.B., Amess, P.N., Punwani, S., Clemence, M. and Wyatt, J.S. (1998) Temporal and anatomical variations of brain water apparent diffusion coefficient in perinatal cerebral hypoxic-ischemic injury: relationships to cerebral energy metabolism. *Magn. Reson. Med.*, **39**, 920-927.

Tierney, L. (1996) Introduction to general state-space Markov chain theory. In *Markov chain Monte Carlo in practice* (Ed. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J.) pp. 59-74. Chapman and Hall, London.

Toet, M.C., Hellström-Westas, L., Groenendaal, F., Eken, P. and de Vries, L.S. (1999) Amplitude integrated EEG 3 and 6 hours after birth in full term neonates with hypoxic-ischaemic encephalopathy. *Arch. Dis. Child Fetal Neonatal Ed.*, **81**, F19-F23.

Träuble, H. and Eibl, H. (1974) Electrostatic effects on lipid phase transitions: membrane structure and ionic environment. *Proc. Natl. Acad. Sci. USA*, **71**, 214-219.

Träuble, H., Teubner, M., Woolley, P. and Eibl, H. (1996) Electrostatic interactions at charged lipid membranes. I. Effects of pH and univalent cations on membrane structure. *Biophys. Chem.*, **4**, 319-342.

van der Toorn, A., Syková, E., Dijkhuizen, R.M., Voříšek, I., Vargová, L., Škobisová, E., van Lookeren-Campagne, M., Reese, T. and Nicolay, K. (1996) Dynamic changes in water ADC,

energy metabolism, extracellular space volume, and tortuosity in neonatal rat brain during global ischemia. *Magn. Reson. Med.*, **36**, 52-60.

Vardeman, S.B. (1987) Testing a point null hypothesis: the irreconcilability of *P* values and evidence (comment). *J. Am. Statist. Assoc.*, **82**, 130-131.

Vejchapipat, P. (2001) Intestinal and hepatic metabolism after intestinal ischaemia-reperfusion. Ph.D. Thesis. Institute of Child Health, University College, University of London.

Vejchapipat, P., Proctor, E., Ramsay, A., Petros, A., Gadian, D.G., Spitz, L. and Pierro, A. (2002) Intestinal energy metabolism after ischemia-reperfusion: effects of moderate hypothermia and perfluorocarbons. *J. Pediatr. Surg.*, **37**, 786-790.

Vejchapipat, P., Williams, S.R., Proctor, E., Lauro, V., Spitz, L. and Pierro, A. (2001) Moderate hypothermia ameliorates liver energy failure after intestinal ischaemia-reperfusion in anaesthetised rats. *J. Pediatr. Surg.*, **36**, 269-275.

Vines, S.K., Gilks, W.R. and Wild, P. (1996) Fitting Bayesian multiple random effects models. *Statistics and Computing*, **6**, 337-346.

Vink, R. and McIntosh, T.K. (1997) Traumatic brain injury. In *Magnetic resonance spectroscopy and imaging in neurochemistry* (Ed. Bachelard, H.) pp. 91-116. Plenum Press, New York.

Vonesh, E.F. (1992) Non-linear models for the analysis of longitudinal data. *Stat. Med.*, **11**, 1929-1954.

Vonesh, E.F. and Carter, R.L. (1992) Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics*, **48**, 1-17.

Wakefield, J. (1996) The Bayesian analysis of population pharmacokinetic models. *J. Am. Statist. Assoc.*, **91**, 62-75.

Walker, M.C., White, H.S. and Sander, J.W.A.S. (2002) Disease modification in partial epilepsy. *Brain*, **125**, 1937-1950.

Warren, G.B., Toon, P.A., Birdsall, N.J.M., Lee, A.G. and Metcalfe, J.C. (1974) Reconstitution of a calcium pump using defined membrane components. *Proc. Natl. Acad. Sci. USA*, **71**, 622-626.

Welbourn, C.R.B., Goldman, G., Paterson, I.S., Valeri, C.R., Shepro, D. and Hechtman, H.B. (1991) Pathophysiology of ischaemia reperfusion injury: central role of the neutrophil. *Br. J. Surg.*, **78**, 651-655.

Wick, M., Nagatomo, Y., Prielmeier, F. and Frahm, J. (1995) Alteration of intracellular metabolite diffusion in rat brain *in vivo* during ischemia and reperfusion. *Stroke*, **26**, 1930-1933.

Williams, D.A. (1970) Discrimination between regression models to determine the pattern of enzyme synthesis in synchronous cell cultures. *Biometrics*, **26**, 23-32.

Wolfe, J. and Bagnall, D. (1979) Statistical tests to decide between straight line segments and curves as suitable fits to Arrhenius plots or other data. In *Low temperature stress in crop plants. The role of the membrane*. (Ed. Lyons, J.M., Graham, D. and Raison, J.K.) pp. 527-533. Academic Press, New York.

Wolfe, J. and Bagnall, D.J. (1980) Arrhenius plots - curves or straight lines? *Ann. Bot.*, **45**, 485-488.

Wolfinger, R.D. and Kass, R.E. (2000) Nonconjugate Bayesian analysis of variance component models. *Biometrics*, **56**, 768-774.

Wolfinger, R.D. and Rosner, G.L. (1996) Bayesian and frequentist analyses of an *in vivo* experiment in tumor hemodynamics. In *Bayesian biostatistics* (Ed. Berry, D.A. and Stangl, D.) pp. 389-409. Marcel Dekker, New York.

Wynn-Williams, A.T. (1976) An explanation of apparent sudden change in the activation energy of membrane enzymes. *Biochem. J.*, **157**, 279-281.

Zakim, D., Kavecansky, J. and Scarlata, S. (1992) Are membrane enzymes regulated by the viscosity of the membrane environment? *Biochemistry*, **31**, 11589-11594.

Zeger, S.L. and Liang, K.-Y. (1992) An overview of methods for the analysis of longitudinal data. *Stat. Med.*, **11**, 1825-1839.

Zeger, S.L., Liang, K.-Y. and Albert, P.S. (1988) Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, **44**, 1049-1060.

Zellner, A. (1987) Testing precise hypotheses. Comment. *Statistical Sci.*, **2**, 339-341.