University of Southampton

Faculty of Engineering, Science and Mathematics

School of Mathematics

Empirical Models for Cyclic Voltammograms

by

Jeffrey J. Samuel

Thesis for the degree of Doctor of Philosophy

August 2010

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

SCHOOL OF MATHEMATICS

Doctor of Philosophy

Empirical Models for Cyclic Voltammograms

by Jeffrey Joseph Samuel

Technological devices such as mobile phones and laptop computers have created an immense demand for efficient and long lasting power sources such as Lithium-ion batteries. Key to improving the current generation of batteries is the understanding of Lithium based materials that are suitable for use in batteries. Researchers investigating battery materials often plot the output from their experiments as a cyclic voltammogram. A voltammogram is simply a plot of Current against Potential.

In this thesis we investigate a range of empirical models for cyclic voltammograms with a Bayesian perspective, using data from experiments carried out in the School of Chemistry, University of Southampton. This work is motivated by the lack of well formulated mathematical models for cyclic voltammograms involving a Lithium-ion compound. By setting the models within a Bayesian framework, we are able to obtain posterior predictive distributions for characteristics of the voltammogram of interest to chemists.

Markov Chain Monte Carlo sampling methods are used to explore the posterior distribution of the model parameters and to estimate the posterior predictive distributions. We investigate four methods of modelling the experimental data: multiple regression models for summary statistics, autoregressive models, sinusoidal models and stochastic volatility models. The application of Bayesian model choice techniques showed that the sinusoidal model provided the best description of the data.

Contents

List of Tables

List of Figures

DECLARATION OF AUTHORSHIP

I, Jeffrey Joseph Samuel, declare that the thesis entitled

Empirical Models for Cyclic Voltammograms

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research.
I confirm that:

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

Signed:
Date:

Acknowledgements

There are many people who I would like to express gratitude to. Firstly, I would like to thank my supervisor Dr. Sujit Sahu and my advisor Prof. Sue Lewis for all their support, advice and time throughout the course of the research and writing of this thesis. I would also like to express a sincere note of gratitude to Dr. David Woods for all his support, advice and time. I am especially grateful to Prof. John Owens and his solid state electrochemistry group, especially Girt Vitens, Alan Spong and Matthew Roberts for conducting the experiments from which the data were obtained and for their advice and input to this collaborative project. I am also like grateful to the Engineering and Physical Sciences Research Council for funding this research.

I would like to thank the many student colleagues who supported me throughout my PhD. I wish to thank the various teachers at The John Fisher School who taught me in my formative years, especially Fred Flaherty, Colin Smith, Sadie Sheehan, the late Fr. Fawcett and Patrick Liddiard who ensured that The John Fisher School was a pleasant place to study.

A final note of gratitude goes to various friends and family, especially my Mum, my Grandad, my late Nana (who is missed dearly), my Brother and Larry Mackenzie. My greatest thanks goes to the Holy Trinity, Mother Mary, Saint Joseph, the angels and the saints through whom all things are possible.

Chapter 1

Introduction

1.1 Motivation

Advances in technological devices have created an immense demand for efficient battery power sources as demonstrated by the development of mobile phones and laptop computers. If battery power sources had not been improved, mobile phones would still be of a similar size as those in the 1980s. An example of an advancement is the development of the Lithium-ion (Li-ion) battery. Without this development, the invention of the laptop computer would not have been brought to fruition. However, there are still improvements to be made. An example of this is to extend the length of time a mobile phone or laptop computer can be used before the battery requires re-charging. The Li-ion battery is the type of battery use in the majority of mobile phones and computer laptops.

Chemists have a keen interest in furthering their knowledge about the chemical compositions used in batteries, especially Lithium based compositions. It would therefore be beneficial to the chemists if they had a model that could predict various characteristics about the Current obtained from a given chemical composition. At present there is no well formulated mathematical model available for Current as

Potential varies over time. An alternative is to use experimental data to produce an empirical model, for example using time series analysis with a Bayesian perspective. Once we have chosen a model to use, it will then be possible to estimate particular characteristics of the Current output curve. These characteristics are described in Section 1.4.

1.2 The Experiment

The experiment described in this section is part of the Combechem *e*-Science project (see www.combechem.org for further details). The experiment was carried out by the Solid State Electrochemistry Group (referred to as the chemists henceforth) at the University of Southampton. The purpose of this experiment was to explore how Carbon affects the electrical output of battery power sources. The experiment involved an 8×8 array of electrodes (or channels) shown in the left-hand panel in Figure 1.1. The middle panel in Figure 1.1 shows a side view of the array and the cell fully assembled is shown in the right-hand panel.

Figure 1.1: Equipment used in the experiment.

Sixty-three electrodes were coated with a chemical composition and the remaining electrode was used as a reference electrode. Sixteen arrays can be run in parallel which gives the capacity to study sixteen arrays each using sixty-three electrodes, giving the capacity to examine 1008 individual electrodes of different chemical com-

positions. The composition has three components, namely Acetylene Black (a pure fine powdered Carbon), which conducts the Current and is called Carbon henceforth, Active (Lithium Manganese Oxide, $LiMn_2O_4$) which stores the energy, and Binder which binds the Carbon and Active together.

The array of electrodes forms part of a circuit in which the Potential is varied throughout the experiment. The Potential, measured in Volts (V), is applied to the entire array and is varied between 3.20V and 4.50V since this is the chemists' range of interest. The experiment is started by setting the Potential at 3.20V and increasing it until it reaches 4.50V. The Potential is then decreased from 4.50V back to 3.20V, which defines one complete cycle. The rate at which the Potential is increased or decreased is called the scan rate and is measured in millivolts per second (mVs^{-1}). In the experiment discussed in this thesis, three different scan rates were used: 0.05 mVs^{-1}, 0.10mVs^{-1} and 0.20mVs^{-1} (referred to as scan rates 1, 2 and 3 respectively).

The Current is a measure of the amount of electric charge passed through the electrode per second and is measured in milliamps (mA). The Current characterises how fast the electrode can adapt to the new Potential supplied. The Potential and Current are measured every 10 seconds. A burn-in cycle is the length of time required for the equipment to settle down before the output produced from the experiment is stable. Therefore, the burn-in cycle is removed before carrying out any analysis.

In the experiment, the percentage of Carbon was varied across the array. The Carbon levels used were 0%, 1%, 2%, 3%, 5%, 7%, 10% and 20%. Seven channels were allocated a Carbon level of 0% and eight channels were allocated to each of the remaining Carbon levels. The level of Binder was set at 10% and the amount of Active was set at $100(0.9\text{-}c)\%$ where c is the proportion of Carbon used.

For scan rates 1, 2 and 3 we have 2, 3 and 4 replications, respectively. Each replication consisted of two complete cycles which can be used for data analysis. Thus the

total experiment consisted of 19 cycles. The order in which the the scan rates were run was randomised (see Table 1.1) so that any systematic variation in the output from the experiment was avoided.

Run	1	2	3	4	5	6	7	8	9
Scan Rate (s)	0.10	0.20	0.10	0.20	0.05	0.20	0.05	0.10	0.20

Table 1.1: Order in which the scan rates were run.

As the scan rate increases, the time taken to complete one cycle decreases. This is because we are increasing or decreasing the Potential from 3.20V to 4.50V via different scan rates. A cycle for scan rate 1 takes approximately 14 hours, for scan rate 2 it takes approximately 7 hours and for scan rate 3 it takes approximately 3.5 hours.

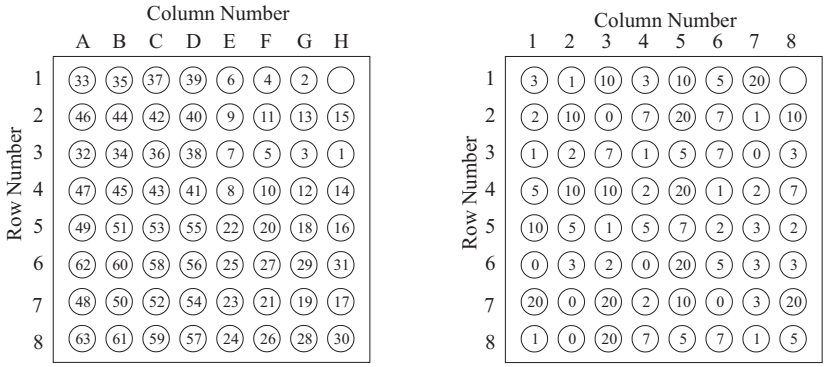


Figure 1.2: Channel numbers (right) and their Carbon allocation (left). The empty circle represents the reference electrode.

The row and column number for each electrode in the array were recorded and allocated a Carbon level at random. This randomisation is to avoid any systematic error occurring. For example, if all the channels with Carbon set at 1% were along one edge of the array and some edge effect existed, then the results from these

channels could produce incorrect results. The grids of the channel numbers and Carbon allocation are shown in Figure 1.2.

The output values of the Current for each time index from each channel are extremely small and were scaled up by a factor of 10^6 so that they are measured in nano-amps. From herein, Current will be the original Current multiplied by 10^6. This also makes the analysis easier as it avoids underflow in the computation.

The time index is denoted by T. Let T_a denote the time elapsed in seconds since the start of the run (see Table 1.2). An observation on Current is made every 10 seconds.

Let $y_{s,q,k}(T)$ denote the observed Current at the T^{th} time point, for Carbon level $k = 1, 2, \ldots, 8$, scan rate $s = 1, 2, 3$ and channel $q = 1, 2, \ldots, 63$. Note each run has a different scan rate s, see Table 1.1. The Potential at the T^{th} observation for scan rate s is denoted by $p_s(T)$.

Scan Rate					
Rate 1		Rate 2		Rate 3	
T	T_a	T	T_a	T	T_a
1	10	1	10	1	10
2	20	2	20	2	20
3	30	3	30	3	30
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	1300	13000
⋮	⋮	2600	26000		
⋮	⋮				
5200	52000				

Table 1.2: Time elapsed, T_a, in seconds, since the start of the run with corresponding time index, T, for each of the three scan rates.

k	1	2	3	4	5	6	7	8
Percentage of Carbon	0	1	2	3	5	7	10	20

Table 1.3: Percentages of Carbon with index k.

1.3 Voltammogram

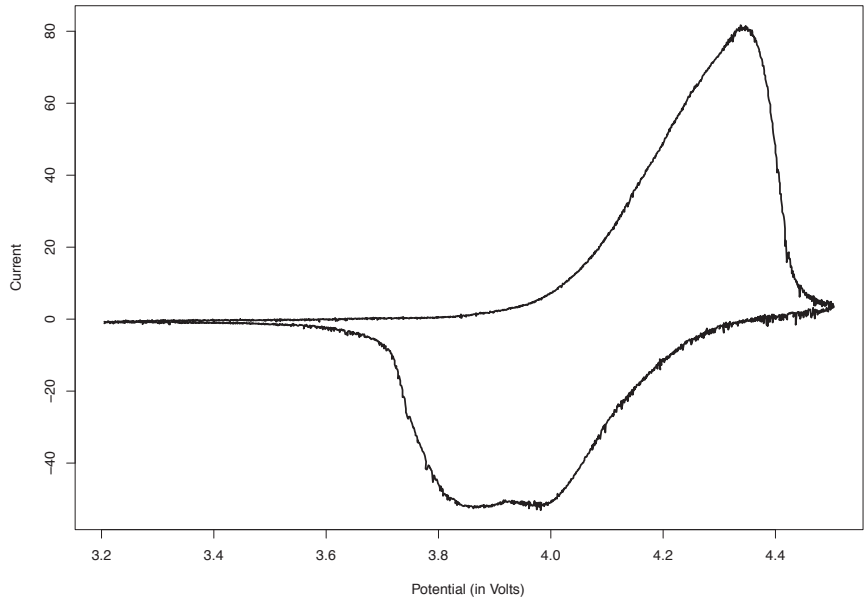


Figure 1.3: Voltammogram from the experiment using scan rate 2 and Carbon at 10%.

A plot of Current against Potential, known as a voltammogram, is often used to visualise output from many electrochemistry experiments such as the one described above. An example of a voltammogram is given in Figure 1.3 which shows one

complete cycle. Recall that for one complete cycle the Potential is started at 3.20V, is increased to 4.50V and is decreased to 3.20V according to the scan rate. For the voltammogram shown in Figure 1.3, we start at the point where the Potential is 3.20V and the Current is approximately zero. We follow the graph in a clockwise direction until we return to the starting point. It should be noted that at the start of a cycle the Current will not necessarily start at approximately zero due to possible experimental error.

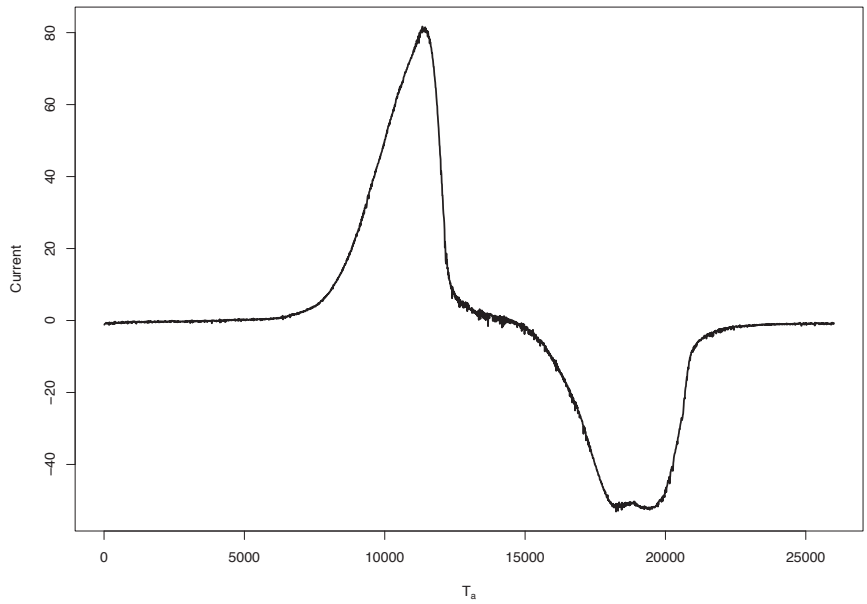


Figure 1.4: Plot of Current as a time series from the experiment using scan rate 2 and Carbon at 10%, where elapsed time T_a is measured in seconds.

Instead of plotting Current against Potential, an alternative is to plot the Current as a time series, as shown in Figure 1.4. When viewing the data in this fashion the task of building a model becomes easier.

We analyse the data from each of the three scan rates separately because the three

time series (as shown in Table 1.2) have different lengths (5200, 2600 and 1300).

The ideal situation would be to have a joint multivariate analysis of the data from the three different scan rates. It may be feasible to construct a hierarchical model for the data from all the different scan rates. However, in this thesis we analyse the data from the three different scan rates separately to avoid the issue of temporal misalignment. We concentrate mainly on the data from scan rate 2 because there were fewer malfunctioning channels (as discussed with the chemists).

1.4 Characteristics of Voltammograms

In this section we describe the characteristics of a voltammogram which are of scientific interest to the chemists. These are defined for a typical voltammogram which arises from a single channel, for a given scan rate, s, and a given Carbon level, k. The features are of value to the chemists because they provide information on the conditions where an experimental battery provides, for example, the greatest output of Current.

The definitions are given in Table 1.4 in simple terms, and are illustrated in Figures 1.5 and 1.6. We return to the characteristics in Section 2.4, where they are defined more formally and applied to the experiment.

Characteristic	Symbol	Definition
Minimum Potential	$P_{s,k}^{(\min)}$	The Potential value that produces the *Minimum Current*.
Minimum Time	$T_{s,k}^{(\min)}$	The time index where the *Minimum Current* occurs.
Minimum Current	$I_{s,k}^{(\min)}$	The smallest value of the observed Current.
Peak Potential	$P_{s,k}^{(\max)}$	The Potential value that produces the *Peak Current*.
Peak Time	$T_{s,k}^{(\max)}$	The time point where the *Peak Current* occurs.
Peak Current	$I_{s,k}^{(\max)}$	The maximum value of the observed Current.
Peak Separation in Potential	$P_{s,k}^{(\text{sep})}$	The difference between the *Peak Potential* and the *Minimum Potential*.
Peak Separation in Time	$T_{s,k}^{(\text{sep})}$	The difference between the *Peak Time* and the *Minimum Time*.
Peak Separation in Current	$I_{s,k}^{(\text{sep})}$	The difference between the *Peak Current* and the *Minimum Current*.
Peak Width in Potential	$P_{s,k}^{(\text{wid})}$	The difference between the first and the last value of Potential when the Current is at half its observed maximum value.
Peak Width in Time	$T_{s,k}^{(\text{wid})}$	The difference between the first and the last time when the Current is at half its observed maximum value.

Table 1.4: Characteristics and their symbols for a voltammogram where subscripts s and k are the scan rate and the level (proportion) of Carbon respectively. The superscript min, max, sep or wid is according to whether the characteristic is a minimum, a maximum, a separation or a width respectively. These characteristics are shown graphically in Figures 1.5 and 1.6.

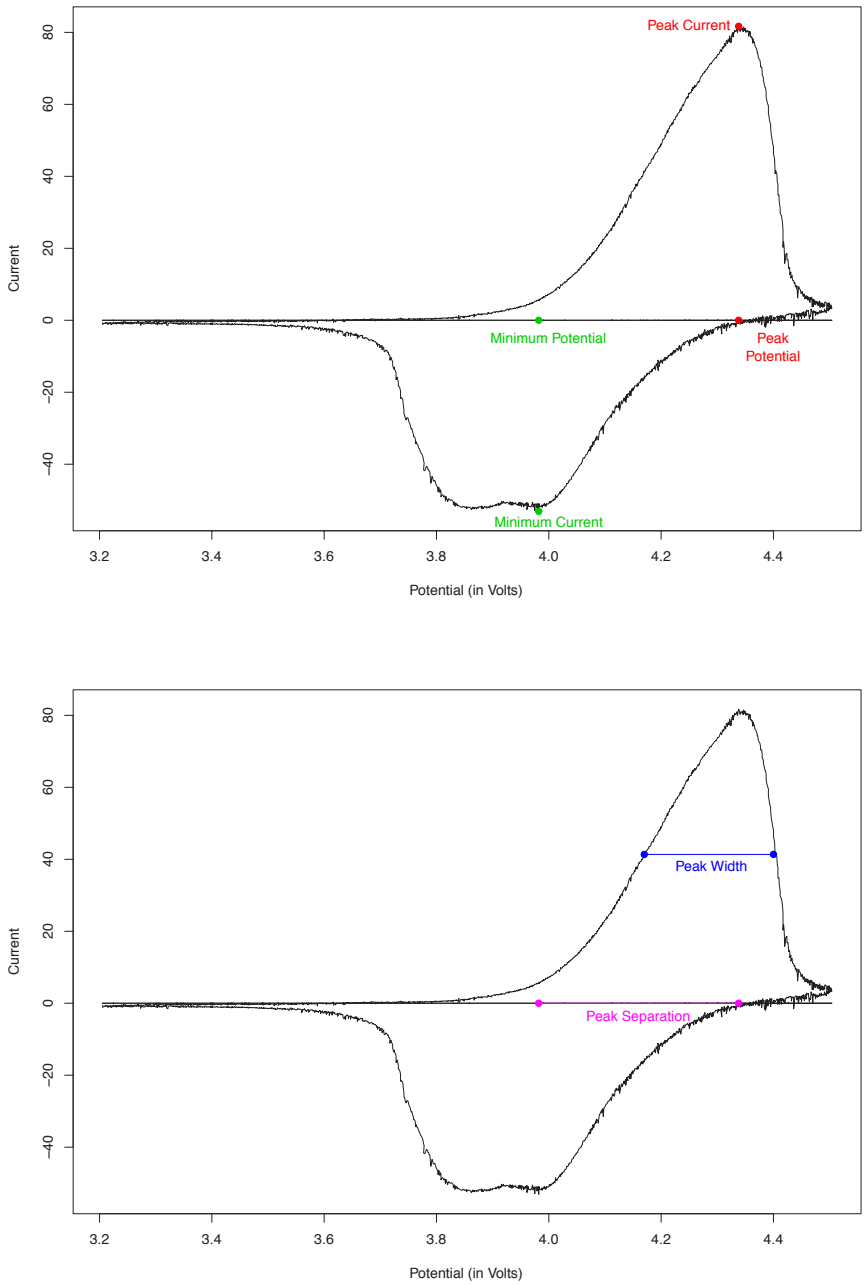


Figure 1.5: Characteristics of interest in a voltammogram.

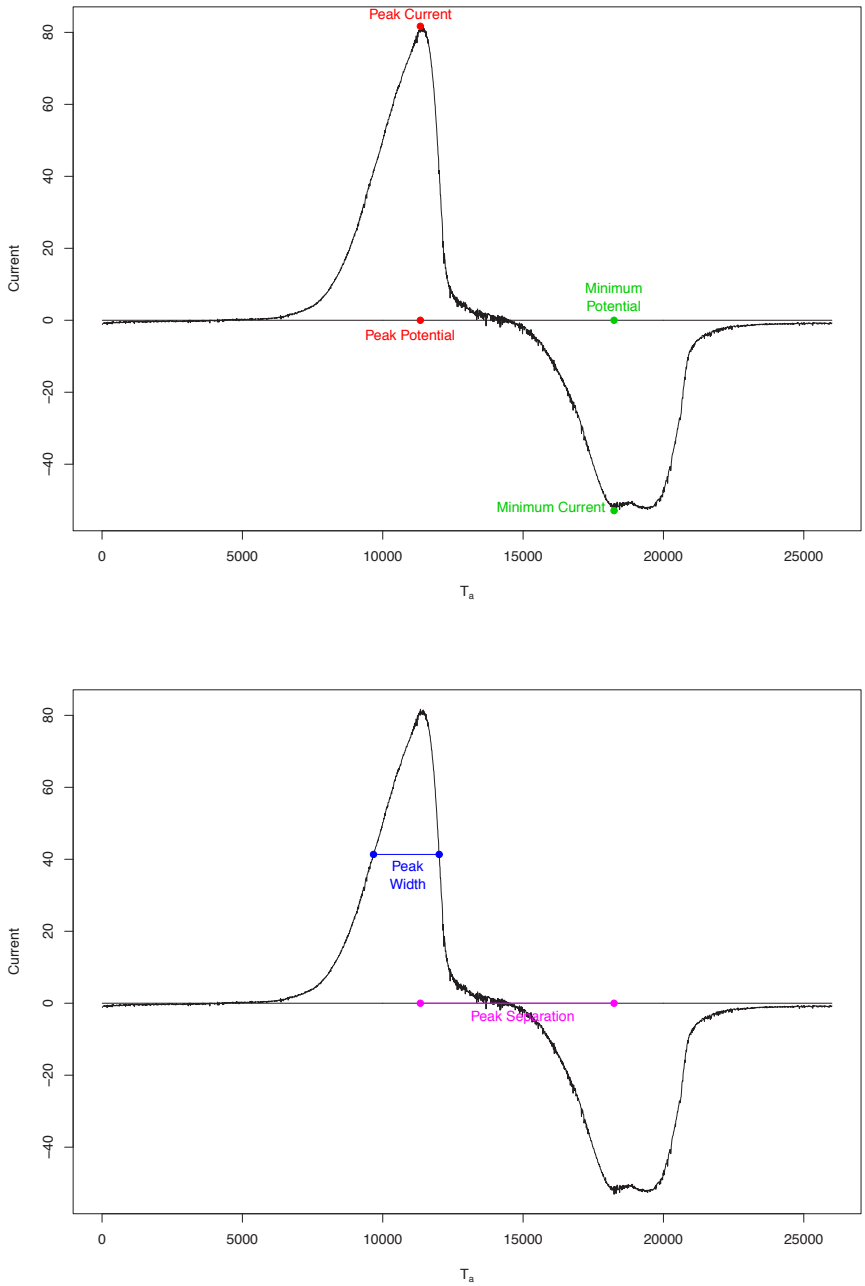


Figure 1.6: Characteristics of interest in a time series representation of a voltammogram, where T_a is the elapsed time measured in seconds.

1.5 Literature Review of Cyclic Voltammogram Models

There are a number of articles that have proposed models for cyclic voltammograms which have arisen from a variety of different experiments, for example, see Lundquist et al. (2001), Lovric and Scholz (2003), Novak et al. (2001), Myland and Oldham (2002a) and Myland and Oldham (2002b). We were informed by the chemists that these were the most relevant papers for their experimental data. However, as will be highlighted, the models proposed in these articles are not suitable for statistical analysis for a variety of reasons.

Before we comment on the model proposed in Myland and Oldham (2002a), we present the model, highlighting the key equations. Further, we note that we follow the notation used in Myland and Oldham (2002a). The authors point out that the key relationship is embodied by the equations

$$E(t) \;=\; E(<0) + v(t_{\text{rev}} - |t_{\text{rev}} - t|) \tag{1.1}$$

$$E(t) - E(<0) \;=\; \eta_{\text{conc}}(t) + \eta^{\text{org}}_{\text{ohm}}(t) + \eta^{\text{aq}}_{\text{ohm}}(t) + \eta^{\text{org}}_{\text{asym}}(t) + \eta^{\text{aq}}_{\text{asym}}(t), \tag{1.2}$$

where v is the scan rate, t is the time in seconds and t_{rev} is the reversal time. The formal definition for the reversal time is given by (1.15). At each time t, the true Potential is $E(t)$ which is given by (1.1), and $E(<0)$ is the resting electrode Potential, that is the Potential of the electrode before the experiment has started. Equation (1.2) shows the difference between the true Potential at time t and the resting Potential, $E(<0)$. $E(<0)$ is given by the Nernst's relationship

$$E(<0) = E^{\circ} + \frac{RT}{F} \ln\left\{\frac{c^2_{\text{PA}}}{c_{\text{N}} c_{\text{CA}}}\right\} \tag{1.3}$$

where the following are taken to have known values:

E° is standard Potential and, when multiplied by Faraday's Constant, it is equal to the Gibbs energy that accompanies a chemical reaction described in Myland and Oldham (2002a, Section 3),

T is Temperature in Kelvin, R denotes the gas constant and F denotes Faraday's constant.

The concentration coefficients c_{N} and c_{PA} used in Equation (1.3) are given by

$$c_{\mathrm{N}} = c_{\mathrm{N}}^{\mathrm{orig}} - c_{\mathrm{PA}}, \tag{1.4}$$

$$\tag{1.5}$$

and

$$c_{\mathrm{PA}} = \frac{c_{\mathrm{CA}}^{\mathrm{b}}\xi}{2}\left(-1+\sqrt{1+\frac{4c_{\mathrm{N}}^{\mathrm{orig}}}{\xi c_{\mathrm{CA}}^{\mathrm{b}}}}\right), \tag{1.6}$$

where $\xi = \exp\left\{\frac{F(E(<0)-E^\circ)}{RT}\right\}$. The original concentration coefficients $c_{\mathrm{N}}^{\mathrm{orig}}$ and $c_{\mathrm{CA}}^{\mathrm{b}}$ and hence c_{PA} and c_{N} are assumed to have known values.

We are now able to return to Equations (1.1), (1.2) and define the notation used in these equations as set out in Myland and Oldham (2002a).

$$\eta_{\mathrm{conc}}(t) = \frac{RT}{F}\ln\left\{\frac{c_{\mathrm{PA}}(0,t)c_{\mathrm{PA}}(Z,t)c_{\mathrm{N}}(0,0)c_{\mathrm{CA}}(Z,0)}{c_{\mathrm{PA}}^2(0,0)c_{\mathrm{N}}(0,t)c_{\mathrm{CA}}(Z,t)}\right\} \tag{1.7}$$

$$\eta_{\mathrm{ohm}}^{\mathrm{org}}(t) = \frac{I(t)RT}{F^2A(D_{\mathrm{P}}+D_{\mathrm{A}}^{\mathrm{org}})}\int_0^Z \frac{1}{c_{\mathrm{PA}}(z,t)}\,\mathrm{d}z \tag{1.8}$$

$$\eta_{\mathrm{ohm}}^{\mathrm{aq}}(t) = \frac{I(t)RT}{F^2A(D_{\mathrm{C}}+D_{\mathrm{A}}^{\mathrm{aq}})}\int_Z^{Z_{\mathrm{ref}}} \frac{1}{c_{\mathrm{CA}}(z,t)}\,\mathrm{d}z \tag{1.9}$$

$$\eta_{\mathrm{asym}}^{\mathrm{org}} = \frac{RT(D_{\mathrm{P}}-D_{\mathrm{A}}^{\mathrm{org}})}{F(D_{\mathrm{P}}+D_{\mathrm{A}}^{\mathrm{org}})}\ln\left\{\frac{c_{\mathrm{PA}}(Z,t)}{c_{\mathrm{PA}}(0,t)}\right\} \tag{1.10}$$

$$\eta_{\mathrm{asym}}^{\mathrm{aq}} = \frac{RT(D_{\mathrm{C}}-D_{\mathrm{A}}^{\mathrm{aq}})}{F(D_{\mathrm{C}}+D_{\mathrm{A}}^{\mathrm{aq}})}\ln\left\{\frac{c_{\mathrm{CA}}(Z_{\mathrm{ref}},t)}{c_{\mathrm{PA}}(Z,t)}\right\} \tag{1.11}$$

where:

$$c_{\mathrm{N}}(0,t) = c_{\mathrm{N}} - \frac{I(t) \times \Theta_4\{\frac{1}{2}; \frac{D_{\mathrm{N}}t}{Z^2}\}}{FAZ}, \tag{1.12}$$

$$c_{\mathrm{PA}}(z,t) = c_{\mathrm{PA}}(0,0) + \frac{D_{\mathrm{PA}}}{2FAZD_{\mathrm{A}}^{\mathrm{org}}}\left[I(t) \times \Theta_4\left\{\frac{z}{2Z}; \frac{D_{\mathrm{PA}}t}{Z^2}\right\}\right], + \frac{D_{\mathrm{PA}}}{2FAZD_{\mathrm{P}}}\left[I(t) \times \Theta_4\left\{\frac{Z-z}{2Z}; \frac{D_{\mathrm{PA}}t}{Z^2}\right\}\right], \tag{1.13}$$

$$c_{\mathrm{CA}}(z,t) = c_{CA}^{\mathrm{b}} - \frac{\sqrt{D_{\mathrm{CA}}}}{2FAD_A^{\mathrm{aq}}}\left[\frac{I(t)}{\sqrt{\pi t}}\exp\left\{\frac{-(z-Z)^2}{4D_{\mathrm{CAt}}}\right\}\right], \tag{1.14}$$

$$\Theta_4(\zeta,\tau) = 1 + 2\sum_{j=1}^{\infty}(-1)^j \cos(2j\pi\zeta)\exp(-j^2\pi^2\tau).$$

Note that $c_{\mathrm{PA}}(0,0) = c_{\mathrm{PA}}$ and $c_{\mathrm{N}}(0,0) = c_{\mathrm{N}}$. The constants listed below, as defined in Myland and Oldham (2002a), used in Equations (1.7)-(1.14) are assumed to have known values.

The diffusion coefficients D_{N}, D_{P} and $D_{\mathrm{A}}^{\mathrm{org}}$. As a consequence D_{PA} can also be calculated since:

$$D_{\mathrm{PA}} = \frac{2D_{\mathrm{P}}D_{\mathrm{A}}^{\mathrm{org}}}{D_{\mathrm{P}} + D_A^{\mathrm{org}}}.$$

The diffusion coefficients D_{C} and $D_{\mathrm{A}}^{\mathrm{aq}}$. As a consequence D_{CA} can also be calculated since:

$$D_{\mathrm{CA}} = \frac{2D_{\mathrm{C}}D_{\mathrm{A}}^{\mathrm{aq}}}{D_{\mathrm{C}} + D_A^{\mathrm{aq}}}.$$

The area of the electrode denoted by A and the width of the organic layer (see Figure 1.7).

Z is defined to be the distance between the working electrode and the end of the organic layer (see Figure 1.7).

Z_{dep} is defined to be the distance between the working electrode and the end of the depletion zone (see Figure 1.7).

Z_{ref} is defined to be the gap between the working electrode and the reference electrode (see Figure 1.7).

z is defined to be the distance from the working electrode (see Figure 1.7).

The reversal Potential is denoted by E_{rev} and the reversal time is given by

$$\frac{E_{\text{rev}} - E(< 0)}{v}. \tag{1.15}$$

$I(t)$ is a guess of the Current at time t.

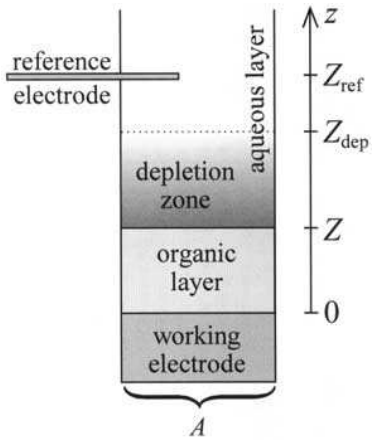


Figure 1.7: Schematic from Myland and Oldham (2002a) showing the construction of their cell and the definitions of the different distances used in their model.

The difference between $E(t)$ and $E(< 0)$ is denoted by ε, and is given by

$$\varepsilon = \eta_{\text{conc}}(t) + \eta^{\text{org}}_{\text{ohm}}(t) + \eta^{\text{org}}_{\text{asym}}(t) + \eta^{\text{asym}}_{\text{aq}}(t) - v\left(t_{\text{rev}} - |t_{\text{rev}} - t|\right). \tag{1.16}$$

The sign and magnitude of ε are used to refine the guess of $I(t)$. If the mth guess for the Current, I_m, gives a difference ε_m and the $(m+1)$th guess, I_{m+1}, is still unsatisfactory, giving an offset of ε_{m+1}, then the $(m+2)$th guess is set to

$$I_{m+2} = \frac{I_m\varepsilon_{m+1} - I_{m+1}\varepsilon_m}{\varepsilon_{m+1} - \varepsilon_m}.$$

In general, the difference ε in Equation (1.16) has to be less than 1 micro-Volt for the difference to be acceptable.

The experiment in Myland and Oldham (2002a) differs from ours as they are using a solution and our experiment is solid state based. In solution, reactions occur much more quickly and experiments take approximately a minute to execute. In a solid, reactions occur much more slowly, hence experiments take much longer to run. As a consequence, the diffusive and concentration coefficients will differ significantly. We would therefore have to treat the diffusive and concentration coefficients as unknowns in our application. We were also informed by the chemists that, even if we had estimates of these coefficients, the model would be inaccurate because of the quicker reaction times in liquid compared to solids. One criticism of the model is that there is no formal way of using empirical evidence for the initial guess of $I(t)$. Further, as this model has no error structure, modelling the underlying error that is present in most, if not all experiments, leads to the belief that this model is inappropriate.

Lundquist et al. (2001) only consider the discharge part of the cycle. This approach is fine if the objective is to analyse the characteristics that occur in the discharge process. However we are concerned with a variety of characteristics that occur throughout a full cycle. In addition, the authors are trying to adapt what happens in solution to solid materials. This approach is not necessarily the best as information from one type of experiment is used to model data from another type of experiment.

The model proposed in Novak et al. (2001) is used to estimate the diffusion coefficient of Lithium in $LiMn_2O_4$ (which measures the movement of Lithium). The authors only consider using their model to simulate the discharge part of the cycle. In addition, the authors only consider a Carbon loading (i.e proportions) of 18.4%.

In conclusion, we have been unable to find a model that incorporates different Carbon loadings into the current array of cyclic voltammogram models. Also, there

is no possibility of incorporating previous scientific knowledge into these models. The chemists also informed us that none of the models presently available in the literature are suitable for adaptation to our experiment. A further criticism, is that there is no consideration to account for the random error that is present in all experiments. It should also be noted that it is not possible to predict the characteristics of interest using these models. In fact, we have been unable to find any article that considers finding predictive distributions for different characteristics for the battery experiments.

The statistical models that we investigate in this thesis are purely empirical and are not based on physical properties of Lithium-ion batteries. In theory, models based on physical characteristics will perform better than the proposed empirical ones. However, the models based on physical characteristics reviewed in this section do not account for experimental error, forcing us to take the empirical modelling route.

1.6 Context of Project

The work presented in this thesis is part of the Combechem *e*-Science project. The raw data was obtained from experiments that were conducted by Alan Spong and post-doc Girts Vitens who provided information about the important characteristics of the voltammogram and definitions of the characteristics. We were also informed by Girts Vitens which characteristics were of particular importance. The concept of being able to calculate the posterior predictive distributions of these characteristics was of particular interest to them. Through regular meetings, especially with Girts Vitens, we discussed results and received feedback. We discussed the validity of the analysis conducted in Chapter 2 and output obtained from the models investigated, as well as the conclusions reached in Chapters 5, 6, 7 and 8. Further to this, they also gave direction on the current literature for modelling cyclic voltammograms. It

should be noted that neither were involved in the formal supervision process.
In our discussions with the chemists, they were unable to provide us with values for the prior distributions used in the models proposed in this thesis (see Section 3.2.1 for a detailed exposition about prior distributions). Also, the models examined in Chapters 6, 7 and 8 are purely empirical and are not based on physical properties of Lithium. They have not been used before to model data generated from experiments such as that described in Section 1.2. As a result, we were unable to obtain knowledge to inform prior distributions for analysing the data in this thesis. This led to the use of vague prior distributions.

1.7 Objectives

There are a number of specific attributes of the voltammogram that are of interest to the chemists which have particular meaning in Chemistry (these characteristics are defined in Table 1.4). The primary objective of the chemists was to find out if a systematic approach could be developed for analysing the data from their experiments that would estimate these characteristics and quantify the uncertainty of these estimates. Our objective was, by adopting the time series plot of Current (Figure 1.4) to develop an empirical statistical model and thus obtain posterior predictive distributions for not only the characteristics of interest defined in Table 1.4 but also any other attributes. Therefore, we can satisfy the primary objective of the chemists by developing an appropriate empirical model from which we can obtain posterior predictive distributions for the characteristics. From these posterior predictive distributions we can provide estimates for the characteristics of interest and probabilistic statements quantifying the uncertainty of these estimates. It should be noted that the methodology developed in this thesis can be adapted to obtain posterior predictive distributions for any other characteristics of interest which may be proposed in the future.

1.8 Choice of Model

The task of modelling the data becomes easier by thinking of the data as a time series. There is a vast literature on the choice and complexity of time series models that can be used in any practical application. Often, a very simple model is fitted at the start of an investigation. Additional features are introduced if the simple model is not found to be adequate for the data; and this process is repeated. Thus, a model that is found to fit the data poorly is replaced by an entirely different type of model. Alternatively, a modification to the model that is being applied could also be considered. When considering how many parameters to include in the model, there are various model choice criteria that can be used, see Section 3.7.

The initial time series model that we will use is an autoregressive process (AR process), see Section 6.2 for more details. The reason for considering an autoregressive process is based on the chemists' belief that the present value of the Current is dependent upon the recent previous values of the Current. The choice of the number of past values of the Current (or previous values of the Current) to be used in the autoregressive process will also be discussed in Section 6.2. The complexity of the model will be increased by incorporating additional variables into the model. We will also fit a model based on a Fourier series as this should model the sinusoidal behaviour of the data shown in Figure 1.4. The complexity of the model will be increased by attempting to model the variance. Our aim is to produce a parsimonious model which will be used to make predictive inferences.

1.9 Overview of the Thesis

In this chapter, we have discussed the motivation for this thesis and provided a synopsis of the experiment that gave rise to the data set. In Chapter 2, we carry out exploratory analyses of the data sets.

In Chapter 3 we review Bayesian methodology and techniques that will be used to make predictive inferences. Essential to Bayesian methodology is the posterior distribution which will be used in the analysis carried out in Chapters 5, 6, 7 and 8. In Chapter 4, we combine Bayesian methodology with the definitions of the characteristics in Chapter 2 to derive posterior predictive distributions for each of the characteristics of interest.

In Chapter 5, we analyse the summaries of the characteristics by fitting regression models. In Chapters 6, 7 and 8 we propose three different modelling strategies for the full current output curve. In each of these chapters we obtain the best model and discuss the posterior predictive distributions of the characteristics of interest under the best model.

In Chapter 9 we compare the various models developed in Chapters 5-8. In this comparison we will include a discussion of the advantages and disadvantages of the different techniques used to model the data. Finally, we conclude with some discussions for possible improvements and future work.

Chapter 2

Exploratory Data Analysis

2.1 Introduction

The purpose of this chapter is to explore the data set. For brevity, we will show the full exploratory analysis for the second replication of scan rate 2. A similar analysis was also conducted for the second replication of scan rate 1 and scan rate 3. The data from these replications have been chosen since there is no evidence of the instrumentation malfunctioning in these cases.

The size of the data set and the second replication of the scan rate is still large for modelling purposes. One of the aims of this chapter is to reduce the size of the data set further to lessen the computational burden. If we did not reduce the size of the data set we would have to analyse 63 different time series where each time series has length between 1250 and 5200. This would require a substantial amount of computation time to fit the models and perform inference for the characteristics of interest. Reduction in computation time would also make it more practical to fit more complex models.

In the exploratory analysis, any malfunctioning channels are removed from the data set so that the results and inferences are not biased.

2.2 Exploratory Analysis

Plots of the Current against time for each channel, for replicate 2 of scan rate 2 can be found in Figures 2.1-2.8. From these time plots, it is clear that as the percentage of Carbon increases, the profile of the Current output curve becomes much more pronounced. According to the chemists, the Current output curve is expected to look like the curve shown in the plot shown in Figure 1.4. A similar descriptive analysis for each of the scan rates 1 and 3 was completed and similar results were observed.

One of the anomalies from the data is the occasional spikes that occur. The spikes occur when the instrument used to measure the Potential becomes unstable. The home made instrumentation used in the experiment is complex and not very stable. Another reason for the spikes is that the thin Lithium branches on the electrode surface may cause a short in the circuit which could cause a sudden increase in the Current. Therefore, it has been advised by our collaborative chemists that the data must be smoothed in some manner so that these anomalies do not affect the analysis.

When Carbon is set at 0%, we expect to see a flat line with Current having a value of approximately zero, see Figure 2.1. This is the expected pattern as there is no Carbon present to form a path to enable the Current to pass through the electrode. At 1% Carbon, we have observed much more variation. The time series plots of channels 10, 35, 38 and 63 in Figure 2.2 show a fairly flat pattern with very small deviations about zero. The other four channels show some characteristics of the time series curve that we expect with higher proportions of Carbon. When the Carbon percentage is increased to 2% the variation is reduced with only channels 12, 16 and 41 showing departure from the expected pattern exhibited by the time series of the other channels shown in Figure 2.3. When Carbon is set at 3% there are no outliers but the spikes still occur occasionally. The plots shown in Figure 2.5

show that channel 55 is the only channel that does not follow the pattern exhibited by the other channels when the proportion of Carbon is 5%. Analysing the time series plots for 7% Carbon shown in Figure 2.6, it can be gleamed that there are two channels showing abnormal behaviour, namely channels 22 and 40. The plots with 10% Carbon show that channel 15 is the only channel exhibiting departure from the behaviour indicated by the rest of the plots in Figure 2.7. From the time series plots with 20% Carbon, shown in Figure 2.8, we can clearly see that channel 48 exhibits behaviour different from the other channels shown. In Section 2.3, we discuss how we dealt with malfunctioning channels, such as channel 48.

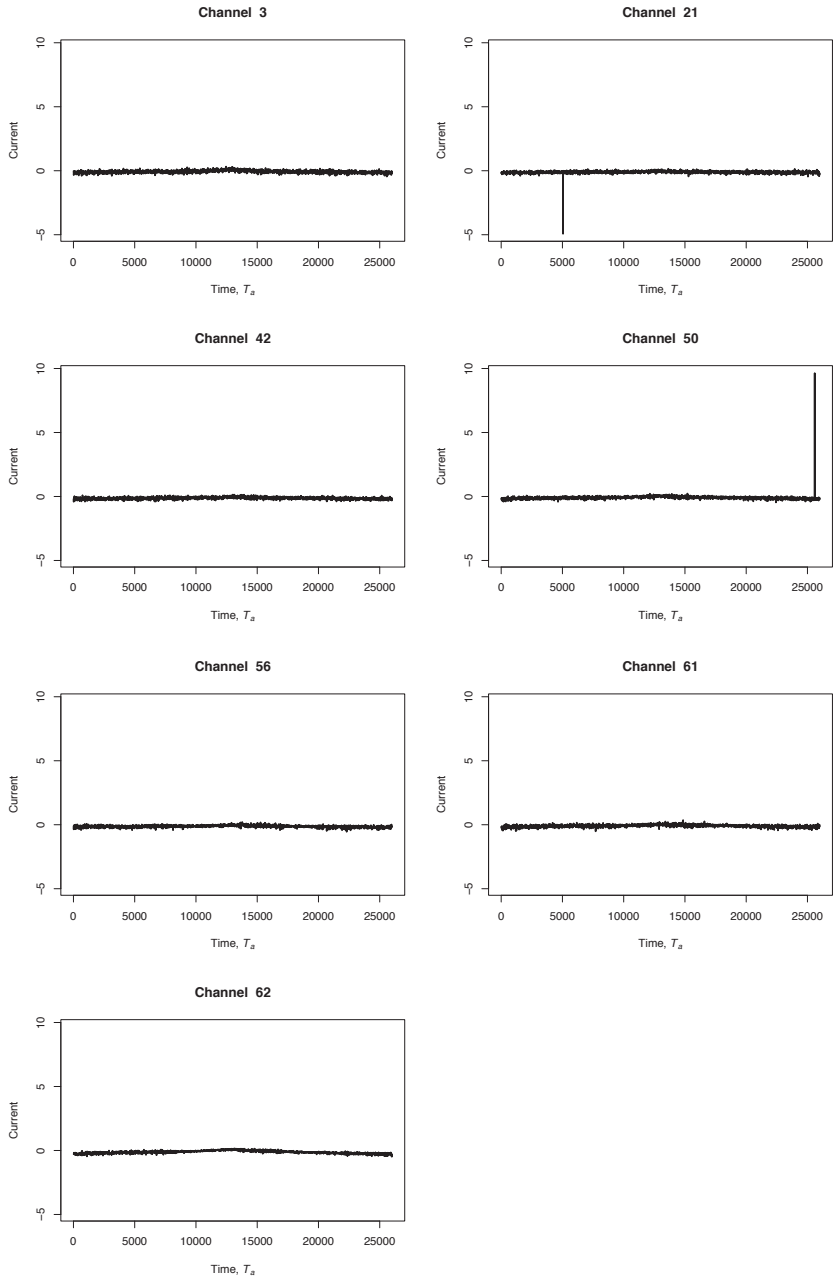


Figure 2.1: Time series plots of Current (original Current$\times 10^6$), with Carbon set at 0% and scan rate 2, where elapsed time T_a is measured in seconds.

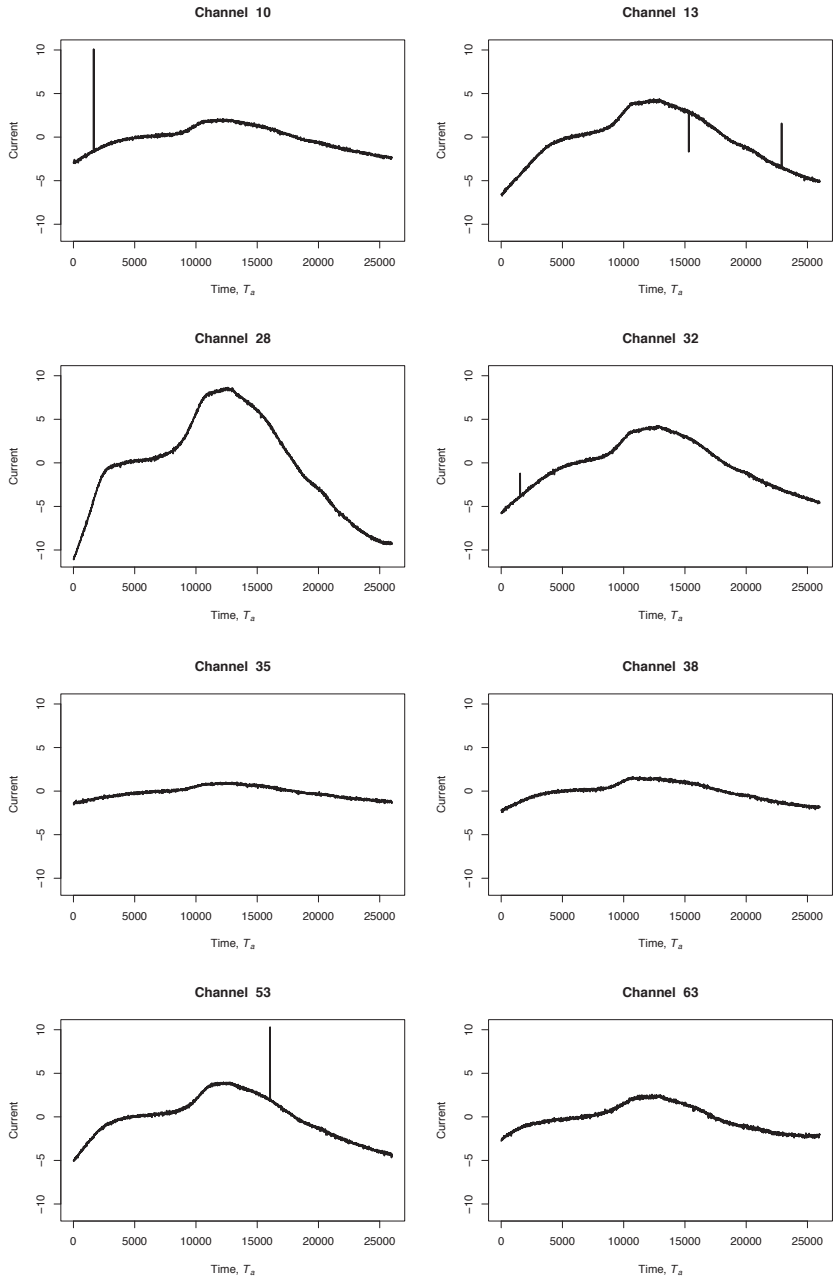


Figure 2.2: Time series plots of Current (original Current$\times 10^6$), with Carbon set at 1% and scan rate 2, where elapsed time T_a is measured in seconds.

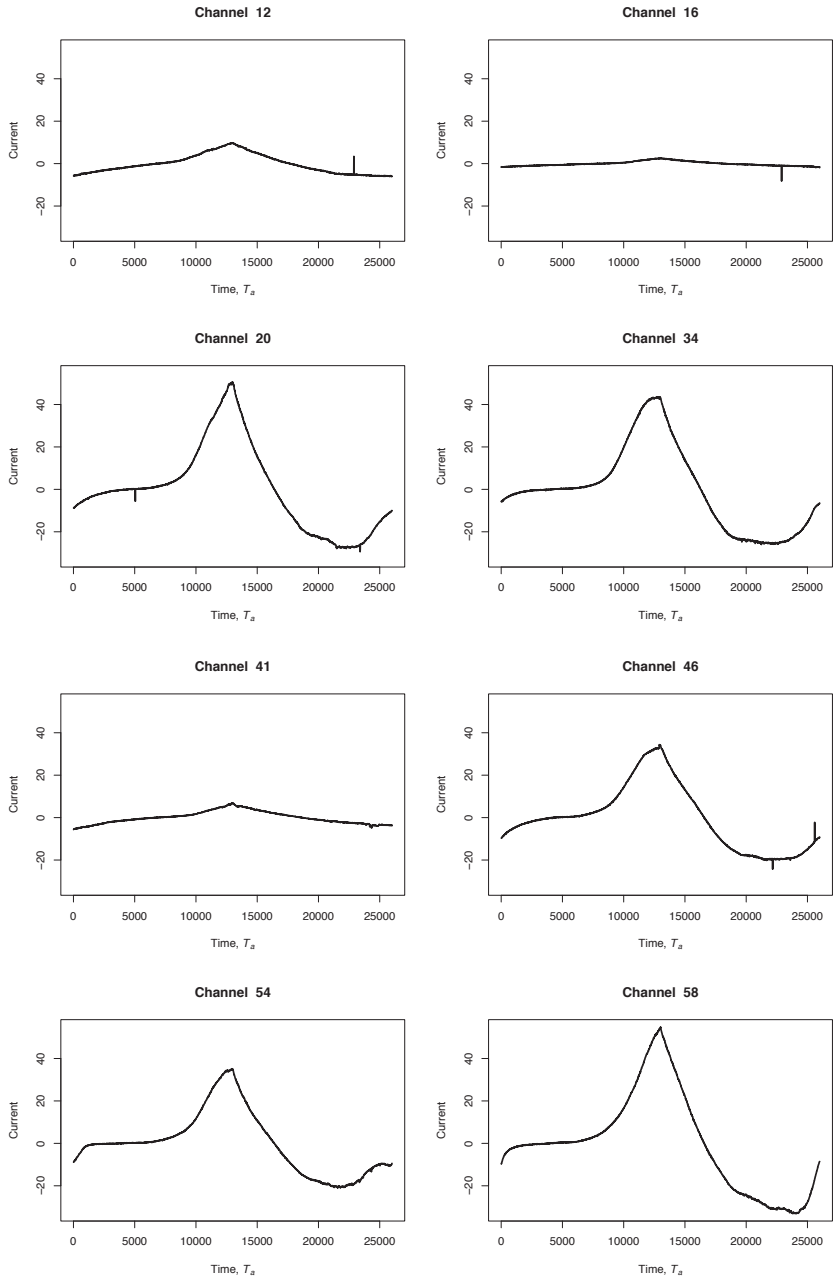


Figure 2.3: Time series plots of Current (original Current$\times 10^6$), with Carbon set at 2% and scan rate 2, where elapsed time T_a is measured in seconds.

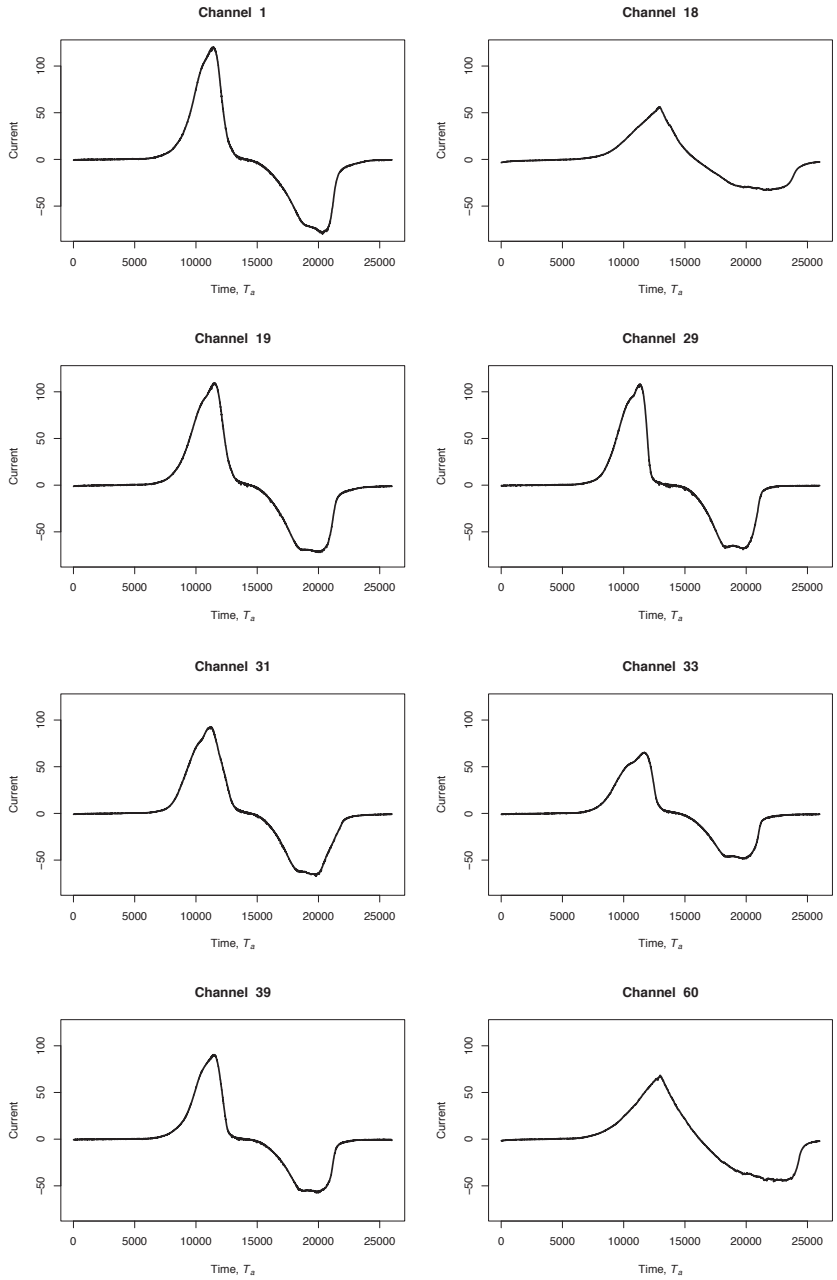


Figure 2.4: Time series plots of Current (original Current$\times 10^6$), with Carbon set at 3% and scan rate 2, where elapsed time T_a is measured in seconds.

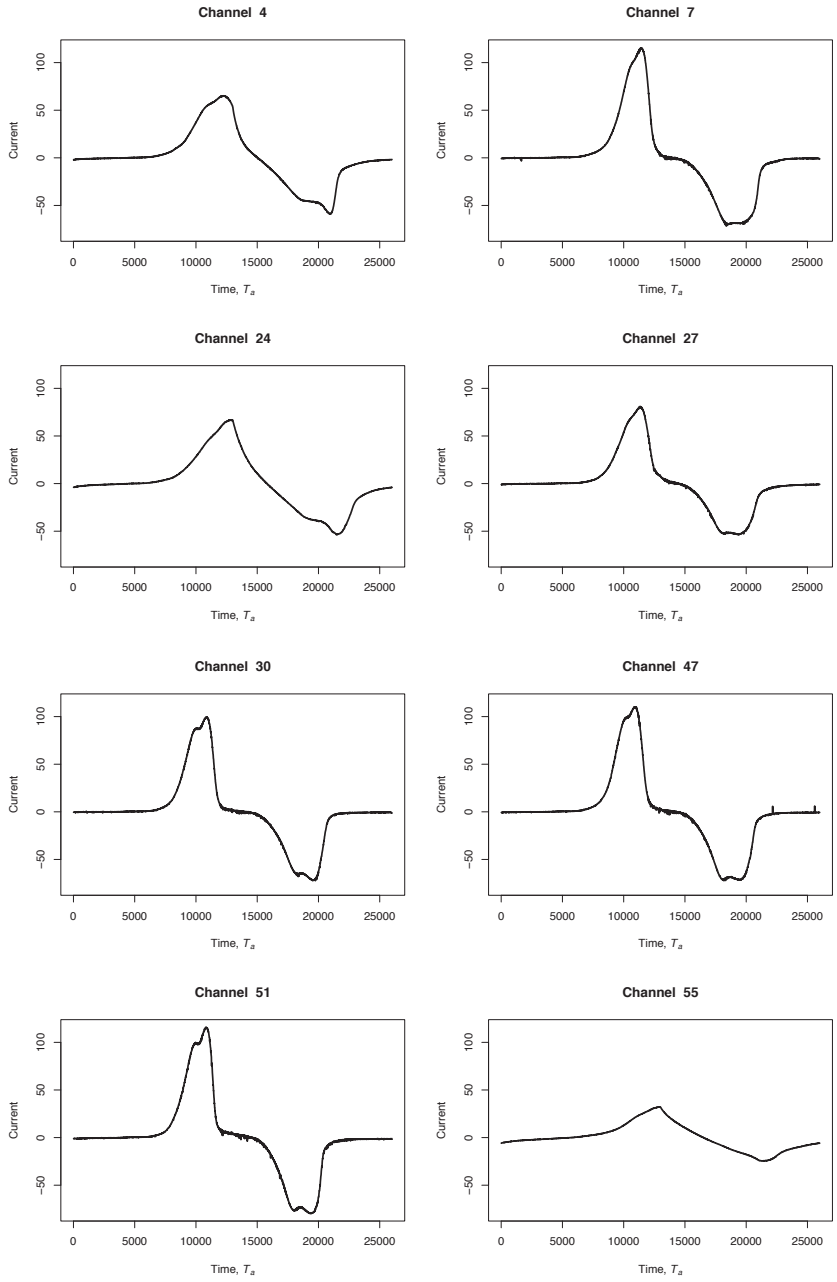


Figure 2.5: Time series plots of Current (original Current $\times 10^6$), with Carbon set at 5% and scan rate 2, where elapsed time T_a is measured in seconds.

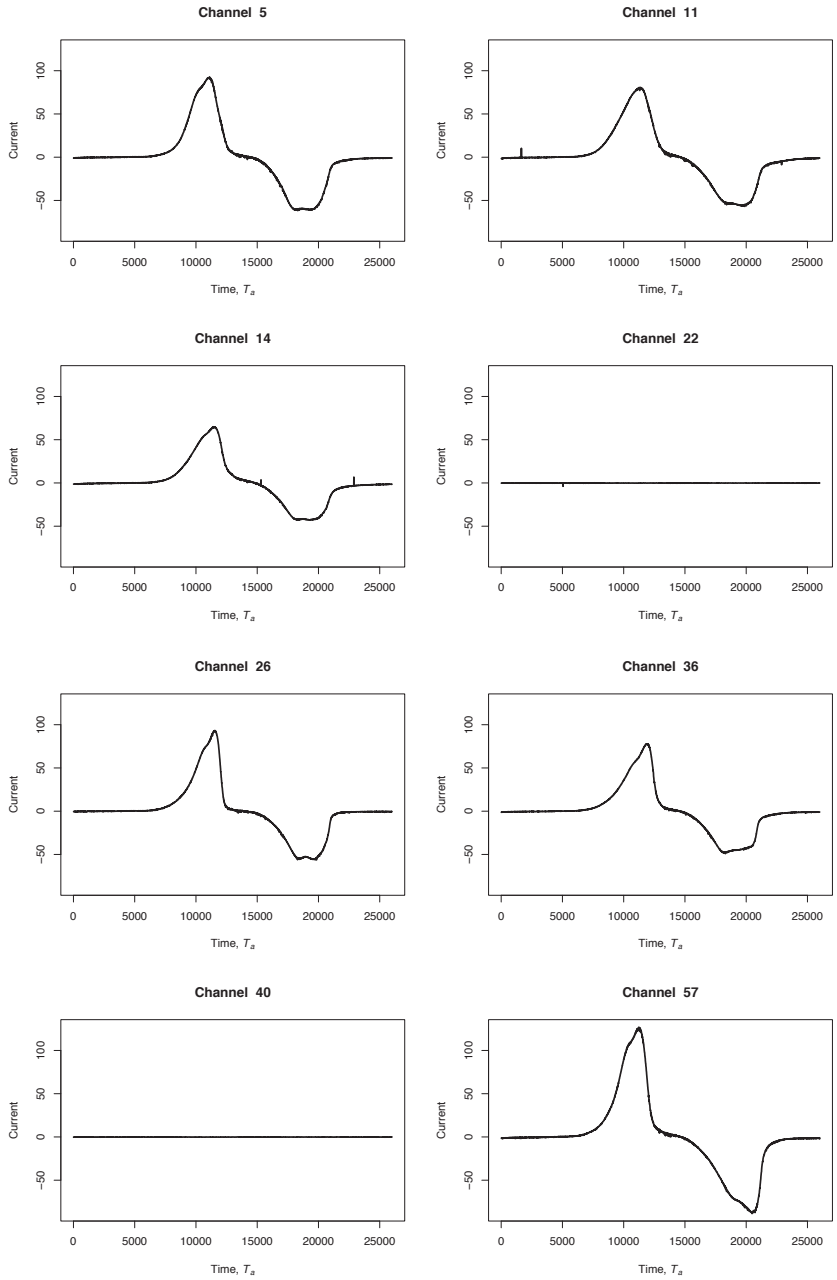


Figure 2.6: Time series plots of Current(original Current$\times 10^6$), with Carbon set at 7% and scan rate 2, where elapsed time T_a is measured in seconds.

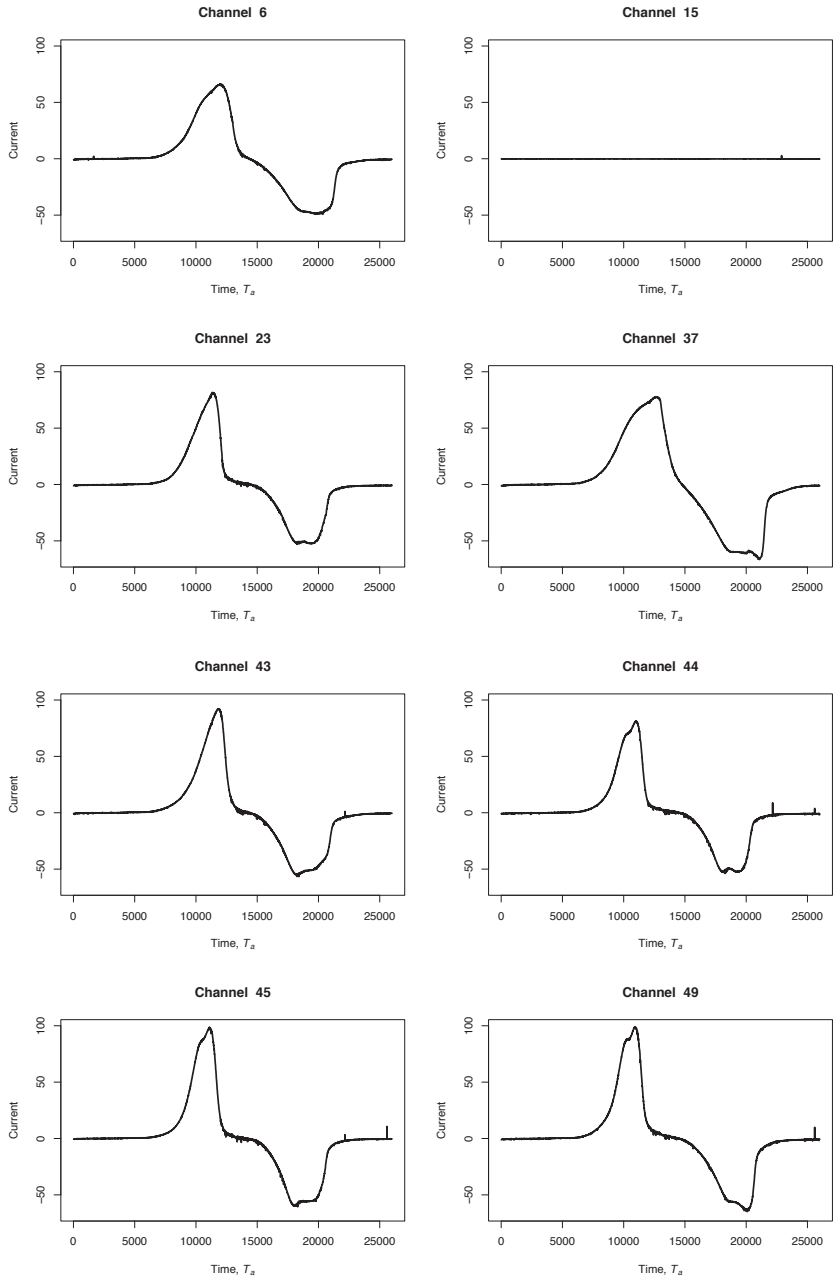


Figure 2.7: Time series plots of Current (original Current$\times 10^6$), with Carbon set at 10% and scan rate 2, where elapsed time T_a is measured in seconds.

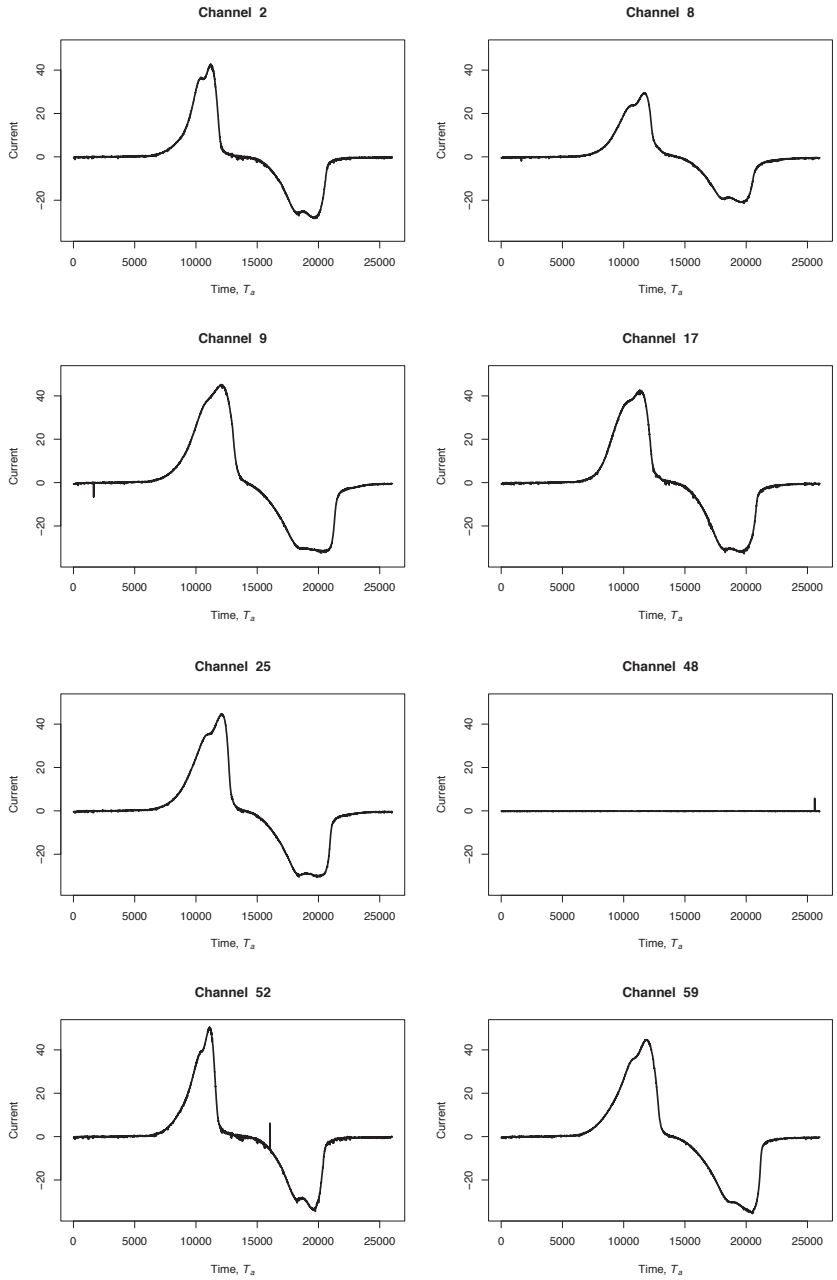


Figure 2.8: Time series plots of Current (original Current$\times 10^6$) with Carbon set at 20% and scan rate 2, where elapsed time T_a is measured in seconds.

2.3 Data Aggregation

As highlighted in Section 2.2, the Current output contains randomly occurring spikes. It is important to note that the spikes will not necessarily occur at a single time point. The effect of the spike could affect several consecutive time points, although this is unlikely to happen. The length of time that the spike will last or the probability of the spikes occurring cannot be calculated or quantified. As far as we are concerned, the occurrence of these spikes is unpredictable. To remove these fluctuations a smoothing process may be applied to the time series. It should be noted that a smoothing process is not the only possibility. The data can be transformed in a variety of ways, for example using the logarithmic scale or square root transformations, for example, see Chatfield (2003, Chapter 2).

While smoothing the data, it is vital that important characteristics of the data are retained in the transformed time series. We will calculate the mean of every five time points to form a new smoother time series, see Equation (2.1). This shortens the length of the time series to one fifth of the original time series. This will reduce the computation time required to analyse the data. In Figure 2.9 it can be seen that the effect of the spikes has been mostly removed. In addition, Figure 2.9 also shows that the new time series follows almost the same pattern of the original time series without the spikes. The transformed data shown in Figure 2.9 can be expressed as

$$x_{s,q,k}(t) = \frac{1}{5} \sum_{u=-2}^{2} y_{s,q,k}(t+u). \tag{2.1}$$

Another possibility would have been to take the median of every 5 time points. This is not guaranteed to remove the spikes as a spike could potentially influence more than 5 five time points. In Table 2.1, we show the set of possible values for the time index (t) and the time elapsed since start of run (t_a) for the aggregated data.

Scan Rate					
Rate 1		Rate 2		Rate 3	
t	t_a	t	t_a	t	t_a
1	50	1	50	1	50
2	100	2	100	2	100
3	150	3	150	3	150
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	260	13000
⋮	⋮	520	26000		
⋮	⋮				
1040	52000				

Table 2.1: Time elapsed, t_a, in seconds, since the start of the run with corresponding time index, t, for each of the three scan rates (Aggregated Data).

In Section 2.2, it was highlighted that some of the channels exhibited abnormal behaviour. The time series plots in Figure 2.1 display a flat line about a Current value of zero. Hence, there is little information to be gained by including these channels in the analysis. Similarly, the time series plots shown in Figure 2.2 will add very little information. In addition, the chemists' scientific knowledge leads us to believe that little will be gained by including the channels with 1% and 2% of Carbon in the analysis hence these channels will be removed from our analysis. It has been advised by the chemists that any malfunctioning channels should also be removed from the analysis. By including these malfunctioning channels in the analysis incorrect results and conclusions may be produced. It should be noted that it is not possible to define a malfunctioning channel according to a set of axioms such that they can be removed automatically. A malfunctioning channel needs to be identified by careful inspection. In addition, we consulted the chemists about

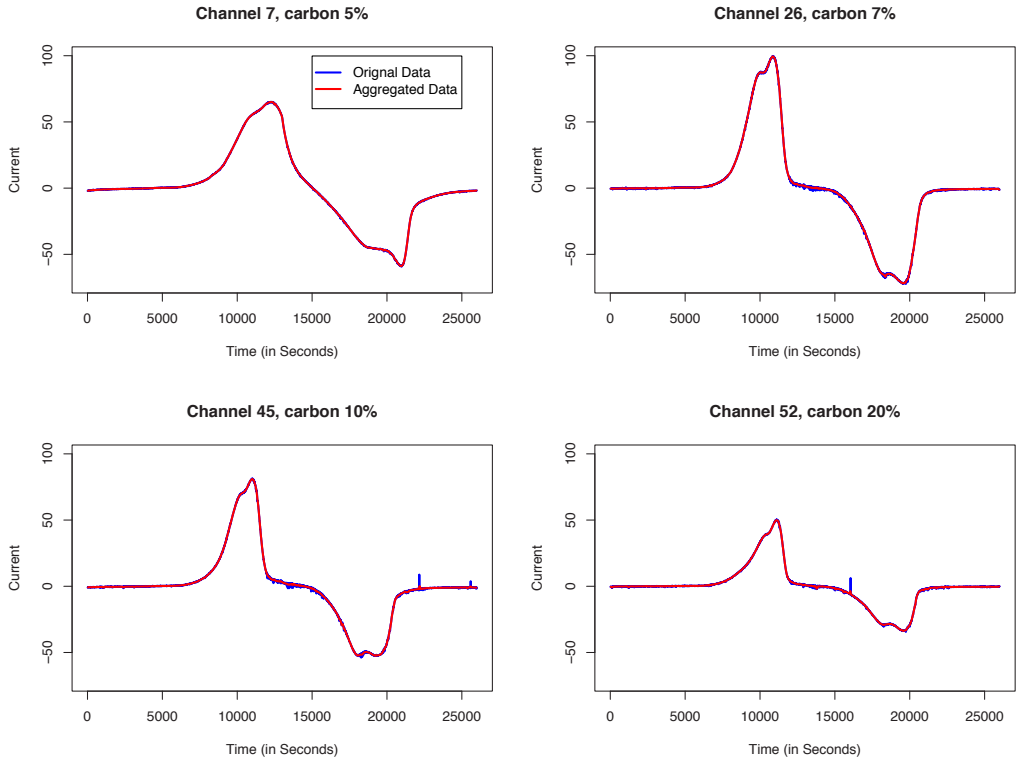


Figure 2.9: Comparison of raw and smoothed data for scan rate 2. The time elapsed for the original data is T_a and for the aggregated data is t_a. As T_a and t_a are both measured in seconds, this allows time to be measured in seconds for both sets of data on the x-axis.

which channels were malfunctioning according to the profile of the Current output curve to ensure that no functioning channels were removed accidentally.

A further transformation of the data that will be used takes the mean over the channels to be included in the analysis. This aggregation is given by

$$x_{s,k}(t) = \frac{1}{r_{s,k}} \sum_{\forall q \in Q_{s,k}} \sum_{u=-2}^{2} x_{sqk}(t+u) \tag{2.2}$$

where $x_{s,q,k}(t)$ is defined in Equation (2.1), $Q_{s,k}$ is the set of channels used for analysis with Carbon level k and scan rate s, $r_{s,k}$ is the number of channels in set $Q_{s,k}$, that is the malfunctioning channels have been excluded. We note that $s = 1$, 2, and 3 correspond to scan rates 0.05mVs^{-1}, 0.10mVs^{-1} and 0.20mVs^{-1} respectively. The sets $Q_{s,k}$ are shown below; $Q_{s,1}$, $Q_{s,2}$ and $Q_{s,3}$ will be empty sets since we are not including channels with 0%, 1% and 2% Carbon in our analysis. For scan rate 1, the sets $Q_{1,k}$ are given by:

$$\begin{aligned}
r_{1,4} &= 8,\ Q_{1,4} = \{1, 18, 19, 29, 31, 33, 39, 60\} \\
r_{1,5} &= 8,\ Q_{1,5} = \{4, 7, 24, 27, 30, 47, 51, 55\} \\
r_{1,6} &= 6,\ Q_{1,6} = \{5, 11, 14, 26, 36, 57\} \\
r_{1,7} &= 7,\ Q_{1,7} = \{6, 23, 37, 43, 44, 45, 49\} \\
r_{1,8} &= 7,\ Q_{1,8} = \{2, 8, 9, 17, 25, 52, 59\}.
\end{aligned}$$

For scan rate 2, the sets $Q_{2,k}$ are given by:

$$\begin{aligned}
r_{2,4} &= 8,\ Q_{2,4} = \{1, 18, 19, 29, 31, 33, 39, 60\} \\
r_{2,5} &= 7,\ Q_{2,5} = \{4, 7, 24, 27, 30, 47, 51\} \\
r_{2,6} &= 6,\ Q_{2,6} = \{5, 11, 14, 26, 36, 57\} \\
r_{2,7} &= 7,\ Q_{2,7} = \{6, 23, 37, 43, 44, 45, 49\} \\
r_{2,8} &= 7,\ Q_{2,8} = \{2, 8, 9, 17, 25, 52, 59\}.
\end{aligned}$$

For scan rate 3, the sets $Q_{3,k}$ are given by:

$$\begin{aligned}
r_{3,4} &= 8,\ Q_{3,4} = \{1, 18, 19, 29, 31, 33, 39, 55, 60\} \\
r_{3,5} &= 7,\ Q_{3,5} = \{4, 7, 24, 27, 30, 47, 51\} \\
r_{3,6} &= 6,\ Q_{3,6} = \{5, 11, 14, 26, 36, 57\} \\
r_{3,7} &= 6,\ Q_{3,7} = \{23, 37, 43, 44, 45, 49\} \\
r_{3,8} &= 7,\ Q_{3,8} = \{2, 8, 9, 17, 25, 52, 59\}.
\end{aligned}$$

In this thesis we never model the raw Current output denoted so far by $y_{s,q,k}(t)$. Instead we shall only consider the aggregated data denoted by the notation x. Hence, we will use the notation y to denote other quantities.

The aggregated data are shown in Figure 2.10. These are the data that we use for fitting various models making predictive inferences.

The plots in Figures 2.10(a), 2.10(b), 2.10(c) clearly indicate that as the proportion of Carbon increases the Peak Current continues to increase until the level of Carbon reaches 10%. The Peak Current is lower at 20% of Carbon compared to that at 10%. This indicates that the maximum Peak Current is achieved when the percentage of Carbon is between 10% and 20%. The Minimum Current continues to decrease up to 10% of Carbon and then increases at 20% of Carbon. When we compare the plots of the aggregated data simultaneously it can be observed that the Peak Current increases and the Minimum Current decreases as the scan rate increases. This comparison also reveals that with slower scan rates there is a double peak that becomes more pronounced. In addition, with the faster scan rates, the peaks and troughs are much more smooth. Thus the aggregation of the data has made the time series smoother and easier to manage in comparison to the original data. The corresponding voltammograms are shown in Figure 2.11. We will refer back to these voltammograms in the analysis and conclusion sections of later chapters.

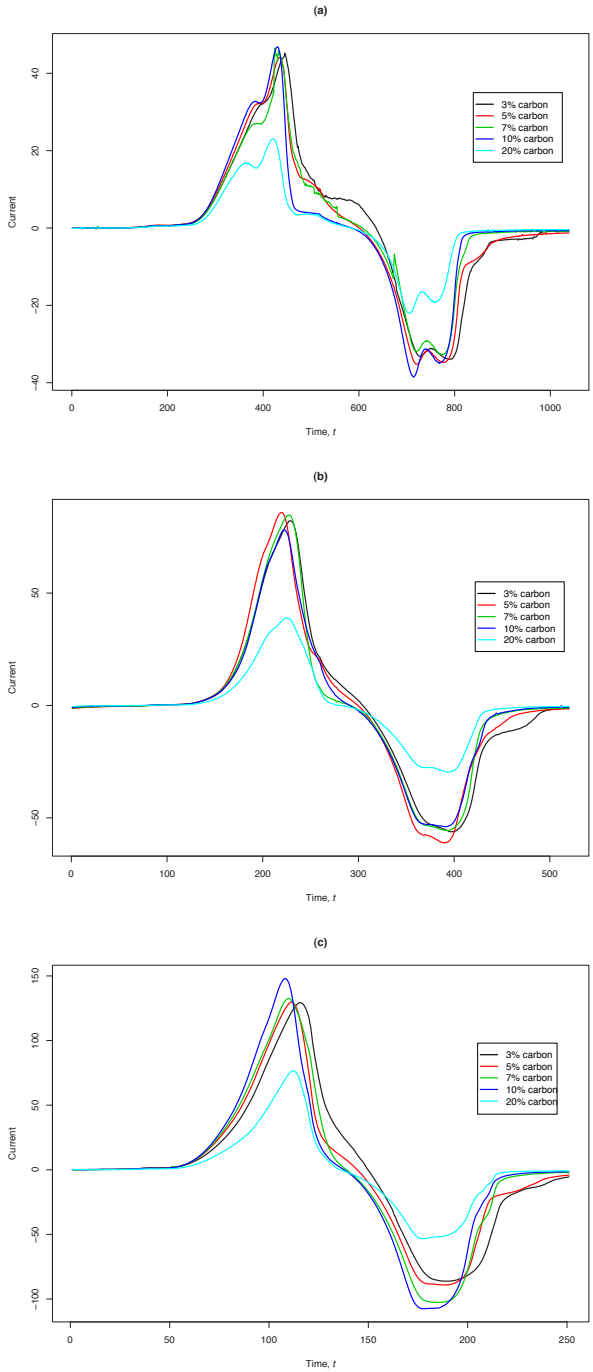


Figure 2.10: Time series plots of aggregated data for various percentages of Carbon for: (a) scan rate 1, (b) scan rate 2 and (c) scan rate 3, where t denotes the time index defined in Table 2.1.

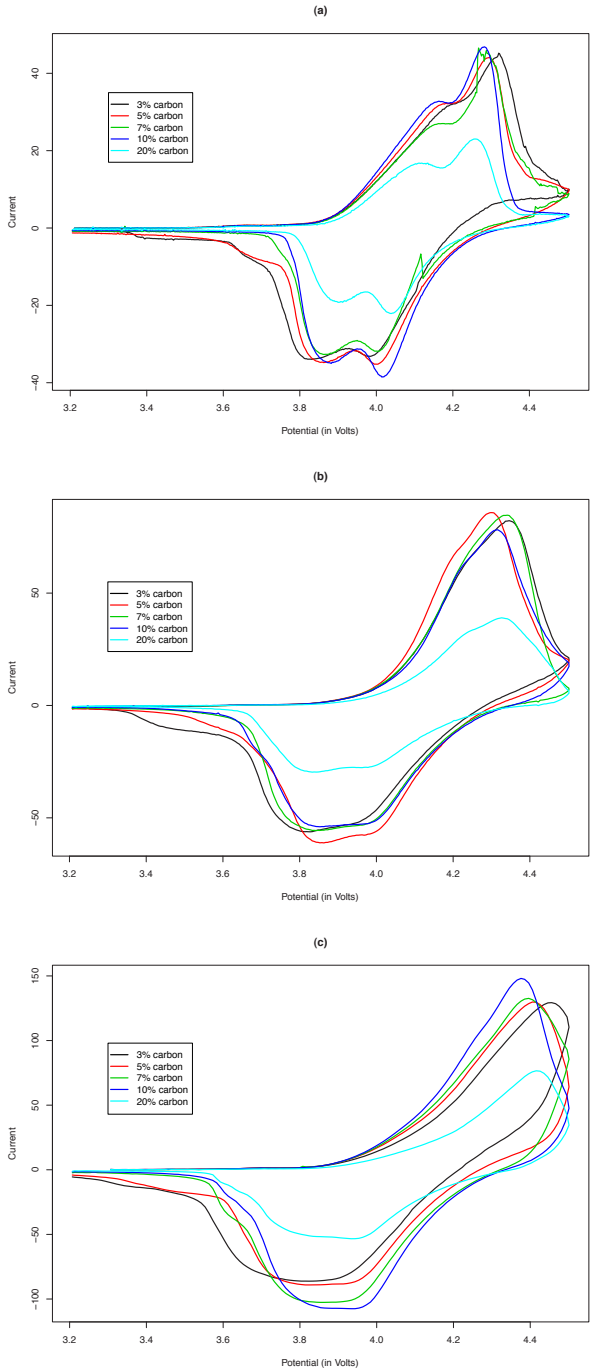


Figure 2.11: Aggregated voltammograms for various percentages of Carbon for: (a) scan rate 1, (b) scan rate 2 and (c) scan rate 3.

2.4 Definitions of Characteristics of Voltammograms

In this section we define formally the characteristics that are associated with the voltammogram which were introduced in Table 1.4 and Figures 1.5 and 1.6. The Peak Current, denoted by $I_{s,k}^{(\max)}$, is defined to be the maximum observed value of the Current, that is:

$$I_{s,k}^{(\max)} = \max\{x_{s,k}(t);\ t = 1, \ldots, N_s\}$$

where N_s is the number of observations at scan rate s, and $x_{s,k}(t)$ is the observed Current at time t for scan rate s and Carbon level k (see Equation (2.2)). We note that $s = 1, 2$ and 3 corresponds to the values of the scan rate 0.05mVs^{-1}, 0.10mVs^{-1} and 0.20mVs^{-1} respectively.

Let $I_{s,k}^{(\text{wid})}$ denote the quantity $I_{s,k}^{(\max)}/2$. We are now in a position to define the remaining characteristics of interest. Let $t_{1,s,k}$ be the first time that the Current reaches $I_{s,k}^{(\text{wid})}$ and $t_{2,s,k}$ be the last time the Current reaches $I_{s,k}^{(\text{wid})}$ for the kth Carbon level and scan rate s. In a similar way, let $P_{1,s,k}$ be the first value of Potential at which the Current reaches $I_{s,k}^{(\text{wid})}$ and $P_{2,s,k}$ to be the last value of Potential at which the Current is equal to $I_{s,k}^{(\text{wid})}$.

The reason for the chemists' interest in these characteristics is that the characteristics defined in Table 2.2 provide a way of summarising the profile curve. The values associated with the Peak Current and Minimum Current provide information that can be used to understand how the Peak Current or Minimum Current changes according to different scan rates. In addition, the Peak Time and Peak Potential identify the point in the experiment when the Peak Current occurs, and the associated Potential value. The reasoning and interest in the Minimum Current is the same as that for the Peak Current.

Characteristic	Symbol	Definition
Minimum Potential	$P_{s,k}^{(\min)}$	The Potential value that produces $I_{s,k}^{(\min)}$.
Minimum Time	$t_{s,k}^{(\min)}$	The time point where $I_{s,k}^{(\min)}$ occurs.
Minimum Current	$I_{s,k}^{(\min)}$	$I_{s,k}^{(\min)} = \min\{x_{s,k}(t);\ t = 1, \ldots, N_s\}$
Peak Potential	$P_{s,k}^{(\max)}$	The Potential value that produces $I_{s,k}^{(\max)}$.
Peak Time	$t_{s,k}^{(\max)}$	The time point where $I_{s,k}^{(\max)}$ occurs.
Peak Current	$I_{s,k}^{(\max)}$	$I_{s,k}^{(\max)} = \max\{x_{s,k}(t);\ t = 1, \ldots, N_s\}$
Peak Separation in Potential	$P_{s,k}^{(\text{sep})}$	$P_{s,k}^{(\max)} - P_{s,k}^{(\min)}$
Peak Separation in Time	$t_{s,k}^{(\text{sep})}$	$t_{s,k}^{(\max)} - t_{s,k}^{(\min)}$
Peak Separation in Current	$I_{s,k}^{(\text{sep})}$	$I_{s,k}^{(\max)} - I_{s,k}^{(\min)}$
Peak Width in Potential	$P_{s,k}^{(\text{wid})}$	$P_{2,s,k} - P_{1,s,k}$
Peak Width in Time	$t_{s,k}^{(\text{wid})}$	$t_{2,s,k} - t_{1,s,k}$

Table 2.2: Definitions for peak characteristics where subscripts s and k are the scan rate and the level (proportion) of Carbon respectively. The superscript is min, max, sep or wid according to whether the characteristic is a minimum, a maximum, a separation or a width respectively. These characteristics are shown graphically in Figures 1.5 and 1.6.

2.5 Analysis of Characteristics

We now analyse the characteristics of interest. The data in the plots shown in Figure 2.12 (at the end of this section) was obtained from Figure 2.10 and the data in the plots shown in Figure 2.13 obtained from the raw data.

Figure 2.12(a) appears to show no particular clear pattern for the value of the Minimum Potential, with respect to Carbon for any of the scan rates. To achieve the Peak Current, Figure 2.12(d) indicates that higher values of Potential are required with faster scan rates to obtain the highest possible Current output.

Figure 2.12(b) indicates that the Minimum Current occurs at approximately the same time for all proportions of Carbon for scan rates 2 and 3. For scan rate 1, this relationship appears to become slightly more unstable, but the Minimum Current still appears to occur at roughly the same time. The same relationship appears to hold for the Peak Time, see Figure 2.12(e).

Figure 2.12(c) shows that for scan rates 1 and 2 the Minimum Current appears to be roughly the same for Carbon proportions less than or equal to 10%. When the level of Carbon is increased to 20% the Minimum Current appears to increase dramatically. For scan rate 3, the Minimum Current seems to follow the characteristic of a parabola. From Figure 2.12(f), the Peak Current appears to be roughly the same when the Carbon level is less than or equal to 10% for scan rates 1 and 2. When the level of Carbon is increased to 20% there is a substantial decrease in the value of the Peak Current.

The Peak Separation in Current, shown in Figure 2.12(i), increases with the scan rate. In addition, the Peak Separation in Current appears to rise and then fall as the proportion of Carbon increases. The Peak Separation in Time, shown in Figure 2.12(h), does not seem to be affected by the proportion of Carbon and is lower for lower scan rates. Figure 2.12(g) does not appear to display any clear behaviour.

The Peak Width in Potential, shown in Figure 2.12(j), appears to decrease and then

increase for scan rates 1 and 2 as the proportion of Carbon increases. This pattern seems to follow the shape of a parabola. We note that the Peak Width in Potential for scan rates 1 and 2, with Carbon set at 20%, is approximately the same. For scan rate 3, the Peak Width in Potential appears to increase and then decrease as the proportion of Carbon increases. The plot in Figure 2.12(k) indicates that the Peak Width in terms of time appears to remain roughly the same for levels of Carbon within the same scan rate. It should also be noted that for slower scan rates the Peak Width in Time is longer.

The characteristics of interest defined in Table 1.4 were calculated for each of the three scan rates. Table 2.3 provides a rough idea of the values of the characteristics of interest for scan rate 2. It should be noted that the number of significant figures will vary depending on the nature of the characteristic of interest that we are analysing. Consider the Minimum Time and the Minimum Current. The Minimum Time is the observation number at which the Minimum Current occurs, hence the Minimum Time will be an integer. However, the Minimum Current is a continuous measure and can therefore be recorded to the desired degree of accuracy. Therefore, we have decided to record the values for each characteristic to an appropriate degree of accuracy depending on its nature.

A plot of the data used for the regression models in Chapter 5 for scan rate 2 (using the raw data) is shown Figure 2.13. The values for each of the characteristics were calculated using the definitions of the characteristics of interest (shown in Tables 1.4 and 2.2) and the raw data from the experiment, which consists of all 9 runs (see Table 1.1), and all the replicates for Carbon. If we were simply presented with these summaries, the plots shown in Figure 2.13 would be the only plots we would be able to obtain. The variability in the plots in Figure 2.13 arises because we have not removed the malfunctioning channels. If we look at Figures 2.4 - 2.8, we can see that there were quite a few channels that malfunctioned in the experiment.

In our discussions with the chemists, they were unwavering in their interest in the characteristics defined in Table 1.4. Further to this, they pointed out that these characteristics are important and sensible measures to be considering even in light of the variability of the data shown in Figure 2.13.

The values of the Minimum Potential (Figure 2.13(a)) indicate that there are outliers present in the data. This pattern also occurs in the Peak Time values. In addition, Figure 2.13(d) appears to show that the Peak Potential is approximately the same for all proportions of Carbon. There is a similar pattern exhibited by the Minimum Time (shown in Figure 2.13(b)). From Figure 2.13(e), it can be observed that while there are a few outliers the Peak Time appears to be approximately the same for all proportions of Carbon.

The values of the Minimum Current shown in Figure 2.13(c) appears to indicate no outliers. The plot also indicates that as the Carbon increases the Minimum Current value initially dips slightly and then increases. Figure 2.13(f) indicates that variance is small for all the different proportions of Carbon and also suggests that the Peak Current value significantly changes for different proportions of Carbon.

The Peak Separation values, shown in Figure 2.13(i), have small variation for each proportion of Carbon. Also, the Peak Separation in Current increases slightly then decreases as the proportion of Carbon increases. Peak Separation in Time (Figure 2.13(h)) shows that there are some outliers which originated from the malfunctioning channels. However, the Peak Separation in Time appears to be approximately the same for all proportions of Carbon. The Peak Separation in Potential appears to decrease in linear fashion as the proportion of Carbon decreases.

Figures 2.13(j) and 2.13(k) show that there are a few outliers for the Peak Width in Potential and Peak Width in Time respectively. We discovered that the data points that appear to be outliers in the plots in Figures 2.13(j) and 2.13(k) are from channels that malfunctioned.

The characteristics of interest were also calculated for scan rates 1 and 3. Further, to this, we also note that a critical analysis for scan rates 1 and 3 was conducted.

When generating the posterior predictive distributions, we have chosen to only look at the predictive distributions for the Potential instead of time as it is the values of the Potential that the chemists are interested in. Further, we note that we will only concentrate on the Peak Potential, Peak Current and Peak Separation in Potential. The first reason for this is that these were the initial characteristics of interest to the chemists. The second reason is that these are the main characteristics that are used to describe or analyse data that has arisen from the experiment described in Section 1.2.

In Chapter 5, we perform a regression models based analysis of these data without removing the values from the malfunctioning channels. This allows us to view the results from the crude analysis in the presence of outliers. The model based methods presented in Chapters 6-8 have these outliers removed.

Characteristic	Percentage of Carbon				
	3	5	7	10	20
Minimum Potential	3.8280	3.8630	3.8430	3.8528	3.8328
Minimum Time	396	389	393	391	395
Minimum Current	-56.2125	-61.1457	-55.6400	-53.9771	-29.6629
Peak Potential	4.3462	4.3010	4.3412	4.3162	4.3262
Peak Time	229	220	228	223	225
Peak Current	82.2200	85.8114	84.6600	78.1857	39.0000
Peak Separation in Potential	0.5182	0.4380	0.4982	0.4634	0.4934
Peak Separation in Time	167	169	165	168	170
Peak Separation in Current	138.4325	146.9571	140.3000	132.1629	68.66286
Peak Width in Potential	0.2750	0.2596	0.2550	0.2600	0.2896
Peak Width in Time	55	52	51	52	58

Table 2.3: Peak characteristics for scan rate 2 using the aggregated data shown in Figure 2.10.

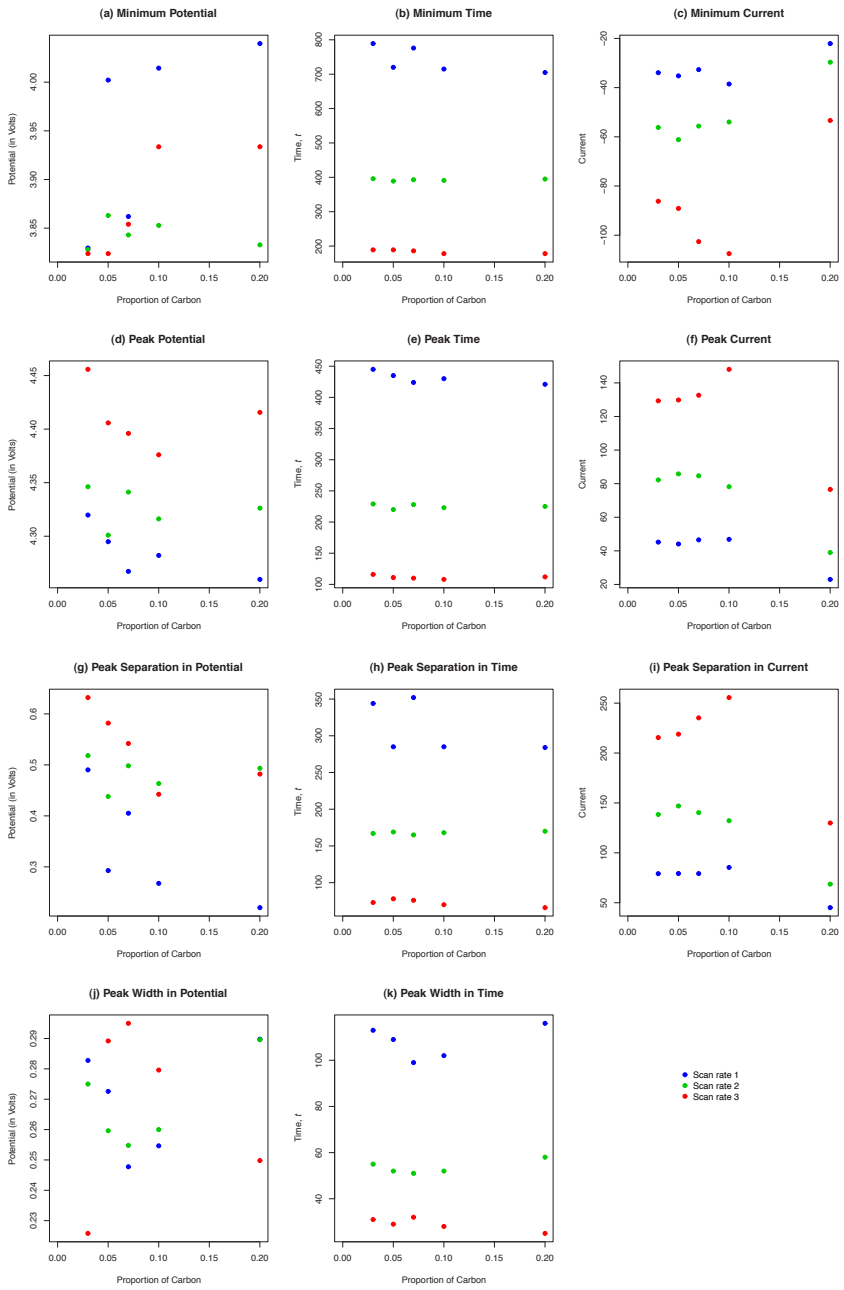


Figure 2.12: Characteristics of interest for each scan rate and Carbon level using the aggregated data shown in Figure 2.10, that is outliers have been removed, where t is as defined in Table 2.1.

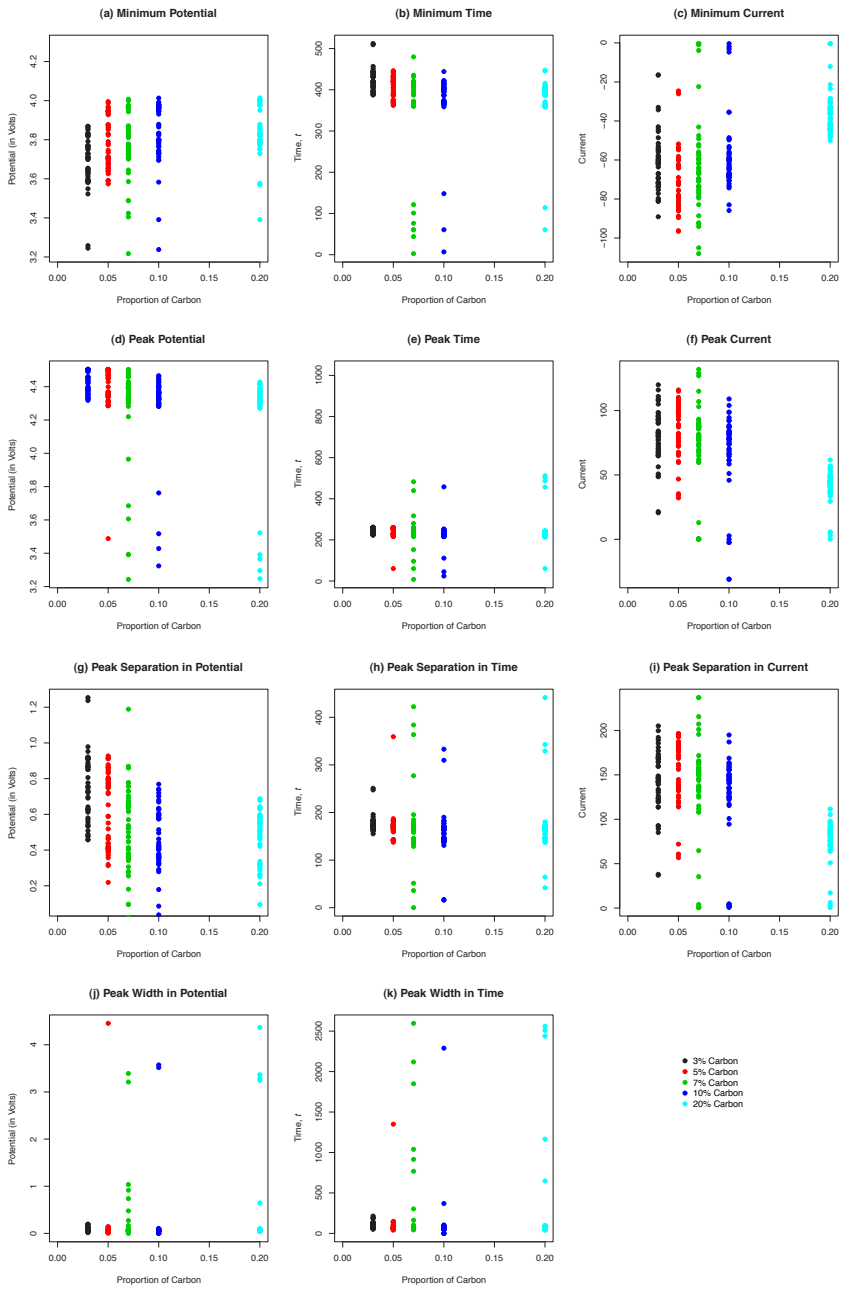


Figure 2.13: Characteristics of interest for scan rate 2 at each Carbon level using the raw data, that is outliers have not been removed, where t is as defined in Table 2.1.

Chapter 3

Bayesian Methods

3.1 Introduction

The Bayesian paradigm utilises information from the prior distribution and data to determine the posterior distribution. The prior distribution represents the belief in the parameters before the experiment and the posterior distribution represents the degree of belief in the parameters after the experiment. Even though the idea is not new, it is the development of computational power that has made Bayesian data analysis practical. In particular, if the posterior distribution is analytically intractable or difficult to integrate, the computational power available has made it possible to apply Markov Chain Monte Carlo (MCMC) sampling methods to estimate the posterior distribution of the parameters. We provide a brief synopsis about MCMC methods in Section 3.4. In addition, Bayesian methods can also be applied to assess the model's performance, of which a few are discussed in Section 3.7.

In using the Bayesian approach, we are able to obtain the distribution of the parameters of the model given the data (called the posterior distribution), from which it is possible to quantify the uncertainty about parameters in the model. It is also

possible to obtain the posterior predictive distribution (see Section 3.6) for other characteristics of interest that arise from the data. For examples from the applied problem considered in this thesis, see Table 1.4. We estimate the characteristic of interest by sampling from the posterior predictive distribution. From this sample we can approximate various attributes such as the mean and the variance. It is also possible to determine the probability that the characteristic of interest belongs to any interval (a, b) where a, $b \in \mathbb{R}$.

3.2 Bayesian Paradigm

Let $\boldsymbol{\theta}$ denote the parameters in the model and $\boldsymbol{x}$ denote the data. The Bayes theorem can be expressed as

$$f(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int_\Theta f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})\ \mathrm{d}\boldsymbol{\theta}} \tag{3.1}$$

where $f(\boldsymbol{\theta}|\boldsymbol{x})$ is the probability density function of the posterior distribution, $f(\boldsymbol{\theta})$ is the probability density function of the prior distribution, $f(\boldsymbol{x}|\boldsymbol{\theta})$ is the likelihood and Θ is the set of possible values for $\boldsymbol{\theta}$ (called the parameter space). As noted in Section 3.1, the prior distribution represents belief about the parameters in the model before observing any data and the posterior distribution represents our belief about the parameters after observing the data. The integral

$$f(\boldsymbol{x}) = \int_\Theta f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})\ \mathrm{d}\boldsymbol{\theta} \tag{3.2}$$

is called the marginal likelihood of the data and is only dependent upon $\boldsymbol{x}$ since $\boldsymbol{\theta}$ is integrated out. This integral is simply a constant of proportionality to ensure $\int_\Theta f(\boldsymbol{\theta}|\boldsymbol{x})\ \mathrm{d}\theta = 1$. Hence Bayes theorem can now be expressed as

$$f(\boldsymbol{\theta}|\boldsymbol{x}) \propto f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta}). \tag{3.3}$$

3.2.1 The Prior Distribution

The prior distribution is formed using expert opinion and/or past data. However, in order to assign a prior distribution to the parameters of the model, a subjective definition of probability is required. This is because the postulation of a prior distribution involves personal opinion. O'Hagan and Forster (2004) define subjective probability as a measure of one person's degree of belief. It is this aspect that has attracted considerable criticism from the opponents of Bayesian statistics. The critics argue that two experts with identical prior information may formulate completely different prior distributions which may result in conflicting posterior distributions. The Bayesian counter argument would be that provided the data are strong and the prior distribution is formulated on reasonable grounds, then any inferences made using the posterior distribution should be robust to slight differences in the prior distribution. It is important to realise that prior information is often available and can be extremely useful.

Prior distributions are sometimes chosen for convenience. Suppose we have observed data $\boldsymbol{x}$ with likelihood $f(\boldsymbol{x}|\boldsymbol{\theta})$. Suppose also that the prior distribution for $\boldsymbol{\theta}$ comes from a family of distributions $\mathcal{F}$. If the prior distribution for $\boldsymbol{\theta}$ is conjugate with respect to the likelihood then the posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{x})$ also belongs to the same family $\mathcal{F}$, because of this, it will often be relatively easy to draw inferences from the posterior distribution, especially if the family of distributions $\mathcal{F}$ is well known and understood. This might not have been the case if $f(\boldsymbol{\theta})$ came from a prior distribution that is not in $\mathcal{F}$. See O'Hagan and Forster (2004, Chapter 6) and Robert (2001, Chapter 3) for a more detailed discussion on prior distributions.

3.2.2 The Posterior Distribution

There are various quantitative summaries of the posterior distribution such as a measure of the location or dispersion that can provide answers to questions of interest. Point estimates of the quantity of interest, such as the mean, are often given. In addition, probability intervals for the quantity of interest are also specified. For example, the mean and the 95% probability intervals are often provided.

Suppose we wish to estimate $g(\boldsymbol{\theta})$ which represents the quantity of interest such as the mean. Then the posterior mean for $g(\boldsymbol{\theta})$ is given by

$$E[g(\boldsymbol{\theta})|\boldsymbol{x}] = \frac{\int_\Theta g(\boldsymbol{\theta}) f(\boldsymbol{x}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}}{\int_\Theta f(\boldsymbol{x}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}}. \tag{3.4}$$

We therefore calculate $g(\boldsymbol{\theta})$ using the conditional expectation $E[g(\boldsymbol{\theta})|\boldsymbol{x}]$ shown in Equation (3.4). Just as the posterior mean is a common measure of location, probability intervals are a popular method for measuring dispersion. Let $P(a < \theta_i < b) = (1 - \alpha)$ where a, $b \in \mathbb{R}$, α is a specified significance level and θ_i is some component of $\boldsymbol{\theta}$. Then the probability interval for θ_i is constructed by calculating the real values a and b such that $P(\theta_i < a) = \alpha/2$ and $P(\theta_i > b) = \alpha/2$. One possible hypothesis test would be to calculate the probability that $\theta_i \in (a, b)$ given the observed data $\boldsymbol{x}$, where a, $b \in \mathbb{R}$. This is an easy calculation provided that the posterior density for $\boldsymbol{\theta}$ is known.

3.2.3 Sensitivity Analysis

An essential factor when choosing a vague or diffuse prior distribution is the sensitivity of the model parameters with respect to the prior distribution. Any inferences made via the posterior distribution should generally not be sensitive to any misspecification of the prior distribution. This is because data, not the prior, should drive the inference. A simple method to check the sensitivity of the model parameters with respect to the prior distribution is to vary the values of the parameters in

the prior distribution and compare the inferences from the corresponding posterior distributions. Clearly, if the inferences are very similar, then the model parameters are robust to choice of the prior. However, if the inferences are not similar, then the model parameters are not robust to the choice of the prior and careful thought is required in constructing the prior distribution.

Suppose sensitivity to the prior distribution exists and that the prior distribution is $f_0(\boldsymbol{\theta})$. Now suppose after inferences from the posterior distribution have been drawn that $f_0(\boldsymbol{\theta})$ was not the result of the most careful consideration, and it is now thought that $f_1(\boldsymbol{\theta})$ should be the prior distribution. Suppose also that in changing the prior distribution from $f_0(\boldsymbol{\theta})$ to $f_1(\boldsymbol{\theta})$ significantly changes the inferences drawn from the posterior distribution. We are presented with a sensitivity issue with regards to the prior. One possible way of resolving this issue is to carry out a reliable reassessment of the prior information. If this is not possible, then the sensitivity can only be resolved by direct consideration of the posterior distribution. If the inferences using $f_0(\boldsymbol{\theta})$ and $f_1(\boldsymbol{\theta})$ are clearly different, then one may decide which one disagrees with the prior beliefs about $\boldsymbol{\theta}$ having observed the data.

The above discussion highlights a few issues with regard to robustness to the prior distribution. Further detailed discussion about sensitivity to the prior distribution can be found in O'Hagan and Forster (2004, Chapter 4).

3.3 Hierarchical Models

The structure of a hierarchical model is where the parameters $(\theta_1, \theta_2, \ldots, \theta_p)$ are dependent on some hyperparameters $(\phi_1, \phi_2, \ldots, \phi_m)$ where p and m are the number of parameters and hyperparameters respectively. Just as the distribution of the data is written conditionally on the parameters $(\theta_1, \theta_2, \ldots, \theta_p)$ as $f(\boldsymbol{x}|\boldsymbol{\theta})$, the distribution for $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$ is written conditionally on the hyperparameters as $f(\boldsymbol{\theta}|\boldsymbol{\phi})$.

The Bayesian paradigm can be easily extended to obtain the joint posterior distribution of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ which is given in Equation (3.5). The joint posterior distribution shown in Equation (3.5) is used to make inferences about $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$.

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{x}) &= \frac{f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta}|\boldsymbol{\phi})f(\boldsymbol{\phi})}{\int_{\boldsymbol{\Phi}} \int_{\boldsymbol{\Theta}} f(x|\boldsymbol{\theta})f(\boldsymbol{\theta}|\boldsymbol{\phi})f(\boldsymbol{\phi}) \, \mathrm{d}\boldsymbol{\theta} \, \mathrm{d}\boldsymbol{\phi}} \\ &\propto f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta}|\boldsymbol{\phi})f(\boldsymbol{\phi}) \end{aligned} \tag{3.5}$$

where Θ and Φ are the parameter spaces for $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ respectively. The marginal posterior distribution for $\boldsymbol{\phi}$ is given by

$$\begin{aligned} f(\boldsymbol{\phi}|\boldsymbol{x}) &= \int f(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{x}) \, \mathrm{d}\boldsymbol{\theta} \\ &\propto f(\boldsymbol{\phi}) \int f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta}|\boldsymbol{\phi}) \, \mathrm{d}\boldsymbol{\theta} \\ &\propto f(\boldsymbol{\phi})f(\boldsymbol{x}|\boldsymbol{\phi}) \end{aligned}$$

which can be used to make inferences about $\boldsymbol{\phi}$. We note that the Bayesian paradigm can be easily extended to handle any number of levels that are in the hierarchical model.

3.4 Markov Chain Monte Carlo

A problem that occurs often in Bayesian inference is the calculation of posterior distributions that have high dimensionality. The problem is to evaluate the expectation in Equation (3.4) for some function $g(\boldsymbol{\theta})$. If the posterior distribution is analytically intractable or difficult to integrate a method is required to evaluate the expectation in Equation (3.4). A well known method for dealing with this problem is Markov Chain Monte Carlo (MCMC). A key component of this method is Monte Carlo integration. The general form of MCMC given by the Metropolis-Hastings algorithm will be described in Section 3.5. For a thorough exposition of MCMC methods in Bayesian computation, see, for example, Chen et al. (2000).

3.4.1 Monte Carlo Integration

Monte Carlo integration is a method for estimating the population mean $E[g(\boldsymbol{\theta})|\boldsymbol{x}]$. To estimate the population mean, samples $\boldsymbol{\theta}^{(l)}$, where l is a dummy variable used to reference the samples of $\boldsymbol{\theta}$, are obtained from $f(\boldsymbol{\theta}|\boldsymbol{x})$ via MCMC. The population mean of $g(\boldsymbol{\theta})$ can then be estimated by the mean of the samples

$$E[g(\boldsymbol{\theta})|\boldsymbol{x}] \approx \frac{1}{n}\sum_{l=1}^{n} g(\boldsymbol{\theta}^{(l)}). \tag{3.6}$$

Clearly, as $n \to \infty$ the sample mean will tend to the population mean. It should be noted that the choice of n is under the control of the analyst. Any process can be used to generate the $\boldsymbol{\theta}^{(l)}$ which draws random samples throughout the support of the posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{x})$ in the correct proportions. A popular method is to use a Markov chain (see Gilks et al.,1996, Chapter 1 for further details) that has the posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{x})$ as the stationary distribution.

3.5 Metropolis Hastings Algorithm

A method is now in place to estimate $E[g(\boldsymbol{\theta})|\boldsymbol{x}]$ where $\boldsymbol{x}$ represents the data but a Markov chain needs to be constructed such that the stationary distribution is the distribution of interest, that is the posterior distribution. One method of constructing a suitable Markov chain is the Metropolis-Hastings algorithm. This algorithm was initially proposed by Metropolis et al. (1953) and generalised by Hastings (1970). The Metropolis-Hastings algorithm works in the following way. At each iteration l, the next state $\boldsymbol{\theta}^{(l+1)}$ is chosen by sampling a candidate point $\boldsymbol{\phi}$ from a proposal distribution $q(. \mid \boldsymbol{\theta}^{(l)})$. It should be noted that the proposal distribution could depend on the current point $\boldsymbol{\theta}^{(l)}$. Then the candidate point $\boldsymbol{\phi}$ is accepted with probability $\alpha(\boldsymbol{\theta}^{(l)}, \boldsymbol{\phi})$ which is calculated using

$$\alpha(\boldsymbol{\theta}^{(l)}, \boldsymbol{\phi}) = \min\left(1, \frac{f(\boldsymbol{\phi})q(\boldsymbol{\theta}^{(l)} \mid \boldsymbol{\phi})}{f(\boldsymbol{\theta}^{(l)})q(\boldsymbol{\phi} \mid \boldsymbol{\theta}^{(l)})}\right) \tag{3.7}$$

where $f(.)$ represents the posterior distribution. If $\boldsymbol{\phi}$ is accepted, then the next state becomes $\boldsymbol{\theta}^{(l+1)} = \boldsymbol{\phi}$. If $\boldsymbol{\phi}$ is rejected, then $\boldsymbol{\theta}^{(l+1)} = \boldsymbol{\theta}^{(l)}$, that is the chain does not move.

There are a multitude of sampling methods that are a special case of the Metropolis-Hastings algorithm such as the Gibbs sampler, the Metropolis algorithm and the independence sampler. An exposition of the Metropolis Algorithm and the Gibbs Sampler will be given in Sections 3.5.1 and 3.5.2 respectively. For detailed expositions of the independence sampler see Tierney (1994) and Gilks et al. (1996, Chapter 1).

3.5.1 Metropolis Algorithm

The Metropolis algorithm proposed in Metropolis et al. (1953) only considers proposals from symmetric distributions such that $q(\boldsymbol{\phi}|\boldsymbol{\theta}) = q(\boldsymbol{\theta}|\boldsymbol{\phi})$ for all $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ where $\boldsymbol{\theta}$ is the parameter and $\boldsymbol{\phi}$ is a candidate point from the chosen proposal distribution. For the Metropolis algorithm the acceptance probability in Equation (3.7) becomes

$$\alpha(\boldsymbol{\theta}^{(l)}, \boldsymbol{\phi}) = \min\left(1, \frac{f(\boldsymbol{\phi})}{f(\boldsymbol{\theta}^{(l)})}\right).$$

When choosing the proposal distribution, its scale (for example Σ in the multivariate normal case) needs to be chosen carefully. For the rest of our discussion in this section, we will only consider the case of a single parameter. In this case, the acceptance probability $\alpha(\theta^{(l)}, \phi)$ is given by

$$\alpha(\theta^{(l)}, \phi) = \min\left(1, \frac{f(\phi)}{f(\theta^{(l)})}\right)$$

where θ is the parameter and ϕ is the candidate point. A proposal distribution generating small steps $\phi - \theta^{(l)}$ will have a high acceptance probability but will mix slowly as shown in Figure 3.1(b). If the scale is chosen such that the steps between ϕ and $\theta^{(l)}$ are large, then quite often we will be jumping from the body to the tails

of the distribution resulting in small values of $f(\phi)/f(\theta^{(l)})$ and low probability of acceptance. In this case the chain will not move which will also result in slow mixing as shown in Figure 3.1(c). The ideal scenario is to scale the proposal distribution such that both of these extremes are avoided. It should be noted that similar principles apply when considering more than one parameter.

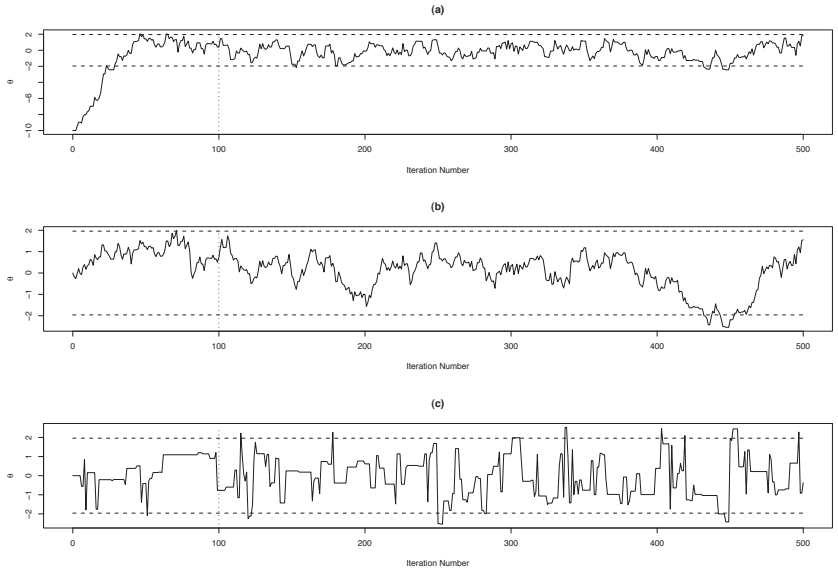


Figure 3.1: Trace plots from Metropolis algorithms with stationary distribution N(0,1) and proposal distributions (a) $q(.|\theta)$ =N(θ,0.5), (b) $q(.|\theta)$ = N(θ,0.1) and (c) $q(.|\theta)$ = N(θ,15.0). The burn-in is taken to be left of the dotted line and the 95% theoretical probability interval is indicated by the dashed line. This example is taken from Gilks et al. (1996, page 6).

3.5.2 Gibbs Sampling

The Gibbs sampler is a special case of the Metropolis-Hastings algorithm and was proposed by Geman and Geman (1984). A clear detailed exposition on Gibbs sam-

pling can be found in Casella and George (1992). Candidates under a Gibbs sampler are always accepted, that is the acceptance probability is always one. Gibbs sampling involves conditioning on each parameter in turn. We start by setting the iteration counter $l = 1$ and give initial values to the parameters $\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)} \dots \theta_p^{(0)}$ where p is the number of parameters and $p > 1$. New values for the parameters are obtained by sampling successively from the conditional posterior distribution for each parameter in turn which are given by:

$$\begin{aligned}
\theta_1^{(l)} &\sim f(\theta_1 \mid \theta_2^{(l-1)}, \theta_3^{(l-1)}, \theta_4^{(l-1)}, \dots, \theta_p^{(l-1)}), \\
\theta_2^{(l)} &\sim f(\theta_2 \mid \theta_1^{(l)}, \theta_3^{(l-1)}, \theta_4^{(l-1)}, \dots, \theta_p^{(l-1)}), \\
&\vdots \\
\theta_p^{(l)} &\sim f(\theta_p \mid \theta_1^{(l)}, \theta_2^{(l)}, \theta_3^{(l)}, \dots, \theta_{p-1}^{(l)}).
\end{aligned}$$

Change the value of l to $l + 1$ and repeat the sampling procedure above until convergence is reached. Gammerman and Lopes (2006) provide further details of the Gibbs sampler.

3.5.3 Monte Carlo Error

When applying MCMC methods, it is important to consider convergence and accuracy of estimation. If the chain does not converge, then clearly we will not be able to obtain a sample of the parameters from the posterior distribution. If the chain does converge then we need to consider the accuracy of the parameter estimates. The accuracy of the estimates can be measured via the mean-squared error of $g(\boldsymbol{\theta})$ from Equation (3.6), which is also referred to as the ergodic mean, see Gilks et al. (1996, Chapters 1 and 3). The mean-squared error of $g(\boldsymbol{\theta})$ is given by

$$\mathrm{MSE}(g(\boldsymbol{\theta})) = \frac{\mathrm{Var}_f(g(\boldsymbol{\theta}))}{n} \left(1 + 2 \sum_{l=1}^{n-1} \rho_d(g(\boldsymbol{\theta}^{(l)})) \right) \tag{3.8}$$

where n is the number of samples, $\text{Var}_f(g(\boldsymbol{\theta}))$ denotes we are taking the variance under the posterior distribution represented by $f(\boldsymbol{\theta}|\boldsymbol{x})$ and $\rho_d(g(\boldsymbol{\theta}^{(l)}))$ is the lag d autocorrelation in $g(\boldsymbol{\theta}^{(l)})$ where l is a dummy variable used to reference the samples of $\boldsymbol{\theta}$ obtained from $f(\boldsymbol{\theta}|\boldsymbol{x})$ via MCMC. It should be noted that we have been unable to find a simpler expression in the literature, however, Besag and Green (1993) note a similar expression using different notation.

If the chain converges geometrically (Meyn and Tweedie, 1993 presents an extensive treatment on geometric convergence), then the mean squared error for $g(\boldsymbol{\theta})$ will be finite. If the mean squared error for $g(\boldsymbol{\theta})$ is finite, then it can be made as small as desired by increasing n, see Geyer (1992) and Besag and Green (1993) for more details.

3.5.4 Issues in Implementing MCMC

When implementing MCMC methods, there are several issues that need to be taken into account. Gilks et al. (1996) has a more detailed account with regards to these issues. Smith and Roberts (1993) also give a detailed exposition of implementation and convergence issues with regards to MCMC methods. Papaspiliopoulos and Roberts (2008) discuss the stability and convergence of the Gibbs sampler for Bayesian hierarchical models. As mentioned in Section 3.1, the development and availability of computational power has created the need to incorporate MCMC methods into statistical software, for example, see Graves (2007) where design ideas for software incorporating MCMC methods are discussed and Chen et al. (2000) discuss MCMC methods in Bayesian computation.

One such issue is the choice of the starting values $\boldsymbol{\theta}^{(0)}$. If the chain mixes rapidly, then the chain will find its way to the stationary distribution fairly quickly from extreme starting values which is illustrated in Figure 3.1(a). For a slow mixing chain, the starting value will need to be chosen with care to avoid a long burn-in

period. Roughly speaking, a chain is said to mix well if samples are drawn from the whole support of the stationary distribution and does not stay at the same value for any length of time.

Determining the length of the burn-in period is an important exercise when using MCMC methods. The length of the burn-in period is dependent upon the initial starting value $\boldsymbol{\theta}^{(0)}$. The next factor to consider when determining the length of the burn-in is the rate of convergence of the distribution of $\boldsymbol{\theta}^{(l)}$ given the initial $\boldsymbol{\theta}^{(0)}$ to the stationary distribution. The final consideration we wish to highlight is how similar the distribution of $\boldsymbol{\theta}^{(l+1)}$ given $\boldsymbol{\theta}^{(l)}$ needs to be to the stationary distribution. The simplest method is to look at a plot of the output of all the samples generated and ascertain how long the chain takes to generate samples from the stationary distribution.

When to stop the chain is another important matter. If the chain is stopped too soon, then $\boldsymbol{\theta}^{(l)}$ will not be a sample from the stationary distribution. One solution to this problem is to run several chains of length n in parallel with different starting values and compare the estimates of $E[g(\boldsymbol{\theta})|\boldsymbol{x}]$. If there is not adequate agreement between the estimates, then it is clear that n must be increased.

As already mentioned, the samples $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(n)}\}$ will be dependent samples. To reduce the dependence in the samples, the simplest method is to take every kth sample, hence $k \times n$ samples must be generated by the chain. It should also be noted that we cannot prove that any of the sampling algorithms described above have converged. We can only empirically assess using some diagnostic plots to check if the sampling algorithm has converged to the stationary distribution.

3.6 Posterior Predictive Distribution

A key concept that is necessary to make predictive inferences about a particular quantity, such as the mean, is the posterior predictive distribution for the mean. For our applied problem, this will enable us to produce posterior predictive distributions for various characteristics of interest for proportions of Carbon that have not been experimented with. It is important to note that we will need the ability to obtain the posterior predictive distribution for the quantity required using different models. Let $\boldsymbol{\theta}$ denote the parameters in the model. Then the posterior distribution for $\boldsymbol{\theta}$, given data $\boldsymbol{x}$, is given by

$$f(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{f(\boldsymbol{\theta}) f(\boldsymbol{x} \mid \boldsymbol{\theta})}{\int_{\Theta} f(\boldsymbol{\theta}) f(\boldsymbol{x} \mid \boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}}$$

where Θ is the parameter space as defined in Section 3.2. Now imagine that the entire experiment is replicated and let $\boldsymbol{z}$ be the vector of possible responses. Then the predictive density for $\boldsymbol{z}$ under the chosen model is

$$f(\boldsymbol{z} \mid \boldsymbol{x}) = \int_{\Theta} f(\boldsymbol{z} \mid \boldsymbol{\theta}) f(\boldsymbol{\theta} \mid \boldsymbol{x}) \, \mathrm{d}\boldsymbol{\theta}. \tag{3.9}$$

If this distribution is analytically intractable or is difficult to integrate, then we can use various MCMC methods described in Section 3.4 to generate samples from the posterior predictive distribution and make inferences based upon these samples. The samples that are drawn from the posterior predictive distribution are obtained in two stages. The first stage is to sample from the posterior distribution of $\boldsymbol{\theta}$ shown in Equation (3.2). Each $\boldsymbol{\theta}^{(l)}$ can then be used to obtain $\boldsymbol{z}^{(l)}$ given $\boldsymbol{\theta}^{(l)}$ and data $\boldsymbol{x}$. We can obtain an estimate of the mean of $\boldsymbol{z}$ using

$$E[\boldsymbol{z}|\boldsymbol{x}] \approx \frac{1}{n} \sum_{l=1}^{n} \boldsymbol{z}^{(l)}.$$

This is the method that will be used to obtain posterior predictive distributions for the quantities described in Tables 1.4 and 2.2. We concentrate on the posterior predictive distributions for specific characteristics in Chapter 4.

3.7 Model Choice Criteria

In Section 3.7.2, the method used to compare the models is presented. The reason for using the chosen criterion is discussed in Section 3.7.3 and comparisons made to other available model choice criteria. However, it is important to look at other diagnostics of model performance and not just numerical values of model choice criterion. For example, plots of the residuals may exhibit any characteristic of the data that has not been accounted for by the model and would not be highlighted by a numerical value calculated by the model choice criterion. This will be discussed further in Section 3.8. In Section 3.7.3, we give a brief evaluation of the model choice criteria we have discussed.

3.7.1 Bayes Factors

The Bayesian approach for hypothesis testing was developed by Jeffreys (1935, 1961). Before we define the Bayes factor, we need to define the marginal likelihood for model $m_{(j)}$. The marginal likelihood for model $m_{(j)}$ is denoted by $f(\boldsymbol{x}|m_{(j)})$ and is defined by Equation (3.2). We define the Bayes factor by

$$B_{21} = \frac{f(\boldsymbol{x}|m_{(2)})}{f(\boldsymbol{x}|m_{(1)})}. \tag{3.10}$$

In the case where we wish to test the null hypothesis H_0 against the alternative hypothesis H_1, the Bayes factor can be expressed as

$$B_{10} = \frac{f(\boldsymbol{x}|m_{(1)})}{f(\boldsymbol{x}|m_{(0)})}$$

where $m_{(0)}$ and $m_{(1)}$ are the models under H_0 and H_1 respectively. It should be noted that the Bayes factor can only be defined when the marginal density of the data under each model is fully defined.

Jeffreys (1961) proposed interpreting B_{10} using half units on the $\log_{10}$ scale. Kass and Raftery (1995) suggest pooling two categories to obtain a method for interpret-

ing B_{10} as shown in Table 3.1.

$\log_{10}(B_{10})$	B_{10}	Evidence against H_0
0 to 0.5	1 to 3.2	Not worth more than a bare mention
0.5 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
> 2	> 100	Decisive

Table 3.1: Interpretation of the Bayes factor as proposed in Kass and Raftery (1995).

For a more detailed modern exposition see Kass and Raftery (1995). In addition, Gelman et al. (2004, Chapter 6) discuss an example of where Bayes factors are useful and an example of where they are a distraction.

3.7.2 Predictive Model Choice Criterion (PMCC)

The PMCC was proposed in Ibrahim and Laud (1995) and developed by Gelfand and Ghosh (1998). For each model $m \in \mathcal{M}$ where $\mathcal{M}$ is the set of models under consideration, the value of the criterion is calculated and the model with the smallest value is the best model. However, it is important to balance the improvement of a model and the amount of extra computation time required to include the extra parameter or parameters. The predictive density in Equation (3.9) is central to the PMCC as this is the distribution from which we obtain $\boldsymbol{z}$. The criteria is to choose a model which has the smallest value of

$$L_m^2 = E[(\boldsymbol{z} - \boldsymbol{x})^T(\boldsymbol{z} - \boldsymbol{x})]$$

where the expectation is taken with respect to the predictive density in Equation (3.9) for model m and $\boldsymbol{x}$ is the observed data. It can be shown that L_m^2 has the

decomposition

$$L_m^2 = \sum_{o=1}^{N} \left\{ (E[z_o|\boldsymbol{x}] - x_o)^2 + \mathrm{Var}(z_o|\boldsymbol{x}) \right\} \tag{3.11}$$

where o is the dummy variable indicating the oth observation and N is the number of observations. In Equation (3.11), the first component measures the goodness of fit, that is how close the predictions produced by the model are to the observed data. The second component is the penalty component, that is the model is penalised for increasing the number of parameters in the model through the variance of the predictions. This is because as the number of parameters increases, the variance of the predictions obtained from the model increases. Hence, the model's performance is measured by a combination of how close the predictions are to the observed data and the variability of the predictions. Ibrahim and Laud (1995) note that a good model should make predictions that are similar to the observations obtained from an experiment. Therefore, the best model from set of models under consideration, should be the model with the lowest L_m^2.

3.7.3 Evaluation of Model Choice Criteria

Some other popular criterion for comparing models are Akaike's Information Criterion (AIC; originally proposed in Akaike, 1973 and also published in Koehler and Murphree, 1988). The AIC from Koehler and Murphree (1988) is given by Equation (3.13). Let p denote the number of parameters and $\hat{\boldsymbol{\theta}}$ is the maximum likelihood of $\boldsymbol{\theta}$. The Bayes Factor and the AIC are given by Equations (3.12) and (3.13), respectively.

$$B_{21} = \frac{f(\boldsymbol{x}|m_{(2)})}{f(\boldsymbol{x}|m_{(1)})} \tag{3.12}$$

$$\mathrm{AIC} = -2\log f(\boldsymbol{x}|\hat{\boldsymbol{\theta}}) + 2p \tag{3.13}$$

Model selection via the AIC is based upon selecting the minimum value. When using the AIC, asymptotic considerations are required for a formal comparison be-

tween competing models. Sahu (2004) notes that model selection criterion involving asymptotic arguments are often invalid for small data sets. It should be noted that the AIC is not robust to outliers. Ibrahim and Laud (1995) point out that PMCC probably suffers from the same problem. One solution that the authors suggest would be to calculate the PMCC with and without the outliers which should highlight their effects.

Sahu (2004) points out that Bayesian model choice methods are attractive because they do not rely on asymptotic arguments. When using decision theoretic method to choose a model, the PMCC is the most appropriate criterion under normal likelihood and a symmetric loss function. Sahu (2004) highlights that it is possible to use different loss functions. There are other Bayesian methods that have been proposed to discriminate between models. One such method is the reversible jump MCMC which is proposed by Green (1995).

Kass and Raftery (1995) point out that the integral required to calculate $f(\boldsymbol{x})$ shown in Equation (3.2) can be analytically intractable and must be computed via numerical methods. In the same article, the authors also highlight that the statistical software available is inefficient due to large sample sizes. In this case, the integrand becomes highly peaked around its maximum. A second reason for the integrand in Equation (3.2) being intractable is due to the dimensions. In this case, Markov Chain Monte Carlo methods can be applied with some adaptation. A review of these techniques is provided in Evans and Swartz (1995). As we have highlighted here, the Bayes Factor also has its technical difficulties.

A more detailed exposition on Bayesian model selection and applications in practical problems are presented in Sahu (2004) and Robert (2001, Chapter 7). Another Bayesian model selection criterion is the Deviance Information Criterion, proposed by Spiegelhalter et al. (2002).

3.8 Model Adequacy via Residual Analysis

Once a model has been fitted, the residuals should be examined to check that the model describes the data adequately. If the chosen model provides a good fit, the residuals should show a random cluster around zero. For the time series models considered in this thesis (see Chapters 6, 7 and 8), we expect the plot of the residuals against time to have a mean of zero and a constant variance.

In time series analysis, it is possible to define a number of different types of residuals, for example see Mauricio (2008). In the analysis presented in this thesis, we define the residuals to be the difference between the observed and the fitted values. The fitted values are obtained by replacing parameters by their Bayes estimates. These fitted values are often the one-step ahead predictions. This is a consequence of using the time series analysis, see West and Harrison (1999, Chapter 10) for several examples of using these type of residuals in practical data analyses.

Chapter 4

Predictive Distributions

4.1 Introduction

As mentioned in Section 1.6, the main objectives of this thesis is to study the characteristics of interest defined in Table 1.4. In this chapter, we develop Bayesian methodologies to achieve this objective. We also develop the computational methods required to implement these Bayesian methods.

Throughout this thesis, we will treat Potential and Current as continuous variables. This is because Current is the response variable and can theoretically take any value within a given range for a particular experiment. In the literature it is implicitly assumed that Potential is a continuous variable, for examples, see Lovric and Scholz (2003) and Lundquist et al. (2001). Further to this, we will treat time as discrete (except in Chapter 5 where time is treated as continuous), since the Current is recorded at a particular time point and not between time points. These issues are discussed further in Section 4.5.

We define the following notations used in this chapter. Unless stated otherwise, for ease of notation we drop the scan rate subscript; the developments in this chapter will be applied to each of the scan rates individually. The rest of the generic nota-

tion that is common to all sections of this chapter is given by: $\boldsymbol{\theta}$ is the vector of parameters for the model, Θ is the parameter space of $\boldsymbol{\theta}$, c_k is the proportion of Carbon corresponding to the k^{th} level, t is the time index as defined in Table 2.1 and n is the number of samples from the sampling algorithm chosen, for example, the Gibbs sampler. Further to this, let $x_k(t)$ represent the Current at time t for Carbon level k and

$$\begin{aligned} \boldsymbol{x}_k &= (x_k(1), x_k(2), \ldots, x_k(N)), \\ \boldsymbol{p} &= (p(1), p(2), \ldots, p(N)) \end{aligned}$$

where $p(t)$ is the Potential at time t and N is the length of the time series (which is dependent on the scan rate). In addition to this, our generic model for $x_k(t)$ will be of the form

$$x_k(t) = g(x_k(t-1), \ldots) + \varepsilon_k(t)$$

where $t = 1, \ldots, N$ for a suitable function $g(\ldots)$ which may depend on additional parameters.

The voltammogram characteristics, defined in Table 1.4, are functions of Current, Potential and the associated time indices, say $h(\boldsymbol{x}_k, c_k, \boldsymbol{p})$, where the function $h(\ldots)$ is a characteristic of interest, such as $I_k^{(\max)}$. Under the Bayesian paradigm, we make inferences for $h(\boldsymbol{x}_k, c_k, \boldsymbol{p})$ by using the posterior predictive distribution

$$f(h(\boldsymbol{z}_k, c_k, \boldsymbol{p})|\boldsymbol{x}_k, c_k, \boldsymbol{p}) = \int_\Theta f(h(\boldsymbol{z}_k, c_k, \boldsymbol{p})|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) \, \mathrm{d}\boldsymbol{\theta} \tag{4.1}$$

where $\boldsymbol{z}_k = (z_k(1), z_k(2), \ldots, z_k(N))$ is a set of future observations with the same Potential and other associated covariate values such as Carbon.

The posterior predictive distribution shown in Equation (4.1) can be calculated using Monte Carlo integration (see Section 3.4.1). At each MCMC iteration, we generate a replicate data set $\boldsymbol{z}_k = (z_k(1), z_k(2), \ldots, z_k(N))$ successively in time as the one-step ahead predictions, and the quantity of interest, the $h(\ldots)$ function, is

calculated and then ergodic means are formed at the end of the MCMC run. Thus the time series formulation of the data here leads to the averages of the one-step ahead predictions used to make predictive inference.

The characteristics we will be focusing on are the Peak Potential, Peak Current and the Peak Separation in Potential, since these are the main characteristics of interest. Moreover, the posterior predictive distributions involving time and Potential will be providing the same information as they are just different ways of referencing data points, hence we will focus on the Potential as this is how the chemists reference the data. The final reason is for the purposes of brevity. We will provide brief comments regarding the calculation of the other posterior predictive distributions where appropriate.

4.2 Predictive Inferences for Current

The posterior predictive distribution for Peak Current, $I_k^{(\max)}$, is given by

$$f(I_k^{(\max)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) \;=\; \int_\Theta f(I_k^{(\max)}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}_k, c_k, \boldsymbol{p})\ \mathrm{d}\boldsymbol{\theta}. \tag{4.2}$$

We can estimate the distribution shown in Equation (4.2) using MCMC methods in the following way. A predictive cycle $z_k^{(l)}(t)$ where $t = 1, \ldots, N$ is generated at each iteration of the MCMC sampling algorithm. We can calculate $I_k^{(\max,l)}$ using

$$I_k^{(\max,l)} \;=\; \max\{z_k^{(l)}(t);\ t = 1, \ldots, N\} \tag{4.3}$$

where $I_k^{(\max,l)}$ is a draw from the posterior distribution shown in Equation (4.2). Using these samples, we can now estimate the quantity $E[I_k^{(\max)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$ using

$$\begin{aligned} E[I_k^{(\max)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}] \;&=\; \int I_k^{(\max)} f(I_k^{(\max)}|\boldsymbol{x}_k, c_k, \boldsymbol{p})\ \mathrm{d}I_k^{(\max)} \\ &\backsimeq\; \frac{1}{n}\sum_{l=1}^{n} I_k^{(\max,l)}. \end{aligned} \tag{4.4}$$

The approximation shown in Equation (4.4) can be used to estimate the mean of the $I_k^{(\max)}$ for each level of Carbon using an iterative model fitting procedure such as Gibbs sampler or Metropolis-Hastings algorithm.
Similarly, the posterior predictive distribution for Minimum Current, $I_k^{(\min)}$, can be defined as

$$f(I_k^{(\min)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) \;=\; \int_\Theta f(I_k^{(\min)}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}_k, c_k, \boldsymbol{p})\ \mathrm{d}\boldsymbol{\theta}. \tag{4.5}$$

We can define $E[I_k^{(\min)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$ similarly and obtain an estimate using samples

$$I_k^{(\min,l)} = \min\{z_k^{(l)}(t);\ t = 1, \ldots, N\}$$

from the posterior predictive distribution shown in Equation (4.5) and the estimate is obtained at the end of the MCMC run by forming suitable averages.

4.3 Predictive Inferences for Time

To estimate the Peak Time, $t_k^{(\max)}$, we need to calculate $E[t_k^{(\max)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$, which requires the posterior predictive distribution $f(t_k^{(\max)}|\boldsymbol{x}_k, c_k, \boldsymbol{p})$. This posterior predictive distribution for $t_k^{(\max)}$ is given by

$$f(t_k^{(\max)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) \;=\; \int_\Theta f(t_k^{(\max)}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}_k, c_k, \boldsymbol{p})\ \mathrm{d}\boldsymbol{\theta}. \tag{4.6}$$

We can estimate the posterior predictive distribution shown in Equation (4.6) in exactly the same way we estimated the distribution shown in Equation (4.2). As before, we generate a predictive cycle $z_k^{(l)}(t)$ for each iteration of the MCMC sampling algorithm. We can then obtain $t_k^{(\max,l)}$ for each of the predictive cycles generated by the MCMC sampling algorithm. Each $t_k^{(\max,l)}$ will be a draw from the posterior predictive distribution shown in Equation (4.6). We are now able to calculate the

quantity $E[t_k^{(\max)}|\boldsymbol{x}_k, c_k]$, which is given by

$$\begin{aligned} E[t_k^{(\max)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}] &= \sum_{t_k^{(\max)}} t_k^{(\max)} f(t_k^{(\max)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) \\ &\backsimeq \frac{1}{n}\sum_{l=1}^{n} t_k^{(\max,l)}. \end{aligned} \tag{4.7}$$

Similarly, to estimate the Minimum Time, $t_k^{(\min)}$, we need to calculate $E[t_k^{(\min)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$ which requires the posterior predictive distribution

$$f(t_k^{(\min)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) = \int_{\Theta} f(t_k^{(\min)}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}_k, c_k, \boldsymbol{p})\, \mathrm{d}\boldsymbol{\theta}. \tag{4.8}$$

By using the same methodology to calculate $E[t_k^{(\max)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$, we can calculate $E[t_k^{(\min)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$ using

$$E[t_k^{(\min)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}] \backsimeq \frac{1}{n}\sum_{l=1}^{n} t_k^{(\min,l)}. \tag{4.9}$$

Let $t_k^{(\mathrm{sep},l)}$ be the predicted Peak Separation in Time which is simply the difference between the predicted Peak Time and the predicted Minimum Time, that is $t_k^{(\mathrm{sep},l)} = t_k^{(\max,l)} - t_k^{(\min,l)}$. To estimate $E[t_k^{(\mathrm{sep})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$, we require the posterior predictive density for $t_k^{(\mathrm{sep})}$, which is given by

$$f(t_k^{(\mathrm{sep})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) = \int_{\Theta} f(t_k^{(\mathrm{sep})}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}_k, c_k, \boldsymbol{p})\, \mathrm{d}\boldsymbol{\theta}.$$

We are now in a position to estimate $E[t_k^{(\mathrm{sep})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$, which is given by

$$\begin{aligned} E[t_k^{(\mathrm{sep})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}] &= \sum_{t_k^{(\mathrm{sep})}} t_k^{(\mathrm{sep})} f(t_k^{(\mathrm{sep})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) \\ &\backsimeq \frac{1}{n}\sum_{i=1}^{n} t_k^{(\mathrm{sep},l)}. \end{aligned}$$

We now have a sample from the posterior predictive distribution for the Peak Separation in Time and can therefore estimate $t_k^{(\mathrm{sep})}$.

Before we estimate $E[t_k^{(\text{wid})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$, where $t_k^{(\text{wid})}$ is the Peak Width in Time, we need to know the posterior predictive distribution of $t_k^{(\text{wid})}$, which is given by

$$f(t_k^{(\text{wid})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) = \int_\Theta f(t_k^{(\text{wid})}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) \, \mathrm{d}\boldsymbol{\theta} \tag{4.10}$$

where $t_k^{(\text{wid})} = t_{2,k} - t_{1,k}$. Let $t_{1,k}^{(l)}$ be the first predicted time that the Current reaches $I_k^{(\text{wid},l)}$ and $t_{2,k}^{(l)}$ be the final predicted time the Current reaches $I_k^{(\text{wid},l)}$ where $I_k^{(\text{wid},l)} = I_k^{(\text{max},l)}/2$ and superscript wid denotes that we are considering the Peak Width. Hence, by definition, the predicted Peak Width in Time is given by

$$t_k^{(\text{wid},l)} = t_{2,k}^{(l)} - t_{1,k}^{(l)}. \tag{4.11}$$

Using Equations (4.10) and (4.11), we are able to estimate $E[t_k^{(\text{wid})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$ by

$$\begin{aligned} E[t_k^{(\text{wid})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}] &= \sum_{t_k^{(\text{wid})}} t_k^{(\text{wid})} f(t_k^{(\text{wid})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) \\ &\backsimeq \frac{1}{n} \sum_{l=1}^{n} t_k^{(\text{wid},l)}. \end{aligned}$$

4.4 Predictive Inferences for Potential

To estimate the mean of the predictive Peak Potential for each Carbon level k, $P_k^{(\text{max})}$, we require the posterior predictive distribution

$$f(P_k^{(\text{max})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) = \int_\Theta f(P_k^{(\text{max})}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) \, \mathrm{d}\boldsymbol{\theta}. \tag{4.12}$$

Let $P_k^{(\text{min},l)}$ and $P_k^{(\text{max},l)}$ be the predicted Potential which produces the Minimum or Peak Current for the kth Carbon level and the lth iteration respectively. Each $P_k^{(\text{max},l)}$ will be a draw from the posterior predictive distribution shown in Equation (4.12). Using Equation (4.12), we can now calculate $E[P_k^{(\text{max})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$ which is

given by

$$\begin{aligned} E[P_k^{(\max)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}] &= \int P_k^{(\max)} f(P_k^{(\max)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) dP_k^{(\max)} \\ &\backsimeq \frac{1}{n}\sum_{l=1}^{n} P_k^{(\max,l)}. \end{aligned}$$

In a similar way, we can also calculate $E[P_k^{(\min)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$ using samples from the posterior predictive distribution

$$f(P_k^{(\min)}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) = \int_\Theta f(P_k^{(\min)}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) \, \mathrm{d}\boldsymbol{\theta}.$$

Before we construct an estimate for the Peak Separation in Potential, $P_k^{(\text{sep})} = P_k^{(\max)} - P_k^{(\min)}$, we need to calculate the posterior predictive distribution for $P_k^{(\text{sep})}$, which is given by

$$f(P_k^{(\text{sep})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) = \int_\Theta f(P_k^{(\text{sep})}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) \, \mathrm{d}\boldsymbol{\theta}. \tag{4.13}$$

Using the definitions of $P_k^{(\min,l)}$ and $P_k^{(\max,l)}$, the predictive Peak Separation in Potential, $P_k^{(\text{sep},l)}$, can now be defined as

$$P_k^{(\text{sep},l)} = P_k^{(\max,l)} - P_k^{(\min,l)} \tag{4.14}$$

where the superscript sep denotes that we are concerned with the Peak Separation. Using Equations (4.13) and (4.14) we can estimate $E[P_k^{(\text{sep})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$ using

$$\begin{aligned} E[P_k^{(\text{sep})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}] &= \int P_k^{(\text{sep})} f(P_k^{(\text{sep})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) dP_k^{(\text{sep})} \\ &\backsimeq \frac{1}{n}\sum_{l=1}^{n} P_{\text{sep},k}^{(l)}. \end{aligned}$$

Before we estimate $E[P_k^{(\text{wid})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$ we need to calculate the posterior predictive distribution of $P_k^{(\text{wid})}$ given by

$$f(P_k^{(\text{wid})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) = \int_\Theta f(P_k^{(\text{wid})}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) \, \mathrm{d}\boldsymbol{\theta}. \tag{4.15}$$

Let $P_{1,k}^{(l)}$ be the first predicted Potential value at which the Current is equal to $I_k^{(\text{wid},l)}$. Similarly, let $P_{2,k}^{(l)}$ be the last predicted Potential value at which the Current is equal to $I_k^{(\text{wid},l)}$. Then the predicted Peak Width in Potential, denoted by $P_k^{(\text{wid},l)}$, can be defined as

$$P_k^{(\text{wid},l)} = P_{2,k}^{(l)} - P_{1,k}^{(l)}. \tag{4.16}$$

We are now able to estimate the predictive distribution of the Peak Width in Potential. Using equations (4.15) and (4.16) we are now able to estimate $E[P_k^{(\text{wid})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}]$ by

$$\begin{aligned} E[P_k^{(\text{wid})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}] &= \int P_k^{(\text{wid})} f(P_k^{(\text{wid})}|\boldsymbol{x}_k, c_k, \boldsymbol{p}) dP_k^{(\text{wid})} \\ &\backsimeq \frac{1}{n}\sum_{l=1}^{n} P_k^{(\text{wid},l)} \end{aligned}$$

where $P_k^{(\text{wid})} = P_{2,k} - P_{1,k}$.

4.5 Closing Remarks

In this chapter, we have derived the posterior predictive distributions and discussed how to approximate these via MCMC methods. We note that the characteristics we will be focusing on will be the Peak Potential, Peak Current and Peak Separation in Potential.

We note that characteristics of interest involving time, such as the Peak Time, will be treated as discrete. However, in Section 5.3, the posterior predictive distributions involving time, such as Peak Time, are modelled as a continuous response due to the nature and simplicity of the models. Hence, the predictions for characteristics of interest involving time are on a continuous scale. In contrast, the posterior predictive distributions of the Peak Current as plotted in Figure 6.8 for example, are obtained using the characteristics of the series of replications $z_k(t)$ for the response

Current. Hence the value of t corresponding to the fitted (or equivalently one-step ahead) Peak Current can only take one of the possible discrete values of t. Thus, posterior predictive distributions involving time will essentially be discrete, but these distributions involving time in Section 5.3 will be continuous.

The characteristics of interest involving Potential and Current will be treated as continuous. The reason for treating Current as continuous is because clearly it can take any value and therefore any of the characteristics defined in Table 1.4 involving Current could also take any possible value. As noted in Section 4.1, Potential is assumed to be a continuous variable in the literature, for examples see Lovric and Scholz (2003) and Lundquist et al. (2001). We will follow the approach taken in the literature by treating Potential as a continuous variable.

The methodology developed here is generic in the sense that it can be adopted for any model based analysis. In addition, the computation methods will work with any MCMC algorithm which is able to draw samples from the posterior distribution.

Chapter 5

Regression Models for Summary Characteristics

5.1 Introduction

In this chapter, we calculate the characteristics of interest (see Table 2.2) from the raw data and model each of them using polynomial regression models which the chemists were keen to explore. We adopt a Bayesian approach with vague prior distributions for each of the model parameters, in line with the methods used in the rest of this thesis. The joint posterior and the full conditional posterior distributions will be derived for each model parameter.

We fit linear, quadratic and cubic models and select the best model using the PMCC method (see Section 3.7.2) and assess its predictive value for each characteristic separately, for each scan rate.

Our preferred modelling approaches involve modelling the whole Current output curve rather than the summary statistics and will be presented later in Chapters 6, 7 and 8. These approaches have the advantage that inferences can be made for any characteristics of interest to the chemists from using a single model.

5.2 Regression Models for Summary Characteristics

Let h_k denote the mean value of a particular characteristic, for the kth level of Carbon ($k = 1, \ldots, K$) averaged over the replicate runs and replicate channels (see Chapter 1) for a particular scan rate. For ease of notation, we drop the subscript, s, that has denoted the scan rate ($s = 1, 2, 3$). The characteristics examined are $I_k^{(\max)}$, $P_k^{(\max)}$ and $P_k^{(\text{sep})}$ (see Table 2.2).

It is important to note that the only possible variable that we can use for our model is the proportion of Carbon as we do not know the Potential and time values associated with each of the characteristics. Hence, the only type of model that we will consider is given by

$$h_k = \beta_0 + \sum_{q=1}^{\varpi} \beta_q c_k^q + \varepsilon_k \tag{5.1}$$

where h_k is the characteristic of interest to be modelled for the k^{th} Carbon level, q is dummy variable for the power, ϖ is the degree of the polynomial, c_k is the k^{th} Carbon level, β_0 and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_\varpi)^T$ are the parameters of the model, ε_k represents the residual for the k^{th} Carbon level. As already noted in Section 5.1, this model can be applied to each of the scan rates.

It is possible to use more complex regression models than the model shown in Equation (5.1), such as using fractional powers. Our purpose here is to simply illustrate what is possible using simple regression models by only looking at the data for each characteristic individually. As already mentioned in Section 5.1, we also note that the chemists were keen to explore what this model could achieve.

The likelihood for the model is given by

$$\left(\tau^2\right)^{\frac{K}{2}} \exp\left\{-\frac{\tau^2}{2}\sum_{k=1}^{K}\left(h_k - \beta_0 - \sum_{q=1}^{\varpi}\beta_q c_k^q\right)^2\right\}$$

where $\tau^2 = \sigma^{-2}$ and K is the total number of distinct Carbon levels. We will assume vague priors for the parameters, which are given by:

$$\begin{aligned}\beta_0 &\sim N(0, v_0^2),\\ \beta_q &\sim N(0, v_q^2),\\ \tau^2 &\sim \text{Gamma}(d_1, d_2),\end{aligned}$$

where $v_0, v_1, \ldots, v_\varpi$, d_1 and d_2 are constants to be chosen, and throughout this thesis, Gamma(d_1,d_2) denotes the Gamma distribution with mean $\frac{d_1}{d_2}$. In addition, let $V = \text{diag}(v_1^2, v_2^2, \ldots, v_\varpi^2)$. The posterior distribution for the model in Equation (5.1) is then given by

$$\left(\tau^2\right)^\psi \exp\left\{-\frac{\tau^2}{2}\sum_{k=1}^{K}(\varepsilon_k)^2 - \frac{\beta_0^2}{2v_0^2} - \frac{1}{2}\boldsymbol{\beta}^{\boldsymbol{T}}V^{-1}\boldsymbol{\beta} - d_2\tau^2\right\}$$

where

$$\varepsilon_k = h_k - \left(\beta_0 + \sum_{q=1}^{\varpi}\beta_q c_k^q\right)$$

and

$$\psi = \frac{K}{2} + d_1 - 1.$$

The conditional posterior distributions for the unknown parameters β_0, $\boldsymbol{\beta}$ and τ^2, are given by:

$$\begin{aligned}\beta_0|\boldsymbol{\beta}, \tau^2 &\sim N\left(\mu_0, \sigma_0^2\right),\\ \boldsymbol{\beta}|\beta_0, \tau^2 &\sim N_\varpi\left(\mu_\beta, \Sigma_\beta\right),\\ \tau^2|\beta_0, \boldsymbol{\beta} &\sim \text{Gamma}(\psi, \phi),\end{aligned}$$

where:

$$\begin{aligned}
\mu_0 &= \tau^2 \left(\tau^2 K + \frac{1}{v_0^2}\right)^{-1} \left(\sum_{k=1}^{K} \left(h_k - \sum_{q=1}^{\varpi} \beta_q c_k^q\right)\right), \\
\sigma_0^2 &= \left(\tau^2 K + \frac{1}{v_0^2}\right)^{-1}, \\
\boldsymbol{\mu}_\beta &= \tau^2 \left(\tau^2 C^T C + V^{-1}\right)^{-1} \sum_{k=1}^{K} \boldsymbol{c}_{\mathrm{vec,k}}(h_k - \beta_0),
\end{aligned}$$

$$\begin{aligned}
\Sigma_\beta &= \left(\tau^2 C^T C + V^{-1}\right)^{-1}, \\
C &= \begin{pmatrix} c_1 & c_1^2 & \dots & c_1^{\varpi} \\ c_2 & c_2^2 & \dots & c_2^{\varpi} \\ \vdots & \vdots & \vdots & \vdots \\ c_K & c_K^2 & \dots & c_K^{\varpi} \end{pmatrix}, \\
\boldsymbol{c}_{\mathrm{vec,k}} &= (c_k, c_k^2, c_k^3, \dots, c_k^{\varpi})^T, \\
\psi &= \frac{K}{2} + d_1, \\
\phi &= \frac{1}{2} \sum_{k=1}^{K} (\varepsilon_k)^2 + d_2.
\end{aligned}$$

5.3 Analysis and Conclusions

An example of the data set to be modelled is shown in Figure 2.13. The estimates of the parameters for the polynomial regression models were found to be fairly robust when we varied the prior variances for β_0, $\boldsymbol{\beta}$ and the values of the hyperparameters d_1 and d_2 for τ^2. For the results presented in this section, we set the value for the prior variances at 10^8 and the values of both d_1 and d_2 were set to 0.001. Vague priors were used as we have no prior information about the model parameters. We compared different regression models using the PMCC and found the best model for each characteristic for each scan rate, see Table 5.1.

	Scan Rate		
	1	2	3
Minimum Potential	3	2	2
Minimum Time	2	3	3
Minimum Current	1	2	1
Peak Potential	3	1	2
Peak Time	2	3	1
Peak Current	1	2	2
Peak Separation in Potential	2	2	2
Peak Separation in Time	2	3	2
Peak Separation in Current	2	1	2
Peak Width in Potential	1	2	3
Peak Width in Time	1	2	3

Table 5.1: The best polynomial degree to use for each characteristic and scan rate.

The MCMC diagnostic plots for the Peak Potential, Peak Current and Peak Separation in Potential are shown in Figures 5.1, 5.2, and 5.3. We thinned the samples obtained from the MCMC algorithm by taking every fifth iteration. The autocorrelation function (acf) plots also provide evidence that the samples of the parameters are not highly correlated. The value of the acf tails off as expected for a MCMC sampling algorithm that has the properties of convergence. The diagnostic plots also indicate that we have obtained an approximate sample from the stationary distribution for each of the parameters.

Parameter	Prior Variances		
	10^8	10^{10}	10^{12}
β_0	4.4162	4.4164	4.4164
	(0.0297)	(0.0295)	(0.0295)
β_1	-0.9802	-0.9796	-0.9796
	(0.2730)	(0.2746)	(0.2746)
σ^2	0.0638	0.0637	0.0637
	(0.0059)	(0.0058)	(0.0058)

Table 5.2: Parameter estimates, posterior mean and standard deviations (within parenthesis) for different prior variances for the linear model for Peak Potential for scan rate 2.

The inferences obtained for the regression model parameters were fairly robust when we varied the prior variances. This is illustrated by the similarity of the parameter estimates under different prior variances, which are shown in Tables 5.2-5.4 for scan rate 2. The results in these tables confirm that the values of the parameter estimates are not changed very much for the different prior variances used. A similar sensitivity analysis was carried for the other characteristics for scan rates 1 and 3, and similar results were obtained (omitted for brevity). For the remainder of the

Parameter	Prior Variances		
	10^8	10^{10}	10^{12}
β_0	89.6317	89.6053	89.6053
	(6.7780)	(6.7656)	(6.7656)
β_1	-192.0526	-192.7984	-192.7984
	(149.2340)	(149.1309)	(149.1309)
β_2	-264.7277	-259.0127	-259.0127
	(612.4815)	(612.5552)	(612.5552)
σ^2	680.9840	680.4227	680.4227
	(62.2958)	(62.6931)	(62.6931)

Table 5.3: Parameter estimates, posterior mean and standard deviations (within parenthesis) for different prior variances for the quadratic model for Peak Current for scan rate 2.

results presented in this section, we use the value for the prior variances at 10^8 and for each of d_1,d_2 we use 0.001.

To plot the distribution densities shown in Figures 5.4-5.6, we used the plot and density commands with a Gaussian kernel in R is employed, see Venables and Ripley (2002) and Silverman (1986).

The posterior predictive distributions for the Peak Potential, Peak Current and Peak Separation in Potential in Figures 5.4, 5.5 and 5.6, respectively, appear to follow the pattern of the observations shown in Figure 2.13. We also note that the variations between the Carbon levels for each of these characteristics are as expected.

We next use the model to predict Peak Potential, Peak Current and Peak Separation in Potential at six values of Carbon within the range of 3-20% which were not run in the experiment, namely 6%, 8%, 12%, 14%, 16%, 18% and 20%. The posterior predictive distributions for the Peak Potential, Peak Current, and Peak

Parameter	Prior Variances		
	10^8	10^{10}	10^{12}
β_0	0.9421	0.9432	0.9432
	(0.0545)	(0.0546)	(0.0546)
β_1	-7.5463	-7.5686	-7.5686
	(1.1998)	(1.2033)	(1.2033)
β_2	26.0416	26.1109	26.1109
	(4.9379)	(4.9537)	(4.9537)
σ^2	0.0441	0.0442	0.0442
	(0.0041)	(0.0041)	(0.0041)

Table 5.4: Parameter estimates, posterior mean and standard deviations (within parenthesis) for different prior variances for the quadratic model for Peak Separation in Potential for scan rate 2.

Separation in Potential for these levels of Carbon are displayed in Figures 5.7, 5.8, and 5.9, respectively. By including the posterior predictive distributions shown in Figures 5.7-5.9, we are simply illustrating the predictive usefulness of the best model proposed in this chapter.

In general, there appears to be very little uncertainty about the posterior predictive distributions for the characteristics shown in detail in Figures 5.7-5.9. This same property was exhibited for the other characteristics of interest (these are shown in Appendix B.2). The mean of the posterior predictive distribution for Peak Current, see Figure 5.8, shows a realistic difference between Carbon levels (guided by the chemists). Similarly, the distributions in Figure 5.7 are centred on values we would expect for the Peak Potential. The mode of the posterior predictive distributions gradually decreases as the proportion of Carbon decreases, following the pattern shown in Figure 2.13(d). The posterior predictive distributions for the Peak Current

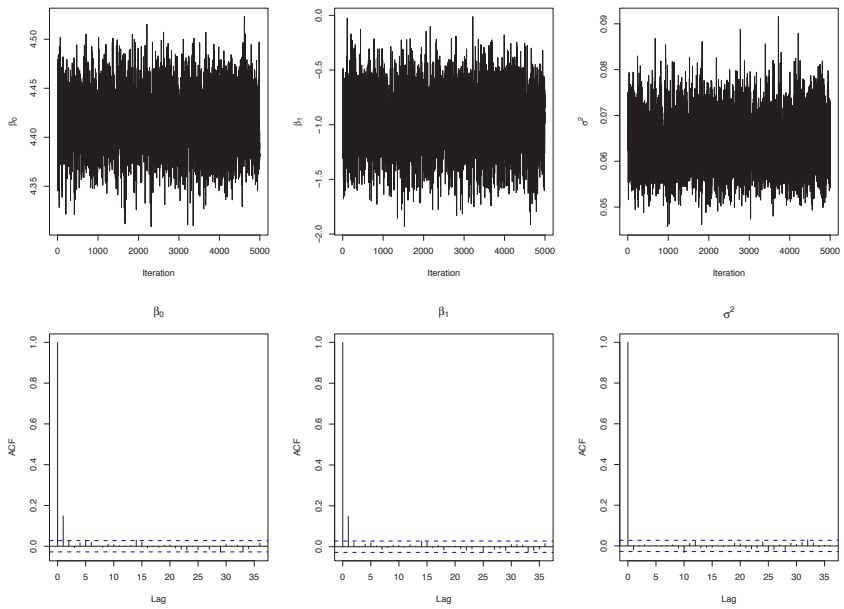


Figure 5.1: Trace and autocorrelation plots of every fifth sample generated by the Gibbs sampler for all the parameters of the linear model for Peak Potential for scan rate 2.

show that as the proportion of Carbon increases the location of the distribution decreases. This is the same pattern of behaviour shown in Figure 2.13(f). When we look at the posterior predictive distributions of the Peak Separation in Potential in Figure 5.9, we see that the location of the distribution decreases as the proportion of Carbon decreases. This follows the pattern exhibited by the data shown in Figure 2.13(g). As in standard regression analysis, it is possible to further refine these models using diagnostic residual plots. However, we do not pursue those here since the regression models are not our preferred modelling approaches as mentioned in Section 5.1; see also further related discussion in Section 9.2.

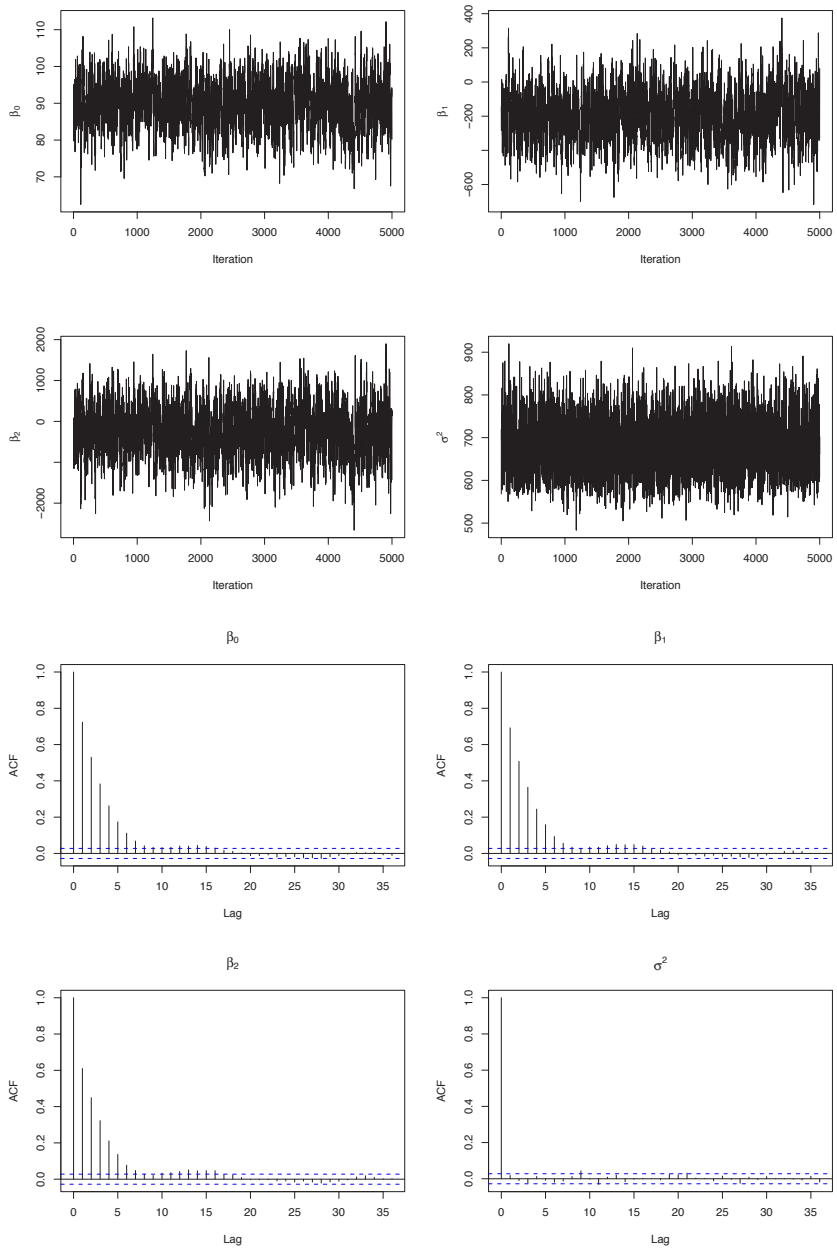


Figure 5.2: Trace and autocorrelation plots of every fifth sample generated by the Gibbs sampler for all the parameters of the quadratic model for Peak Current for scan rate 2.

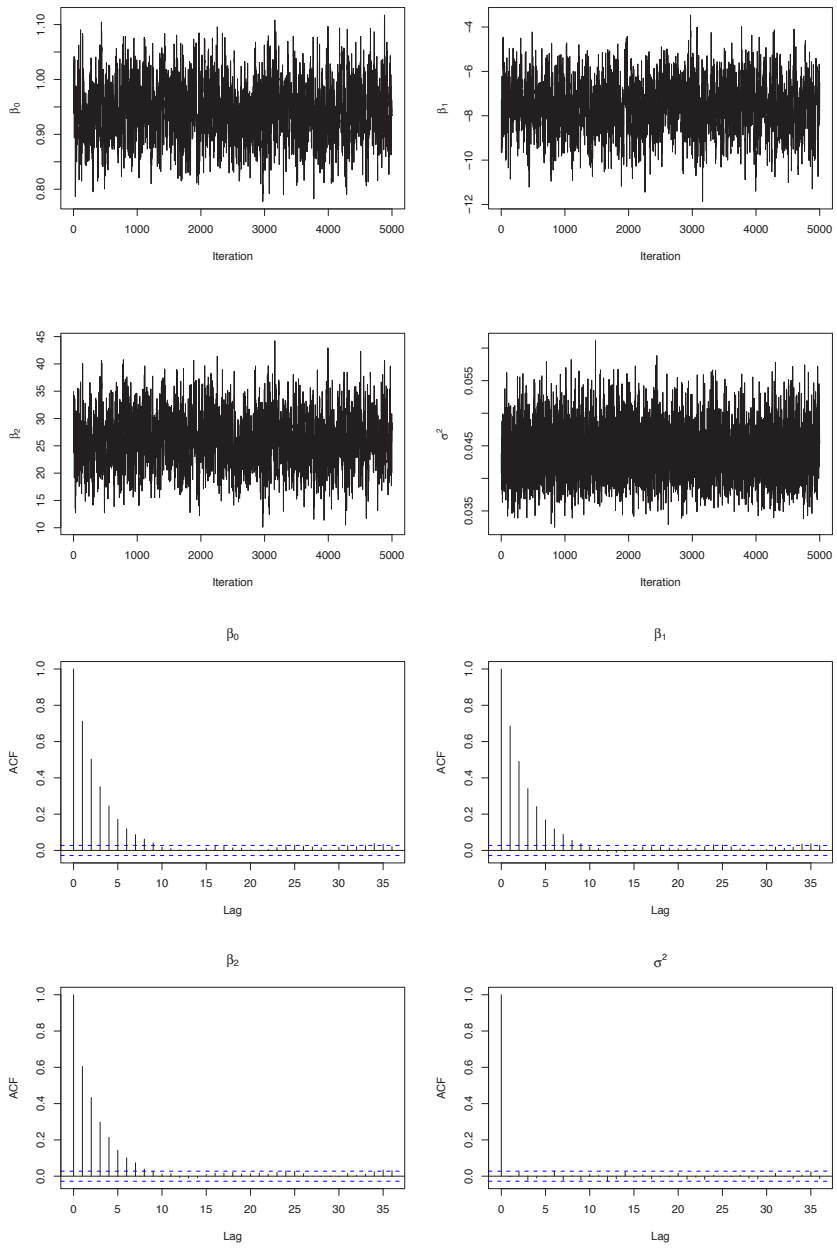


Figure 5.3: Trace and autocorrelation plots of every fifth sample generated by the Gibbs sampler for all the parameters of the quadratic model for Peak Separation in Potential for scan rate 2.

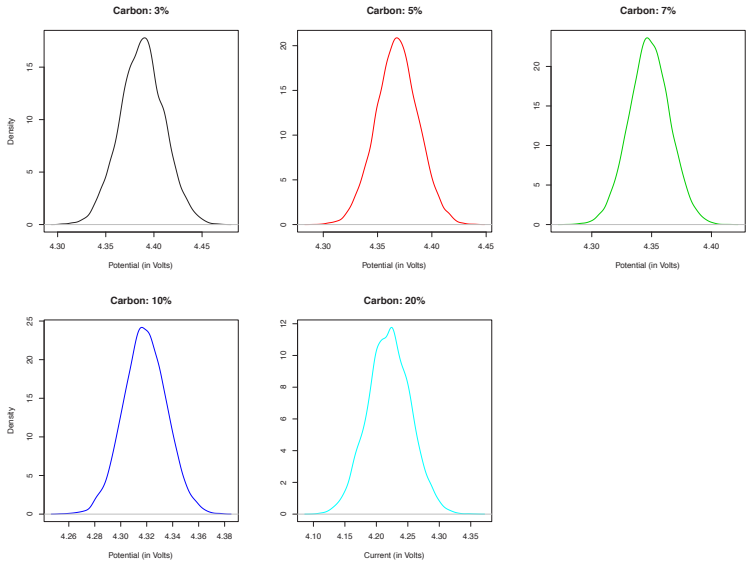


Figure 5.4: Density plots of posterior predictive distributions of Peak Potential for scan rate 2.

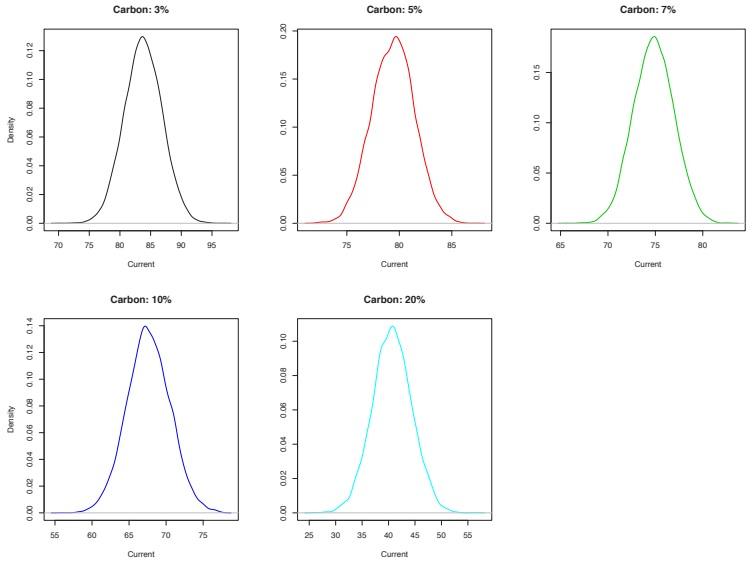


Figure 5.5: Density plots of posterior predictive distributions of Peak Current for scan rate 2.

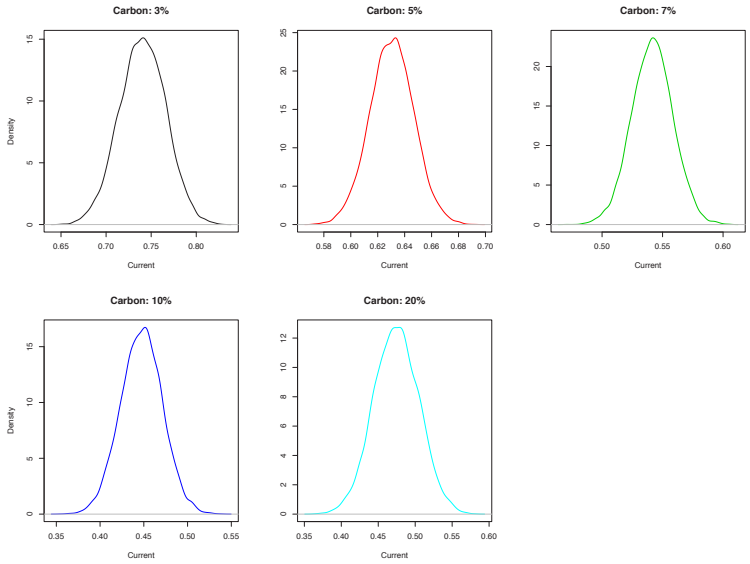


Figure 5.6: Density plots of posterior predictive distributions of Peak Separation in Potential for scan rate 2.

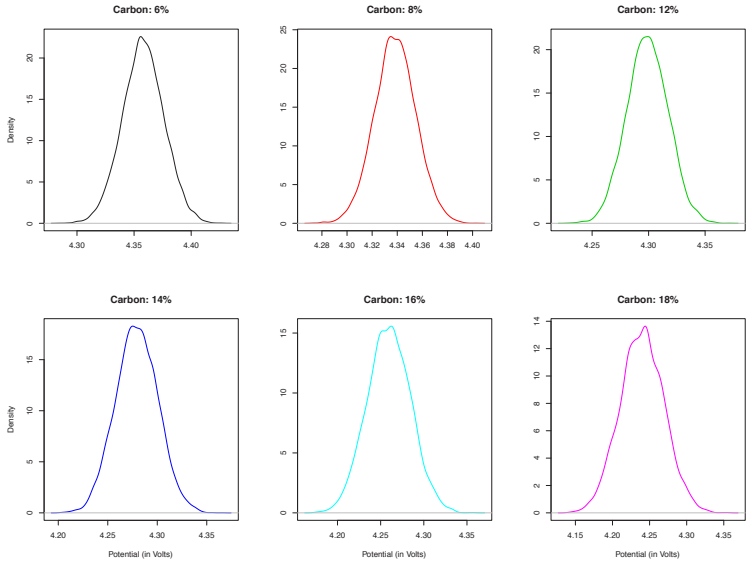


Figure 5.7: Density plots of posterior predictive distributions of Peak Potential for scan rate 2.

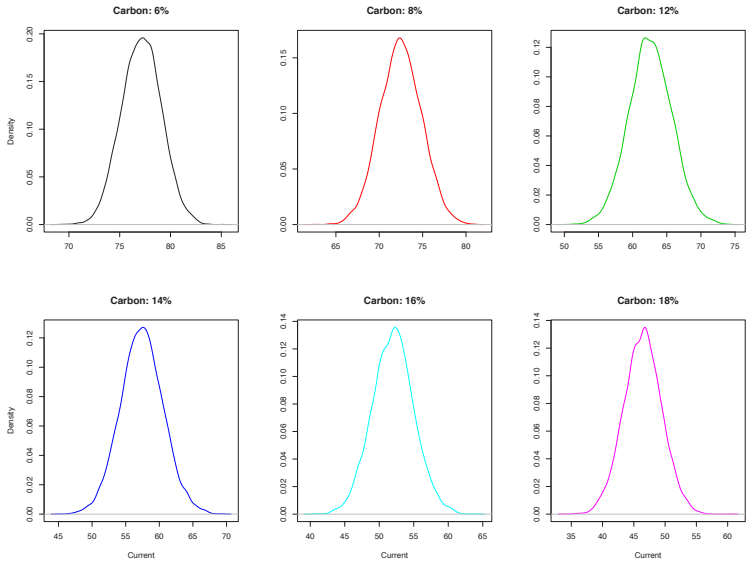


Figure 5.8: Density plots of posterior predictive distributions of Peak Current for scan rate 2.

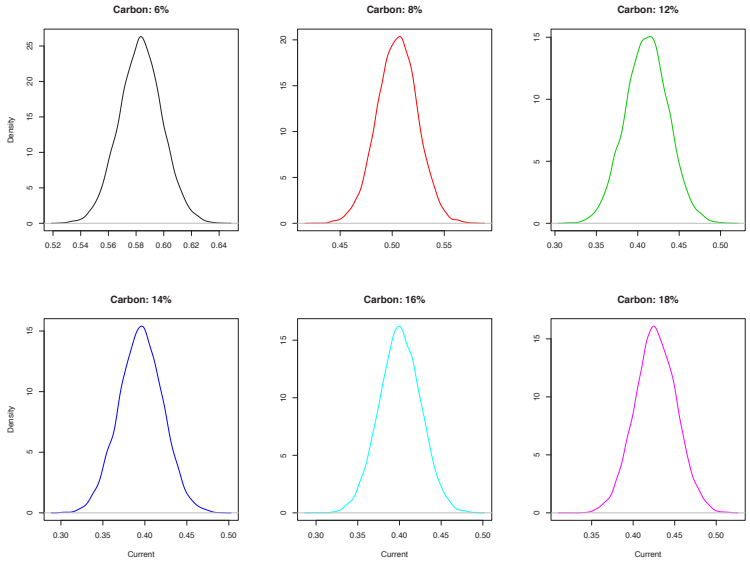


Figure 5.9: Density plots of posterior predictive distributions of Peak Separation in Potential for scan rate 2.

Chapter 6

Autoregressive Models

6.1 Introduction

In statistical modelling, it is desirable to start with the simplest model and then to gradually increase the complexity if appropriate. The first model we will use for the voltammogram is an autoregressive process, as this is the simplest time series model. The complexity of the autoregressive model will be increased by gradually incorporating additional variables into the model. In this chapter, we will develop the autoregressive models under the Bayesian paradigm.

We will derive the full posterior distribution as well as the conditional posterior distribution for each parameter in each model. The conditional posterior distributions will be required to implement the MCMC sampling algorithm. We will concentrate on the aggregated data from derived from array 3 as set out in Chapter 2, although the models developed here can be adjusted for data sets for other replicates and arrays. The output from the MCMC sampling algorithm will be used to make inferences about the characteristics of the Current output curve that are of interest, as developed in Chapter 4.

6.2 Autoregressive Process

An autoregressive model, is very similar to a multiple linear regression model. The difference is that the value we will attempt to predict is regressed on the past values of the data rather than on separate predictor variables. The process $W(t)$ is an autoregressive process of order R (abbreviated to AR(R)) if

$$W(t) = \alpha_1 W(t-1) + \alpha_2 W(t-2) + \ldots + \alpha_R W(t-R) + \varepsilon(t) \tag{6.1}$$

where t is the time index for a generic time series and the process $\varepsilon(t)$ is assumed to be the independently distributed random error. The order of an autoregressive process is often determined using model choice techniques, as discussed in Section 3.7. Other diagnostic methods, such as the autocorrelation function can also be used to determine the order of an autoregressive process, for example see Chatfield (2003, Chapter 4). We will now discuss some basic concepts regarding the AR processes.

6.2.1 Stationarity

From an intuitive point of view, a time series is said to be stationary if there is no systematic change in the mean and the variance over time. More simply, if the properties of one part a stationary time series is identical to any other part of it. Before fitting a time series model, it is often necessary to ensure that it is stationary. There are many techniques that can be used to make a time series stationary such as differencing, which are discussed in Chatfield (2003, Chapter 2).

Formally, there are two types of stationarity: strict stationarity and weak stationarity. A time series is said to be strictly stationary if the joint distribution of $W(t_1), \ldots, W(t_\kappa)$ is the same as the joint distribution of $W(t_1 + \iota), \ldots, W(t_\kappa + \iota)$ for all ι, $t_1, \ldots, t_N$ and $t_1 < t_2 < \ldots t_\kappa$. Hence, shifting the time origin by amount ι has no effect on the joint distributions. In particular, if $\kappa = 1$, strict stationarity implies that for all t, $W(t)$ has the same distribution. Assume that the first

two moments are finite with $E[W(t)] = \mu(t)$ and $\text{Var}(W(t)) = \sigma^2(t)$, then strict stationarity implies that $\mu(t)$ and $\sigma^2(t)$ are both constants independent of t.
To define weak stationarity, we need to define the autocovariance function. If $\{W(t)\}$ is a process where $t = 0, 1, 2, \ldots, N$ such that $\text{Var}(W(t)) < \infty$ for each t, then the autocovariance function of $\{W(t)\}$ is defined by

$$\begin{aligned} \gamma(m,h) &= \text{Cov}(W(m), W(h)) \\ &= E[(W(m) - E[W(m)])(W(h) - E[W(h)])] \end{aligned}$$

where $m, h \in \{0, 1, 2, \ldots, N\}$.
Brockwell and Davis (1991) define weak stationarity as follows. The time series $\{W(t), t \in \mathbb{Z}\}$ where $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, is said to have weak stationarity if :

(i) $E[(W(t))^2] < \infty$ for all $t \in \mathbb{Z}$,

(ii) $E[W(t)] = m$ for all $t \in \mathbb{Z}$ where m is some constant,

(iii) $\gamma(m,h) = \gamma(m+t, h+t)$ for all $m, h \in \mathbb{Z}$.

If the time series $\{W(t)\}$ is weakly stationary, then $\gamma(m-h, 0) = \gamma$ for all $m, h \in \mathbb{Z}$. In this case, it is convenient to redefine the autocovariance function as the function of a single variable, which is given by

$$\begin{aligned} \gamma(\iota) &\equiv \gamma(\iota, 0) \\ &= \text{Cov}(W(t+\iota), W(t)) \quad \text{for all } \iota, t \in \mathbb{Z}. \end{aligned}$$

The function $\gamma(.)$ is often referred to as the autocovariance function of $\{W(t)\}$ and $\gamma(\iota)$ as its value at lag ι. Using this definition of the autocovariance function, the autocorrelation function of $\{W(t)\}$ is defined as the function whose value at lag ι is given by

$$\begin{aligned} \rho(\iota) &\equiv \frac{\gamma(\iota)}{\gamma(0)} \\ &= \text{Corr}(W(t+\iota), W(t)) \quad \text{for all } \iota, t \in \mathbb{Z}. \end{aligned}$$

For more details on stationary time series see Brockwell and Davis (1991, Chapter 1).

6.2.2 Estimating the Parameters of an AR Process

Let $W(t)$ be an AR(R) process with mean μ, then

$$W(t) - \mu = \alpha_1(W(t-1) - \mu) + \ldots + \alpha_R(W(t-R) - \mu) + \varepsilon(t).$$

Given observations $w(1), w(2), \ldots, w(N)$, the parameter estimates $\hat{\mu}, \hat{\alpha_1}, \ldots, \hat{\alpha_R}$ can be obtained by minimising the sum of the squared residuals. A second method is to solve the Yule-Walker equations. There are many sources in the literature that provide a detailed account of the two methods mentioned above, for example, see Chatfield (2003, Chapter 3) and Brockwell and Davis (1991, Chapter 8). Another method of parameter estimation is to use MCMC sampling methods. These are discussed in detail in Section 3.4.

6.2.3 Determining the Order of an AR Process

One method that can be used to determine the order of an AR process is to use the sample autocorrelation function. The sample autocorrelation coefficient at lag d is given by

$$r_d = \frac{\sum_{t=1}^{N-d} (w(t) - \bar{w})\,(w(t+k) - \bar{w})}{\sum_{t=1}^{N} (w(t) - \bar{w})}$$

where N is the length of the time series and $\bar{w} = \frac{1}{N}\sum_{t=1}^{N} w(t)$.
For a first order autoregressive process, AR(1), it can be shown that $\rho(\iota) = \alpha_1^{\iota}$, see Chatfield (2003, pp. 41-42). Thus, the theoretical autocorrelation decreases exponentially and the sample autocorrelation should also follow a similar pattern. For higher order autoregressive processes the autocorrelation function is a mixture

of damped exponential or sinusoidal functions. There are more detailed accounts in the literature on using the autocorrelation function to determine the order of an AR process, for example, see Chatfield (2003, Chapter 4).

As with estimating the parameters, there are also Bayesian methods for dealing with the assessment of the order of the AR process. Various Bayesian model assessment tools can be used to help estimate the order of an AR processes. One such method is called the predictive model choice criteria (PMCC) as discussed in Section 3.7. The model with the lowest PMCC value should be the best model from the set of models under consideration.

6.3 Autoregressive Models

The first model we will fit to the aggregated data from array 3 will be an autoregressive process of order R_α. For the remainder of this chapter, we will denote the vector of parameters by the generic symbol $\boldsymbol{\theta}$, t will denote the time index as defined in Table 2.1 and N will denote the length of the time series, depending on the scan rate. For ease of notation, we drop the subscript, s, which we have used to denote the scan rate. The model is given by

$$x_k(t) = \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) + \varepsilon_k(t) \tag{6.2}$$

where $t = R_\alpha + 1, R_\alpha + 2, \ldots, N$ and $k = 1, 2, \ldots, K$. The likelihood for the model in Equation (6.2) is given by

$$f(\boldsymbol{x}|\boldsymbol{\theta}) \propto (\tau^2)^{\frac{N-R_\alpha}{2}} \exp\left(-\frac{\tau^2}{2} \sum_{k=1}^{K} \sum_{t=R_\alpha+1}^{N} (\varepsilon_k(t))^2 \right)$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \tau^2)^T$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{R_\alpha})^T$ and $\tau^2 = \sigma^{-2}$. The Bayesian model is completed by assuming the prior distributions

$$\alpha_i \sim N(0, v_i^2) \quad \text{and} \quad \tau^2 \sim \text{Gamma}(d_1, d_2), \tag{6.3}$$

where v_i, d_1, d_2 are constants to be chosen. The joint posterior distribution is given by

$$f(\boldsymbol{\theta} \mid \boldsymbol{x}) \propto (\tau^2)^{\psi_2 - 1} \exp\left(-\frac{1}{2}\tau^2 \sum_{k=1}^{K} \sum_{t=R_\alpha+1}^{N} (\varepsilon_k(t))^2 - \frac{1}{2}\boldsymbol{\alpha}^T V_\alpha^{-1} \boldsymbol{\alpha} - d_2\tau^2 \right)$$

where $V_\alpha = \text{diag}(v_1^2, v_2^2, \ldots, v_{R_\alpha}^2)$ and $\psi_1 = \frac{1}{2}(N - R_\alpha) + d_1$. The conditional posterior distributions are obtained from the full posterior distribution by finding the distribution of each parameter given the remaining parameters and data.

6.3.1 Incorporating Potential in the Model

To improve the accuracy of the model in Section 6.3, factors that affect the response need to be incorporated in the model. The simplest factor to incorporate in the model is the Potential via a polynomial in Potential. Incorporating Potential in this way may seem inappropriate, however, it is the first simplest form we have tried. We realise the limitations of this model and propose a more flexible autoregressive model in Section 6.4 and that has been shown to be superior for prediction. We return our attention to the model incorporating Potential via a polynomial which is given by

$$x_k(t) = \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) + \sum_{j_\eta=1}^{R_\eta} \eta_{j_\eta} (p(t))^{j_\eta} + \varepsilon_k(t) \tag{6.4}$$

where $t = R_\alpha + 1, R_\alpha + 2, \ldots, N$, and $p(t)$ is the Potential at time t. The likelihood function of the parameters for this model is given by

$$f(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{p}) \propto \left(\tau^2\right)^{\frac{N-R_\alpha}{2}} \exp\left(-\frac{1}{2}\tau^2 \sum_{k=1}^{K} \sum_{t=R_\alpha+1}^{N} \left(\varepsilon_k(t)\right)^2 \right) \tag{6.5}$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\eta}, \tau^2)^T$, $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_{R_\eta})^T$. we assign non-informative prior distributions to η_{j_η}, $\eta_{j_\eta} \sim N(0, w_{j_\eta}^2)$. Further to this, we note that v_i^2, $w_{j_\eta}^2$, d_1 and d_2 are constants to be chosen. The full posterior distribution is simply the product

of the prior distributions and the likelihood shown in Equation (6.5) which is given by

$$f(\boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{p}) \propto \left(\tau^2\right)^{\psi_1 - 1} \exp\left(-\frac{1}{2}\tau^2 \sum_{k=1}^{K} \sum_{t=R_{\alpha+1}}^{N} \left(\varepsilon_k(t)\right)^2 - \frac{1}{2}\boldsymbol{\alpha}^T V_\alpha^{-1} \boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\eta}^T V_\eta^{-1} \boldsymbol{\eta} - d_2\tau^2 \right) \tag{6.6}$$

where $V_\eta = (w_1^2, w_2^2, \ldots, w_{R_\eta}^2)$ and $\psi_1 = \frac{1}{2}(N - R_\alpha) + d_1$.

It would be possible to use fractional polynomials, that is a polynomial function where the powers are rational numbers rather than just integers. This could be costly exploration as we have very little information as to what sort of fractional powers to use. In addition, a model that incorporates the Potential as a second autoregressive process is likely to be a better model. This is because the value of the Current at time t will be influenced by previous values of the Potential (see Section 6.4 for further details).

6.3.2 Incorporating Potential and Carbon in the Model

To further improve the accuracy of the model we will now deal with the effect of Carbon. A polynomial in Carbon will be used to account for the effect of Carbon. The modified model is given by

$$x_k(t) = \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) + \sum_{j_\eta=1}^{R_\eta} \eta_{j_\eta} (p(t))^{j_\eta} + \sum_{q=1}^{R_\gamma} \gamma_q c_k^q + \varepsilon_k(t) \tag{6.7}$$

where $t = R_\alpha + 1, R_\alpha + 2, \ldots, N$, c_k is the kth level of Carbon and $x_k(t)$ is the Current for Carbon level k at time t. As before, the $\{\varepsilon_k(t)\}$ are a sequence (over time t) of independent and identically distributed Gaussian random variables with mean 0 and variance σ^2. We can clearly obtain the previous models from the general model shown in Equation (6.7) by simply setting the relevant parameters to zero.

The likelihood is given by

$$f(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{c}) \propto (\tau^2)^{\psi_1 - 1} \exp\left(-\frac{1}{2}\tau^2 \sum_{k=1}^{K} \sum_{t=R_\alpha+1}^{N} (\varepsilon_k(t))^2 \right)$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \tau^2)^T$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_{R_\gamma})^T$. To complete the Bayesian model, we use non-informative prior distributions (as stated in Section 6.3.1), together with $\gamma_q \sim N(0, g_q^2)$ where g_q are constants to be chosen and $q = 1, 2, \ldots R_\gamma$. The full posterior distribution is obtained by taking the product of the likelihood and the prior is given by

$$\begin{aligned} f(\boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{p}, \boldsymbol{c}) \quad \propto \quad & (\tau^2)^{\psi_1 - 1} \exp\left(-\frac{1}{2}\tau^2 \sum_{k=1}^{K} \sum_{t=R_\alpha+1}^{N} (\varepsilon_k(t))^2 \right. \\ & \left. -\frac{1}{2}\boldsymbol{\alpha}^T V_\alpha^{-1} \boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\eta}^T V_\eta^{-1} \boldsymbol{\eta} - \frac{1}{2}\boldsymbol{\gamma}^T (V_\gamma)^{-1} \boldsymbol{\gamma} - d_2 \tau^2 \right) \end{aligned}$$

where $V_\gamma = \mathrm{diag}(g_1^2, g_2^2, \ldots, g_{R_\gamma}^2)$ and $\boldsymbol{c} = (c_1, c_2, \ldots, c_K)^T$.
The conditional posterior distributions are given by:

$$\begin{aligned} \boldsymbol{\alpha} \mid \boldsymbol{\eta}, \boldsymbol{\gamma}, \tau^2, X, \boldsymbol{p}, \boldsymbol{c} \quad &\sim \quad N_{R_\alpha}(\boldsymbol{\mu}_\alpha, \Sigma_\alpha), \\ \boldsymbol{\eta} \mid \boldsymbol{\alpha}, \boldsymbol{\gamma}, \tau^2, X, \boldsymbol{p}, \boldsymbol{c} \quad &\sim \quad N_{R_\eta}(\boldsymbol{\mu}_\eta, \Sigma_\eta), \\ \boldsymbol{\gamma} \mid \boldsymbol{\alpha}, \boldsymbol{\eta}, \tau^2, X, \boldsymbol{p}, \boldsymbol{c} \quad &\sim \quad N_{R_\gamma}(\boldsymbol{\mu}_\gamma, \Sigma_\gamma), \\ \tau^2 \mid \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\gamma}, X, \boldsymbol{p}, \boldsymbol{c} \quad &\sim \quad \mathrm{Gamma}(\psi_1, \varphi), \end{aligned}$$

where:

$$\begin{aligned} \boldsymbol{\mu}_\alpha \quad &= \quad \tau^2 \Sigma_\alpha X^T \boldsymbol{Y}_\alpha, \ \Sigma_\alpha = \tau^{-2}(X^T X + \tau^{-2} V_\alpha^{-1})^{-1}, \\ \boldsymbol{Y}_\alpha \quad &= \quad (y_{\alpha,1}(R_\alpha + 1), \ldots, y_{\alpha,1}(N), y_{\alpha,2}(R_\alpha + 1), \ldots, y_{\alpha,2}(N), \ldots, \\ & \quad y_{\alpha,K}(R_\alpha + 1), \ldots, y_{\alpha,K}(N))^T, \\ y_{\alpha,k}(t) \quad &= \quad x_k(t) - \sum_{j_\eta=1}^{R_\eta} \eta_{j_\eta} (p(t))^{j_\eta} - \sum_{q=1}^{R_\gamma} \gamma_q (c_k)^q, \end{aligned}$$

$$
\begin{aligned}
X &= \begin{pmatrix}
x_1(R_\alpha) & x_1(R_\alpha - 1) & \dots & x_1(1) \\
\vdots & \vdots & \vdots & \vdots \\
x_1(N-1) & x_1(N-2) & \dots & x_1(N-R_\alpha) \\
x_2(R_\alpha) & x_2(R_\alpha - 1) & \dots & x_2(1) \\
\vdots & \vdots & \vdots & \vdots \\
x_2(N-1) & x_2(N-2) & \dots & x_2(N-R_\alpha) \\
\vdots & \vdots & \vdots & \vdots \\
x_K(R_\alpha) & x_K(R_\alpha - 1) & \dots & x_K(1) \\
\vdots & \vdots & \vdots & \vdots \\
x_K(N-1) & x_K(N-2) & \dots & x_K(N-R_\alpha)
\end{pmatrix}, \\
\boldsymbol{\mu}_\eta &= \tau^2 \Sigma_\eta P^T Y_\eta,\ \Sigma_\eta = (\tau^2 P^T P + V_\eta^{-1})^{-1}, \\
\boldsymbol{Y}_\eta &= (y_{\eta,1}(R_\alpha+1), \dots, y_{\eta,1}(N), y_{\eta,2}(R_\alpha+1), \dots, y_{\eta,2}(N), \dots, \\
&\quad y_{\eta,K}(R_\alpha+1), \dots, y_{\eta,K}(N))^T, \\
y_{\eta,k}(t) &= x_k(t) - \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) - \sum_{q=1}^{R_\gamma} \gamma_q c_k^q, \\
\boldsymbol{\mu}_\gamma &= \tau^2 \Sigma_\gamma C^T Y_\gamma,\ \Sigma_\gamma = (\tau^2 C^T C + (V_{R_\gamma})^{-1})^{-1}, \\
\boldsymbol{Y}_\gamma &= (y_{\gamma,1}(R_\alpha+1), \dots, y_{\gamma,1}(N), y_{\gamma,2}(R_\alpha+1), \dots, y_{\gamma,2}(N), \dots, \\
&\quad y_{\gamma,K}(R_\alpha+1), \dots, y_{\gamma,K}(N))^T, \\
y_{\gamma,k}(t) &= x_k(t) - \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) - \sum_{j_\eta=1}^{R_\eta} \eta_{j_\eta} (p(t))^{j_\eta}, \\
P &= \begin{pmatrix}
p(R_\alpha+1) & (p(R_\alpha+1))^2 & \dots & (p(R_\alpha+1))^{R_\eta}, \\
\vdots & \vdots & \vdots & \vdots \\
p(N) & (p(N))^2 & \dots & (p(N))^{R_\eta} \\
\vdots & \vdots & \vdots & \vdots \\
p(R_\alpha+1) & (p(R_\alpha+1))^2 & \dots & (p(R_\alpha+1))^{R_\eta} \\
\vdots & \vdots & \vdots & \vdots \\
p(N) & (p(N))^2 & \dots & (p(N))^{R_\eta}
\end{pmatrix},
\end{aligned}
$$

$$
\begin{aligned}
C &= \begin{pmatrix}
c_1 & c_1^2 & \dots & c_1^{R_\gamma} \\
\vdots & \vdots & \vdots & \vdots \\
c_1 & c_1^2 & \dots & c_1^{R_\gamma} \\
c_2 & c_2^2 & \dots & c_2^{R_\gamma} \\
\vdots & \vdots & \vdots & \vdots \\
c_K & c_K^2 & \dots & c_K^{R_\gamma} \\
\vdots & \vdots & \vdots & \vdots \\
c_K & c_K^2 & \dots & c_K^{R_\gamma}
\end{pmatrix}, \\
\varphi &= \frac{1}{2}\sum_{k=1}^{K}\sum_{t=R_\alpha+1}^{N}(\varepsilon_k(t))^2 + d_2.
\end{aligned}
$$

6.4 Models Using Two Autoregressive Processes

A more realistic method is to incorporate the Potential using an autoregressive process as past values of the Potential have an effect on the present value of the Current. In this case the model is given by

$$x_k(t) = \sum_{i=1}^{R_\alpha}\alpha_i x_k(t-i) + \sum_{q=0}^{R_\beta}\beta_j p(t-j) + \sum_{q=1}^{R_\gamma}\gamma_q c_k^q + \varepsilon_k(t)$$

where $t \geq \max(R_\alpha + 1, R_\beta + 1)$ and as noted earlier, $\{\varepsilon_k(t)\}$ denotes a sequence of independent and identically distributed Gaussian random variables with mean 0 and variance σ^2. The likelihood is given by

$$f(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{c}) \propto (\tau^2)^{\frac{-K(n-r)}{2}} \exp\left(-\frac{1}{2}\tau^2\sum_{k=1}^{K}\sum_{t=r'+1}^{N}(\varepsilon_k(t))^2\right)$$

where $r' = \max(R_\alpha, R_\beta)$, $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \tau^2)^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_{R_\beta})^T$ and

$$\varepsilon_k(t) = x_k(t) - \left(\sum_{i=1}^{R_\alpha}\alpha_i x_k(t-i) + \sum_{j=0}^{R_\beta}\beta_j p(t-j) + \sum_{q=1}^{R_\gamma}\gamma_q c_k^q\right).$$

Assign non-informative prior distributions to the parameters, where v_i, w_j, g_q are the variances for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ respectively. The full posterior distribution obtained by taking the product of the likelihood and the prior is given by

$$\begin{aligned} f(\boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{p}, \boldsymbol{c}) \;\; \propto \;\; (\tau^2)^{\psi_2 - 1} \exp\Bigg(& -\frac{1}{2}\tau^2 \sum_{k=1}^{K} \sum_{t=r'+1}^{N} (\varepsilon_k(t))^2 - \frac{1}{2}\boldsymbol{\alpha}^T V_\alpha^{-1} \boldsymbol{\alpha} \\ & -\frac{1}{2}\boldsymbol{\beta}^T V_\beta^{-1} \boldsymbol{\beta} - \frac{1}{2}\boldsymbol{\gamma}^T V_\gamma^{-1} \boldsymbol{\gamma} - d_2 \tau^2 \Bigg) \end{aligned}$$

where $\psi_2 = \frac{1}{2}K(N - r') + d_1$.

The conditional posterior distributions are given by:

$$\begin{aligned} \boldsymbol{\alpha} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \tau^2, X, \boldsymbol{p}, \boldsymbol{c} &\sim N_{R_\alpha}(\boldsymbol{\mu}_\alpha, \Sigma_\alpha), \\ \boldsymbol{\beta} \mid \boldsymbol{\alpha}, \boldsymbol{\gamma}, \tau^2, X, \boldsymbol{p}, \boldsymbol{c} &\sim N_{R_\beta}(\boldsymbol{\mu}_\beta, \Sigma_\beta), \\ \boldsymbol{\gamma} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \tau^2, X, \boldsymbol{p}, \boldsymbol{c} &\sim N_{R_\gamma}(\boldsymbol{\mu}_\gamma, \Sigma_\gamma), \\ \tau^2 \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, X, \boldsymbol{p}, \boldsymbol{c} &\sim \text{Gamma}(\psi_2, \varphi), \end{aligned}$$

where:

$$\begin{aligned} \boldsymbol{\mu}_\alpha &= \tau^2 \Sigma_\alpha X^T \boldsymbol{Y}_\alpha, \; \Sigma_\alpha = \tau^{-2}(X^T X + V_\alpha^{-1})^{-1}, \\ \boldsymbol{Y}_\alpha &= (y_{\alpha,1}(r'+1), \ldots, y_{\alpha,1}(N, y_{\alpha,2}(r'+1), \ldots, y_{\alpha,2}(N), \ldots, \\ &\quad y_{\alpha,K}(r'+1), \ldots, y_{\alpha,K}(N))^T, \\ y_{\alpha,k}(t) &= x_k(t) - \sum_{j=0}^{R_\beta} \beta_j p(t-j) - \sum_{q=1}^{R_\gamma} \gamma_q (c_k)^q, \end{aligned}$$

$$
X = \begin{pmatrix}
x_1(r') & x_1(r'-1) & \dots & x_1(r'-R_\alpha+1) \\
\vdots & \vdots & \vdots & \vdots \\
x_1(N-1) & x_1(N-2) & \dots & x_1(N-R_\alpha+1) \\
x_2(r') & x_2(r'-1) & \dots & x_2(r'-R_\alpha+1) \\
\vdots & \vdots & \vdots & \vdots \\
x_2(N-1) & x_2(N-2) & \dots & x_2(N-R_\alpha+1) \\
\vdots & \vdots & \vdots & \vdots \\
x_K(r') & x_K(r'-1) & \dots & x_K(r'-R_\alpha+1) \\
\vdots & \vdots & \vdots & \vdots \\
x_K(N-1) & x_K(N-2) & \dots & x_K(N-R_\alpha+1)
\end{pmatrix},
$$

$$
\begin{aligned}
\boldsymbol{\mu}_\beta &= \tau^2 \Sigma_\beta P^T Y_\beta,\ \Sigma_\beta = (\tau^2 P^T P + V_\beta^{-1})^{-1}, \\
\boldsymbol{Y}_\beta &= (y_{\beta,1}(r'+1), \dots, y_{\beta,1}(N), y_{\beta,2}(r'+1), \dots, y_{\beta,2}(N), \dots, \\
&\quad y_{\beta,K}(r'+1), \dots, y_{\beta,K}(N))^T, \\
y_{\beta,k}(t) &= x_k(t) - \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) - \sum_{q=1}^{R_\gamma} \gamma_q c_k^q,
\end{aligned}
$$

$$
P = \begin{pmatrix}
p(r'+1) & p(r') & p(r'-1) & \dots & p(r'-R_\beta+1) \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
p(N) & p(N-1) & p(N-2) & \dots & p(N-R_\beta+1) \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
p(r'+1) & p(r') & p(r'-1) & \dots & p(r'-R_\beta+1) \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
p(N) & p(N-1) & p(N-2) & \dots & p(N-R_\beta+1)
\end{pmatrix},
$$

$$
\begin{aligned}
\boldsymbol{\mu}_\gamma &= \tau^2 \Sigma_\gamma C^T Y_\gamma,\ \Sigma_\gamma = (\tau^2 C^T C + V_\gamma^{-1})^{-1}, \\
\boldsymbol{Y}_\gamma &= (y_{\gamma,1}(r'+1), \dots, y_{\gamma,1}(N), y_{\gamma,2}(r'+1), \dots, y_{\gamma,2}(N), \dots, \\
&\quad y_{\gamma,K}(r'+1), \dots, y_{\gamma,K}(N))^T,
\end{aligned}
$$

$$
\begin{aligned}
y_{\gamma,k}(t) &= x_k(t) - \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) - \sum_{j=1}^{R_\beta} \beta_j p(t-j), \\
\varphi &= \frac{1}{2} \sum_{k=1}^{K} \sum_{t=r'+1}^{N} (\varepsilon_k(t))^2 + d_2.
\end{aligned}
$$

6.5 Analysis and Conclusions

Equation (6.8), see below, defines the way in which we will refer to each model and to which models we are restricting ourselves to. We note that in trying to develop a model in applied problems, it is possible to continually adjust models to obtain a better model. Hence, it is necessary to restrict the models that we are considering in some way. Further, the models we have chosen cover the model space developed in this chapter quite extensively with relatively little increase in the complexity of the analysis. We have chosen to purposely include Carbon as the chemists, based on their experience, believe that Carbon has a fundamental effect on the Current output.

$$
\begin{aligned}
M_1(R_\alpha, R_\gamma) &: x_k(t) = \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) + \sum_{q=1}^{R_\gamma} \gamma_q c_k^q + \varepsilon_k(t) \\
M_2(R_\alpha, R_\eta, R_\gamma) &: x_k(t) = \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) + \sum_{j_\eta=1}^{R_\eta} \eta_{j_\eta} (p(t))^{j_\eta} + \sum_{q=1}^{R_\gamma} \gamma_q c_k^q + \varepsilon_k(t) \\
M_3(R_\alpha, R_\beta, R_\gamma) &: x_k(t) = \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) + \sum_{j=0}^{R_\beta} \beta_j p(t-j) + \sum_{q=1}^{R_\gamma} \gamma_q c_k^q + \varepsilon_k(t)
\end{aligned}
\tag{6.8}
$$

The values for R_α, R_β (or R_η) and R_γ shown in Tables 6.1, 6.2 and 6.3 were chosen by cycling around each variable in a nested loop structure. Hence, a thorough model search was conducted for various combinations of R_α, R_β (or R_η) and R_γ. The combination of values shown in Tables 6.1, 6.2 and 6.3 were those that produced the

lowest PMCC values. It can be observed from these tables that for each scan rate, the 3 models presented have similar performances, under PMCC.

Model Specification	Goodness of fit	Penalty	PMCC
$M_1(R_\alpha = 30, R_\gamma = 1)$	194.4657	195.0966	389.5623
$M_2(R_\alpha = 34, R_\eta = 3, R_\gamma = 1)$	197.5542	192.6735	390.2277
$M_3(R_\alpha = 25, R_\beta = 20, R_\gamma = 1)$	193.8198	195.9621	389.7819

Table 6.1: Predictive model choice criteria for selected models for scan rate 1.

Model Specification	Goodness of fit	Penalty	PMCC
$M_1(R_\alpha = 11, R_\gamma = 1)$	38.8481	40.7423	79.5903
$M_2(R_\alpha = 11, R_\eta = 3, R_\gamma = 2)$	38.5283	45.2663	83.7945
$M_3(R_\alpha = 11, R_\beta = 11, R_\gamma = 1)$	36.3307	38.8855	75.2162

Table 6.2: Predictive model choice criteria for selected models for scan rate 2.

Model Specification	Goodness of fit	Penalty	PMCC
$M_1(R_\alpha = 12, R_\gamma = 2)$	217.7710	220.2873	438.0583
$M_2(R_\alpha = 13, R_\eta = 3, R_\gamma = 1)$	214.9017	223.8122	438.7138
$M_3(R_\alpha = 12, R_\beta = 11, R_\gamma = 1)$	170.5442	30.1754	351.5307

Table 6.3: Predictive model choice criteria for selected models for scan rate 3.

To complete the rest of our analysis for scan rate 2, we use model $M_3(R_\alpha = 11, R_\beta = 11, R_\gamma = 1)$ which has PMCC value at least 5.9% smaller than the values of the

other two models, see Table 6.2. The diagnostic plots for this model is shown in Figures 6.1 and 6.2. These plots indicate that the MCMC algorithm is sampling accurately from the posterior distribution. The algorithm appears to be covering the full space for each of the parameters. Figure 6.2 shows that there is relatively little autocorrelation present in the sample.

The parameter estimates for the chosen model were fairly robust when we varied the prior variances for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and the hyperparameters d_1, d_2 for τ^2, see Table 6.4. We also calculated the standard deviations of the posterior distribution of the parameters. These turned out to be similar to each other and are omitted for brevity. This study shows that the choice of the prior variance has little effect on the parameter estimates, as also seen in Section 5.3. As in Section 5.3, we set the value for each of the prior variances to 10^8 and the values of d_1 and d_2 to 0.001.

We now examine the residuals which were calculated by working out the difference between the one-step ahead predictions and the observed values shown in Figures 2.10 and 2.11. This is a standard technique used in time series analysis, which is also discussed in Section 3.8. The residual plot shown in Figure 6.3 indicates that the variance of the residuals is not constant. The behaviour of the residuals becomes much more varied around the times when the Current is near its peak value. This pattern repeats itself when the Current is near its minimum value. As we are interested in predicting various characteristics about the peak and minimum points of the Current output, it is necessary to improve the model. We note that, although we have only shown the plots from scan rate 2 (to illustrate our findings), the corresponding plots for scan rates 1 and 3 exhibited similar behaviour.

The location of the distributions shown in Figure 6.4 appears to be near the value of the Potential that we would expect, see Figure 2.11. This is also true for the distributions of the Peak Separation in Potential, see Figure 6.6.

As noted in Section 5.3, we used the plot and density commands in R to plot the

posterior predictive distribution densities shown in Figures 6.4-6.6. This is also the case for the other posterior predictive distribution densities shown in Figures 6.7-6.9.

We next use the model to predict Peak Potential, Peak Current and Peak Separation in Potential at six values of Carbon within the range of 3-20% which were not run in the experiment, namely 6%, 8%, 12%, 14%, 16%, 18% and 20%. The posterior predictive distributions for the Peak Potential, Peak Current and Peak Separation in Potential for these levels of Carbon are displayed in Figures 6.7, 6.8 and 6.9, respectively. We are unable to compare these posterior predictive distributions to observed data. Hence, we need to use the chemists' knowledge of these experiments to analyse the posterior predictive distributions in Figures 6.7, 6.8 and 6.9. In addition to this, as noted in Section 5.3, by including the posterior predictive distributions shown in the aforementioned figures, we are illustrating the predictive usefulness of the best model proposed in this chapter.

Based on the chemists' feedback, we realised that the AR model is rather poor at predicting the Peak Potential. Note that the posterior predictive distribution is multi-modal (see Figure 6.7) and, guided by the chemists, we expected the mode of this distribution to be near the second mode. This multi-model nature of the posterior predictive distribution for the Peak Potential will, in turn, affect the posterior predictive distribution for the Peak Width in Potential. As the model appears to be rather poor at predicting the various characteristics involving Potential, it is not surprising that the same behaviour was found with respect to the posterior predictive distributions for Peak Time and Peak Separation in Time. We re-iterate that this is due to Potential and time are just different ways of referring to a particular observation.

Parameter	Prior Variances		
	10^8	10^{10}	10^{12}
α_1	1.5663	1.5659	1.5659
α_2	0.0774	0.0784	0.0784
α_3	-0.5127	-0.5134	-0.5134
α_4	-0.1492	-0.1489	-0.1489
α_5	-0.2032	-0.2042	-0.2042
α_6	0.1107	0.1112	0.1112
α_7	-0.0195	-0.0189	-0.0189
α_8	0.1780	0.1776	0.1776
α_9	-0.0537	-0.0528	-0.0528
α_{10}	0.0757	0.0742	0.0742
α_{11}	-0.0705	-0.0699	-0.0699
β_0	33.0286	33.1308	33.1308
β_1	-59.9870	-60.2273	-60.2273
β_2	-0.7939	-0.6421	-0.6421
β_3	16.2813	16.2315	16.2315
β_4	21.0300	21.0805	21.0805
β_5	11.4082	11.4892	11.4892
β_6	-11.2009	-11.1565	-11.1565
β_7	-8.6967	-8.9484	-8.9484
β_8	3.5305	3.4945	3.4945
β_9	-17.9063	-17.6956	-17.6956
β_{10}	16.1107	16.0585	16.0585
β_{11}	-2.8047	-2.8155	-2.8155
γ_1	0.0011	0.0024	0.0024
σ^2	0.0152	0.0152	0.0152

Table 6.4: Parameter estimates (posterior mean) for different prior variances for the best autoregressive model for scan rate 2.

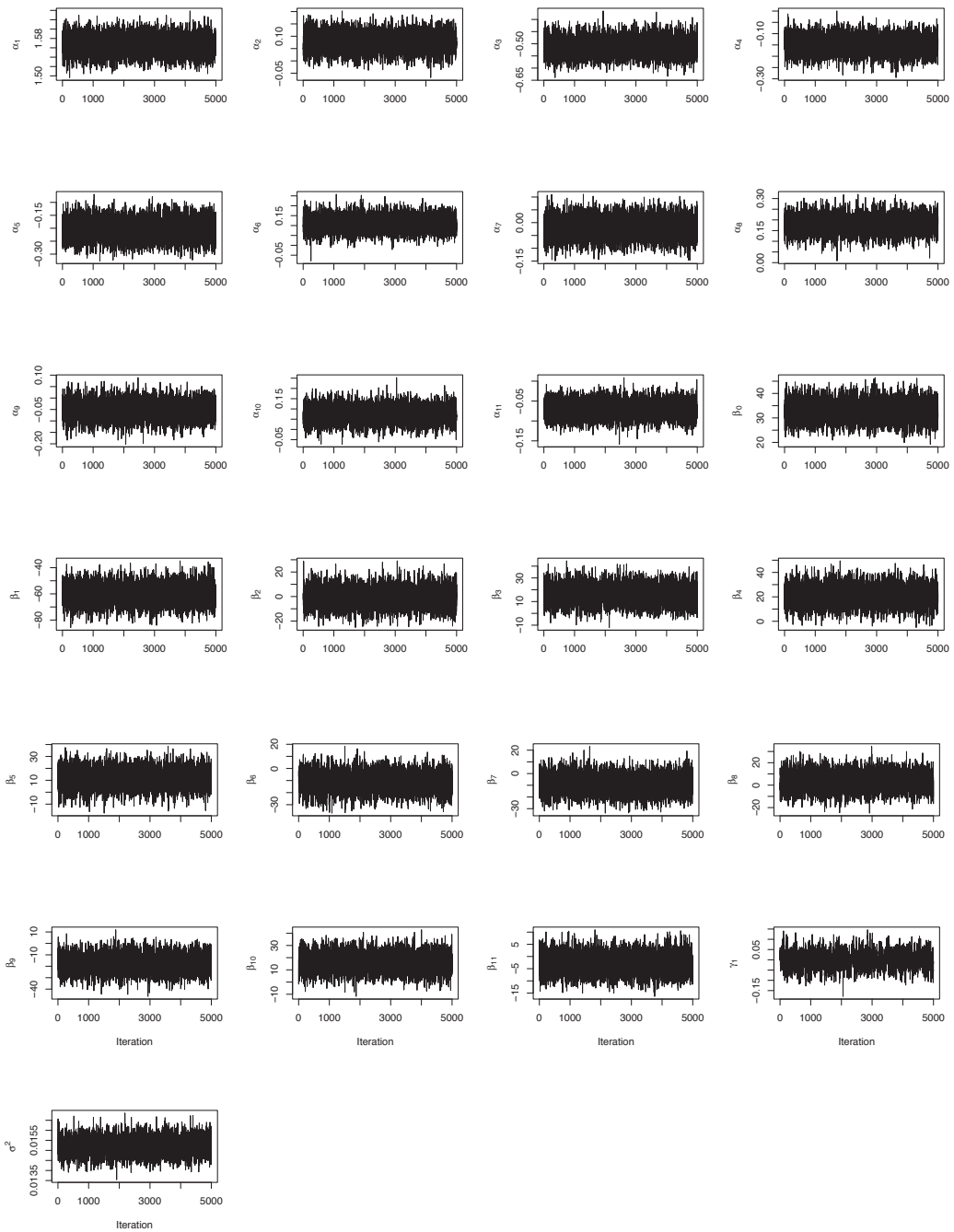


Figure 6.1: Trace plots generated by the Gibbs sampler for all the parameters of the best autoregressive model for scan rate 2.

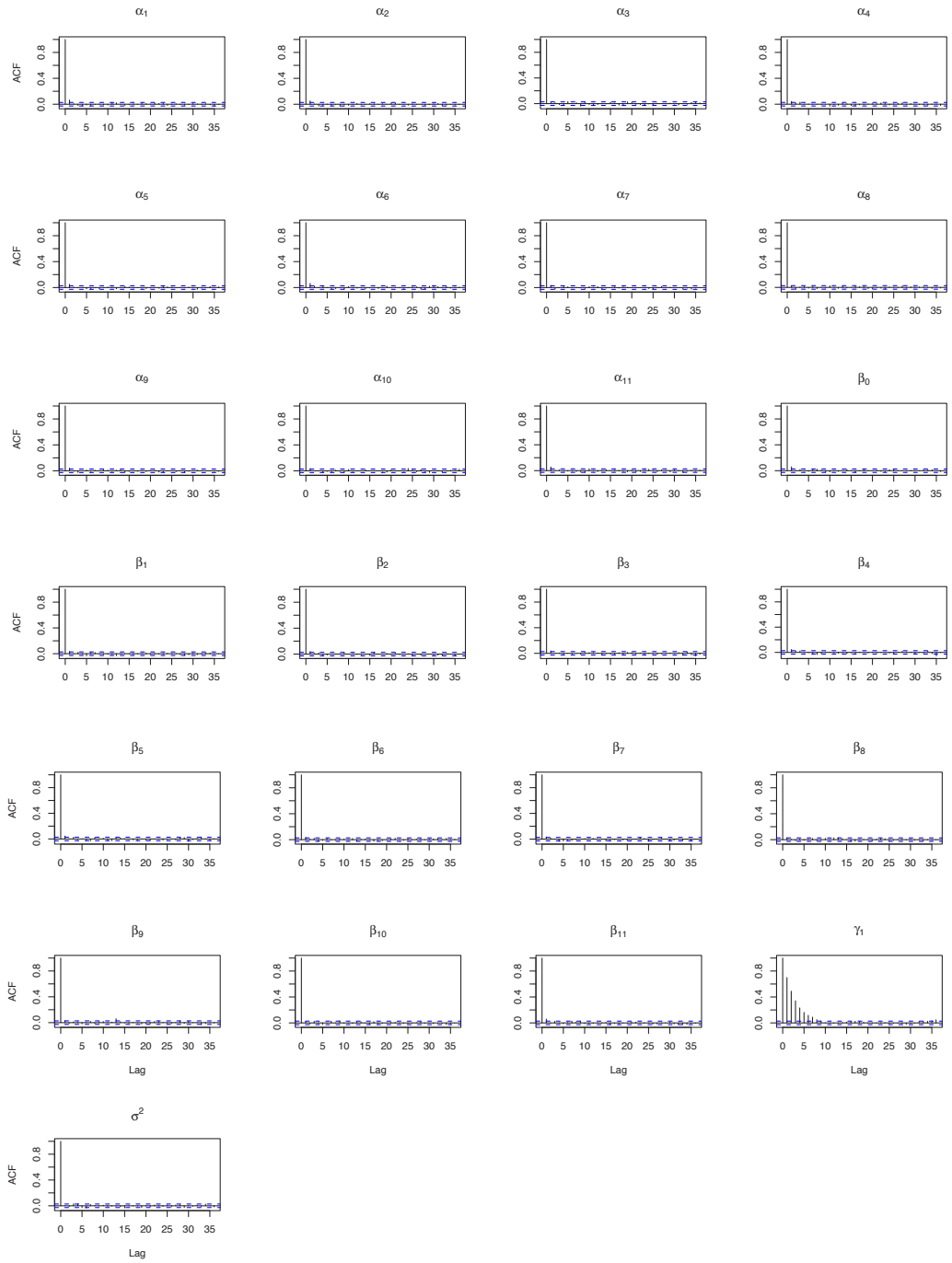


Figure 6.2: Autocorrelation plots generated by the Gibbs sampler for all the parameters of the best autoregressive model for scan rate 2.

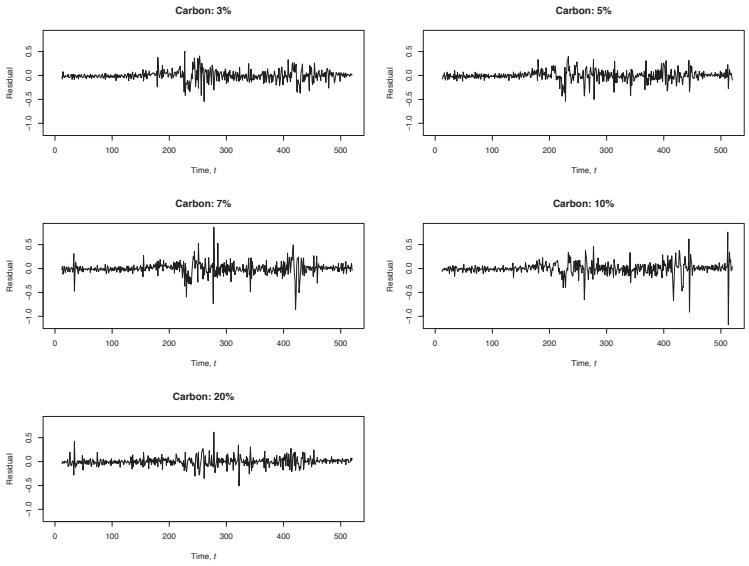


Figure 6.3: Differences between data and one-step ahead predictions, where t denotes the time index defined in Table 2.1.

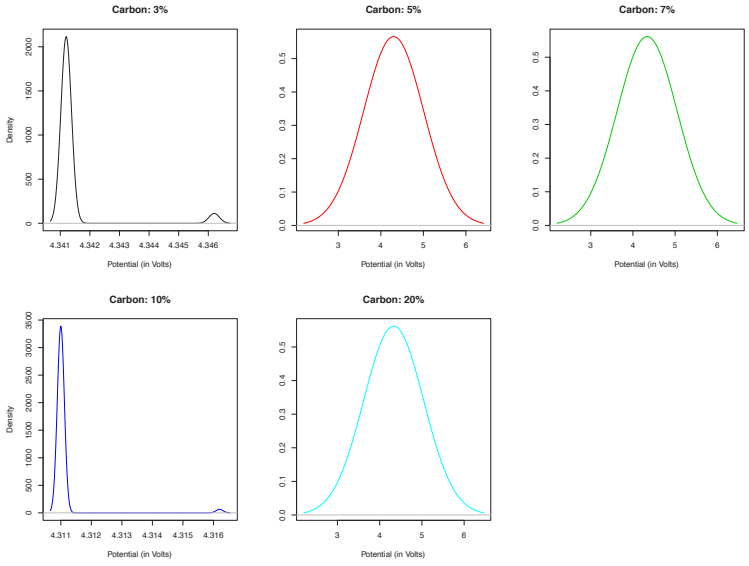


Figure 6.4: Density plots of posterior predictive distributions of Peak Potential for scan rate 2.

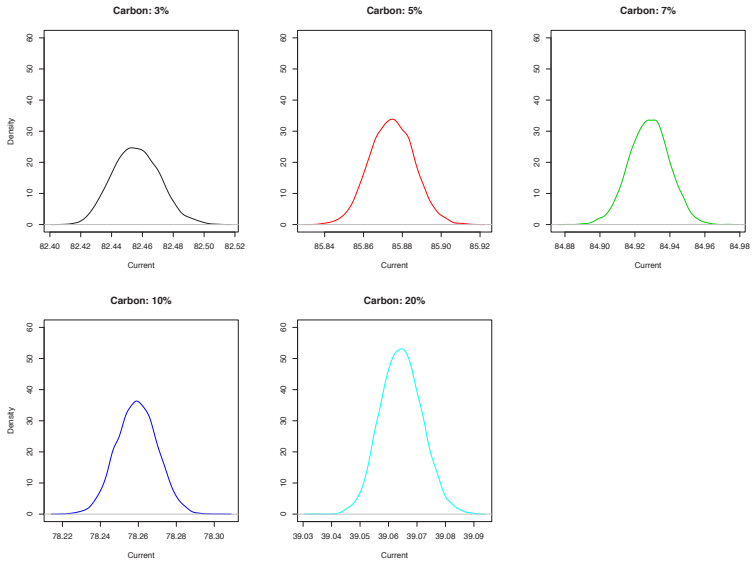


Figure 6.5: Density plots of posterior predictive distributions of Peak Current for scan rate 2.

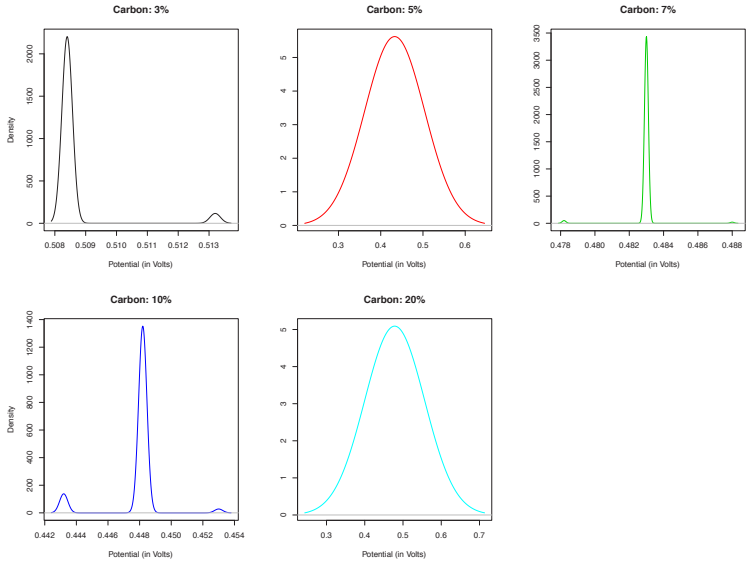


Figure 6.6: Density plots posterior predictive distributions of Peak Separation in Potential for scan rate 2.

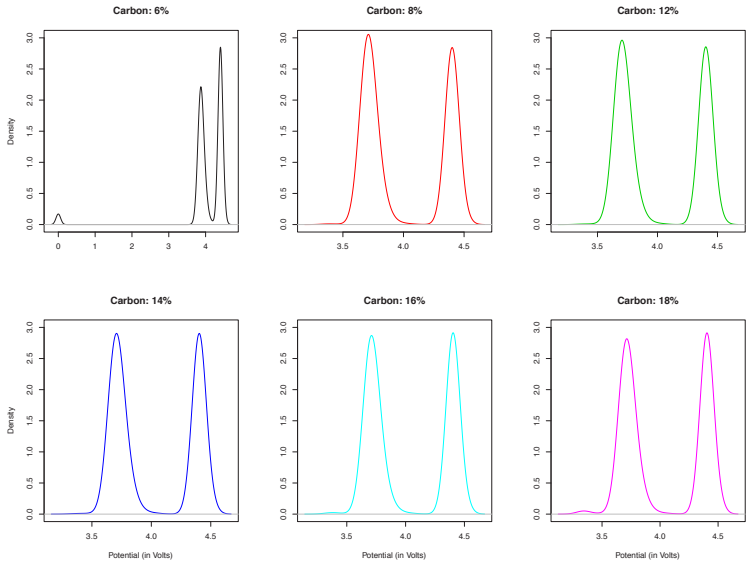


Figure 6.7: Density plots of posterior predictive distributions of Peak Potential for scan rate 2.

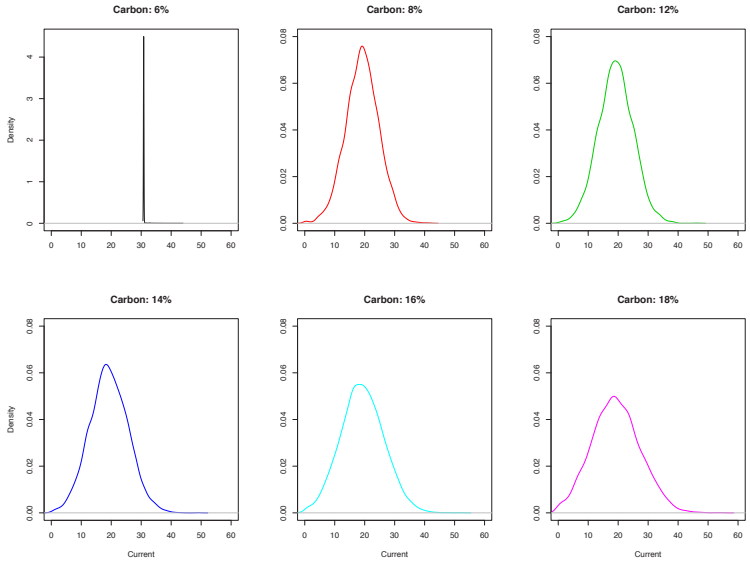


Figure 6.8: Density plots posterior predictive distributions of Peak Current for scan rate 2.

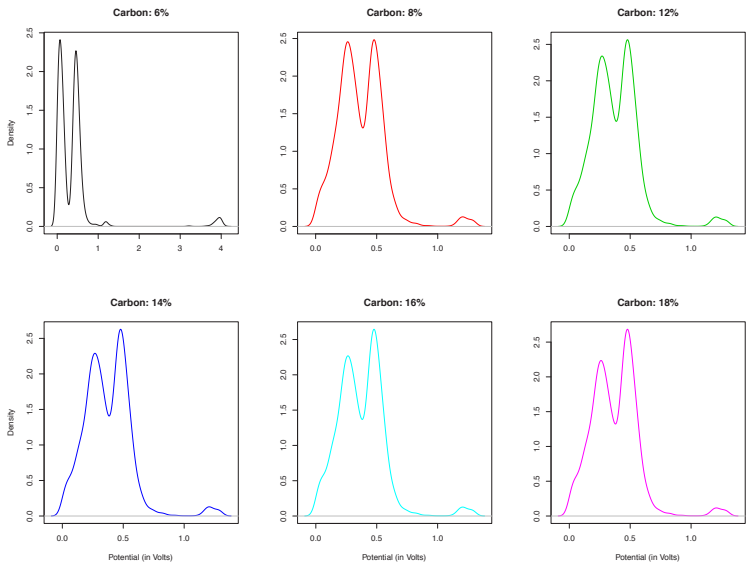


Figure 6.9: Density plots of posterior predictive distributions of Peak Separation in Potential for scan rate 2.

Chapter 7

Sinusoidal Models

7.1 Introduction

A well established method for modelling periodical features is to use trigonometric functions (referred to as sinusoidal models or Fourier Form). Sinusoidal models have been applied in a variety of applications, for example: Muller and Phillips (2007) applies sinusoidal models to Ozone Air pollution, Crellin et al. (1998) and Srivastava et al. (2003) apply these type of models in image analysis, and Dubnov and Rodet (1997) and Desainte-Catherine and Hanna (2000) are examples of applications to the modelling of sound. There has also been much research on estimating the parameters of sinusoidal models, for example Hainsworth and Macleod (2003) and Barone and Ragona (1997).

If a time series exhibits a periodical nature, then it is logical to build a model that accounts for this. The data we wish to analyse, shown in Figure 2.10, exhibits a periodical nature and hence we need to adapt the model we have developed in Chapter 6. In this chapter, we further develop our model to account for the periodical nature.

7.2 Sinusoidal Model

Assume that we have a time series that contains a deterministic sinusoidal model which is given by

$$\lambda_t = \mu + a\cos(\omega t) + b\sin(\omega t) + \varepsilon_t \tag{7.1}$$

where ω is some known frequency, the random error term is denoted by ε_t, λ_t denotes the observations, t denotes the time index for a generic time series and μ, a and b are parameters to be estimated. By writing the model in Equation (7.1) in matrix form, we obtain

$$E[\mathbf{\Lambda}] = A\boldsymbol{\theta}$$

where

$$\mathbf{\Lambda} = (\lambda_1, \ldots, \lambda_N)^T, \quad \boldsymbol{\theta} = (\mu, a, b)^T$$

and

$$A = \begin{pmatrix} 1 & \cos(\omega) & \sin(\omega) \\ 1 & \cos(2\omega) & \sin(2\omega) \\ \vdots & \vdots & \vdots \\ 1 & \cos(N\omega) & \sin(N\omega) \end{pmatrix}$$

where N is the length of a generic time series. This model is an example of a general linear model as it is linear in μ, a and b. The least squares estimate of $\boldsymbol{\theta}$ is obtained by minimising $\sum_{t=1}^{N}(\lambda_t - \mu - a\cos(\omega t) - b\sin(\omega t))^2$ from which we obtain

$$\hat{\boldsymbol{\theta}} = (A^T A)^{-1} A^T \mathbf{\Lambda}. \tag{7.2}$$

The solution in Equation (7.2) is valid for any frequency ω. However, the model only makes practical sense for values of ω that are not too high or low. The Nyquist frequency (see Section 7.3 and West and Harrison, 1999), given by $\omega = \pi$, completes

one cycle every two observations. The lowest frequency is where one complete cycle is the whole length of the time series. The highest and lowest possible values of the frequency ω will be explained in more detail in Section 7.3. If we equate the cycle length $2\pi/\omega$ to N, the lowest frequency is then given by $2\pi/N$. The least squares estimate of $\hat{\boldsymbol{\theta}}$ turn out to be simple if ω is restricted to one of the values

$$\omega_m = 2\pi m/N, \text{ where } m = 1, \ldots, N/2. \tag{7.3}$$

From Equation (7.3) we can see that the frequency increases in equal steps from the lowest frequency $2\pi/N$ up to the Nyquist frequency π. In the case when $\omega = \pi$, $(A^T A)$ is a diagonal matrix as a result of the trigonometric results shown in equations (7.4)-(7.7) below:

$$\sum_{t=1}^{N} \cos(\omega_m t) = \sum_{t=1}^{N} \sin(\omega_m t) = 0, \tag{7.4}$$

$$\sum_{t=1}^{N} \cos(\omega_m t)\cos(\omega_n t) = \begin{cases} 0, & m \neq n, \\ N, & m = n = N/2, \\ N/2, & m = n \neq N/2, \end{cases} \tag{7.5}$$

$$\sum_{t=1}^{N} \sin(\omega_m t)\sin(\omega_n t) = \begin{cases} 0, & m \neq n, \\ 0, & m = n = N/2 \\ N/2, & m = n \neq N/2, \end{cases} \tag{7.6}$$

$$\sum_{t=1}^{N} \cos(\omega_m t)\sin(\omega_n t) = 0, \quad \forall \ m, n. \tag{7.7}$$

If $(A^T A)$ is diagonal, then $(A^T A)^{-1}$ is also diagonal which makes it easy to calculate the least squares estimate of $\boldsymbol{\theta}$. In this case, for some frequency ω_m, where $m \neq N/2$,

we obtain:

$$\begin{aligned}\hat{\mu} &= \sum_{t=1}^{N} \lambda_t/N = \bar{\lambda},\\ \hat{a} &= 2\Big[\sum_{t=1}^{N} \lambda_t \cos(\omega_m t)\Big]/N,\\ \hat{b} &= 2\Big[\sum_{t=1}^{N} \lambda_t \sin(\omega_m t)\Big]/N.\end{aligned}$$

If $m = N/2$ we obtain

$$\begin{aligned}\hat{\mu} &= \bar{\lambda},\\ \hat{\alpha} &= \sum_{t=1}^{N} (-1)^t \lambda_t/N\end{aligned}$$

and the term in $b\sin(\omega t)$ is zero for all t.

7.3 The Nyquist Frequency and the Lowest Fundamental Frequency

We define the Nyquist frequency as follows as the upper bound. Suppose that observations are taken at intervals equal time intervals δt, then the Nyquist frequency is given by $\omega_N = \pi/\delta t$. In terms of cycles per unit time we have $\omega_N/2\pi = 1/2\delta t$. We note, throughout this section, that t and N are as defined in Section 7.2.

Consider the following example which involves thinking about the sampling rate required and the Nyquist frequency. Suppose that the situation where temperature readings are taken daily at midday for a particular period of time in a certain town. From these observations, it would not be possible to ascertain whether nights are warmer or colder than the days during the period of time of interest. In this situation, the Nyquist frequency is $\omega_N = \pi$ per day, that is 1 cycle is completed every 2 days. If we wanted to investigate variation within a day, we need a frequency of

2π per day, that is one complete cycle is completed per day. However, to do this a higher sampling rate would have to be used.

We will now demonstrate why there exists a lowest fundamental frequency below which it is not sensible to fit a sinusoidal model to the data. Suppose that we have 6 months of daily temperature readings from winter to summer. Using this data alone, it would not be possible to differentiate between an upward trend or if winters are colder than summers. In contrast, with 1 year's data, it would then become obvious that winters are colder than summers. Thus, if we are interested in temperature variation between the seasons, we require at least 1 year's worth of data and thus the lowest frequency we can fit is 1 cycle per year. For example, if we had weekly observations, for 1 year we would have $N = 52$, $\delta t = 1$ week and the lowest angular frequency given by $2\pi/N\delta t$ corresponds to $1/N\delta t$ cycles per week, which corresponds to $1/52$ cycles per week.

The lowest fundamental frequency which is $2\pi/N\delta$ is also referred to as the fundamental Fourier frequency since the Fourier series representation of the data calculated when $\omega_m = 2\pi m/N\delta$ for $m = 1, \ldots, N/2$ are all integer multiples of the fundamental frequency which are also referred to as harmonics. The term fundamental frequency is more appropriately used when some function $f(t)$ is periodic with period Π such that $f(t+m\Pi) = f(t)$ where $m \in \{1, 2, \ldots, N\}$. Then $\omega = 2\pi/\Pi$ is the fundamental frequency and the Fourier series representation of $f(t)$ is a sum over the harmonics of the fundamental frequency.

From the discussion above, we can see that the Nyquist frequency does not depend on N but the sampling frequency. In contrast to this, the lowest frequency does not depend on the sampling rate but is dependent upon N. This means that for lower frequencies we have to collate data over longer periods of time and for higher frequencies observations need to be taken more frequently. For a more detailed synopsis see Chatfield(2003, Chapter 7) and West and Harrison(1999, Chapter 8).

7.4 Fourier Form Representation of Cyclical Functions

The model in Equation (7.1) is the simplest example of a sinusoidal model. A more complex and practical representation of periodic functions that arise in various scenarios such as astronomy, geophysical studies and electrical systems, is the Fourier form of the model. The basic idea behind this representation of some cyclical function $g(t)$ where $t = 0, \ldots, N-1$, that is any N real numbers can be expressed as a linear combination of trigonometric terms which is dependent on the frequency ω. Let $R_f = N/2$ if N is even and $R_f = (N-1)/2$ if N is odd. The Fourier series representation of some time series λ_t is given by

$$\lambda_t = a_0 + \sum_{r=1}^{R_f} \left[a_r \cos\left(\frac{2\pi rt}{N}\right) + b_r \sin\left(\frac{2\pi rt}{N}\right) \right] \tag{7.8}$$

where $R_f \leq N/2$ and coefficients a_r, b_r are given by:

$$a_0 = \frac{1}{N}\sum_{t=0}^{N-1} \lambda_t, \; a_{N/2} = \frac{1}{N}\sum_{t=0}^{N-1} (-1)^t \lambda_t, \; b_{N/2} = 0,$$

$$\left.\begin{array}{rcl} a_r & = & \frac{2}{N}\sum_{t=0}^{N-1} \lambda_t \cos(2\pi rt/N) \\ b_r & = & \frac{2}{N}\sum_{t=0}^{N-1} \lambda_t \sin(2\pi rt/N) \end{array}\right\}, 1 \leq r < N/2. \tag{7.9}$$

An analysis using this model setup is referred to as a Fourier analysis or harmonic analysis. The Fourier series representation in Equation (7.8) has exactly N parameters to describe N observations, hence there is no error term. The results shown

in equation (7.9) can be proved (see Appendix C for proof) using the identities:

$$\begin{aligned}
\sum_{t=0}^{N-1} \cos\left(\frac{2\pi mt}{N}\right) \sin\left(\frac{2\pi nt}{N}\right) &= 0, \\
\sum_{t=0}^{N-1} \cos\left(\frac{2\pi mt}{N}\right) \cos\left(\frac{2\pi nt}{N}\right) &= \begin{cases} 0, & m \neq n, \\ N, & m = n = N/2, \\ N/2, & m = n \neq N/2, \end{cases} \\
\sum_{t=0}^{N-1} \sin\left(\frac{2\pi mt}{N}\right) \sin\left(\frac{2\pi nt}{N}\right) &= \begin{cases} 0, & m \neq n, \\ 0, & m = n = N/2, \\ N/2, & m = n \neq N/2, \end{cases} \\
\sum_{t=0}^{N-1} \cos\left(\frac{2\pi mt}{N}\right) = \sum_{t=0}^{N-1} \sin\left(\frac{2\pi mt}{N}\right) &= 0,
\end{aligned}$$

where m and n are integers.

We now return our attention to Equation (7.8). The parameters a_r and b_r are often referred to as the Fourier coefficients. For $r = 1, \ldots, R_f$, define the function $S_r(t)$ (known as the r^{th} harmonic) by:

$$\begin{aligned}
S_r(t) &= a_r \cos\left(\frac{2\pi rt}{N}\right) + b_r \sin\left(\frac{2\pi rt}{N}\right) \\
&= A_r \cos\left(\frac{2\pi rt}{N} + \rho_r\right), \\
A_r &= (a_r^2 + b_r^2)^{1/2}, \\
\rho_r &= \arctan\left(\frac{-b_r}{a_r}\right),
\end{aligned}$$

where $t = 0, \ldots, N-1$, A_r and ρ_r are referred to as the amplitude and phase of the r^{th} harmonic respectively. The maximum value of $S_r(t)$ is equal to the amplitude and the position of the maximum value of the r^{th} harmonic is determined by the phase. The r^{th} harmonic occurs when $\rho = 2R_f$ and is known as the Nyquist harmonic. As $b_{N/2} = 0$, then $A_{N/2} = |a_{N/2}|$ and $\rho_{N/2} = 0$. The frequency of the r^{th} harmonic

is given by $2\pi r/N$ and the cycle length is ρ/r. The first harmonic is also known as the fundamental harmonic with frequency α and length ρ. The r^{th} harmonic completes r full cycles for each complete cycle of the fundamental harmonic. For a more detailed exposition including how Fourier series are used in Dynamic Models, see Pole et al. (1994, Chapter 3) and West and Harrison (1999, Chapter 8).

7.5 Adaptation of the Fourier Model

The Fourier model described in Section 7.4 will be adapted slightly for application to our data sets. Instead of having the full set of Fourier coefficients we will determine the number of coefficients via the PMCC, hence there will be an error term in the model. The adjusted Fourier model will be

$$x_k(t) = a_0 + \sum_{r=1}^{R_f} \Big[a_r \cos(\omega r(t-1)) + b_r \sin(\omega r(t-1)) \Big] + \varepsilon_k(t) \tag{7.10}$$

where $t = 1, \ldots N$, $\omega = \frac{2\pi}{N}$, $\varepsilon_k(t) \sim N(0, \sigma^2)$,$R_f \leq N/2$ for N even and $R_f \leq (N-1)/2$ for N odd. For the remainder of this chapter, t denotes the time index defined in Table 2.1 and N denotes the number of observations which is dependent upon the scan rate. The likelihood for the model in Equation (7.10) is given by

$$\left(\frac{\tau^2}{2\pi}\right)^{\frac{NK}{2}} \exp\left(-\frac{\tau^2}{2} \sum_{k=1}^{K} \sum_{t=1}^{N} \varepsilon_k^2(t) \right) \tag{7.11}$$

where $\tau^2 = \sigma^{-2}$. As with our earlier models, we will use vague priors for the parameters which are given by:

$$\begin{aligned} a_0 &\sim N(0, v_0^2), \\ a_r &\sim N(0, v_{a,r}^2), \\ b_r &\sim N(0, v_{b,r}^2), \\ \tau^2 &\sim \text{Gamma}(d_1, d_2), \end{aligned} \tag{7.12}$$

where v_0, $v_{a,r}$, $v_{b,r}$, d_1 and d_2 are constants to be chosen. Let $\boldsymbol{a} = (a_1, a_2, \ldots, a_{R_f})^T$ and $\boldsymbol{b} = (b_1, b_2, \ldots, b_{R_f})^T$. The posterior distribution for the model in equation (7.10) is given by

$$(\tau^2)^{\psi_3 - 1} \exp\left(-\frac{\tau^2}{2} \sum_{k=1}^{K} \sum_{t=1}^{N} \varepsilon_k^2(t) - \frac{a_0^2}{2v_0^2} - \frac{1}{2}\boldsymbol{a}^T (V_a)^{-1} \boldsymbol{a} - \frac{1}{2}\boldsymbol{b}^T (V_b)^{-1} \boldsymbol{b} - d_2\tau^2 \right)$$

where $\psi_3 = \frac{NK}{2} + d_1$, $V_a = \text{diag}(v_{a,1}^2, v_{a,2}^2, \ldots, V_{a,R_f}^2)$ and $V_b = \text{diag}(v_{b,1}^2, v_{b,2}^2, \ldots, V_{b,R_f}^2)$. We will now consider additional features exhibited in our data, shown in Figure 2.10, we wish to model. If we look at Figure 2.10, we can see that different values are observed for the different proportions of Carbon. A simple way to account for this in the model would be to incorporate a polynomial in Carbon. In addition, to model the effect of the Potential, we will also incorporate an AR process in Potential and previous values of Current. The new model is given by

$$\begin{aligned} x_k(t) &= a_0 + \sum_{r=1}^{R_f} \Big[a_r \cos(\omega r(t-1)) + b_r \sin(\omega r(t-1)) \Big] + \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) \\ &\quad + \sum_{j=0}^{R_\beta} \beta_j p(t-j) + \sum_{q=1}^{R_\gamma} \gamma_q c_k^q + \varepsilon_k(t) \end{aligned}$$

where $t \geq (R_\alpha + 1, R_\beta + 1)$. The likelihood for this model is of the same form as the likelihood given in Equation (7.11). The $\varepsilon_k(t)$ is given by

$$\begin{aligned} \varepsilon_k(t) &= x_k(t) - \Bigg(a_0 + \sum_{r=1}^{R_f} \Big[a_r \cos(\omega r(t-1)) + b_r \sin(\omega r(t-1)) \Big] \\ &\quad + \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) + \sum_{j=0}^{R_\beta} \beta_j p(t-j) + \sum_{q=1}^{R_\gamma} \gamma_q c_k^q \Bigg). \end{aligned}$$

As before, we will use vague priors for the parameters which are stated in Equation (7.12) and Section 6.4. The posterior distribution for the model in equation (7.10)

is given by

$$\begin{aligned} f(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{p},\boldsymbol{c}) \;\propto\; & (\tau^2)^{\psi_2-1}\exp\left(-\frac{\tau^2}{2}\sum_{k=1}^{K}\sum_{t=r'+1}^{N}\varepsilon_k(t)-\frac{a_0^2}{2v_0^2}-\frac{1}{2}\boldsymbol{a}^T V_a^{-1}\boldsymbol{a}\right. \\ & -\frac{1}{2}\boldsymbol{b}^T V_b^{-1}\boldsymbol{b}-\frac{1}{2}\boldsymbol{\alpha}^T V_\alpha^{-1}\boldsymbol{\alpha}-\frac{1}{2}\boldsymbol{\beta}^T V_\beta^{-1}\boldsymbol{\beta} \\ & \left.-\frac{1}{2}\boldsymbol{\gamma}^T V_\gamma^{-1}\boldsymbol{\gamma}-d_2\tau^2\right) \end{aligned}$$

where $\boldsymbol{x}$, $\boldsymbol{p}$ and $\boldsymbol{c}$ are vectors for the Current, Potential and Carbon levels respectively, $\psi_2=\frac{K(N-r')}{2}+d_1$, $r'=\max(R_\alpha,R_\beta)$. The conditional posterior distributions are given by:

$$\begin{aligned} a_0|\boldsymbol{a},\boldsymbol{b},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\tau^2,X,\boldsymbol{p},\boldsymbol{c} &\sim N(\mu_0,\sigma_0) \\ \boldsymbol{a}|a_0,\boldsymbol{b},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\tau^2,X,\boldsymbol{p},\boldsymbol{c} &\sim N_{R_f}(\boldsymbol{\mu}_a,\Sigma_a) \\ \boldsymbol{b}|a_0,\boldsymbol{a},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\tau^2,X,\boldsymbol{p},\boldsymbol{c} &\sim N_{R_f}(\boldsymbol{\mu}_b,\Sigma_b) \\ \boldsymbol{\alpha}|a_0,\boldsymbol{a},\boldsymbol{b},\boldsymbol{\beta},\boldsymbol{\gamma},\tau^2,X,\boldsymbol{p},\boldsymbol{c} &\sim N_{R_\alpha}(\boldsymbol{\mu}_\alpha,\Sigma_\alpha) \\ \boldsymbol{\beta}|a_0,\boldsymbol{a},\boldsymbol{b},\boldsymbol{\alpha},\boldsymbol{\gamma},\tau^2,X,\boldsymbol{p},\boldsymbol{c} &\sim N_{R_\beta}(\boldsymbol{\mu}_\beta,\Sigma_\beta) \\ \boldsymbol{\gamma}|a_0,\boldsymbol{a},\boldsymbol{b},\boldsymbol{\alpha},\boldsymbol{\beta},\tau^2,X,\boldsymbol{p},\boldsymbol{c} &\sim N_Q(\boldsymbol{\mu}_\gamma,\Sigma_\gamma) \\ \tau^2|a_0,\boldsymbol{a},\boldsymbol{b},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},X,\boldsymbol{p},\boldsymbol{c} &\sim \text{Gamma}(\psi_2,\varphi), \end{aligned}$$

where:

$$\begin{aligned} \mu_0 &= \left(\frac{\tau}{\sigma_0}\right)^2\left(\sum_{k=1}^{K}\sum_{t=r'+1}^{N}y_k(t)\right), \\ \sigma_0^2 &= \left(\tau^2K(N-r')-v_0^{-1}\right)^{-1}, \\ \boldsymbol{\mu}_a &= \tau^2\Sigma_a A^T\boldsymbol{Y}_a,\ \Sigma_a=\left(\tau^2A^TA+V_a^{-1}\right)^{-1}, \\ \boldsymbol{Y}_a &= \left(\sum_{k=1}^{K}y_{a,k}(r'+1),\sum_{k=1}^{K}y_{a,k}(2),\ldots,\sum_{k=1}^{K}y_{a,k}(N)\right)^T, \end{aligned}$$

$$
\begin{aligned}
y_{a,k}(t) &= x_k(t) - \left(a_0 + \sum_{r=1}^{R_f} b_r \sin(r\omega t - r\omega) + \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) \right. \\
&\quad \left. + \sum_{j=0}^{R_\beta} \beta_j p(t-j) + \sum_{q=1}^{R_\gamma} \gamma_q c_k^q \right), \\
A &= \begin{pmatrix} \cos(\omega r') & \cos(2\omega r') & \dots & \cos(R_f \omega r') \\ \cos(\omega(r'+1)) & \cos(2\omega(r'+1)) & \dots & \cos(R_f\omega(r'+1)) \\ \vdots & \vdots & & \vdots \\ \cos(\omega(N-1)) & \cos(2\omega(N-1)) & \dots & \cos(R_f\omega(N-1)) \end{pmatrix}, \\
\boldsymbol{\mu}_b &= \tau^2 \Sigma_b B^T \boldsymbol{Y}_b, \\
\Sigma_b &= \left(\tau^2 B^T B + V_b^{-1}\right)^{-1}, \\
\boldsymbol{Y}_b &= \left(\sum_{k=1}^{K} y_{b,k}(r'+1), \sum_{k=1}^{K} y_{b,k}(2), \dots, \sum_{k=1}^{K} y_{b,k}(N) \right)^T, \\
y_{b,k}(t) &= x_k(t) - \left(a_0 + \sum_{r=1}^{R_f} a_r \cos r\omega(t-1) + \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) \right. \\
&\quad \left. + \sum_{j=0}^{R_\beta} \beta_j p(t-j) + \sum_{q=1}^{R_\gamma} \gamma_q c_k^q \right), \\
B &= \begin{pmatrix} \sin(\omega r') & \sin(2\omega r') & \dots & \sin(R_f \omega r') \\ \sin(\omega(r'+1)) & \sin(2\omega(r'+1)) & \dots & \sin(R_f\omega(r'+1)) \\ \vdots & \vdots & & \vdots \\ \sin(\omega(N-1)) & \sin(2\omega(N-1)) & \dots & \sin(R_f\omega(N-1)) \end{pmatrix}, \\
\boldsymbol{\mu}_\alpha &= \tau^2 \Sigma_\alpha X^T \boldsymbol{Y}_\alpha, \\
\Sigma_\alpha &= \left(\tau^2 X^T X + V_\alpha^{-1}\right)^{-1}, \\
\boldsymbol{Y}_\alpha &= \left(y_{\alpha,1}(r'+1), \dots, y_{\alpha,1}(N), \dots, y_{\alpha,K}(r'+1), \dots, y_{\alpha,K}(N)\right)^T, \\
y_{\alpha,k}(t) &= x_k(t) - \left(a_0 + \sum_{r=1}^{R_f} \left(a_r \cos\left(R_f\omega(t-1)\right) + b_r \sin\left(R_f\omega(t-1)\right)\right) + \right. \\
&\quad \left. + \sum_{j=0}^{R_\beta} \beta_j p(t-j) + \sum_{q=1}^{R_\gamma} \gamma_q c_k^q \right),
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\mu}_\beta &= \tau^2 \Sigma_\beta P^T \boldsymbol{Y}_\beta, \\
\Sigma_\beta &= \left(\tau^2 P^T P + V_\beta^{-1}\right)^{-1}, \\
\boldsymbol{Y}_\beta &= (y_{\beta,1}(r'+1), \ldots, y_{\beta,1}(N), y_{\beta,2}(r'+1), \ldots, y_{\beta,2}(N), \ldots, \\
&\quad y_{\beta,K}(r'+1), \ldots, y_{\beta,K}(N))^T, \\
y_{\beta,k}(t) &= x_k(t) - \left(a_0 + \sum_{r=1}^{R_f} \left(a_r \cos\left(r\omega(t-1)\right) + b_r \sin\left(r\omega(t-1)\right)\right) + \right. \\
&\quad \left. + \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) + \sum_{q=1}^{R_\gamma} \gamma_q c_k^q \right), \\
\boldsymbol{\mu}_\gamma &= \tau^2 \Sigma_\gamma C^T \boldsymbol{Y}_\gamma, \\
\Sigma_\gamma &= \left(\tau^2 C^T C + (V_\gamma)^{-1}\right)^{-1}, \\
\boldsymbol{Y}_\gamma &= (y_{\gamma,1}(r'+1), \ldots, y_{\gamma,1}(N), y_{\gamma,2}(r'+1), \ldots, y_{\gamma,2}(N), \ldots, \\
&\quad y_{\gamma,K}(r'+1), \ldots, y_{\gamma,K}(N))^T, \\
y_{\gamma,k} &= x_k(t) - \left(a_0 + \sum_{r=1}^{R_f} \left(a_r \cos r\omega(t-1) + b_r \sin r\omega(t-1)\right) + \right. \\
&\quad \left. + \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) + \sum_{j=0}^{R_\beta} \beta_j p(t-j) \right), \\
\varphi &= \frac{1}{2} \sum_{k=1}^{K} \sum_{t=r'+1}^{N} \varepsilon_k^2(t) + d_2.
\end{aligned}
$$

7.6 Analysis and Conclusions

The parameter estimates for the sinusoidal models were fairly robust when we varied the prior variances for a_0, $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and the hyperparameters d_1 and d_2 for τ^2. This is illustrated by the similarity of the parameter estimates under different prior variances, which are shown in Table 7.2 for the best sinusoidal model for scan rate 2. We also calculated the standard deviations for the parameter estimates and found that these were very similar under different prior variances and hence are omitted

for brevity. As in Section 6.5, we set the value for the prior variances to 10^8 and for the hyperparameters d_1 and d_2 to 0.001. The analysis presented in the next paragraph finds this best model.

As in Section 6.5, the values for R_α, R_β, R_f and R_γ shown in Table 7.1, were chosen by cycling around each variable in a nested loop structure. Hence, a thorough model search was conducted for various combinations of R_α, R_β, R_f and R_γ. The combination of values for R_α, R_β, R_f and R_γ for each scan rate, shown in Table 7.1, is the combination that produced the lowest PMCC value.

Scan Rate	Model Specification	Goodness of fit	Penalty	PMCC
1	$M_4(R_\alpha = 21, R_\beta = 15, R_\gamma = 1, R_f = 20)$	192.1585	182.3373	374.4958
2	$M_4(R_\alpha = 8, R_\beta = 10, R_\gamma = 1, R_f$=11)	36.2345	33.74563	69.9802
3	$M_4(R_\alpha = 8, R_\beta = 5, R_\gamma = 1, R_f = 11)$	149.7985	161.6283	311.4268

Table 7.1: Table of PMCC values for the best sinusoidal model for each scan rate.

From Table 7.1, according to the PMCC, we can see that the sinusoidal model performs better than the AR models used in Chapter 6. When we looked at the time series diagnostic plots for the parameters we found that the algorithm was covering the parameter space and that the algorithm did not appear to get stuck in a particular location for any of the parameters. However, the acf plot indicated that there was strong autocorrelation (significant as the value of the acf was above the dotted line). To reduce the dependence in the parameter samples, we took every 40^{th} sample generated by the algorithm. This gave us the well behaved diagnostic plots shown in Figures 7.1 and 7.2.

As we have already noted in Section 6.5, the residuals were calculated by working out the difference between the one-step ahead predictions and the observed values shown in Figures 2.10 and 2.11. The residual plots, shown in Figure 7.3, indicate that the variance of the residuals is not constant. The behaviour of the residuals becomes much more varied around the times when the Current is around the peak and minimum values. This same behaviour was exhibited in the residual plots in Figure 6.3. As we noted in Section 6.5, our main objective is predicting various characteristics about the peak and minimum points of the Current output, hence it is necessary to improve the model. We have only shown the plots from scan rate 2 to illustrate our findings. The corresponding plots for scan rates 1 and 3 exhibited similar behaviour.

Figures 7.4 and 7.6 provide the plots of the posterior predictive distribution included in the analysis for the Peak Potential and the Peak Separation in Potential, for each of the five levels of Carbon included in the analysis. The location of the posterior predictive distributions shown in Figure 7.5 appears to be located near the values that we would expect for the Peak Current. This is similarly true for the posterior predictive distributions of the Peak Potential which are shown in Figure 7.4.

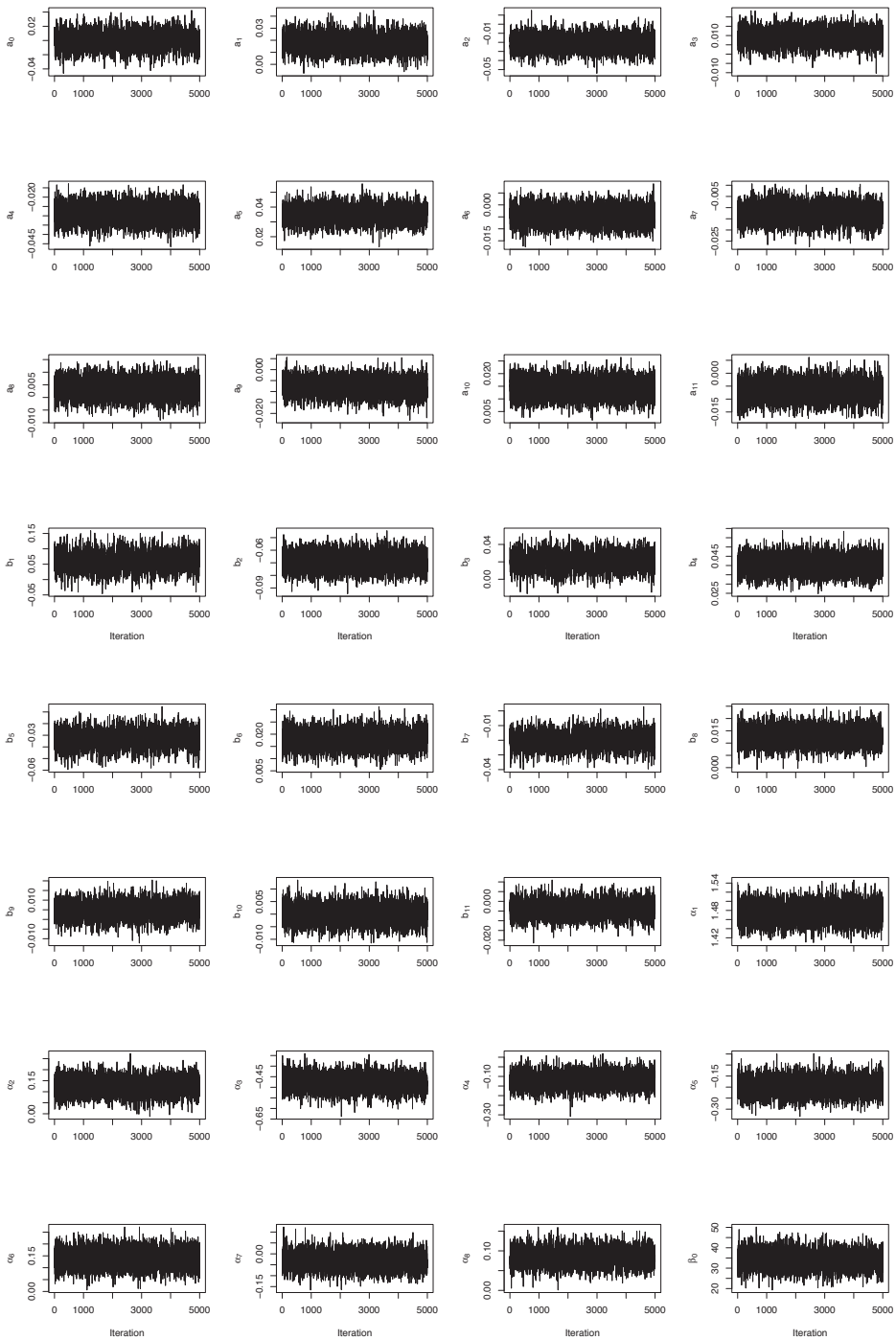
We note that the posterior predictive distributions shown in Figures 7.7, 7.8 and 7.9, are for Carbon levels not used in the experiment. As already noted in Section 5.3, we are unable to compare the posterior predictive distributions in the aforementioned plots to any observations. By including the posterior predictive distributions shown in Figures 7.7-7.9, we are simply illustrating the predictive usefulness of the sinusoidal model proposed in this chapter. The posterior predictive distributions for the Peak Potential in Figure 7.7 and Peak Current in Figure 7.8 are very similar (or almost identical) and there is little difference between Carbon levels. This is also the case for the posterior predictive distributions shown in Figure 7.9. For the posterior predictive distributions shown in Figures 7.7-7.9 we would expect there to

be slightly more variation between different Carbon levels. A similar critical analysis was also conducted for the other scan rates, but as mentioned earlier, we illustrate with data obtained for scan rate 2.

Parameter	Prior Variances		
	10^8	10^{10}	10^{12}
a_0	-0.0002	-0.0002	-0.0004
a_1	0.0177	0.0176	0.0176
a_2	-0.0206	-0.0207	-0.0207
a_3	0.0090	0.0089	0.0089
a_4	-0.0288	-0.0286	-0.0285
a_5	0.0352	0.0349	0.0349
a_6	-0.0046	-0.0046	-0.0045
a_7	-0.0125	-0.0126	-0.0126
a_8	0.0036	0.0037	0.0038
a_9	-0.0079	-0.0079	-0.0079
a_{10}	0.0141	0.0141	0.0141
a_{11}	-0.0066	-0.0067	-0.0068
b_1	0.0595	0.0599	0.0603
b_2	-0.0691	-0.0689	-0.0689
b_3	0.0204	0.0205	0.0205
b_4	0.0404	0.0404	0.0404
b_5	-0.0332	-0.0331	-0.0330
b_6	0.0178	0.0178	0.0179
b_7	-0.0199	-0.0198	-0.0197
b_8	0.0129	0.0128	0.0129
b_9	0.0048	0.0048	0.0049
b_{10}	-0.0001	-0.0001	-0.0001
b_{11}	-0.0039	-0.0040	-0.0039
α_1	1.4733	1.4743	1.4734
α_2	0.1293	0.1280	0.1292

Parameter	Prior Variances		
	10^8	10^{10}	10^{12}
α_3	-0.4775	-0.4787	-0.4784
α_4	-0.1206	-0.1188	-0.1189
α_5	-0.1985	-0.1982	-0.1992
α_6	0.1398	0.1394	0.1395
α_7	-0.0305	-0.0308	-0.0300
α_8	0.0835	0.0837	0.0832
β_0	34.1056	34.0896	34.0636
β_1	-57.6140	-57.4719	-57.3746
β_2	-2.6375	-2.8461	-2.9301
β_3	13.8800	14.0571	13.9931
β_4	18.7478	18.5584	18.5396
β_5	9.9999	10.0387	10.0224
β_6	-8.8091	-8.5617	-8.3225
β_7	-7.9568	-8.1181	-8.1941
β_8	8.0766	7.8820	7.7699
β_9	-20.0307	-19.9551	-19.8304
β_{10}	12.2381	12.3272	12.2630
γ_1	0.0019	0.0018	0.0032
σ^2	0.0142	0.0142	0.0142

Table 7.2: Parameter estimates (posterior mean) for different prior variances for the best sinusoidal model for scan rate 2.


Iteration
Iteration
Iteration
Iteration
Iteration
Iteration
Iteration
Iteration

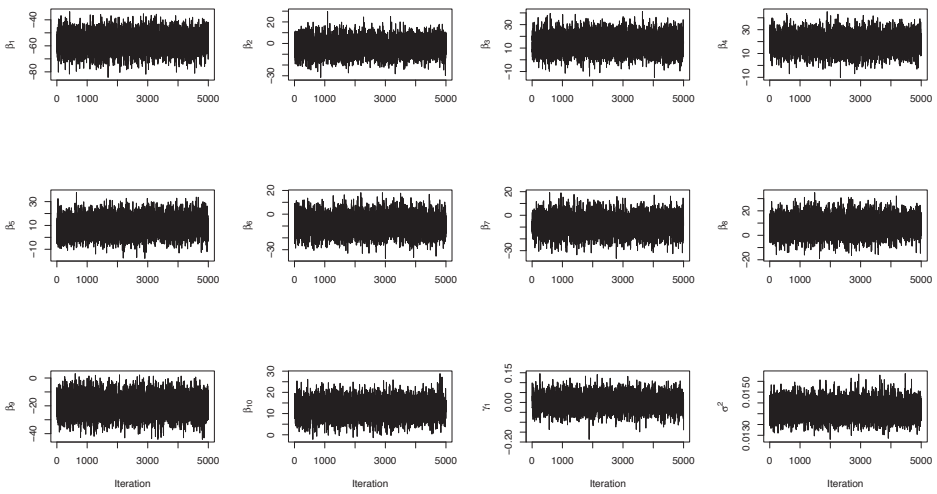
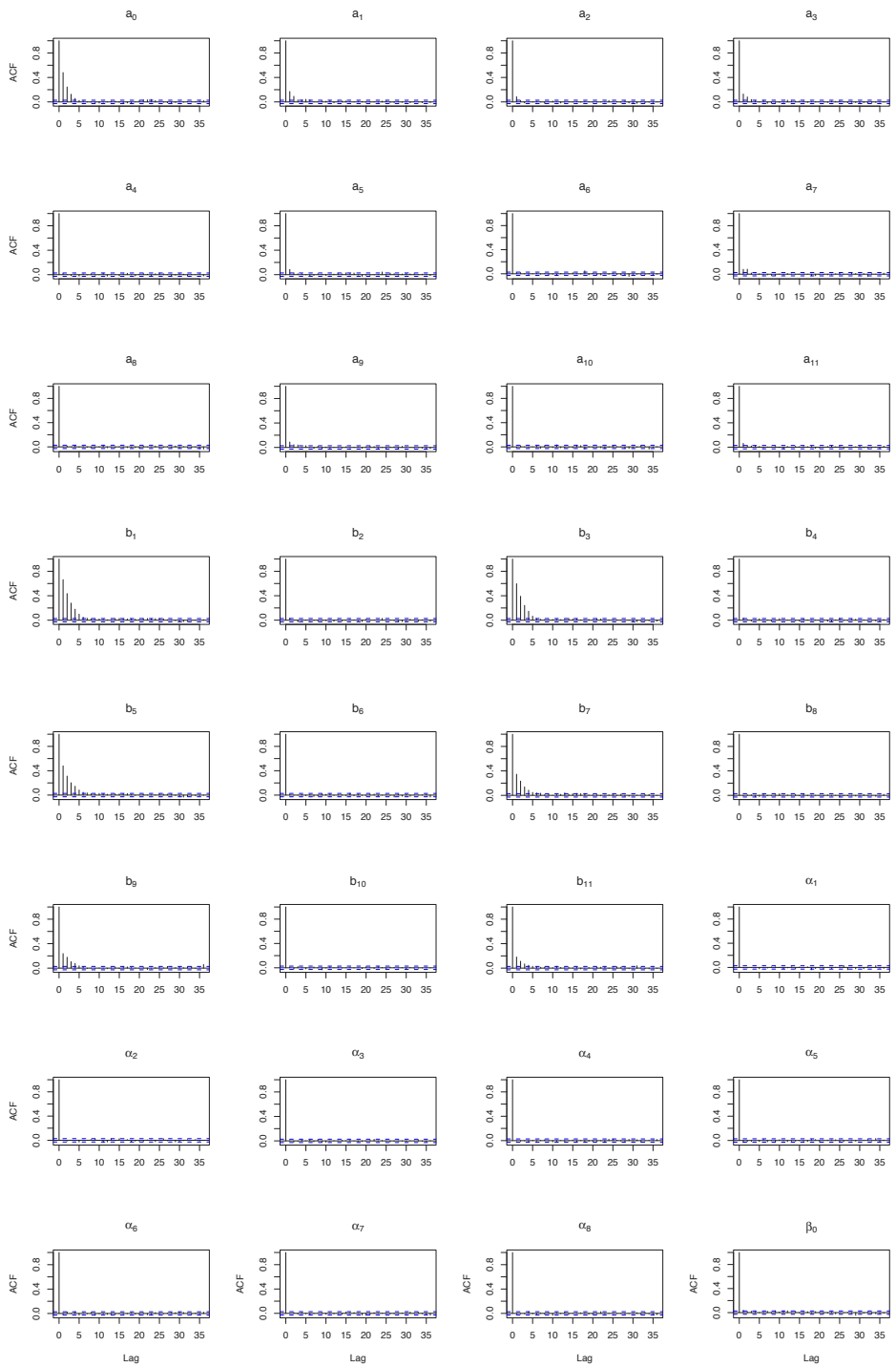


Figure 7.1: Trace plots of every 40^{th} sample generated by the Gibbs sampler for all the parameters of the best sinusoidal model for scan rate 2.


a0
a1
a2
a3
a4
a5
a6
a7
a8
a9
a10
a11
b1
b2
b3
b4
b5
b6
b7
b8
b9
b10
b11
α1
α2
α3
α4
α5
α6
α7
α8
β0
ACF
Lag
0.0 0.4 0.8
0 5 10 15 20 25 30 35

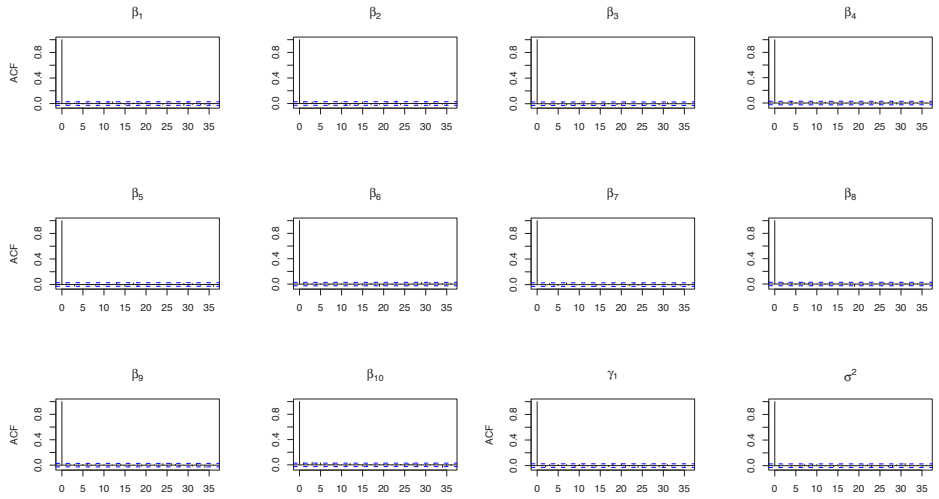


Figure 7.2: Autocorrelation plots of every 40^{th} sample generated by the Gibbs sampler for all the parameters of the best sinusoidal model for scan rate 2.

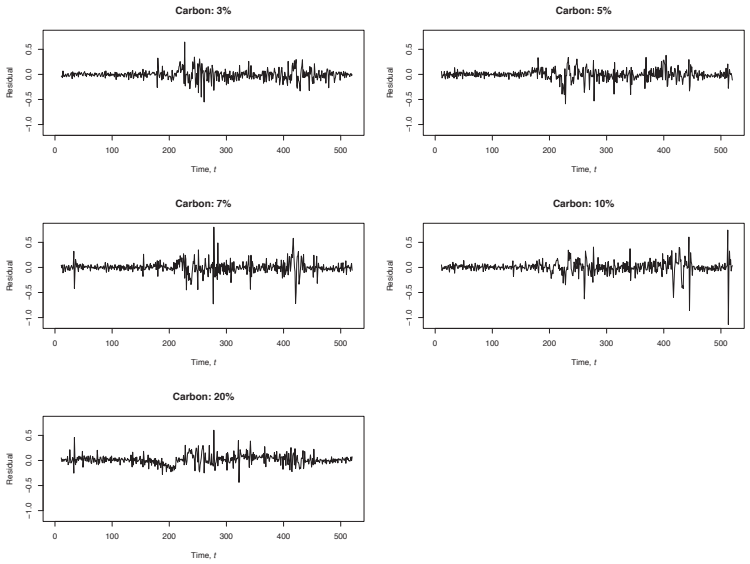


Figure 7.3: Differences between data and one-step ahead predictions, where t denotes the time index defined in Table 2.1.

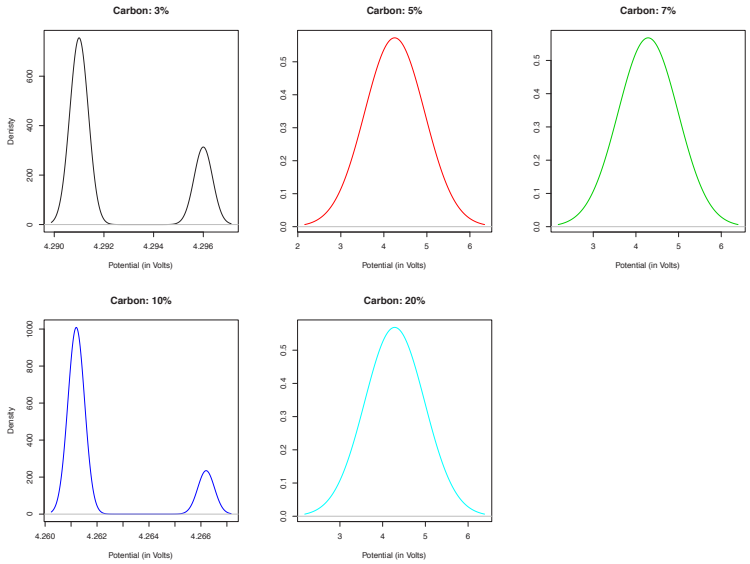


Figure 7.4: Density plots of posterior predictive distributions of Peak Potential for scan rate 2.

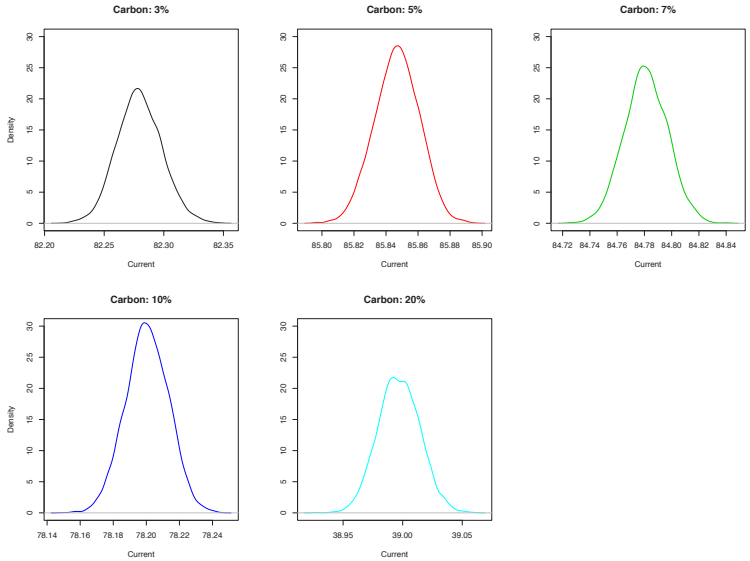


Figure 7.5: Density plots of posterior predictive distributions of Peak Current for scan rate 2.

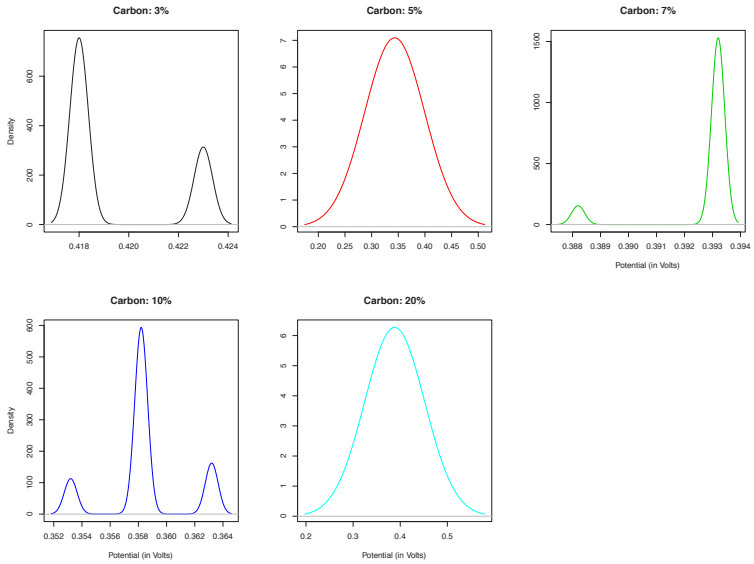


Figure 7.6: Density plots of posterior predictive distributions of Peak Separation in Potential for scan rate 2.

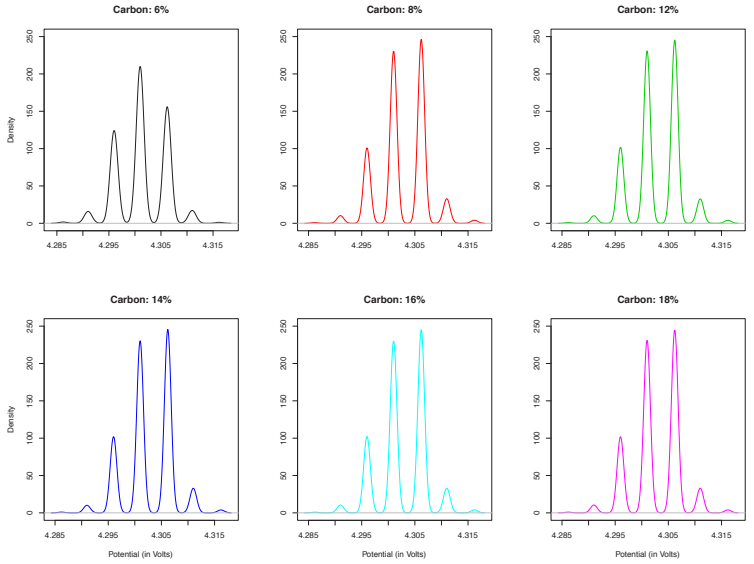


Figure 7.7: Density plots of posterior predictive distributions of Peak Potential for scan rate 2.

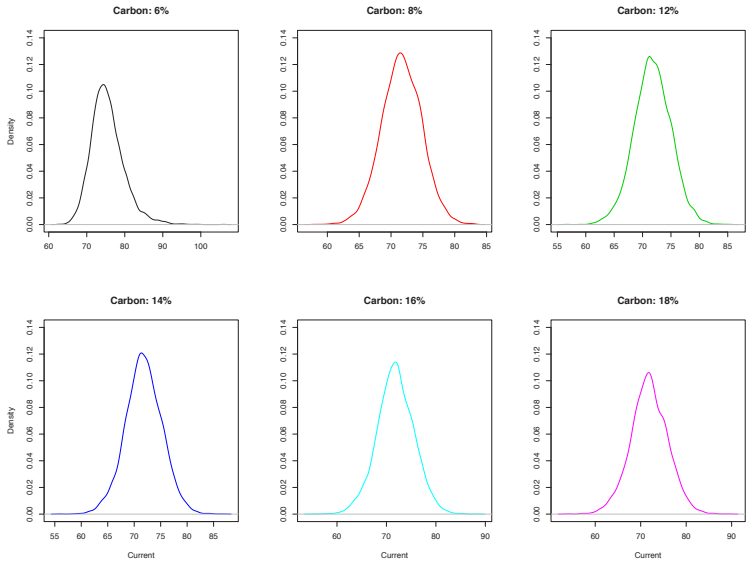


Figure 7.8: Density plots of posterior predictive distributions of Peak Current for scan rate 2.

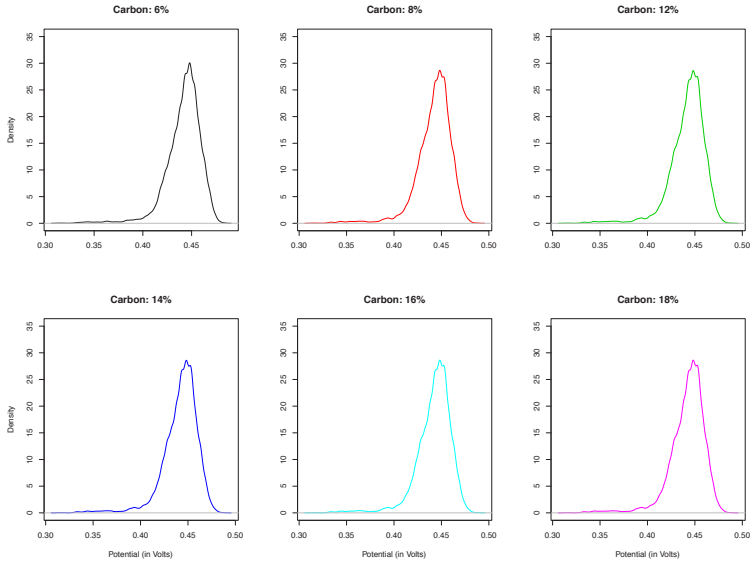


Figure 7.9: Density plots of posterior predictive distributions of Peak Separation in Potential for scan rate 2.

Chapter 8

Stochastic Volatility Models

8.1 Introduction

The residual plots for the models developed in Chapters 6 and 7 show more variation where the peaks and troughs occur. Although the residuals are quite small, we need to model the peaks and troughs more accurately. This is because we are interested in predicting characteristics about the peaks and troughs using their posterior predictive distributions. We will attempt to resolve this issue by modelling the variance instead of assuming the variance to be constant. These types of models are more generally known as stochastic volatility models. As before we will use a MCMC sampling algorithm to make predictive inferences about the characteristics of interest. We review the current literature on stochastic volatility models with a Bayesian perspective before analysing the data. As stated in Sections 6.1 and 7.1, we will be modelling the aggregated data shown in Figure 2.10.

There are a number of different stochastic volatility models that have been proposed such as the Black-Scholes (Black and Scholes, 1973) model and the Heston model (Heston, 1993) which are widely used in finance. Another type of stochastic volatility model was introduced by Engle (1982) known as autoregressive conditional het-

eroscedastic (ARCH) models. Before reviewing ARCH (autoregressive conditional heteroscedastic) and GARCH (general autoregressive conditional heteroscedastic) models, we consider modelling conditional means and variances in Section 8.3. In Section 8.2 we provide a brief literature review on stochastic volatility models. Our adopted models for conditional means and variances are described in Section 8.8.

8.2 Literature Review

Although the subject of Bayesian stochastic volatility models is relativity young in comparison to some of the other areas of statistical modelling, the body of research on stochastic volatility models is vast. We provide a brief review by discussing the contributions from the key articles in this subject.

As mentioned in Section 8.1, the main purpose of stochastic volatility models has been to model various aspects of the financial markets, for example, see Jacquier et al. (1994). In the aforementioned article, the authors propose new techniques and a simplified approach for the analysis of stochastic volatility models in which the logarithm of a conditional variance follows an autoregressive model. Jacquier et al. (1994) compare stochastic volatility models and ARCH models. The authors of the aforementioned article conclude that in their view stochastic volatility models are a promising alternative to various ARCH models. The vast majority of this article concentrates on how the Metropolis algorithm is used to construct a Markov chain simulation tool and how this can be used to draw inferences about the parameters and construct multi-step-ahead predictive densities. The authors compare their results to those obtained from the method of moments (see Melino and Turnbull, 1990) and quasi-maximum likelihood methods (see Ruiz, 1994 and Harvey et al.,1994) that have been proposed. Jacquier et al. (1994) concludes that their proposed method outperforms the method of moments and quasi-maximum likelihood techniques. A

number of articles have appeared since the publication of Jacquier et al. (1994), for example, see Chib et al. (2001), Kim et al. (1998), Broto and Ruiz (2004) and Liesenfeld and Richard (2006).

While ARCH models are useful for a variety of applications, especially finance, this type of model would be more powerful if we could generalise to the multivariate case. Harvey et al. (1994) point out that the generalisation of this model to the multivariate case means that it can be difficult to obtain parameter estimates and interpret. They suggest an alternative method of modelling the variance as an unobserved stochastic process. The logarithm of this component is modelled directly as a linear stochastic process, such as an autoregressive process. According to Harvey et al. (1994), one of the advantages is that their properties can be obtained from the properties of the process generating the variance component. The principal disadvantage of this method is that the maximum likelihood method is difficult to apply. The authors conclude that the multivariate stochastic variance model (or stochastic volatility model) has a natural interpretation and is relatively parsimonious, although the authors do not quantify how parsimonious in comparison to a suitable alternative model. Harvey et al. (1994) applied their methods to model daily dollar-pound exchange rates and show that the parameters can be estimated without too much difficulty via a quasi-maximum likelihood approach. The model fits well to the exchange rates and is able to capture common movements in volatility.

Chib et al. (2001) also discuss the fitting and comparison of high dimensional multivariate time series models with time varying conditions. The class of stochastic volatility models the authors used are more complex compared to the models used in the articles highlighted above. The sampling algorithm used to obtain estimates for the parameters relies on MCMC methods which incorporate a special method for sampling the parameters of the univariate stochastic volatility process. Chib et al. (2001) also provide methods for estimating the log-likelihood functions. The

authors conclude their work by highlighting that their model is robust to the choice of the prior distribution as well as the starting value for the MCMC algorithm. The authors believed that their approach was the first practical approach for modelling financial assets such as exchange rates. It should be noted that many contributions to developing various ways of using stochastic volatility models in finance such as Harvey et al. (1994), were made throughout the nineties.

We conclude this section by commenting on Broto and Ruiz (2004). The authors point out that while stochastic volatility models have an intuitive appeal, their application has been limited due to the inability of estimating their parameters. However, as we have shown and as is noted by Broto and Ruiz (2004), there have been several new techniques for estimating the parameters in a stochastic volatility model which are reviewed in this article. Liesenfeld and Richard (2006) present an estimation technique which is very close to what has been proposed by Broto and Ruiz (2004). Liesenfeld and Richard (2006) use a type of importance sampling algorithm (see Spiegelhalter et al., 2002 for further details on importance sampling) to perform a classical and a Bayesian analysis of univariate and multivariate stochastic volatility models. The authors point out that their sampling procedure is highly generic and hence changes in the model being analysed can be accommodated. Broto and Ruiz (2004) conclude that the several estimation techniques that they have considered for the parameters in a stochastic volatility model seem to match the benchmark established by the MCMC procedure of Jacquier et al. (1994).

8.3 Modelling Conditional Means and Variances

Throughout this section, and Sections 8.4-8.6, we use t, $t = 1, 2, \ldots, N$, and N to denote the time index and the number of observations respectively for a generic time series. Further, let Y_t be the observation and $u_{1,t}, \ldots u_{p,t}$ be the explanatory

variables at time t, where p denotes the number of explanatory variables. Consider the situation where we are modelling with a constant conditional variance, that is $\text{Var}(Y_t|u_{1,t},\ldots,u_{p,t})=\sigma^2$. The regression model of Y_t on $u_{1,t},\ldots,u_{p,t}$ is given by

$$Y_t = f(u_{1,t}, u_{2,t}, \ldots, u_{p,t}) + \varepsilon_t \tag{8.1}$$

where ε_t has zero mean and constant variance σ^2. The conditional mean of Y_t given $u_{1,t},\ldots,u_{p,t}$ is given by

$$E[Y_t|u_{1,t}, u_{2,t}, \ldots, u_{p,t}] = f(u_{1,t}, u_{2,t}, \ldots, u_{p,t}).$$

Equation (8.1) can be modified such that a non-constant variance is allowed, that is conditional heteroscedasticity. Let $\sigma^2(u_{1,t}, u_{2,t}, \ldots, u_{p,t})$ be the conditional variance, that is

$$\text{Var}(Y_t|u_{1,t}, u_{2,t}, \ldots, u_{p,t}) = \sigma^2(u_{1,t}, u_{2,t}, \ldots, u_{p,t}).$$

The model is given by

$$Y_t = f(u_{1,t}, u_{2,t}, \ldots, u_{p,t}) + \sigma(u_{1,t}, u_{2,t}, \ldots, u_{p,t})\varepsilon_t.$$

The function $\sigma(u_{1,t}, u_{2,t}, \ldots, u_{p,t})$ represents the standard deviation and should therefore be non-negative. This implies that if $\sigma(.)$ is a linear function then the parameters must be constrained such that $\sigma(.) \geq 0$. Modelling non-constant variances in regression is treated in detail in Ruppert (1988). Models that have a conditional variance are sometimes referred to as variance function models. The GARCH model is a special class of these types of models and are discussed in Section 8.6.

8.4 ARCH Processes

Let $\{\varepsilon_t\}$ be independent $N(0,1)$ where $t = 1,\ldots,N$. Then

$$E[\varepsilon_t|\varepsilon_{t-1},\ldots] = 0 \quad \text{and} \quad \text{Var}(\varepsilon_t|\varepsilon_{t-1},\ldots) = 1.$$

The process Y_t is an ARCH(1) process if

$$Y_t = \varepsilon_t \sqrt{\alpha_0 + \alpha_1 Y_{t-1}^2} \tag{8.2}$$

for $\alpha_0 \geq 0$ and $\alpha_1 \geq 0$. We can express Equation (8.2) as

$$Y_t^2 = (\alpha_0 + \alpha_1 Y_{t-1}^2)\varepsilon_t^2. \tag{8.3}$$

We can see from Equation (8.3) that an ARCH(1) process is similar to a AR(1) process in Y_t^2 and with a multiplicative white noise in place of a additive white noise process. Let σ_t^2 represent the conditional variance of Y_t given past values, that is $\sigma_t^2 = \text{Var}(Y_t|Y_{t-1}, \ldots)$. Since ε_t is independent of Y_{t-1} and $E[\varepsilon_t^2] = \text{Var}(\varepsilon_t) = 1$, then the mean and variance are given by

$$E[Y_t|Y_{t-1}, \ldots] = 0, \tag{8.4}$$

and

$$\begin{aligned} \text{Var}(Y_t|Y_{t-1}, \ldots) &= E[(\alpha_0 + \alpha_1 Y_{t-1}^2)\varepsilon_t^2|Y_{t-1}, Y_{t-2}, \ldots] \\ &= (\alpha_0 + \alpha_1 Y_{t-1}^2)E[\varepsilon_t^2|Y_{t-1}, Y_{t-2}, \ldots] \\ &= \alpha_0 + \alpha_1 Y_{t-1}^2. \end{aligned} \tag{8.5}$$

It is a requirement that $\alpha_0 \geq 0$ and $\alpha_1 \geq 0$ since the standard deviation cannot be negative. A further restriction is that $\alpha_1 < 1$ for Y_t to have a finite variance. When $\alpha_1 \geq 1$, the variance becomes infinite. We start proving this by taking the expectation of Equation (8.3), which is given by

$$\begin{aligned} E[Y_t^2] &= E[\varepsilon_t^2(\alpha_0 + \alpha_1 Y_{t-1}^2)] \\ &= E[\varepsilon_t^2]E[\alpha_0 + \alpha_1 Y_{t-1}^2]. \end{aligned}$$

Since $E[\epsilon_t^2] = 1$ and ε_t is independent of Y_{t-1}, we obtain

$$\begin{aligned} E[Y_t^2] &= E[\alpha_0 + \alpha_1 Y_{t-1}^2] \\ E[Y_t^2] &> \alpha_1 E[Y_{t-1}^2] \\ \frac{E[Y_t^2]}{E[Y_{t-1}^2]} &> \alpha_1. \end{aligned}$$

If $\alpha_1 > 1$, then $E[Y_t^2] > E[Y_{t-1}^2]$, that is the sequence Y_t is continually increasing, then σ_t^2 will continue to increase as σ_t^2 is dependent upon α_1 and the sequence of Y_t. Hence, the variance of Y_t will be infinite. Consider the equations

$$E[\sigma_t^2|Y_{t-1}] = \alpha_0 + \alpha_1 Y_{t-1}^2$$

and

$$E[\sigma_t^2|Y_{t-2}, \ldots] = \alpha_0 + \alpha_1(\alpha_0 + \alpha_1 Y_{t-2}^2).$$

If $\alpha_1 = 1$ then the quantity $\alpha_1(\alpha_0+\alpha_1 Y_{t-2}^2)$ will continue to grow, and hence resulting in an infinite variance. Therefore we require $\alpha < 1$ for Y_t to have a finite variance.

From Equation (8.5) we can see that if Y_{t-1} has an unexpected large deviation from the mean such that Y_{t-1} is large, then the variance of Y_{t-1} will have an unusually large variance. It follows that Y_t is expected to have a large deviation from the mean of zero. This volatility continues to have an affect, that is if Y_t has a large deviation, then σ_{t+1}^2 will be large, which means that Y_{t+1} will be large and so the volatility continues to propagate. Conversely, if Y_{t-1}^2 is smaller than expected, then σ_t^2 is small and Y_t^2 is expected to be small and so the pattern continues. This behaviour of unusual volatility continues its presence in the Y_t but not forever.

In this section, we have only considered an ARCH(1) process. However this can easily be extended to an ARCH(q) process. We say that Y_t is an ARCH(q) process if

$$Y_t = \sigma_t \varepsilon_t,$$

where the conditional standard deviation given past values $Y_{t-1}, Y_{t-2}, \ldots$ is given by

$$\sigma_t = \sqrt{\alpha_0 + \sum_{i=1}^{q} \alpha_i Y_t^2},$$

where q is the number of parameters in the ARCH process. The properties regarding the conditional and unconditional mean for an ARCH(1) process is same for an

ARCH(q) process, that is the conditional and unconditional mean are constant. Just like an ARCH(1) process, an ARCH(q) process has a constant unconditional variance and non-constant conditional variance.

8.5 Combining ARCH and AR Processes

An AR(1) has a non-constant mean and a constant conditional variance. In Section 8.4, we showed that an ARCH(1) process has a conditional mean that is constant and a non-constant conditional variance. If it is believed that the mean and variance of a process are dependent on the past then we can combine the AR and ARCH models. We will now concentrate on the simple task of combining an AR(1) model with an ARCH(1) model.

Let ζ_t be an ARCH(1) process such that $\zeta_t = \varepsilon_t\sqrt{\alpha_0 + \alpha_1\zeta_{t-1}^2}$ where ε_t is Gaussian white noise with zero mean and unit variance. Suppose that

$$Y_t = \phi Y_{t-1} + \zeta_t.$$

The process Y_t looks like an AR(1) process, except the noise term is an ARCH(1) process which replaces the independent white noise process. To ensure that Y_t is stationary with finite variance it is necessary that $|\phi| < 1$ and $\alpha_1 < 1$. In addition, it is naturally assumed that $\alpha_0 \geq 0$ and $\alpha_1 \geq 0$, as discussed in Section 8.4. In Figure 8.1, a simulation of an AR(1)ARCH(1) process with the individual parts is shown. It can be seen that when the ARCH(1) term is more volatile (shown in the bottom left hand panel of Figure 8.1), then the AR(1)ARCH(1) process moves more rapidly.

The process Y_t has a conditional mean and variance that are dependent on the past and non-constant. By combining an AR process with an ARCH process it is possible to model a wide variety of time series. Indeed, it is possible to combine any ARMA

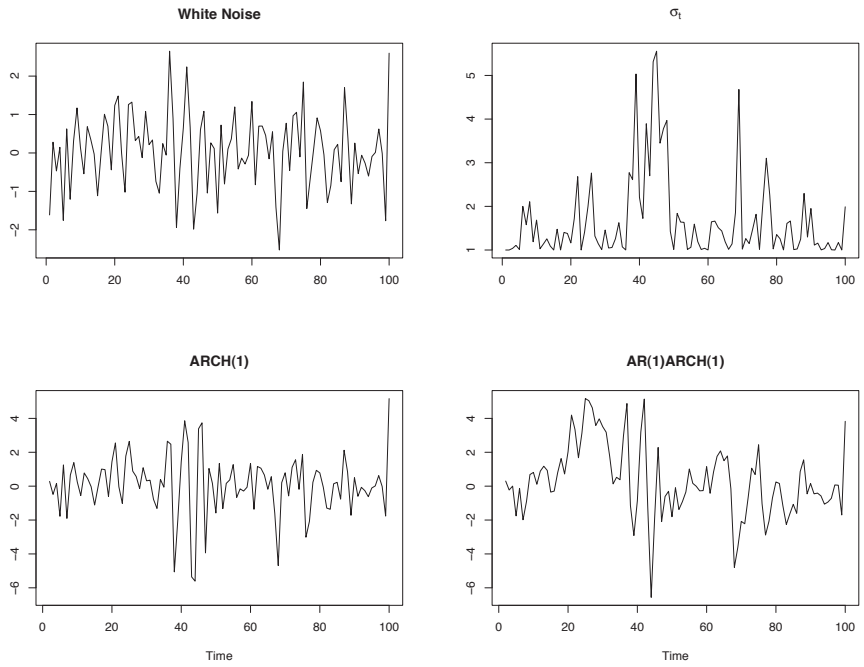


Figure 8.1: Simulation of 100 observations from an AR(1)ARCH(1) process. The values of the parameters are $\alpha_0 = 1.00$, $\alpha_1 = 0.95$, $\mu = 0.10$ and $\phi = 0.80$. This example is taken from Ruppert (2004, page 369).

model with any GARCH model which increases the variety and complexity of the models that can be used.

8.6 GARCH Models

The ARCH(q) process has a deficiency in that the volatility comes in short bursts. This is illustrated by the bottom left hand panel in Figure 8.1. If we want a model with volatility that is sustained for a longer period of time, then a GARCH model would be a better choice. The GARCH(p,q) model is given by

$$Y_t = \varepsilon_t \sigma_t$$

and

$$\sigma_t = \sqrt{\alpha_0 + \sum_{i=1}^{q} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2}, \tag{8.6}$$

where $\alpha_i \geq 0$ for $i = 1, \ldots, q$ and $\beta_j \geq 0$ for $j = 1, \ldots, p$. From Equation (8.6), it can be seen that $\sigma_{t-1}, \sigma_{t-2}, \ldots, \sigma_{t-p}$ are fed back into σ_t, hence the conditional standard deviation exhibit longer periods of high or low volatility than that of an ARCH process. An ARCH model is a special case of a GARCH model.

If we compare simulations from an GARCH(1,1) and an AR(1)GARCH(1,1) processes shown in Figure 8.2, it can be seen that the GARCH(1,1) process is less volatile than the AR(1)GARCH(1,1) process. The large value of the parameter β_1 will mean that σ_t will have a high correlation with σ_{t-1}. This is the force behind the longer lasting effect of the volatility in comparison to the ARCH(1) process shown in Figure 8.1.

8.7 Time Dependent Variance

A simple extension of the AR(R_α) model is to use a time dependent variance instead of assuming a constant variance. In this case, the model will have a variance for each time point. This will mean that for each scan rate we will have a different number of time dependent variances. In this case, the model for the aggregated data in Figures 2.10 and 2.11 is given by

$$x_k(t) = \sum_{i=1}^{R_\alpha} \alpha_i x_k(t-i) + \varepsilon_k(t),$$

where $\varepsilon_k(t) \sim N(0, \sigma_t^2)$, $\tau_t^2 = \sigma_t^{-2}$ and from herein, t denotes the time index defined in Table 2.1. As with the other models, we will use vague priors which are given by

$$\alpha_i \sim N(0, \omega_i) \quad \text{and} \quad \tau_t^2 \sim \text{Gamma}(d_1, d_2),$$

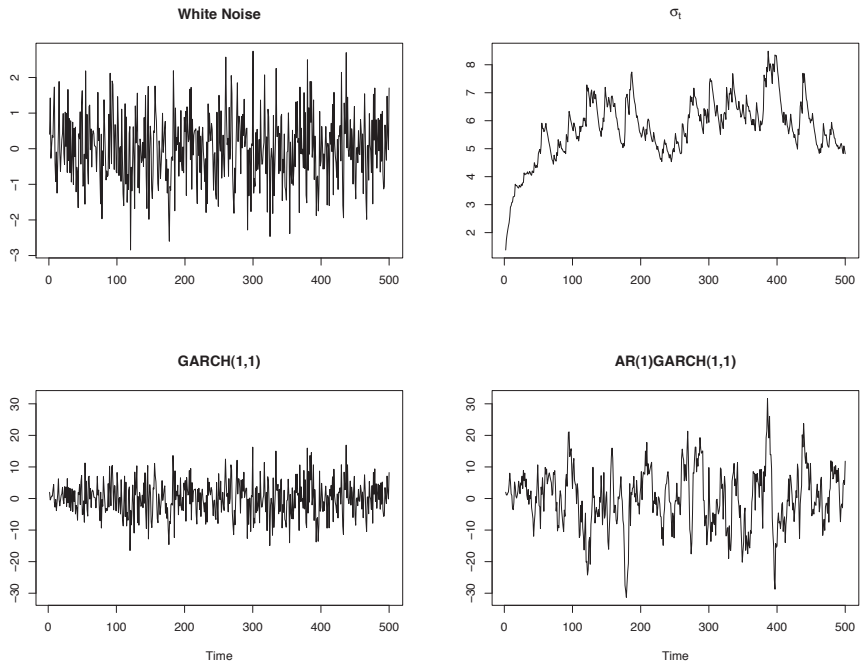


Figure 8.2: Simulation of 500 observations from an AR(1)GARCH(1,1) process. The values of the parameters are $\alpha_0 = 1.00$, $\alpha_1 = 0.08$, $\beta_1 = 0.90$, and $\phi = 0.80$. This example is taken from Ruppert (2004, page 371).

where ω_i is the prior variance for α_i and d_1, d_2 are constants. We note that τ_t^2 are identically and independently distributed a-priori. The posterior distribution for this model is given by

$$\prod_{t=r+1}^{N} (\tau_t)^{K-2a-2} \exp\left(-\frac{1}{2} \sum_{k=1}^{K} \sum_{t=r+1}^{N} \tau_t^2 \varepsilon_k^2(t) - \frac{1}{2} \boldsymbol{\alpha}^T \boldsymbol{\Omega}_\alpha^{-1} \boldsymbol{\alpha} - b \sum_{t=r+1}^{N} \tau_t^2 \right),$$

where $\Omega_\alpha = \text{diag}(\omega_1, \omega_2, \ldots, \omega_{R_\alpha})$ and from herein, N denotes the number of observations which is dependent on the scan rate. From a theoretical point of view, this model should perform poorly as the number of parameters is going to be large. This model can also be seen as a slight departure from our parsimonious approach as the number of parameters will increase dramatically. When we used this model on a

single level of Carbon at scan rate 2, we found that the performance of this model was extremely poor under the PMCC in comparison to models already tested and developed in Chapters 6 and 7.

8.8 Variance Function

A better approach is to use a simple functional form for the variance. If we use too many parameters in the variance function, the model may not perform well in comparison to the competing models developed thus far. We are going to use a simple AR(1)ARCH(1) model which will incorporate a sinusoidal component. This model will have a conditional mean and variance that are both non-constant. In addition, this model also produces short bursts of volatility which is what happens about the peaks and minimum points in the data.

There are obviously an infinite number of different variance functions that we could try and hence will be unable to test them all. We will restrict our approach to incorporating simple polynomial of the Carbon and a AR process in Potential in the variance as these are believed to be the driving forces behind the chemical process which affects the Current. As we can see from the plots in Figure 2.10 it is the Carbon that has the biggest effect on the characteristics of interest. The model we will use is of the form

$$\begin{aligned} x_k(t) &= \alpha_0 + \alpha_1 x_k(t-1) + \beta_0 p(t) + \beta_1 p(t-1) + \sum_{q=1}^{R_q} \gamma_q c_k^q \\ &\quad + \sum_{r=1}^{R_f} \left(a_r \cos(\omega r(t-1)) + b_r \sin(\omega r(t-1))\right) + \varepsilon_k(t). \end{aligned} \tag{8.7}$$

We assume that $\varepsilon_k(t) \sim N(0, \sigma_k^2(t))$, where

$$\sigma_k^2(t) = \exp\left(-\left(\phi_0 + \sum_{i=1}^{R_\phi} \phi_j c_k^i + \sum_{m=1}^{R_\nu} \nu_m p(t-m) + \eta x_k(t-1)\right)\right). \tag{8.8}$$

We will consider one specific model in detail, that is, we will show how the posterior distribution is constructed and how to estimate the parameters. The model we will consider in detail is the general model that is shown in Equation (8.7) in conjunction with Equation (8.8). The likelihood for this model is given by

$$\left(\prod_{k=1}^{K}\prod_{t=2}^{N}\left(\mathrm{Var}\left(x_k(t)|\boldsymbol{\theta}\right)\right)^{-\frac{1}{2}}\right)\exp\left(-\frac{1}{2}\sum_{k=1}^{K}\sum_{t=2}^{N}\frac{\left(x_k(t)-E\left[x_k(t)|\boldsymbol{\theta}\right]\right)^2}{\mathrm{Var}\left(x_k(t)|\boldsymbol{\theta}\right)}\right), \tag{8.9}$$

where $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{a}, \boldsymbol{b}, \phi_0, \boldsymbol{\phi}, \nu, \eta)^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_{R_\gamma})^T$, $\boldsymbol{a} = (a_1, a_2, \ldots, a_{R_f})^T$, $\boldsymbol{b} = (b_1, b_2, \ldots, b_{R_f})^T$ and $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_{R_\phi})^T$. As we are including an autoregressive process of order 1 in our model, it follows that $t \geq 2$. The conditional expectation $E\left[x_k(t)|\boldsymbol{\theta}\right]$ is given by

$$\alpha_0 + \alpha_1 x_k(t-1) + \beta_0 p(t) + \beta_1 p(t-1) + \sum_{q=1}^{R_q}\gamma_q c_k^q$$
$$+\sum_{r=1}^{R_f}\left(a_r\cos(\omega r(t-1)) + b_r\sin(\omega r(t-1))\right),$$

since the $\varepsilon_k(t)$ have mean zero. The conditional variance $\mathrm{Var}\left(x_k|\boldsymbol{\theta}\right)$ is given by

$$\exp\left(-\left(\phi_0 + \sum_{i=1}^{R_\phi}\phi_j c_k^i + \sum_{m=0}^{R_\nu}\nu_m p(t-m) + \eta x_k(t-1)\right)\right).$$

As before we will assume non-informative prior distributions. For parameters α_0, α_1, β_0, β_1, a_r and b_r assume non-informative priors as in Section 7.5 and:

$$\begin{aligned} \phi_0 &\sim N(0, d_{\phi_0}), \qquad \phi_i \sim N(0, d_{\phi_i}), \\ \nu &\sim N(0, d_{\nu_m}), \qquad \eta \sim N(0, d_\eta), \end{aligned} \tag{8.10}$$

where d_{1,ϕ_0}, d_{2,ϕ_0}, d_{1,ϕ_i}, d_{2,ϕ_i}, $d_{1,\nu}$, $d_{2,\nu}$, $d_{1,\eta}$, $d_{2,\eta}$ are constants to be chosen. By taking the product of the prior distributions shown in Equation (8.10) and the likelihood

in Equation (8.9) we obtain the posterior distribution given by

$$\exp\left(-\frac{\alpha_0^2}{2v_0}-\frac{\alpha_1^2}{2v_1}-\frac{1}{2}\boldsymbol{\beta}^T V_\beta^{-1}\boldsymbol{\beta}-\frac{1}{2}\boldsymbol{\gamma}^T V_\gamma^{-1}\boldsymbol{\gamma}-\frac{1}{2}\boldsymbol{a}^T V_a^{-1}\boldsymbol{a}-\frac{1}{2}\boldsymbol{b}^T V_b^{-1}\boldsymbol{b}\right.$$
$$\left.-\frac{\phi_0^2}{2d_{2,\phi_0}}-\frac{\nu^2}{2d_{2,\nu}}-\frac{\eta^2}{2d_{2,\eta}}-\sum_{i=1}^{R_\phi}\frac{\phi_i^2}{2d_{2,\phi_i}}\right)\left(\prod_{k=1}^{K}\prod_{t=2}^{N}\left(\mathrm{Var}\left(x_k|\boldsymbol{\theta}\right)\right)^{-\frac{1}{2}}\right)$$
$$\exp\left(-\frac{1}{2}\sum_{k=1}^{K}\sum_{t=2}^{N}\frac{\left(x_k(t)-E\left[x_k(t)|\boldsymbol{\theta}\right]\right)^2}{\mathrm{Var}\left(x_k|\boldsymbol{\theta}\right)}\right). \tag{8.11}$$

The conditional posterior distributions for the parameters obtained from the expression shown in Equation (8.11) are extremely difficult to sample from. The MCMC methods that have been used to obtain parameter estimates for stochastic volatility models such as that shown in Equation (8.11) are the Metropolis-Hastings algorithm (see Section 3.5) and rejection sampling. Jacquier et al. (2003) suggest using a rejection-sampling method or the Metropolis-Hastings independence sampler, (see Gilks et al., 1996, Chapter 5). If the rejection sampling method is used as proposed in Jacquier et al. (2003), then the possibility of a high rejection rate could result in having to run the algorithm for long periods of time to obtain a good approximation of the posterior distribution. When using the independence sampler, the choice of the proposal distribution is critical to the efficiency of the algorithm. This becomes even more important when dealing with large data sets such as the data set we are dealing with. Due to the difficulty of being able to write down the conditional distributions for stochastic volatility models, there are a number of articles such as Yu and Meyer (2006) that have used WinBuGS to obtain estimates for the parameters in the various stochastic volatility models under consideration. One advantage of using WinBUGS is that a proposal distribution is not required. Further to this, WinBUGS is also efficient at generating random samples of the model parameters. This latter approach of using WinBUGS is the one we shall adopt. The sampling methodology implemented by WinBUGS is discussed in Section 8.9.

Now that we have shown how the posterior distribution is constructed and how we

aim to obtain samples of the parameters we will show in slightly more detail the four models we consider here. The four stochastic models we will consider are given below. Note that for each model, $x_k(t)$ is given by Equation (8.7).

Stochastic Volatility Model 1 ($M_5(R_\gamma, R_f, R_\phi)$):

$$\sigma_k^2(t) = \exp\left(-\left(\phi_0 + \sum_{i=1}^{R_\phi} \phi_j c_k^i\right)\right).$$

Stochastic Volatility Model 2 ($M_6(R_\gamma, R_f, R_\phi, R_\nu)$):

$$\sigma_k^2(t) = \exp\left(-\left(\phi_0 + \sum_{i=1}^{R_\phi} \phi_j c_k^i + \sum_{m=0}^{R_\nu} \nu_m p(t-m)\right)\right).$$

Stochastic Volatility Model 3 ($M_7(R_\gamma, R_f, R_\phi)$):

$$\sigma_k^2(t) = \exp\left(-\left(\phi_0 + \sum_{i=1}^{R_\phi} \phi_j c_k^i + \eta x_k(t-1)\right)\right).$$

Stochastic Volatility Model 4 ($M_8(R_\gamma, R_f, R_\phi, R_\nu)$):

$$\sigma_k^2(t) = \exp\left(-\left(\phi_0 + \sum_{i=1}^{R_\phi} \phi_j c_k^i + \sum_{m=0}^{R_\nu} \nu_m p(t-1) + \eta x_k(t-1)\right)\right).$$

8.9 Sampling Methodology Used by WinBUGS

Lunn et al. (2000) state that WinBUGS attempts to use the most appropriate sampling method for each parameter. When the full conditional posterior distribution is available in closed form, WinBUGS can identify the closed form and implement the most appropriate sampling method. When the full conditional posterior distribution is not available in closed form, WinBUGS examines the situation and chooses a suitable general sampling method. Table 8.1 (from Lunn et al., 2000) shows the sampling method hierarchy used by WinBUGS in order of precedence. As we will be using WinBUGS to apply the stochastic volatility models proposed in Section 8.8,

it follows that the sampling method hierarchy shown in Table 8.1 illustrates how the samples of the parameters in our stochastic volatility models will be obtained. For further details on WinBUGS see Lunn et al. (2000).

Target Distribution	Sampling Method
Discrete	Inversion of cumulative distribution function (trivial)
Closed form (conjugate)	Direct sampling using standard algorithms
Log-concave	Derivative-free adaptive rejection sampling (Gilks, 1992)
Restricted range	Slice sampling (Neal, 1997)
Unrestricted range	Metropolis-Hastings (Metropolis et al., 1953 and Hastings, 1970)

Table 8.1: Sampling method hierarchy used by WinBUGS in order of precedence. Each method is only used if no previous method in the hierarchy is appropriate (see Lunn et al., 2000 for further details).

8.10 Analysis and Conclusions

As with the models used in Sections 6.5 and 7.6, the optimal values for R_γ, R_f R_ϕ and R_ν, were found using a nested search method for each in turn. The combinations of values that produced the lowest PMCC values for each scan rate are shown in Table 8.2.

The parameter estimates were fairly robust when we varied the prior variances for all the parameters (see Table 8.3). As in Section 7.6, we calculated the standard deviations for the parameter estimates and found that these were very similar under

Scan Rate	Model	PMCC
1	$M_5(R_\gamma = 1, R_f = 6, R_\phi = 3)$	132556.88
	$M_6(R_\gamma = 1, R_f = 6, R_\phi = 3, R_\nu = 2)$	172043.03
	$M_7(R_\gamma = 1, R_f = 6, R_\phi = 3)$	131783.20
	$M_8(R_\gamma = 1, R_f = 6, R_\phi = 3, R_\nu = 2)$	252200.83
2	$M_5(R_\gamma = 1, R_f = 6, R_\phi = 2)$	10964.21
	$M_6(R_\gamma = 1, R_f = 6, R_\phi = 2, R_\nu = 2)$	17523.50
	$M_7(R_\gamma = 1, R_f = 6, R_\phi = 2)$	9786.03
	$M_8(R_\gamma = 1, R_f = 6, R_\phi = 2, R_\nu = 1)$	26266.48
3	$M_5(R_\gamma = 1, R_f = 6, R_\phi = 2)$	33939.67
	$M_6(R_\gamma = 1, R_f = 6, R_\phi = 2, R_\nu = 2)$	33200.63
	$M_7(R_\gamma = 1, R_f = 6, R_\phi = 2)$	31216.89
	$M_8(R_\gamma = 1, R_f = 6, R_\phi = 2, R_\nu = 2)$	35456.49

Table 8.2: Table of PMCC values for different models for the three different scan rates.

different prior variances and hence are omitted for brevity. As noted in earlier chapters, it is reassuring to find that the parameter estimates are robust to these changes in the prior distributions.

According to the PMCC, the best model is $M_7(R_\gamma = 1, R_f = 6, R_\phi = 2)$ for scan rate 2. For the results presented in this section, we use the aforementioned model and set the value for each prior variance to 10^8.

When we looked at the trace plots for the parameters of the stochastic volatility model we found that the algorithm was covering the parameter space at a very slow rate and that the algorithm did not appear to get stuck in a particular location for any of the parameters. The acf plot indicated that there was strong autocorrelation

Parameter	Prior Variances		
	10^8	10^{10}	10^{12}
α_0	29.4902	29.4901	29.4901
α_1	0.9654	0.9654	0.9654
β_0	-18.5246	-18.5246	-18.5246
β_1	10.8707	10.8707	10.8707
γ_1	0.5605	0.5605	0.5605
a_1	-4.6342	-4.6342	-4.6342
a_2	0.8108	0.8108	0.8108
a_3	-0.5593	-0.5593	-0.5593
a_4	-0.2420	-0.2420	-0.2420
a_5	-0.0264	-0.0264	-0.0264
a_6	-0.0701	-0.0701	-0.0701
b_1	1.0200	1.0200	1.0200
b_2	-0.4111	-0.4111	-0.4111
b_3	0.1344	0.1344	0.1344
b_4	-0.2632	-0.2632	-0.2632
b_5	0.2753	0.2753	0.2753
b_6	-0.0353	-0.0353	-0.0353
ϕ_0	7.2988	7.2988	7.2988
ϕ_1	-148.0661	-148.0673	-148.0673
ϕ_2	537.1717	537.1735	537.1735
η	-0.0143	-0.0143	-0.0143

Table 8.3: Parameter estimates (posterior mean) for different prior variances for the best stochastic volatility model.

(significant as the value of the acf was above the dotted line), in parameter samples generated by the algorithm. To reduce the dependence in the parameter samples, we experimented as to how often to sample from the chain initially choosing relatively small values such as picking every 5^{th}, 10^{th}, 20^{th} sample generated by the algorithm. Further experimentation led us to taking every 200^{th} sample generated by the algorithm. This gave us the well behaved diagnostic plots shown in Figures 8.3 and 8.4, which appear to indicate that the sample obtained covers the full parameter space.

The residual plots in Figure 8.5 appear to indicate that the stochastic volatility model has not adequately modelled the data. Further to this, they do not appear to have a constant variance, as we would expect for a good fitting model. This could partially explain why the PMCC is much higher than that for the sinusoidal model used in Section 7.6.

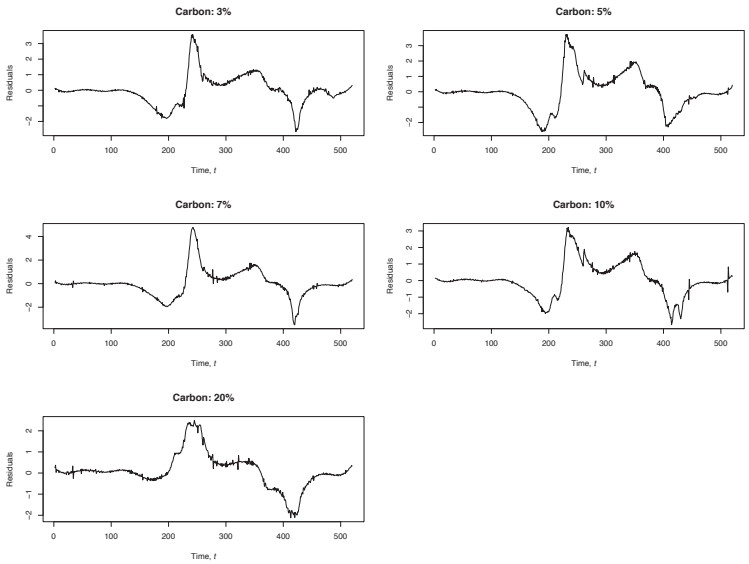


Figure 8.5: Differences between data and one-step ahead predictions.

As in Section 7.6, we will compare the posterior predictive distributions obtained via the best stochastic volatility model to our knowledge obtained from the plots shown

in Figures 2.12. The posterior predictive distributions for the Peak Current, shown in Figure 8.7, appears to produce a much closer fit than either the autoregressive or sinusoidal model. The posterior predictive distributions for the Peak Potential are located where we would expect. We note that these posterior predictive distributions appear to be identical and that we would expect some variation between the different Carbon levels. For the posterior predictive distributions in Figure 8.8, four out of the five Carbon levels also appear to be identical. The model appears to produce posterior predictive distributions for the Peak Separation in Potential located where we would expect.

We note that the posterior predictive distributions shown in Figures 8.9, 8.10 and 8.11, are for Carbon levels not used in the experiment. Therefore, we are unable to compare these posterior predictive distributions to any observed values. As already noted in Section 5.3, by including the posterior predictive distributions shown in Figures 8.9-8.11, we are simply illustrating the predictive usefulness of the best stochastic volatility model proposed in this chapter.

The posterior predictive distributions for the Peak Current shown in Figure 8.10 appear to be centered on values that are much lower than we would expect and compared to the values obtained from the sinusoidal model used in Chapter 7. There is a small amount of variation in the location of the distributions between different Carbon levels. We also note that the distributions in Figure 8.10 are very similar in shape. The posterior predictive distributions for the Peak Potential, shown in Figure 8.9, indicate more uncertainty about where the Peak Potential occurs. The posterior predictive distributions for the Peak Separation in Potential, shown in Figure 8.11, are centred on values that are slightly lower than what we expect based on our experience so far with regards to this particular data set. As we highlighted earlier, we have only concentrated on scan rate 2. A similar analysis was conducted for the other scan rates.

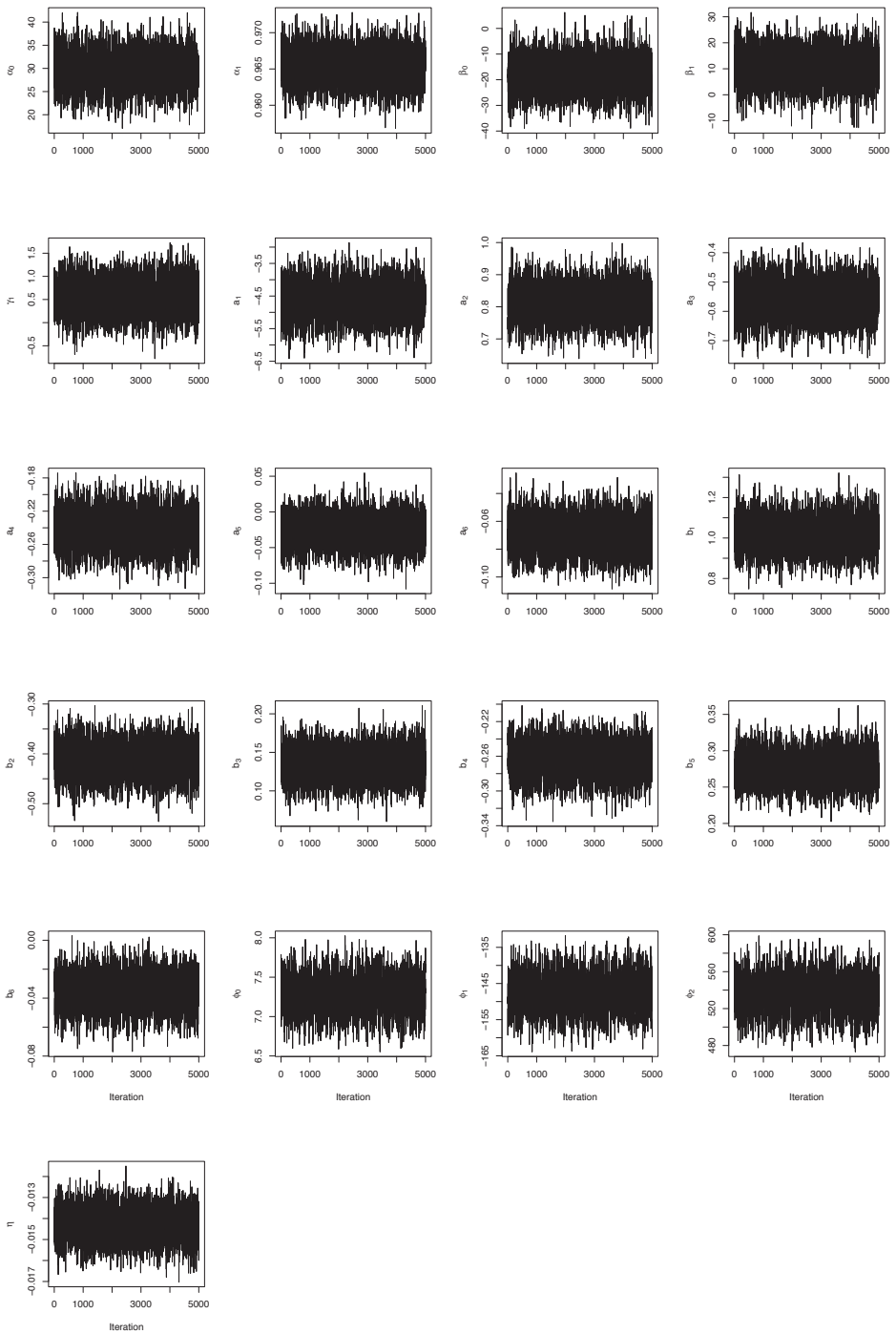


Figure 8.3: Trace plots of every 200^{th} sample generated by WinBUGS for all the parameters of the best stochastic volatility model for scan rate 2.

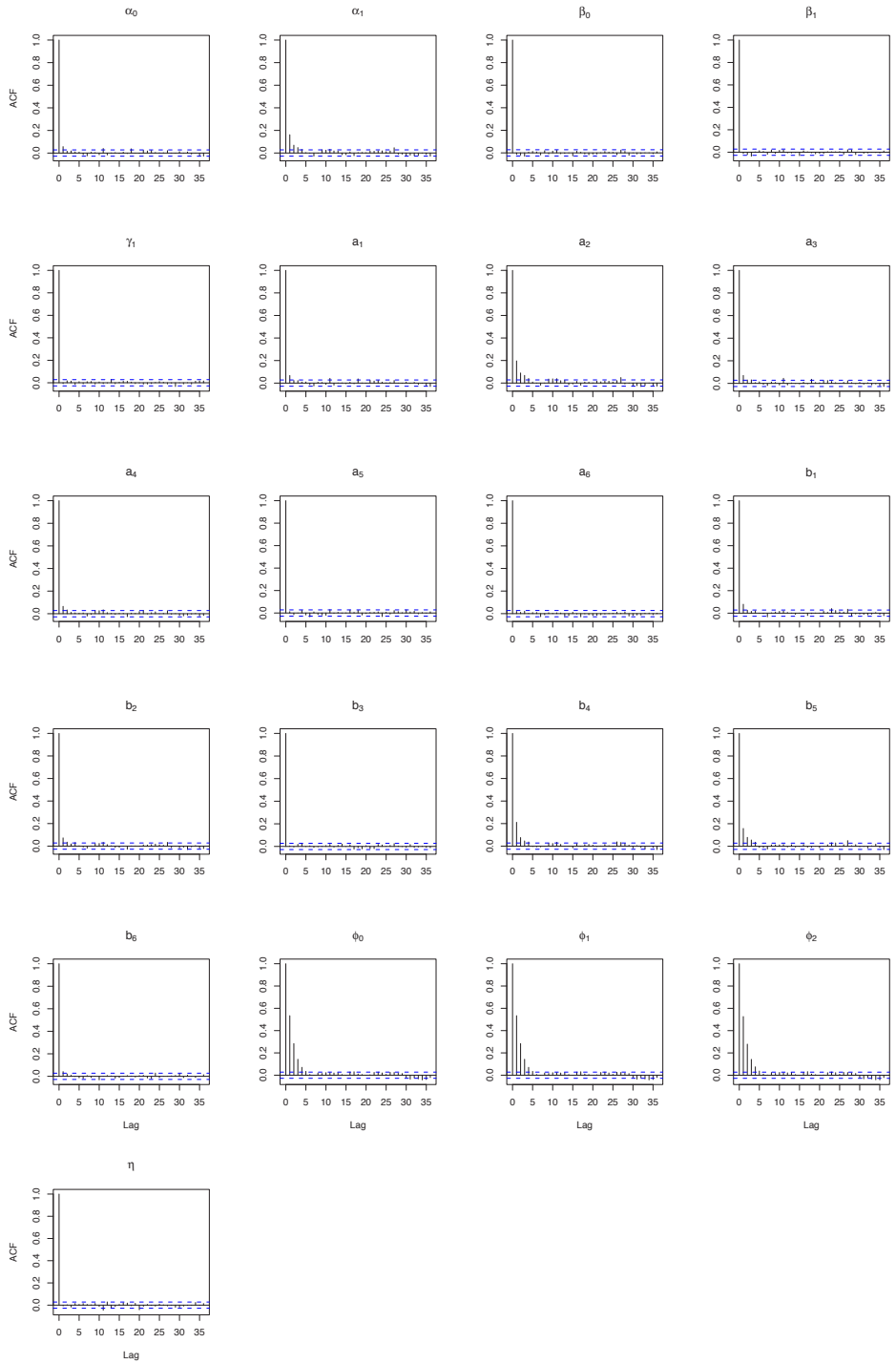


Figure 8.4: Autocorrelation plots of every 200^{th} sample generated by WinBUGS for all the parameters of the best stochastic volatility model for scan rate 2.

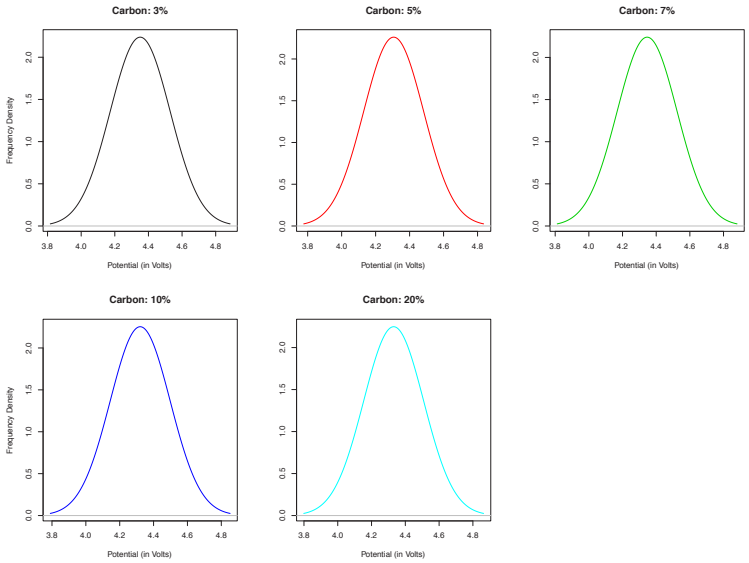


Figure 8.6: Posterior predictive distributions of Peak Potential for scan rate 2.

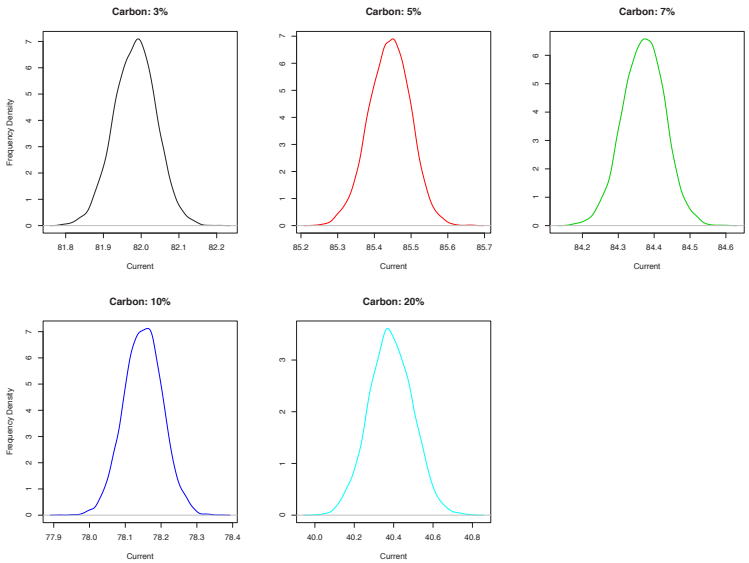


Figure 8.7: Posterior predictive distributions of Peak Current for scan rate 2.

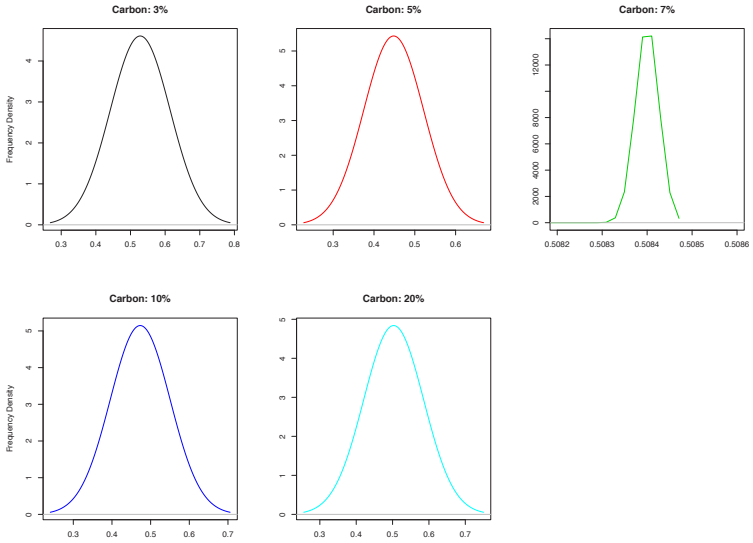


Figure 8.8: Posterior predictive distributions of Peak Separation in Potential for scan rate 2.

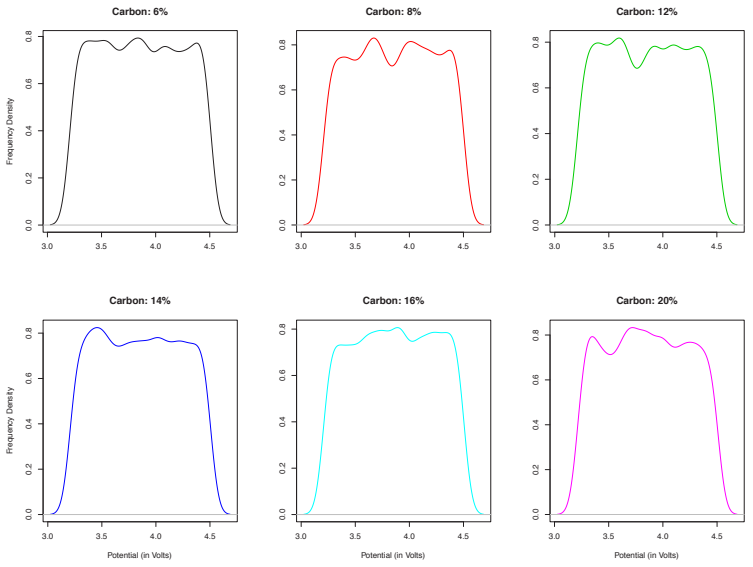


Figure 8.9: Posterior predictive distribution of Peak Potential for scan rate 2.

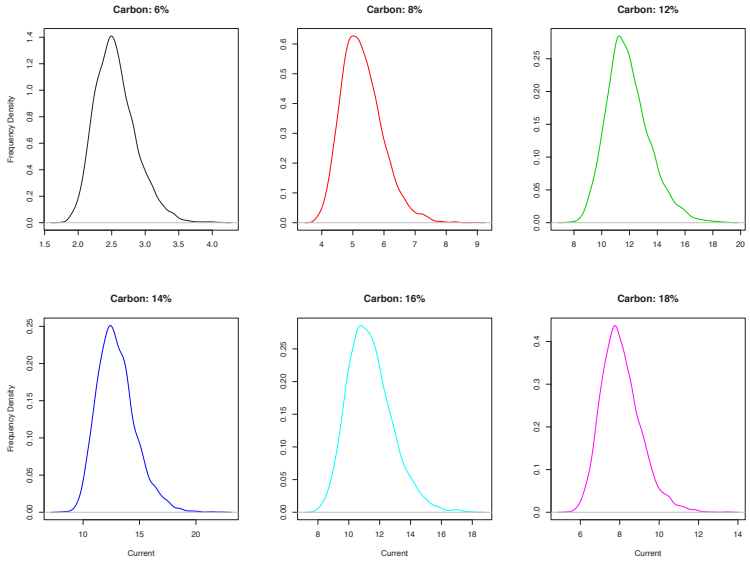


Figure 8.10: Posterior predictive distribution of Peak Current for scan rate 2.

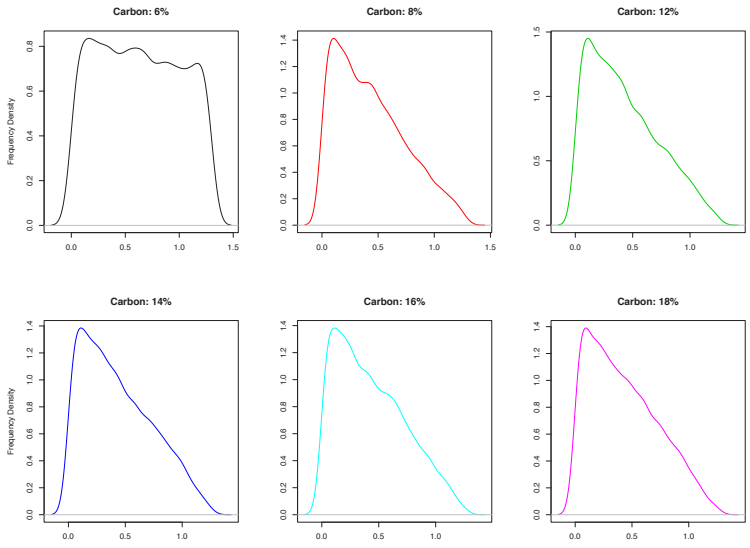


Figure 8.11: Posterior predictive distribution of Peak Separation in Potential for scan rate 2.

Chapter 9

Conclusions and Future Work

9.1 Introduction

This chapter draws a number of overall conclusions by comparing the various modelling strategies presented so far. The comparison using the PMCC allows us to select the best set of models for the data. This chapter also provides some comments pointing out the limitations of the proposed methods, and it ends with a discussion of a few ideas for further developing the analysis and empirical modelling for cyclic voltammograms.

9.2 Model Comparisons

The Bayesian methods used for making inference throughout the thesis also allow us to compare various models presented previously in Chapters 5-8. As mentioned in Section 3.7.2, the best model is the one which provides the minimum value of the PMCC. For the overall comparison, we still use the PMCC since all the models under consideration are based on the assumption of normally distributed error distributions. As stated in Section 3.7.2, the PMCC minimises the expected value of the

squared-error loss function which is most suitable for use with normally distributed error distributions.

In the overall comparison, we do not include the regression models for the summary statistics presented in Chapter 5 since these models do not describe the full data set for a cyclic voltammogram as the other models presented in the subsequent chapters do. As discussed before, these simple to use off-the-shelf regression models can provide a quick and crude analysis of one characteristic at a time, independent of the other characteristics. This may lead to problems in the analysis since the voltammogram characteristics are dependent, for example, the Minimum Current, $I_{s,k}^{(\min)}$, cannot be greater than the Peak Current, $I_{s,k}^{(\max)}$. A single time series model for a voltammogram avoids these types of problems and are compared next.

Scan Rate	Model	PMCC
1	$M_1(R_\alpha = 30, R_\gamma = 1)$	389.56
	$M_4(R_\alpha = 21, R_\beta = 15, R_\gamma = 1, R_f = 20)$	374.50
	$M_7(R_\gamma = 1, R_f = 6, R_\phi = 3)$	131783.20
2	$M_3(R_\alpha = 11, R_\beta = 11, R_\gamma = 1)$	75.22
	$M_4(R_\alpha = 8, R_\beta = 10, R_\gamma = 1, R_f = 11)$	69.98
	$M_7(R_\gamma = 1, R_f = 6, R_\phi = 2)$	9786.03
3	$M_3(R_\alpha = 12, R_\beta = 11, R_\gamma = 1)$	351.53
	$M_4(R_\alpha = 8, R_\beta = 5, R_f = 11, R_\gamma = 1)$	311.43
	$M_7(R_\gamma = 1, R_f = 6, R_\phi = 3, R_\nu = 2)$	31216.89

Table 9.1: The value of the PMCC for the best time series model found for each of the three different scan rates. The first model is the best autoregressive model, the second is the best sinusoidal model and the third is the best stochastic volatility model for each scan rate.

Table 9.1 provides the value of the PMCC for the best time series model found for each of the three different scan rates from the three different classes of models, viz. autoregressive, sinusoidal and stochastic volatility presented in Chapters 6, 7, and 8 respectively. We observe that the stochastic volatility models are the worst in each case. This is expected, since the PMCC penalises a model through its predictive variance and the predictions using the stochastic volatility models are very volatile. The values of the criterion for the autoregressive and sinusoidal models are on a comparable scale, and we can see from Table 9.1 that the sinusoidal models are the best ones for the data for each of the three different scan rates. Although the sinusoidal models turn out to be the best, we recall from Chapter 7 that the sinusoidal models do not describe realistic differences in values of Current due to the differences in Carbon levels. This can be a potential problem if the chemists want to predict values of Current for different levels of Carbon only. However these models are recommended since the primary objective of the thesis is to describe and analyse characteristics of cyclic voltammograms.

We end this section by discussing a few limitations of the modelling approaches presented in the thesis. First, we have only modelled the data for different scan rates independently but scan rates affect the behaviour of the different time series, see Figure 2.11. Ideally a joint model of the data from all three different scan rates should be formulated, see Section 9.3.

Second, we note all the models presented here are well-known time series models. The primary reason for using these models is the need to have easily interpretable models aimed at non-specialist practitioners in Chemistry. Although the simple models have been able to cope with the fitting of data sets that have a large number of observations, we believe there is scope for applying more complex models and alternative techniques such as non-parametric methods, for example, see Fan and Yao (2005) and Chen et al. (2004). In the following section, we discuss some of

these techniques. We note that the Bayesian predictive inference and computation methods for the different characteristics of interest by postulating a single time series model can be adopted for any future more complex model for the data from a cyclic voltammogram.

Lastly, note that throughout the thesis we have used the default vague prior distributions for all the model parameters. These analyses, although successful, can be greatly enhanced by carefully incorporating informative prior distributions. However, that would require a substantial study on prior elicitation and is beyond the scope of this thesis. We simply note that the analysis presented in this thesis will be useful for providing guidance on the choice, scale and location of the prior distributions.

9.3 Future Work

As mentioned before, we have modelled the data independently for the three different scan rates. However, it is desirable to have a single hierarchical model for the data from all three scan rates. Such a model will allow learning across the scan rates by sharing information regarding the behaviour of the characteristics of the data. The main difficulty in developing a single hierarchical model lies in the time series of different lengths that are obtained for different scan rates, see Table 2.1. These unbalanced time series can perhaps be modelled using a variable dimensional parameter space and analysed using the reversible jump Markov Chain Monte Carlo (RJMCMC), for example, see Green (1995) and Dellaportas et al. (2002). Thus by combining a unified hierarchical model with an RJMCMC sampling algorithm a powerful methodology for modelling and analysis of cyclic voltammogram data can be developed, although this would be computationally expensive and would require a long time to generate the results.

The models presented in this thesis can also be improved by well-known variable dimensional B-spline models, for example, see De Boor (2002). Other methods such as principal component analysis, for example, see Johnson and Wichern (2002) could also be considered. This thesis is a start on empirical modelling of cyclic voltammogram data. Such strategies, combined with chemists' knowledge regarding the physical properties of the experiments, can bear further fruitful research in this area.

Appendix A

Data

A.1 Plots of Data for Scan Rates 1 and 3

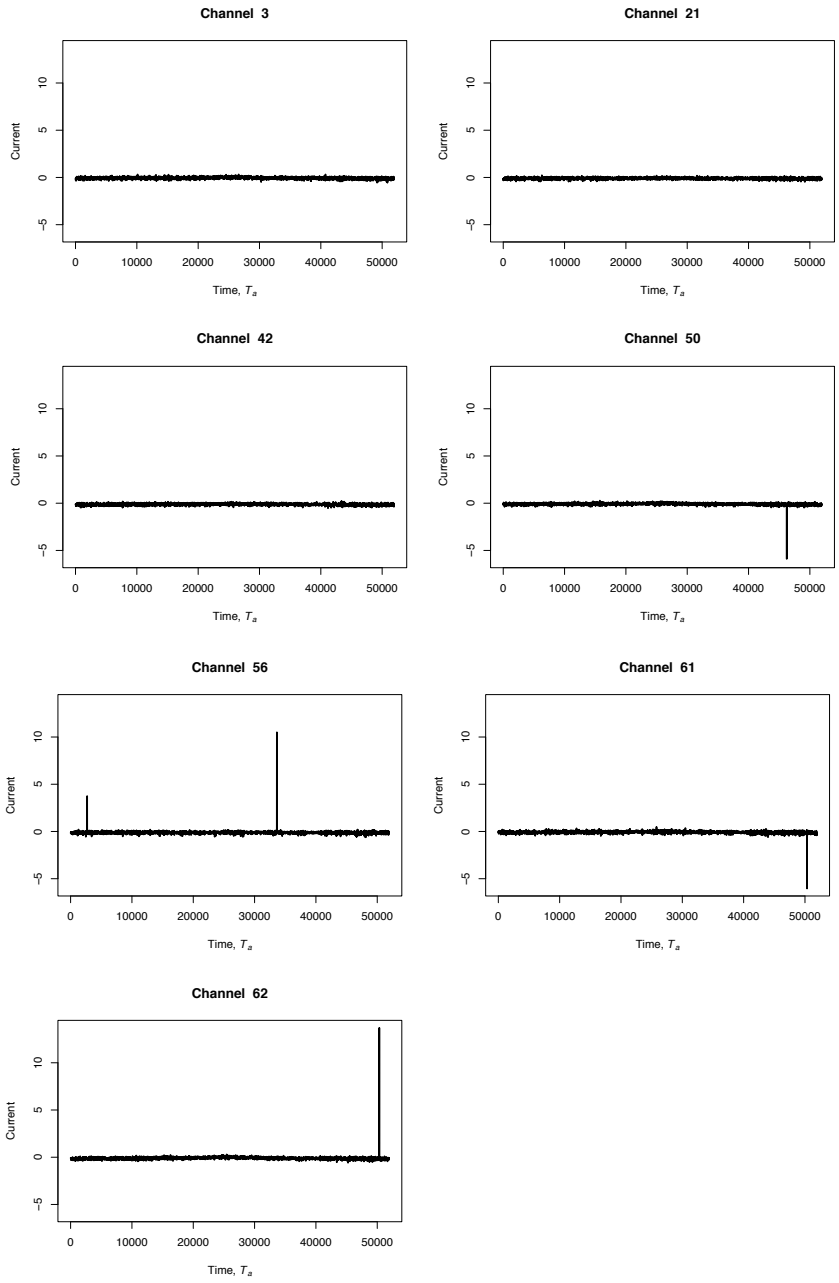


Figure A.1: Time series plots of Current with Carbon set at 0% and scan rate 1, where elapsed time T_a is measured in seconds.

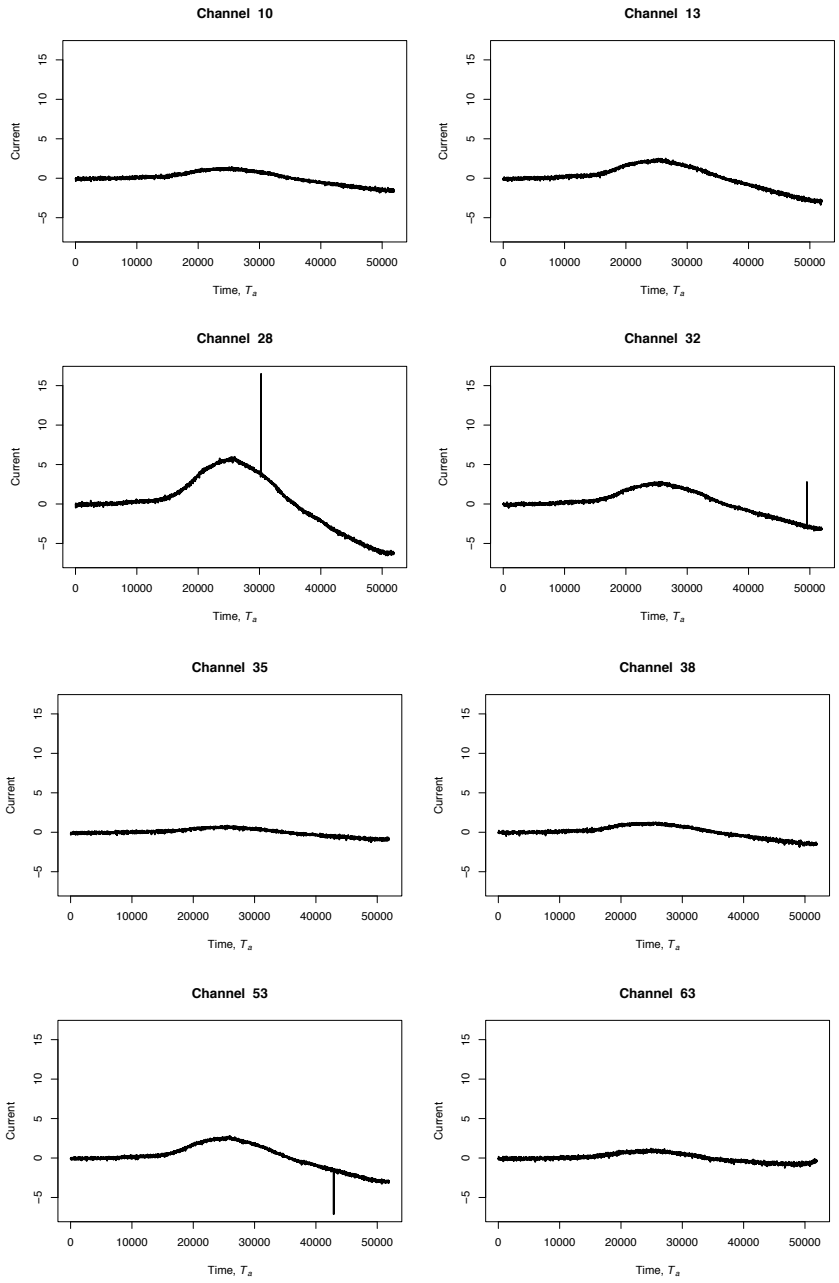


Figure A.2: Time series plots of Current with Carbon set at 1% and scan rate 1, where elapsed time T_a is measured in seconds.

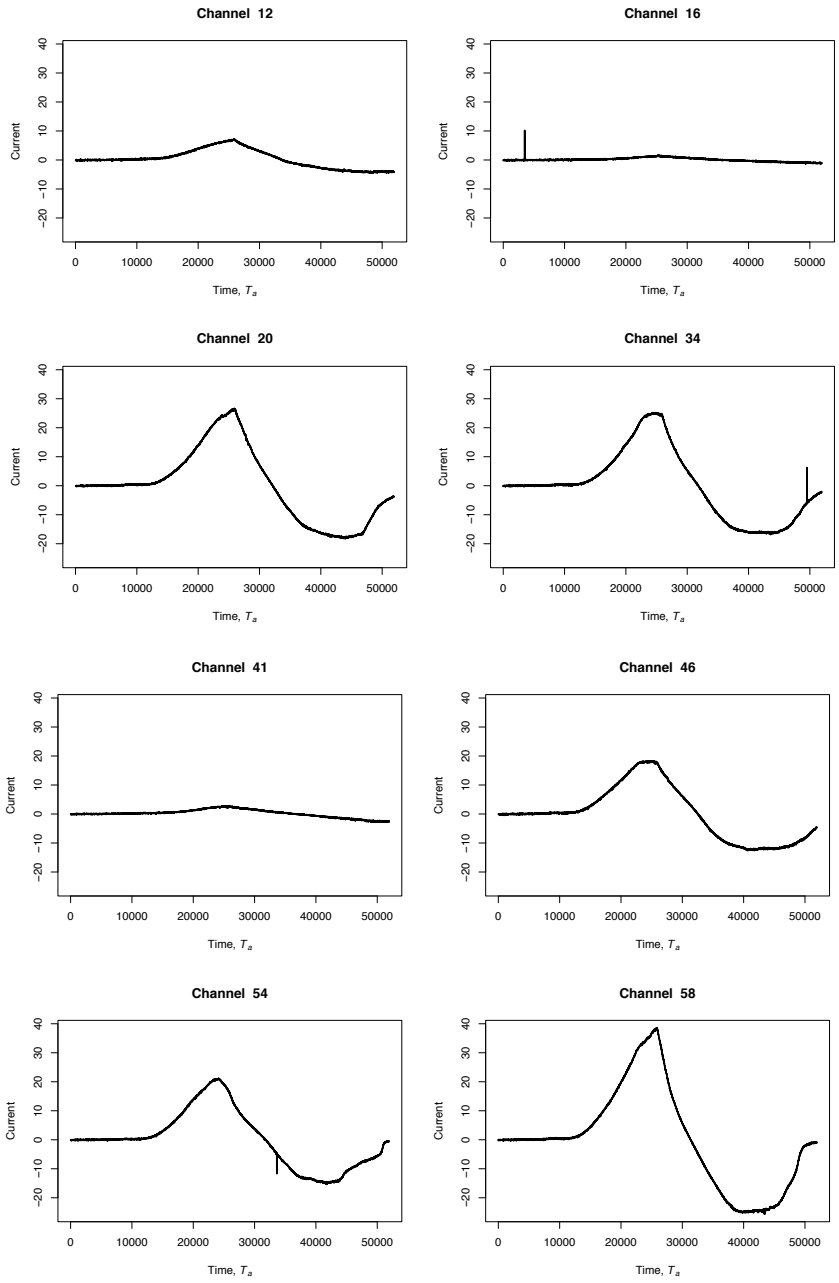


Figure A.3: Time series plots of Current with Carbon set at 2% and scan rate 1, where elapsed time T_a is measured in seconds.

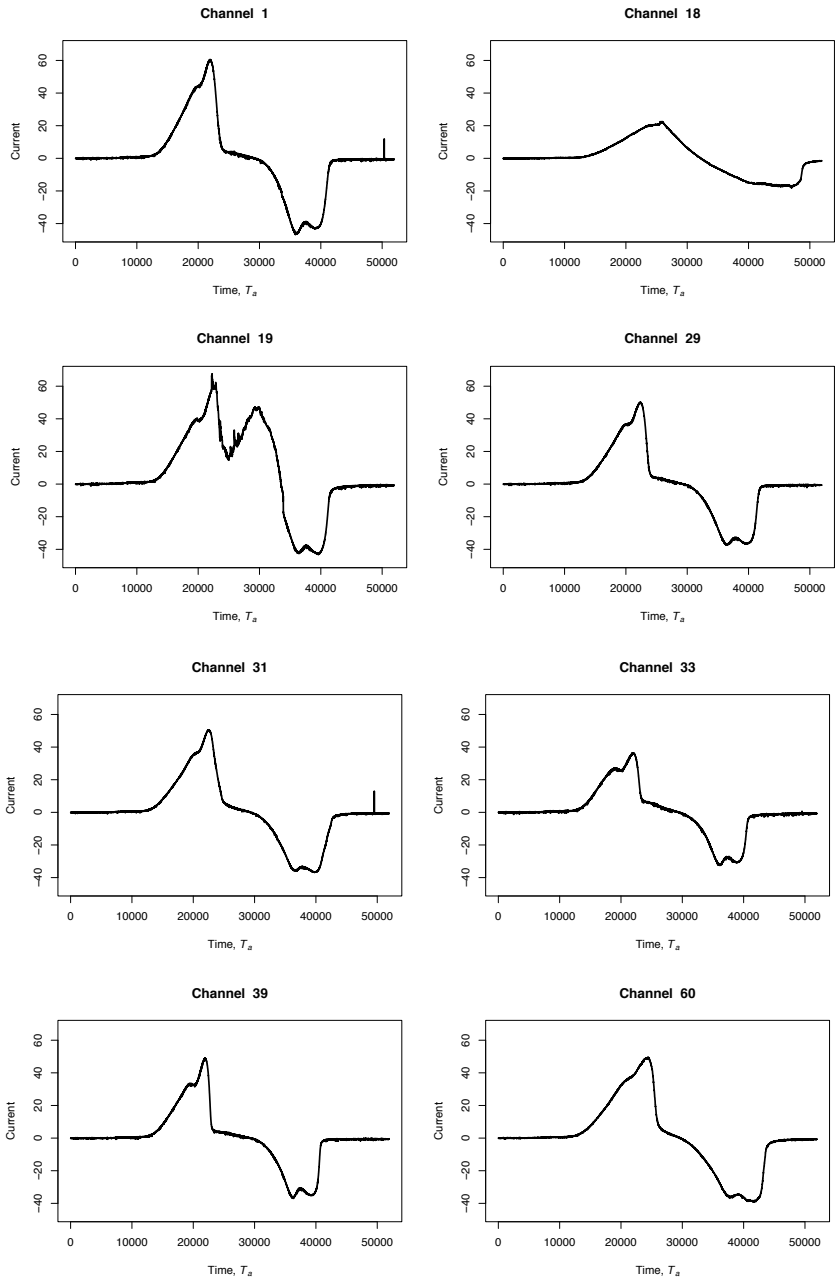


Figure A.4: Time series plots of Current with Carbon set at 3% and scan rate 1, where elapsed time T_a is measured in seconds.

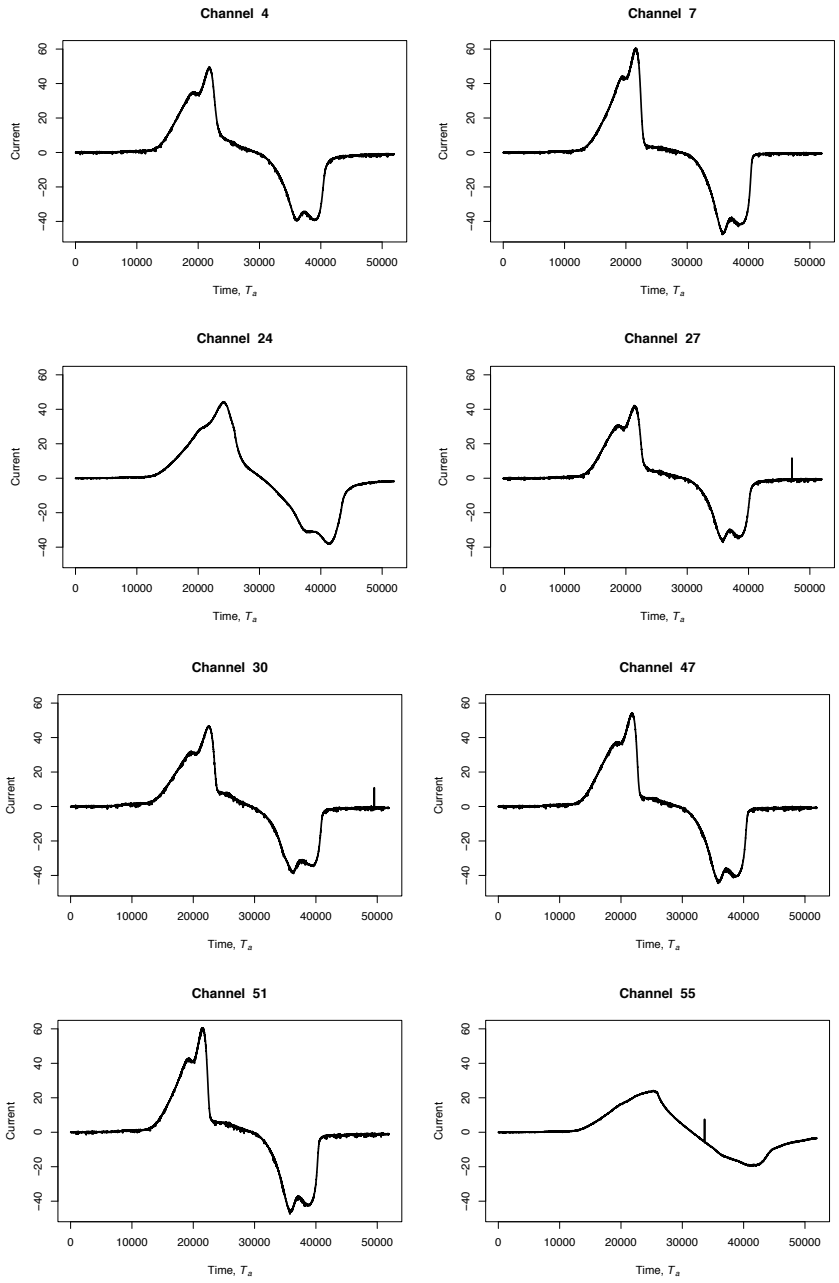


Figure A.5: Time series plots of Current with Carbon set at 5% and scan rate 1, where elapsed time T_a is measured in seconds.

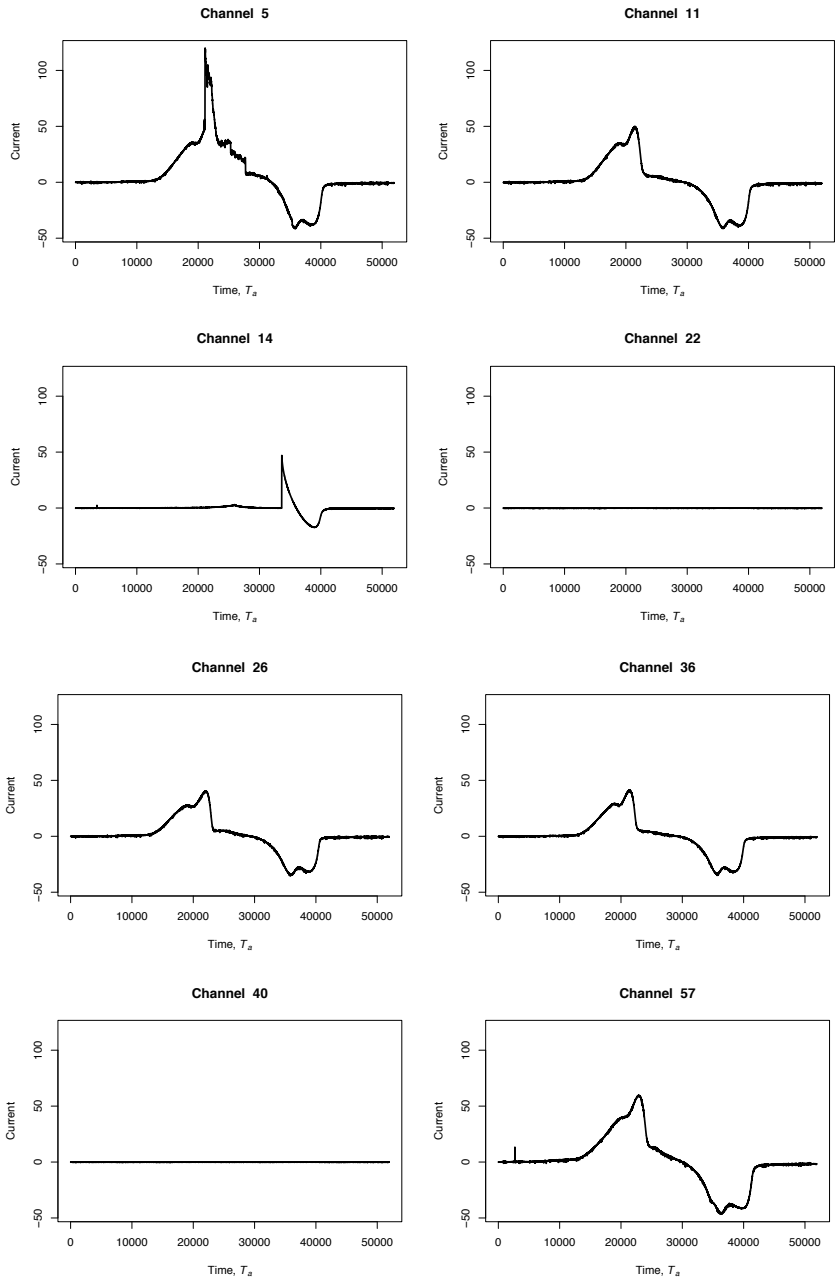


Figure A.6: Time series plots of Current with Carbon set at 7% and scan rate 1, where elapsed time T_a is measured in seconds.

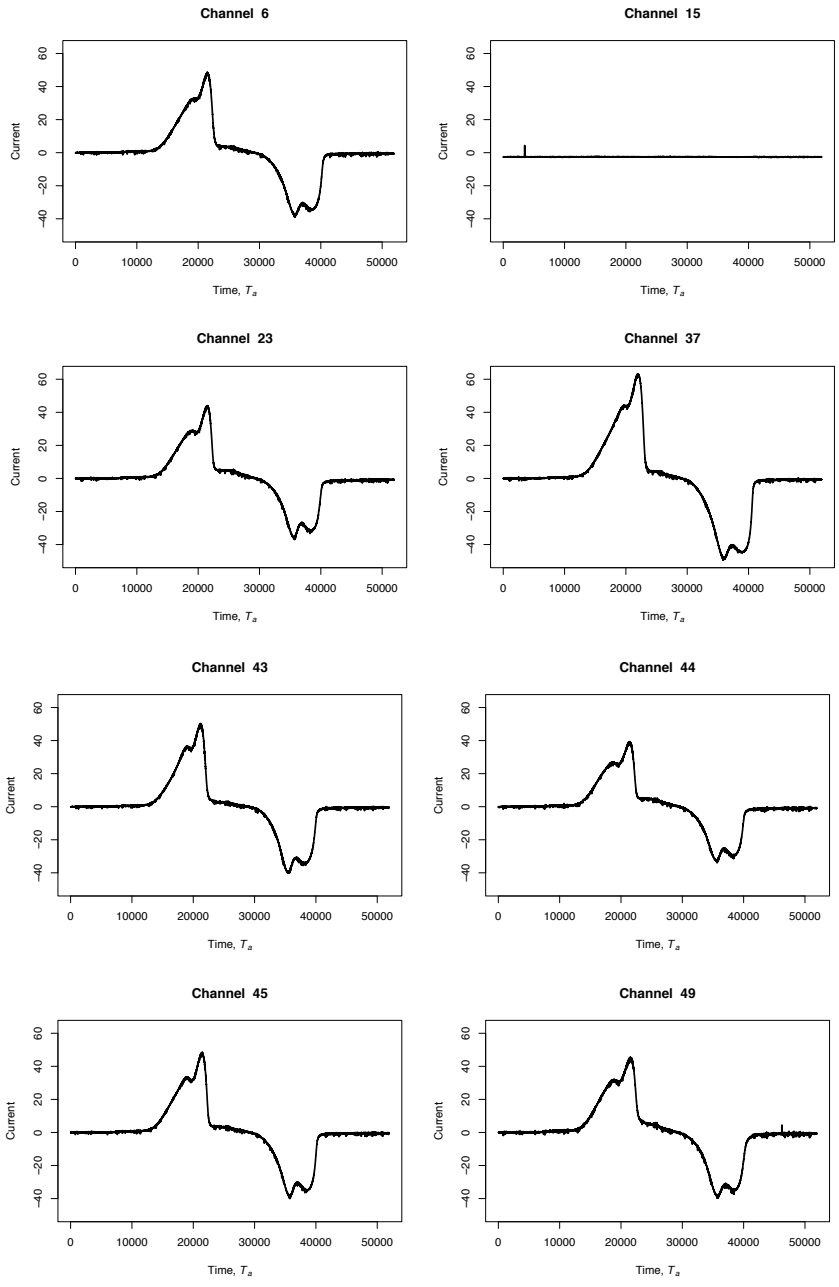


Figure A.7: Time series plots with Carbon set at 10% and scan rate 1, where elapsed time T_a is measured in seconds.

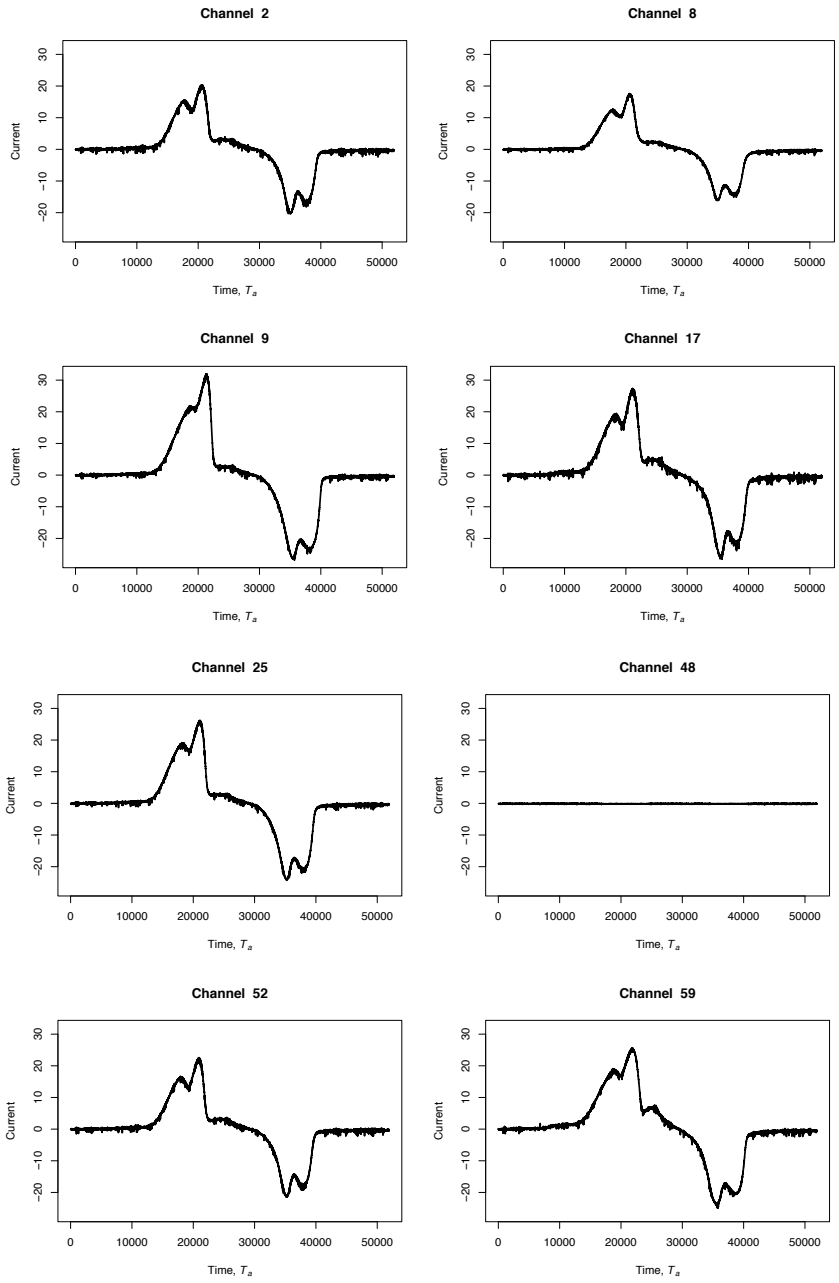


Figure A.8: Time series plots of Current with Carbon set at 20% and scan rate 1, where elapsed time T_a is measured in seconds.

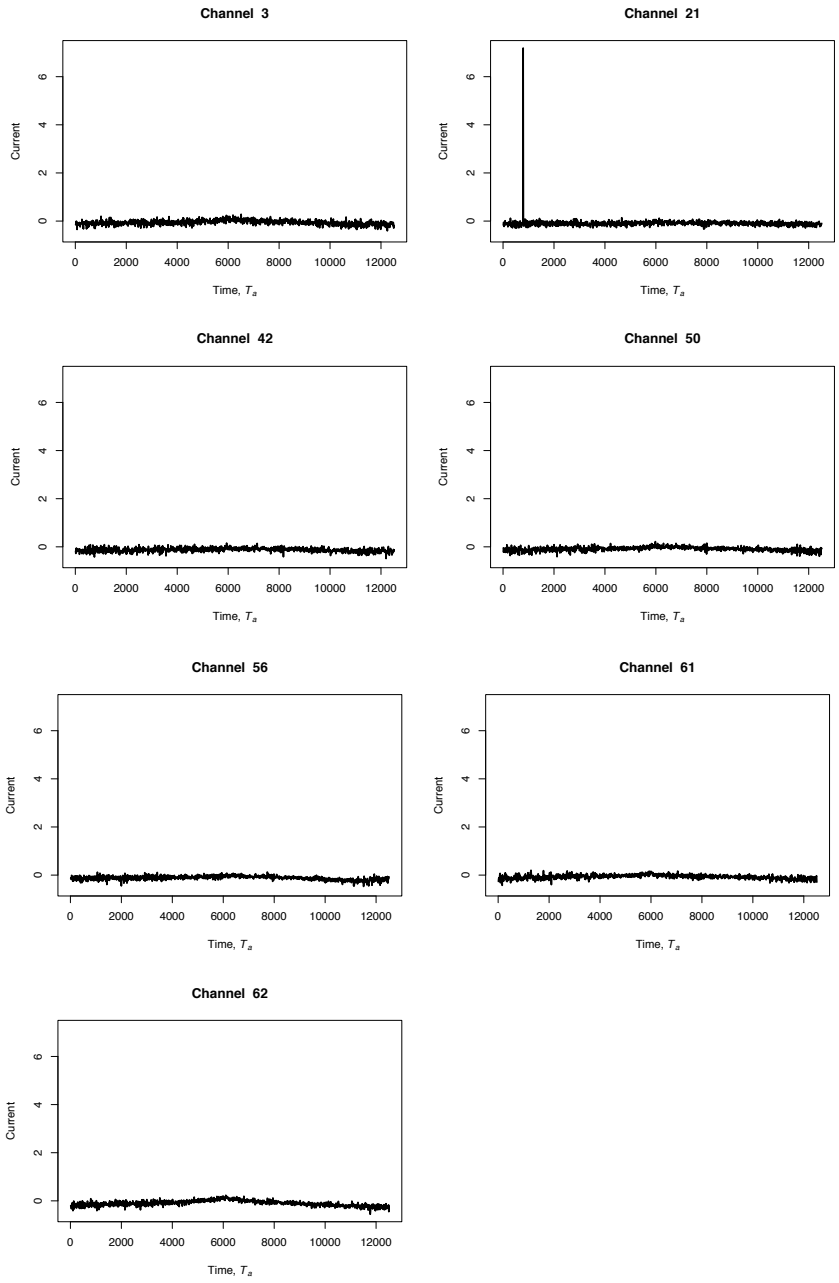


Figure A.9: Time series plots of Current with Carbon set at 0% and scan rate 3, where elapsed time T_a is measured in seconds.

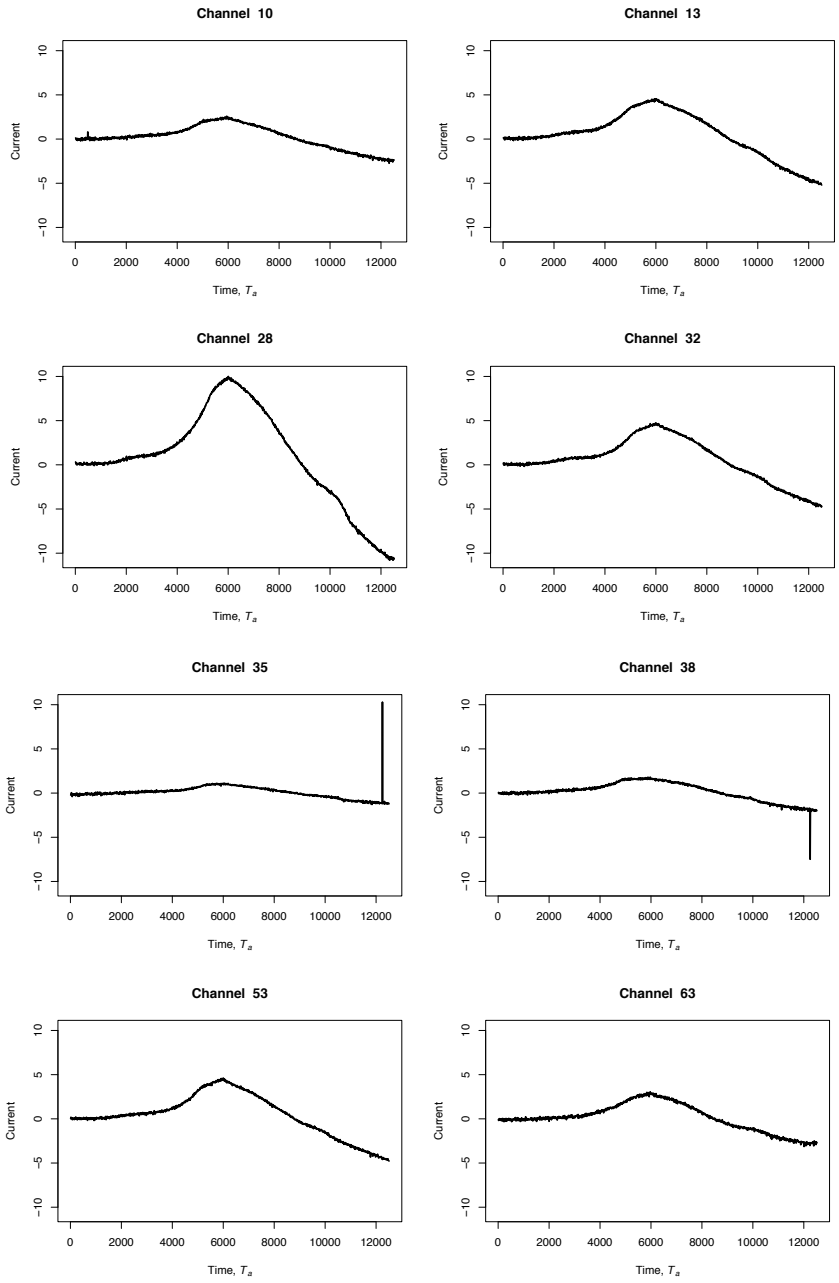


Figure A.10: Time series plots of Current with Carbon set at 1% and scan rate 3, where elapsed time T_a is measured in seconds.

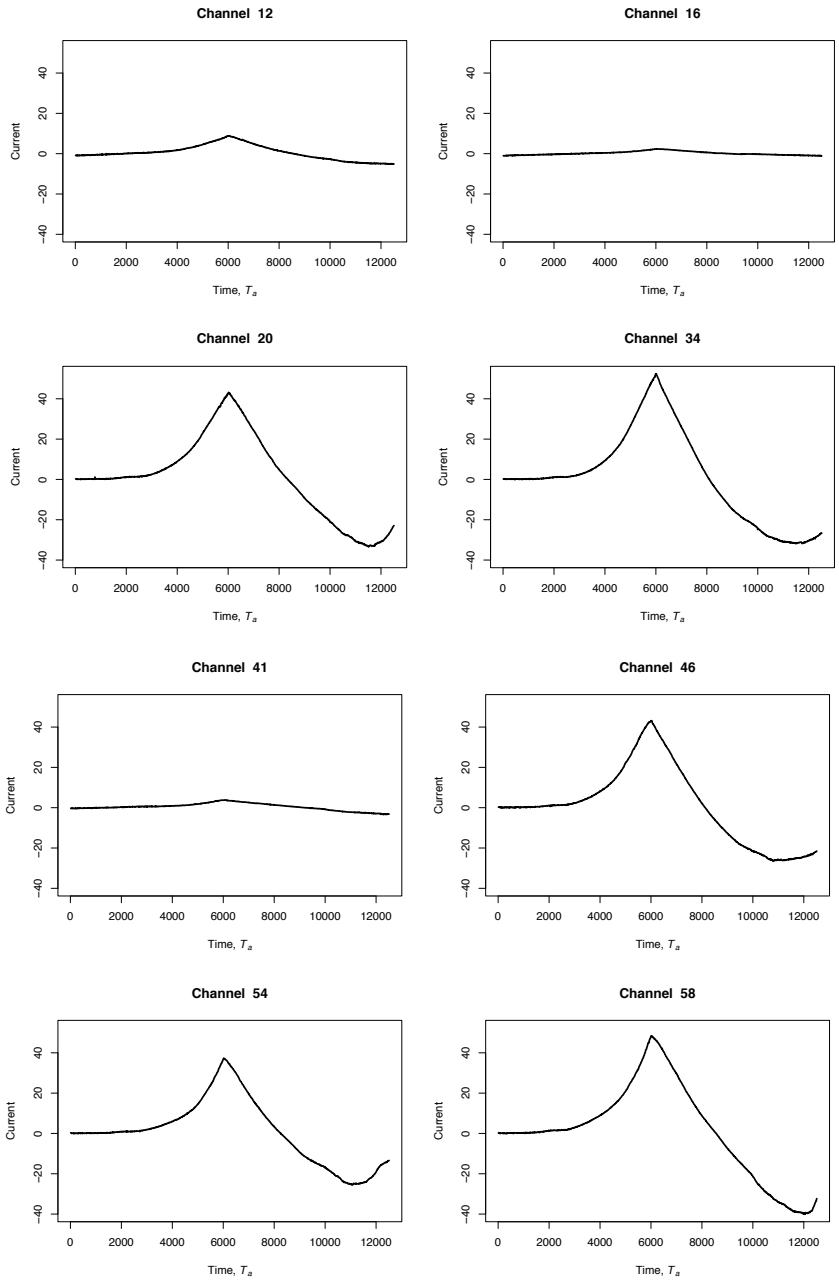


Figure A.11: Time series plots of Current with Carbon set at 2% and scan rate 3, where elapsed time T_a is measured in seconds.

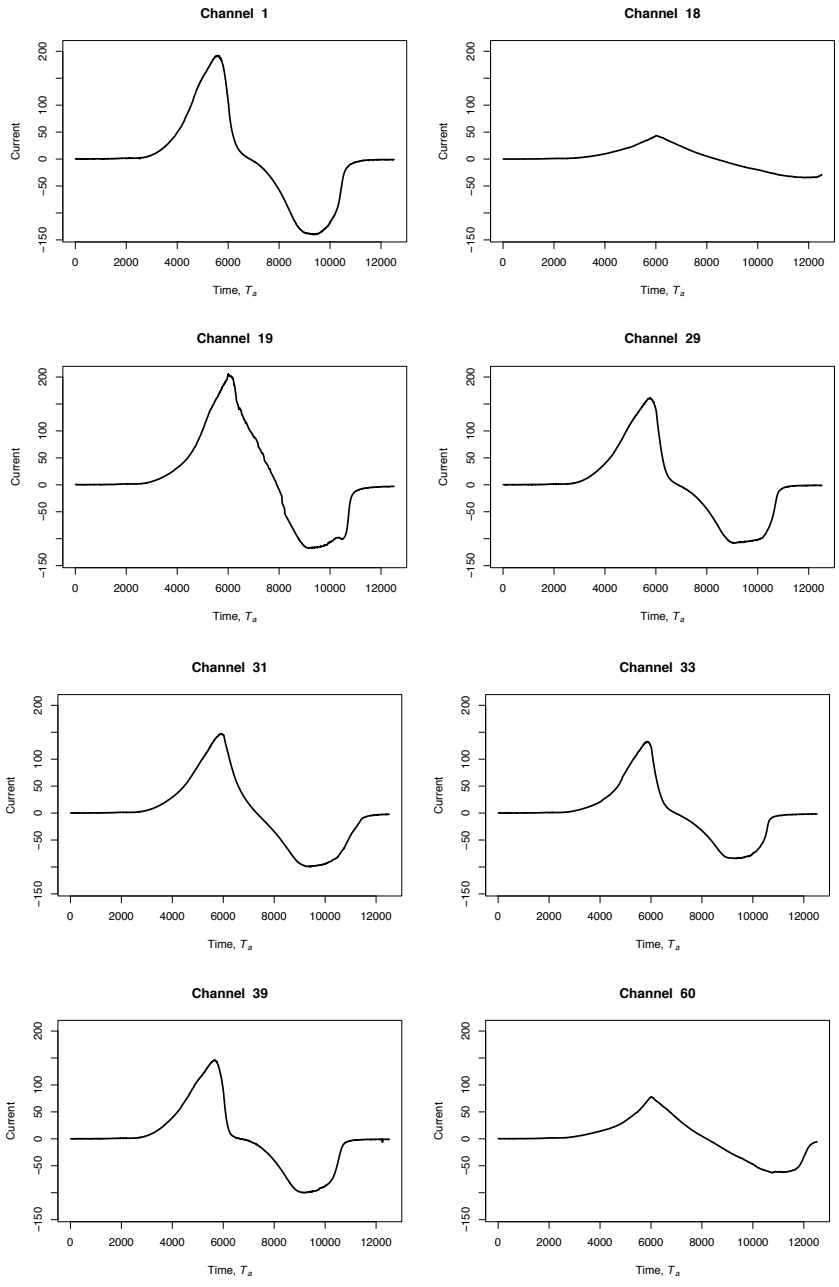


Figure A.12: Time series plots of Current with Carbon set at 3% and scan rate 3, where elapsed time T_a is measured in seconds.

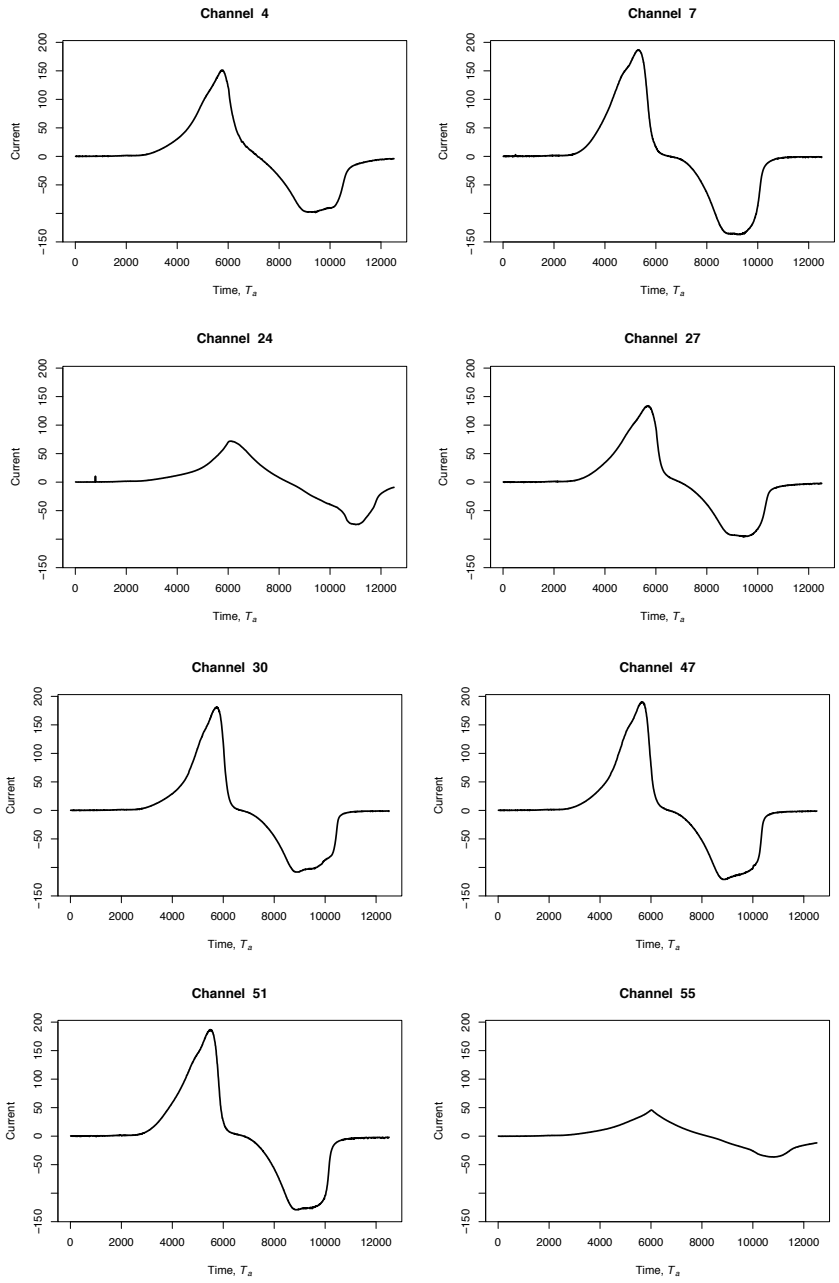


Figure A.13: Time series plots of Current with Carbon set at 5% and scan rate 3, where elapsed time T_a is measured in seconds.

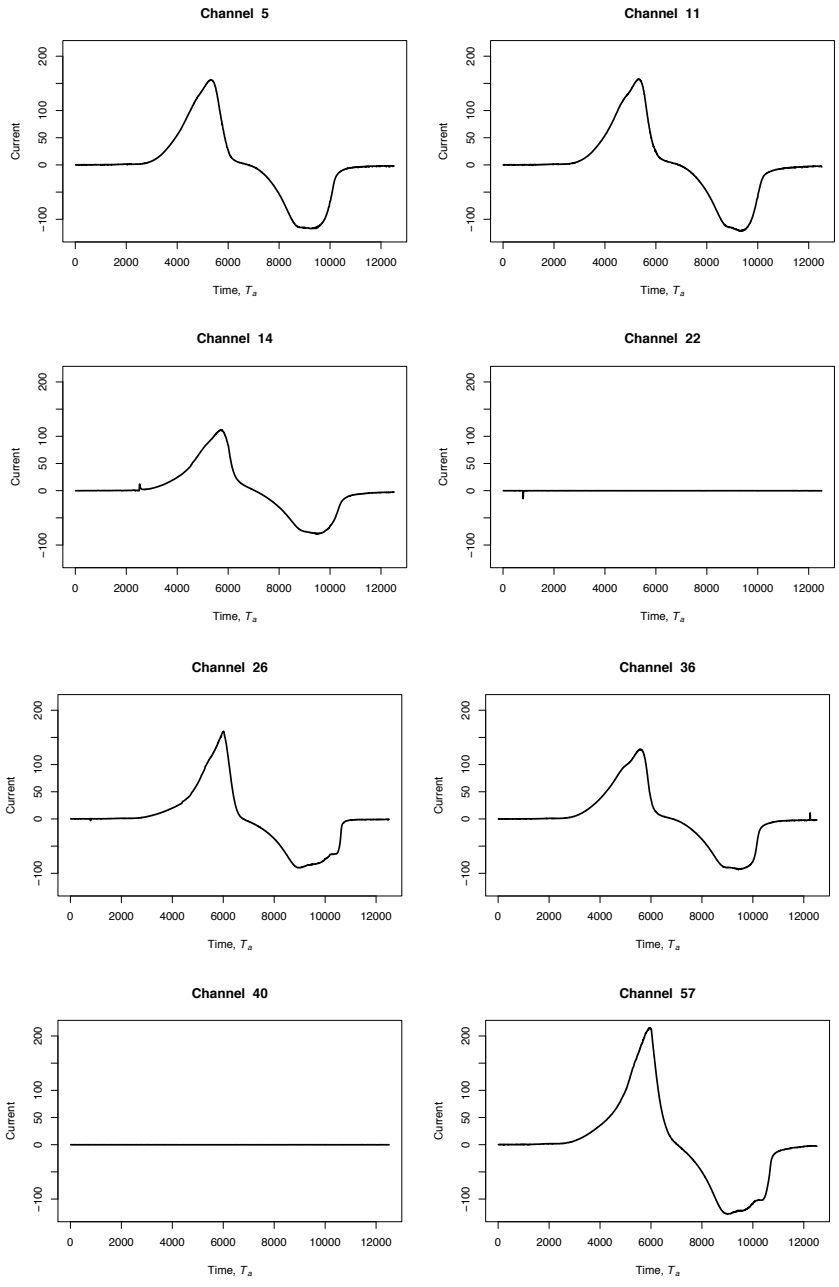


Figure A.14: Time series plots of Current with Carbon set at 7% and scan rate 3, where elapsed time T_a is measured in seconds.

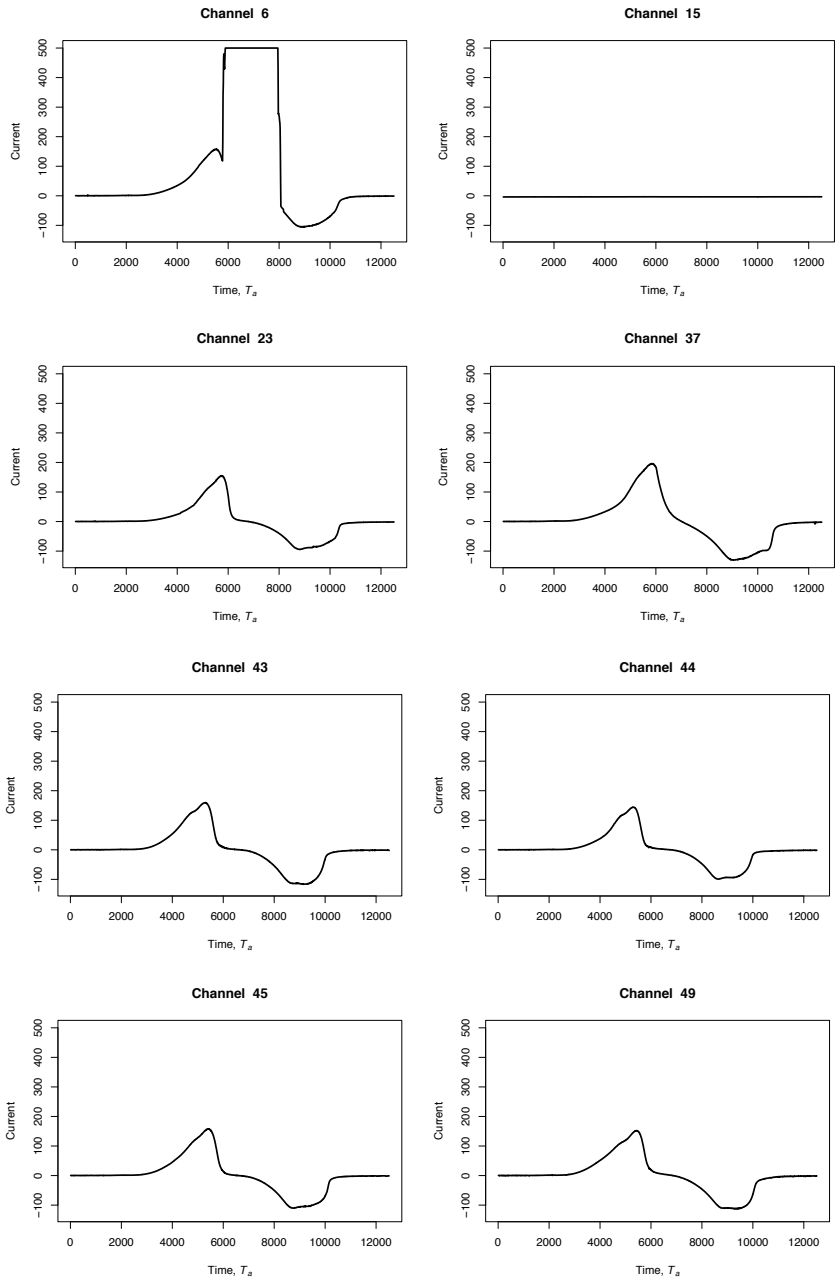


Figure A.15: Time series plots of Current with Carbon set at 10% and scan rate 3, where elapsed time T_a is measured in seconds.

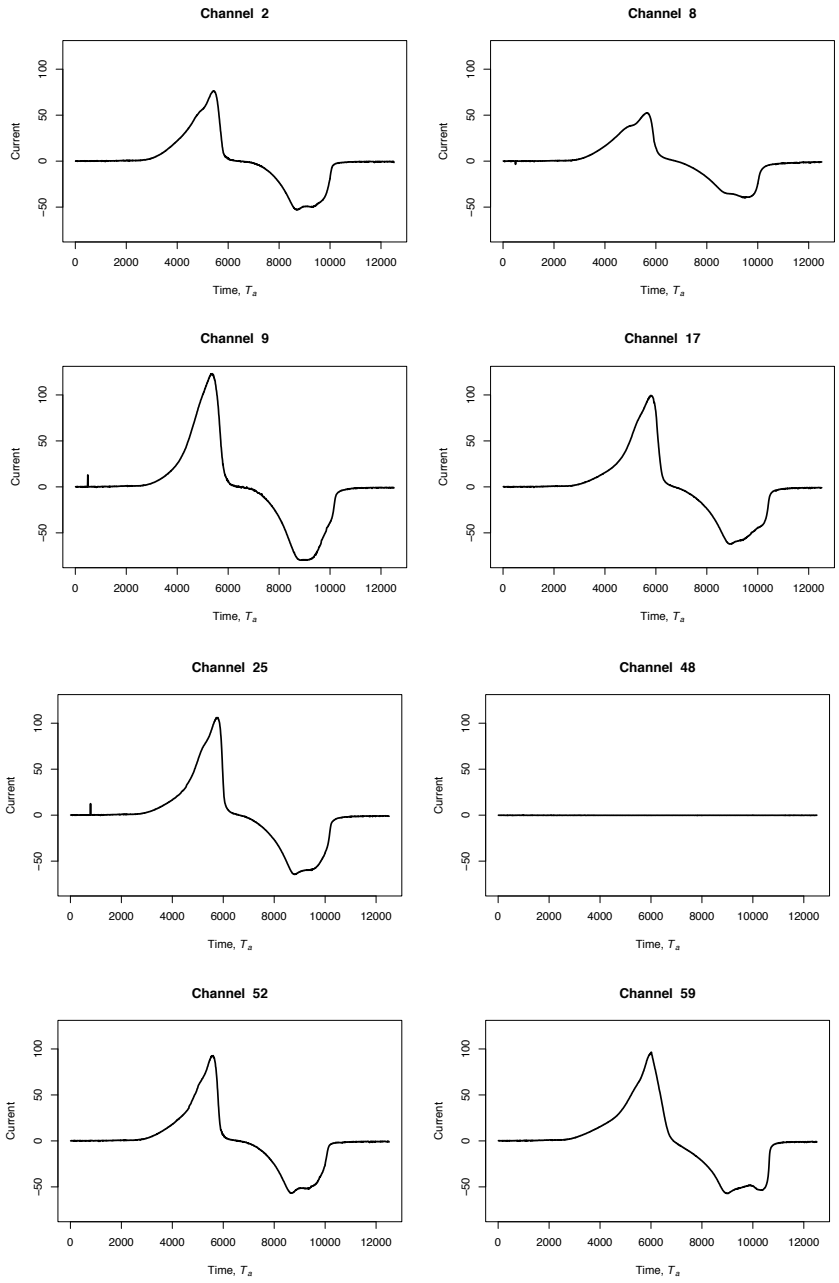


Figure A.16: Time series plots of Current with Carbon set at 20% and scan rate 3, where elapsed time T_a is measured in seconds.

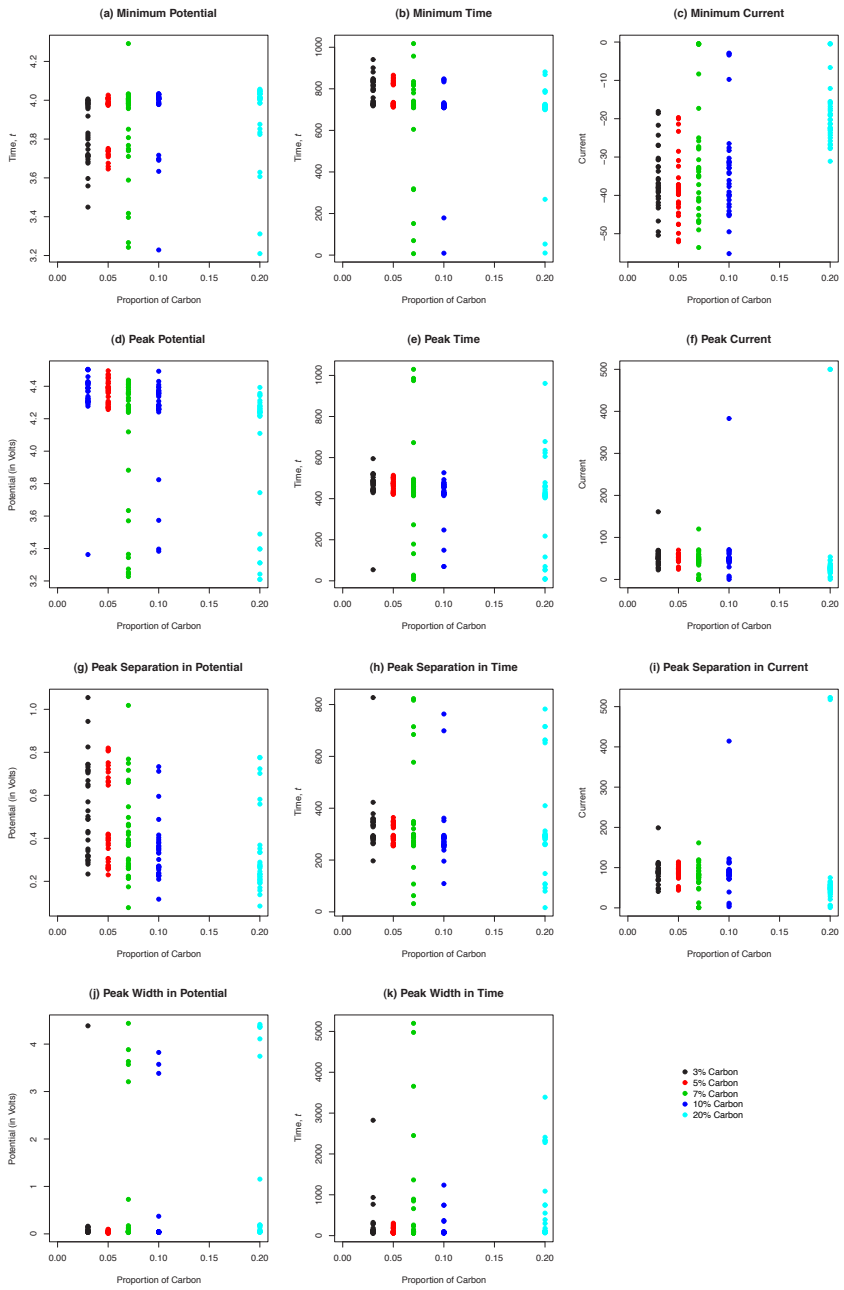


Figure A.17: Plots of characteristics of interest for scan rate 1 at each Carbon level using the raw data, that is, outliers have not been removed, where t is as defined in Table 2.1.

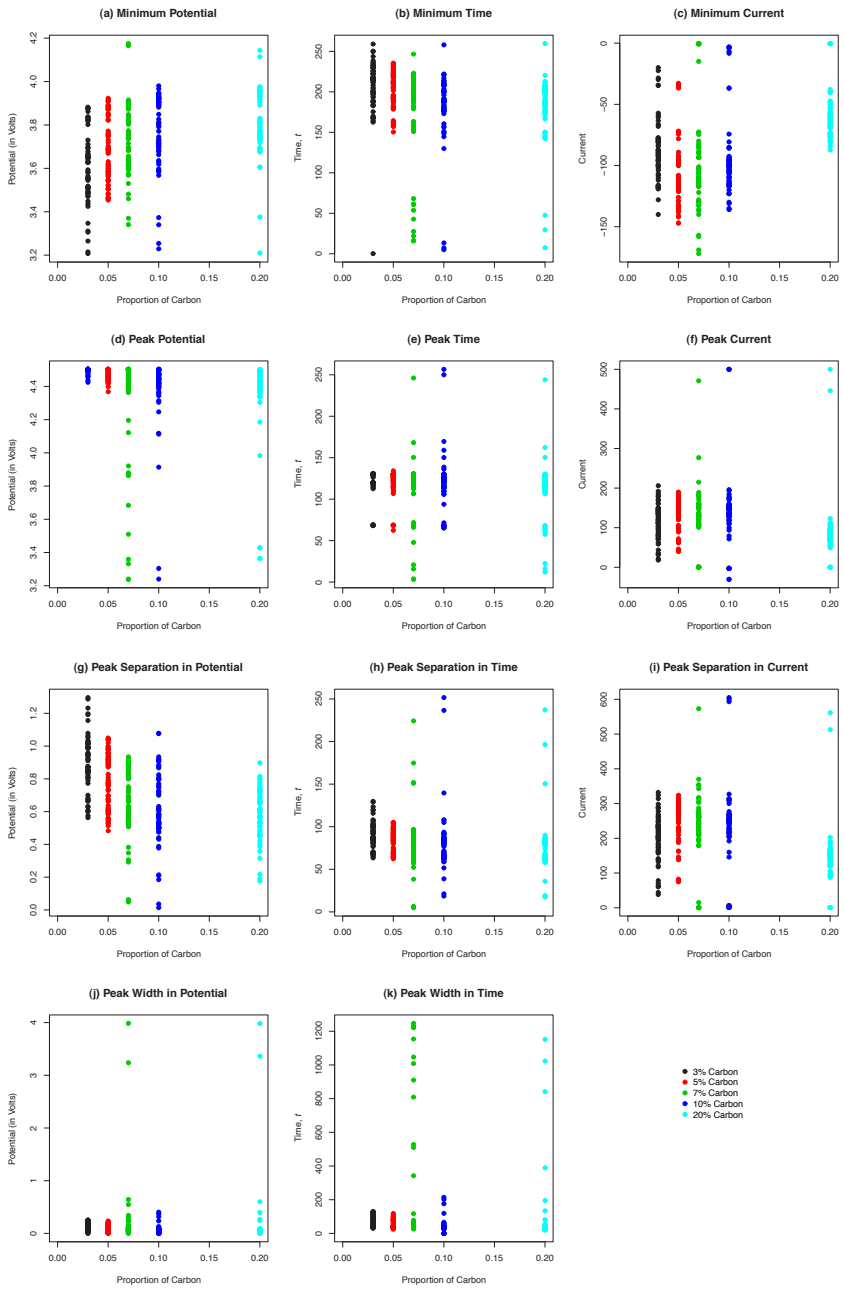


Figure A.18: Plots of characteristics of interest for scan rate 3 at each Carbon level using the raw data, that is, outliers have not been removed, where t is as defined in Table 2.1

Appendix B

B.1 MCMC Diagnostic Plots

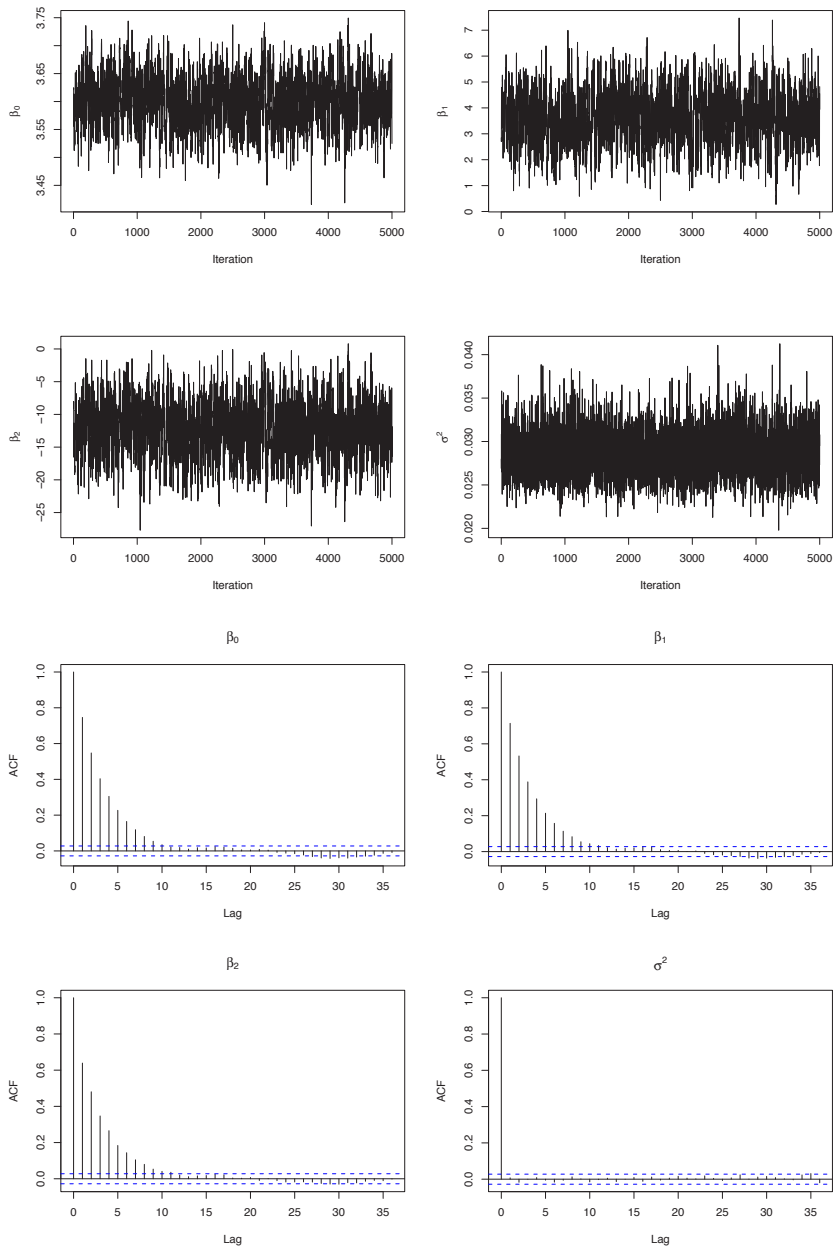


Figure B.1: Trace and autocorrelation plots of every fifth sample generated by the Gibbs sampler for all the parameters of the quadratic model for Minimum Potential.

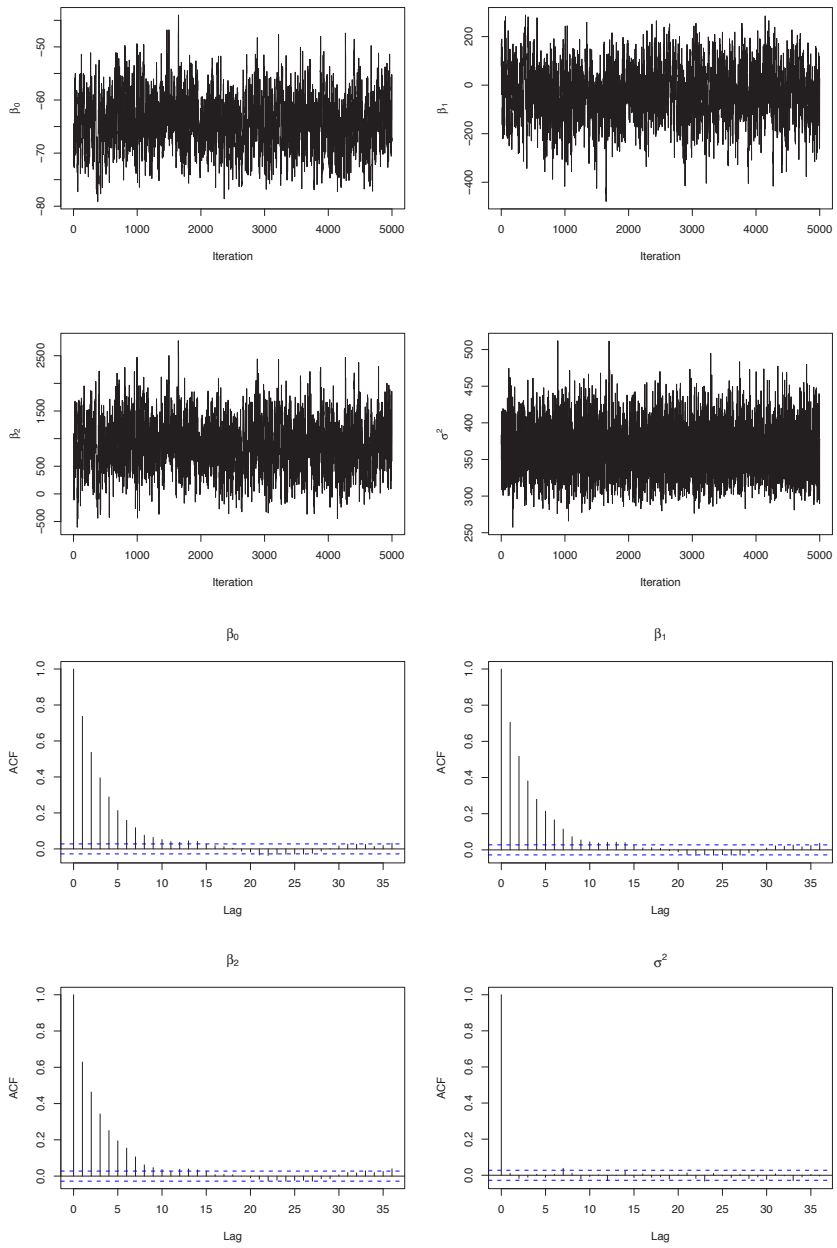


Figure B.2: Trace and autocorrelation plots of every fifth sample generated by the Gibbs sampler for all the parameters of the quadratic model for Minimum Current.

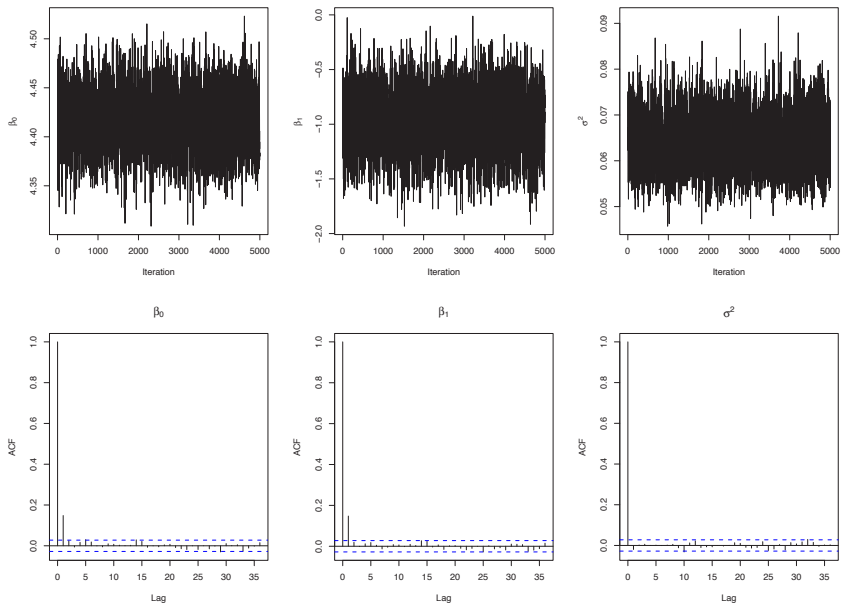


Figure B.3: Trace and autocorrelation plots of every fifth sample generated by the Gibbs sampler for all the parameters of the linear model for Peak Separation in Current.

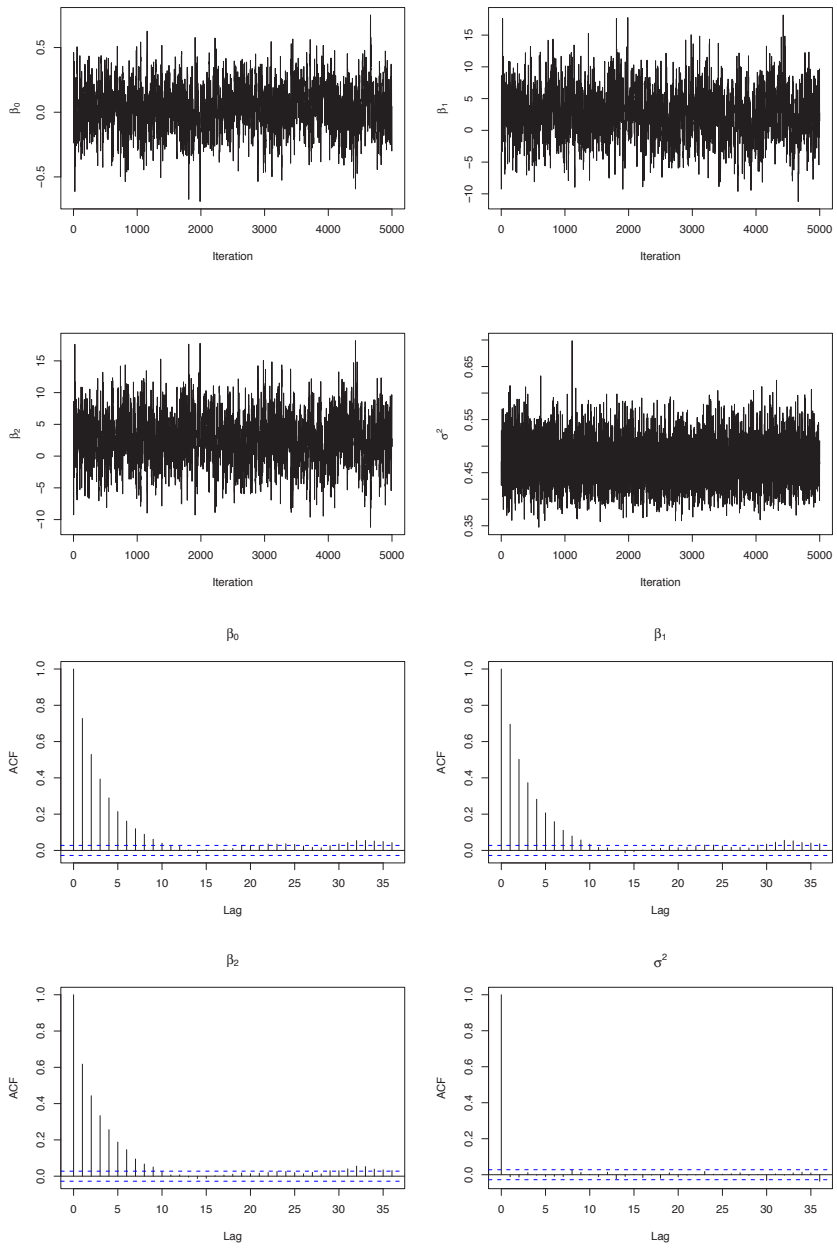


Figure B.4: Trace and autocorrelation plots of every fifth sample generated by the Gibbs sampler for all the parameters of the quadratic model for Peak Width in Potential.

B.2 Density Plots

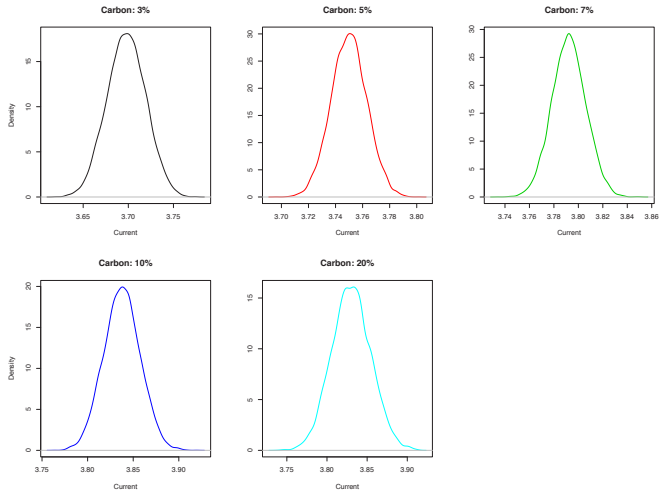


Figure B.5: Density plots of posterior predictive distributions of Minimum Potential for scan rate 2.

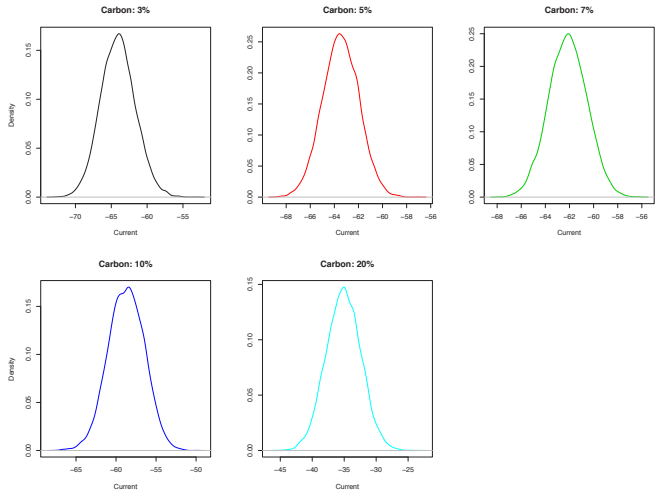


Figure B.6: Density plots of posterior predictive distributions of Minimum Current for scan rate 2.

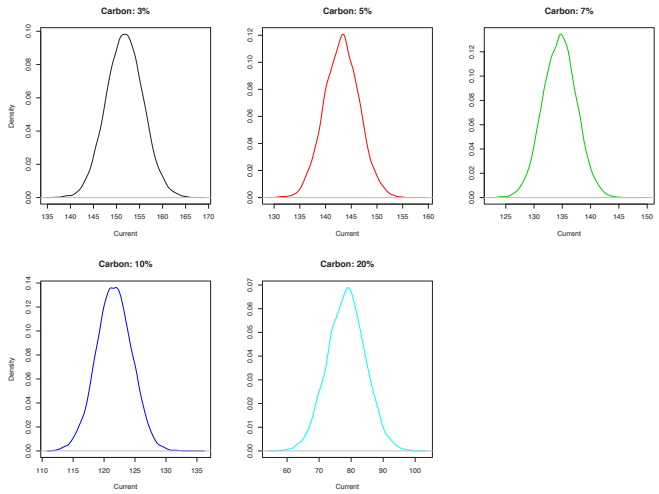


Figure B.7: Density plots of posterior predictive distributions of Peak Separation in Current for scan rate 2.

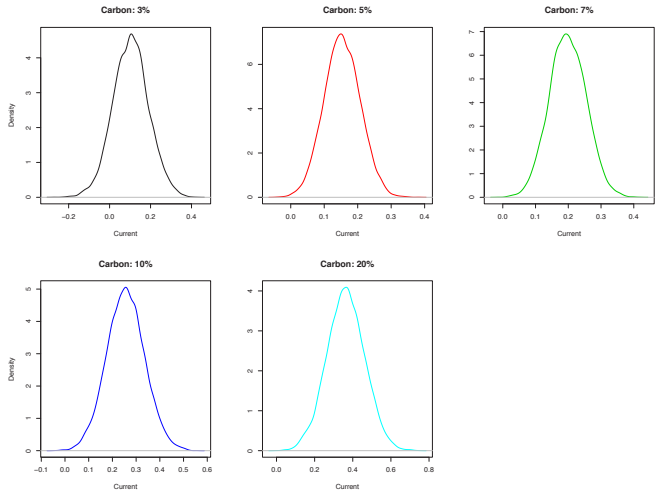


Figure B.8: Density plots of posterior predictive distributions of Peak Width in Potential for scan rate 2.

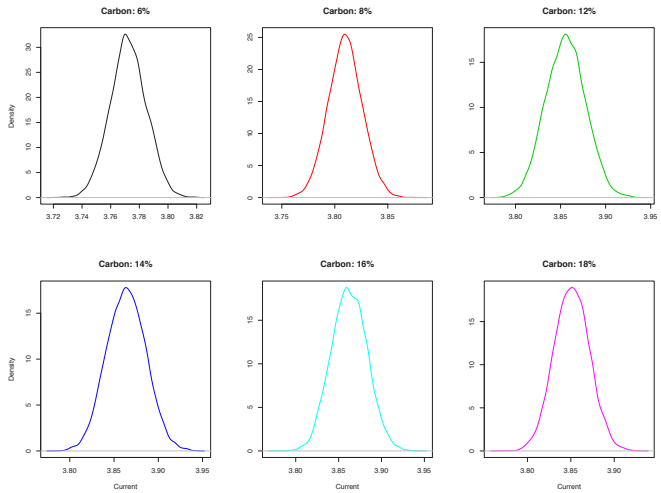


Figure B.9: Density plots of posterior predictive distributions of Minimum Potential for scan rate 2.

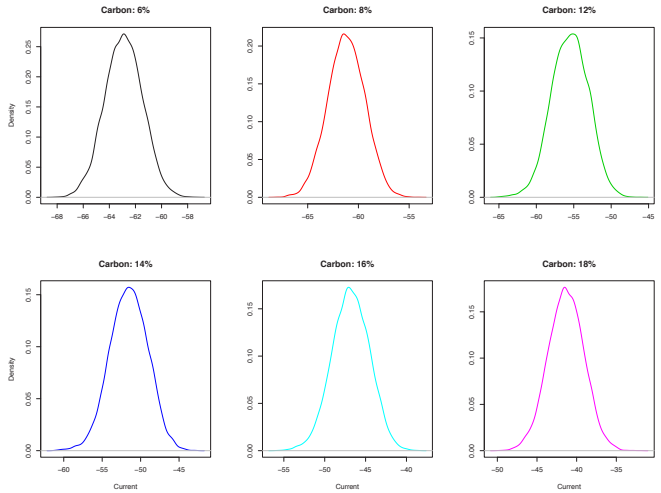


Figure B.10: Density plots of posterior predictive distributions of Minimum Current for scan rate 2.

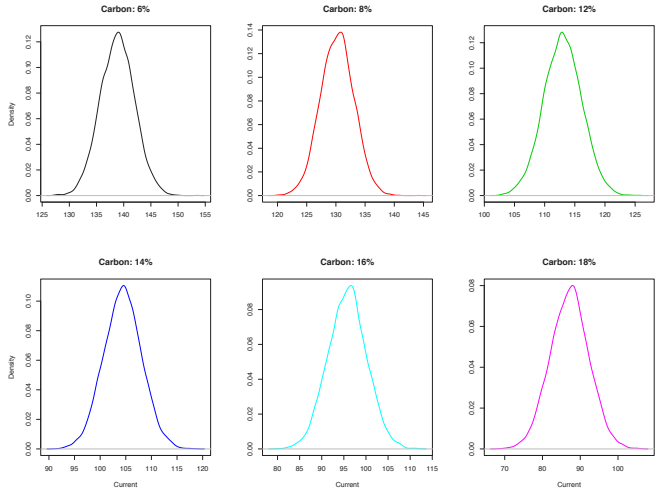


Figure B.11: Density plots of posterior predictive distributions of Peak Separation in Current for scan rate 2.

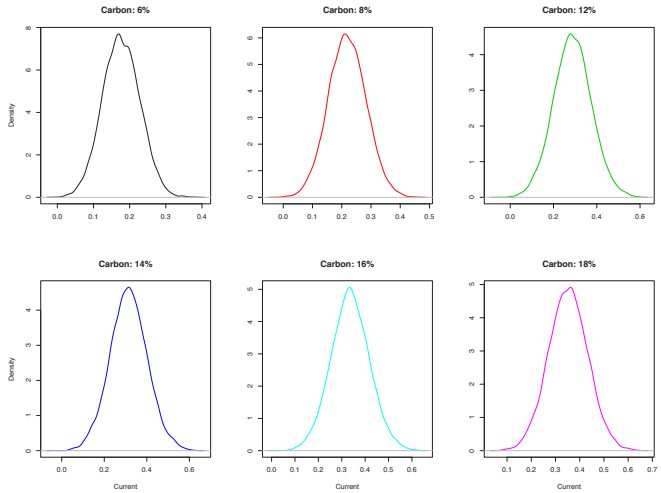


Figure B.12: Density plot of posterior predictive distributions of Peak Width in Potential for scan rate 2.

Appendix C

Fourier Series Models Analysis

Throughout this appendix, N and t denote the number of observations and time index respectively for a generic time series . For this proof we will require the use of the following identities:

$$\sum_{t=0}^{N-1} \cos\left(\frac{2\pi mt}{N}\right) \sin\left(\frac{2\pi nt}{N}\right) = 0, \tag{C.1}$$

$$\sum_{t=0}^{N-1} \cos\left(\frac{2\pi mt}{N}\right) \cos\left(\frac{2\pi nt}{N}\right) = \begin{cases} 0 & m \neq n, \\ N & m = n = \frac{N}{2}, \\ \frac{N}{2} & m = n \neq \frac{N}{2}, \end{cases} \tag{C.2}$$

$$\sum_{t=0}^{N-1} \sin\left(\frac{2\pi mt}{N}\right) \sin\left(\frac{2\pi nt}{N}\right) = \begin{cases} 0 & m \neq n, \\ 0 & m = n = \frac{N}{2}, \\ \frac{N}{2} & m = n \neq \frac{N}{2}, \end{cases} \tag{C.3}$$

$$\sum_{t=0}^{N-1} \cos\left(\frac{2\pi mt}{N}\right) = \sum_{t=0}^{N-1} \sin\left(\frac{2\pi mt}{N}\right) = 0, \tag{C.4}$$

where m and n are integers. From the Fourier series representation, given by

$$\lambda_t = a_0 + \sum_{r=1}^{R_f} \left[a_r \cos\left(\frac{2\pi rt}{N}\right) + b_r \sin\left(\frac{2\pi rt}{N}\right) \right],$$

where $r = 1, \ldots, R_f$, we obtain the equations below for each of the λ_t where $t = 1, \ldots, N$.

$$\begin{aligned}
\lambda_0 &= a_0 + \sum_{r=1}^{R_f} \Big[a_r \cos(0) + b_r \sin(0) \Big] \\
\lambda_1 &= a_0 + \sum_{r=1}^{R_f} \Big[a_r \cos(\alpha r) + b_r \sin(\alpha r) \Big] \\
\lambda_2 &= a_0 + \sum_{r=1}^{R_f} \Big[a_r \cos(2\alpha r) + b_r \sin(2\alpha r) \Big] \\
\vdots &\quad \vdots \\
\lambda_{N-1} &= a_0 + \sum_{r=1}^{R_f} \Big[a_r \cos(\alpha r(N-1)) + b_r \sin(\alpha r(N-1)) \Big]
\end{aligned}$$

Summing the above equations with respect to t, we obtain

$$\sum_{t=0}^{N-1} \lambda_t = \sum_{t=0}^{N-1} a_0 + \sum_{r=1}^{R_f} \sum_{t=0}^{N-1} \Big[a_r \cos(\alpha r t) + b_r \sin(\alpha r t) \Big] \tag{C.5}$$

where $\alpha = 2\pi/N$ throughout the rest of this proof. Re-arranging the summations in Equation (C.5), we obtain

$$\sum_{t=0}^{N-1} \lambda_t = \sum_{t=0}^{N-1} a_0 + \sum_{r=1}^{R_f} a_r \sum_{t=0}^{N-1} \cos(\alpha r t) + \sum_{r=1}^{R_f} b_r \sum_{t=0}^{N-1} \sin(\alpha r t). \tag{C.6}$$

Applying the identity from Equation (C.4) to Equation (C.6), we obtain

$$\begin{aligned}
N a_0 &= \sum_{t=0}^{N-1} \lambda_t \\
a_0 &= \frac{1}{N} \sum_{t=0}^{N-1} \lambda_t
\end{aligned}$$

as required.

We will now take Equation (C.6) and multiply both sides by $\cos(\alpha\varrho t)$ giving

$$\begin{aligned}\sum_{t=0}^{N-1}\lambda_t\cos(\alpha\varrho t) &= a_0\sum_{t=0}^{N-1}\cos(\alpha\varrho t)+\sum_{r=1}^{R_f}a_r\sum_{t=0}^{N-1}\cos(\alpha\varrho t)\cos(\alpha rt)\\ &\quad+\sum_{r=1}^{R_f}b_r\sum_{t=0}^{N-1}\cos(\alpha\varrho t)\sin(\alpha rt),\end{aligned}$$

where ϱ denotes an integer for the remainder of this proof. Using the identities in equations (C.1) and (C.2) we obtain

$$\begin{aligned}\frac{a_rN}{2} &= \sum_{t=0}^{N-1}\lambda_t\cos(\alpha rt)\\ a_r &= \frac{2}{N}\sum_{t=0}^{N-1}\lambda_t\cos(\alpha rt)\end{aligned}$$

when $r=\varrho\neq\frac{N}{2}$, and

$$\begin{aligned}Na_{N/2} &= \sum_{t=0}^{N-1}(-1)^t\lambda_t\\ a_{N/2} &= \frac{1}{N}\sum_{t=0}^{N-1}(-1)^t\lambda_t\end{aligned}$$

when $r=\varrho=\frac{N}{2}$. We will now take Equation (C.6) and multiply through by $\sin(\alpha\varrho t)$ which gives

$$\begin{aligned}\sum_{t=0}^{N-1}\lambda_t\sin(\alpha\varrho t) &= a_0\sum_{t=0}^{N-1}\sin(\alpha\varrho t)+\sum_{r=1}^{R_f}a_r\sum_{t=0}^{N-1}\cos(\alpha\varrho t)\sin\alpha rt\\ &\quad+\sum_{r=1}^{R_f}b_r\sum_{t=0}^{N-1}\sin(\alpha\varrho t)\sin(\alpha rt).\end{aligned}$$

Using the identities in Equations (C.1) and (C.4), we obtain

$$\sum_{t=0}^{N-1}\lambda_t\sin(\alpha rt)=\sum_{r=1}^{R_f}b_r\sum_{t=0}^{N-1}\sin(\alpha\varrho t)\sin(\alpha rt).$$

Applying the identity in Equation (C.3) to the above equation, we obtain

$$\begin{aligned} \frac{Nb_r}{2} &= \sum_{t=0}^{N-1} \lambda_t \sin(\alpha r t) \\ b_r &= \frac{2}{N} \sum_{t=0}^{N-1} \lambda_t \sin(\alpha r t) \end{aligned}$$

when $r = \varrho \neq \frac{N}{2}$ and $b_{N/2} = 0$ when $r = \varrho = \frac{N}{2}$ as required.

Bibliography

Akaike, H. (1973) Information theory and an extension of the likelihood principle. *2nd International Symposium on Information Theory*, 267–281.

Barone, P. and Ragona, R. (1997) Bayesian estimation of parameters of a damped sinusoidal model by a markov chain monte carlo method. *IEEE Transactions on Signal Processing*, **45**, 1806–1814.

Besag, J. and Green, P. J. (1993) Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society, Series B*, **55**, 25–37.

Black, F. and Scholes, M. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 637–654.

Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods*. Springer-Verlag: New York.

Broto, C. and Ruiz, E. (2004) Estimation methods for stochastic volatility models: A Survey. *Journal of Economic Surveys*, **18**, 613–649.

Casella, G. and George, E. I. (1992) Explaining the Gibbs sampler. *The American Statistician*, **46**, 167–174.

Chatfield, C. (2003) *The Analysis of Time Series, sixth edition.* Chapman & Hall/CRC.

Chen, M., Shao, Q. and Ibrahim, J. (2000) *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag: New York.

Chen, R., Yang, L. and Hafner, C. (2004) Nonparametric multistep-ahead prediction in time series analysis. *Journal of the Royal Statistical Society. Series B*, **66**, 669–686.

Chib, S., Nardari, F. and Shephard, N. (2001) Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, **134**, 341–371.

Crellin, N., Hastie, T. and Johnstone, I. (1998) Statistical models for image sequences. *Tech. rep.*

De Boor, C. (2002) *A Practical Guide to Splines*. Springer: New York.

Dellaportas, P., Forster, J. and Ntzoufras, I. (2002) On Bayesian model and variable selection using MCMC. *Statistics and Computing*, **12**, 27–36.

Desainte-Catherine, M. and Hanna, P. (2000) Statistical approach for sound modeling. In *Proceedings of the Digital Audio Effects Workshop (DAFX-00)*, 91–96.

Dubnov, S. and Rodet, X. (1997) Statistical modeling of sound aperodicities. In *in Proc. ICMC, Tessaloniki*, 43–50.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.

Evans, M. and Swartz, T. (1995) Methods for approximating integrals in Statistics with special emphasis on Bayesian integration problems. *Statistical Science*, **10**, 254–272.

Fan, J. and Yao, Q. (2005) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer: New York.

Gammerman, D. and Lopes, H. F. (2006) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, second edition.* Chapman & Hall/CRC.

Gelfand, A. E. and Ghosh, S. K. (1998) Model choice: A minimum posterior predictive loss approach. *Biometrika*, **85**, 1–11.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004) *Bayesian Data Analysis, second edition.* Chapman & Hall/CRC.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Geyer, C. J. (1992) Practical Markov Chain Monte Carlo. *Statistical Science*, **7**, 473–511.

Gilks, W. (1992) Derivative-free adaptive rejection sampling for gibbs sampling. In *Bayesian Statistics 4* (eds. J. M. Bernado, J. O. Berger, A. P. Dawid and A. F. M. Smith), 641–665. Oxford University Press: Oxford.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996) *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC.

Graves, T. L. (2007) Design ideas for Markov Chain Monte Carlo software. *Journal of Computational and Graphical Statistics*, **16**, 24–43.

Green, P. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.

Hainsworth, S. and Macleod, M. (2003) On sinusoidal parameter estimation. In *In Proc. Digital Audio Effects Workshop (DAFx), Queen Marys.*

Harvey, A. C., Ruiz, E. and Shephard, N. (1994) Multivariate stochastic variance models. *Review of Economic Studies*, **61**, 247–264.

Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Heston, S. L. (1993) A closed form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, **6**, 327–343.

Ibrahim, J. G. and Laud, P. W. (1995) Predictive model selection. *Journal of the Royal Statistical Society, Series B*, **57**, 247–262.

Jacquier, E., Polson, N. G. and Rossi, P. E. (1994) Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, **12**, 69–87.

— (2003) Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, **122**, 185–212.

Jeffreys, H. (1935) Some tests of significance, treated by the theory of probability. *Proceedings of the Cambridge Philosophy Society*, **31**, 203–222.

— (1961) *Theory of Probability, third edition.* Oxford University Press.

Johnson, R. and Wichern, D. (2002) *Applied Multivariate Statistical Analysis.* Prentice Hall: New Jersey.

Kass, R. E. and Raftery, A. E. (1995) Bayes Factors. *Journal of the American Statistical Association*, **90**, 773–795.

Kim, S., Shephard, N. and Chib, S. (1998) Stochastic volatility: Likelihood inference and comparison with ARCH Models. *Review of Economic Studies*, **65**, 361–393.

Koehler, A. B. and Murphree, E. S. (1988) A comparison of the akaike and schwarz criteria for selecting model order. *Applied Statistics*, **37**, 187–195.

Liesenfeld, R. and Richard, J. (2006) Classical and Bayesian analysis of univariate and multivariate stochastic volatility models. *Econometric Reviews*, **25**, 335–360.

Lovric, M. and Scholz, F. (2003) Modeling cyclic voltammograms of simultaneous electron and ion transfer reactions at a conic film three-phase electrode. *Journal of Electroanalytical Chemistry*, **540**, 89–96.

Lundquist, A., Hjelm, A. and Lindburgh, G. (2001) Investigation of $LiMn_2O_4$ cathodes for use in rechargeable lithium batteries by linear sweep voltammetry Part 1. Theoretical study. *Journal of Electroanalytical Chemistry*, **506**, 82 – 91.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000) Winbugs - a bayesian modelling framework: Concepts, structure and extensibility. *Statistics and Computing*, **10**, 325–337.

Mauricio, J. A. (2008) Computing and using residuals in time series models. *Computational Statistics and Data Analysis*, **52**, 1746–1763.

Melino, A. and Turnbull, S. (1990) Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, **45**, 7–39.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953) Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, **21**, 1087–1091.

Meyn, S. P. and Tweedie, R. L. (1993) *Markov Chains and Stochastic Stability*. Springer-Verlag: London.

Muller, N. Z. and Phillips, P. C. B. (2007) Sinusoidal modeling applied to spatially variant tropospheric ozone air pollution. *Environmetrics*, **19**, 567–581.

Myers, R. H. (1990) *Classical and Modern Regression with Applications, second edition*. Duxbury.

Myland, J. C. and Oldham, K. B. (2002a) A model of cyclic voltammetry for a thin organic layer sandwiched between an electrode and an aqueous solution. Convolutive modelling in the absence of supporting electrolyte. *Journal of Electroanalytical Chemistry*, **530**, 1–9.

— (2002b) Convolutive modelling in the absence of supporting electrolyte coping with migration and changing resistance in predicting voltammetry. *Journal of Electroanalytical Chemistry*, **529**, 66–74.

Neal, R. M. (1997) Markov chain monte carlo methods based on 'slicing' the density function. *Tech. rep.*, Department of Statistics, University of Toronto.

Novak, P., Haas, O., Deiss, E. and Haringer, D. (2001) Modeling of the charge-discharge dynamics of lithium manganese oxide electrodes for lithium-ion batteries. *Electrochimica Acta*, **46**, 4185–4196.

O'Hagan, A. and Forster, J. (2004) *Kendall's Advanced Theory of Statistics: Bayesian Inference, second edition*. Arnold.

Papaspiliopoulos, O. and Roberts, G. (2008) Stability of the gibbs sampler for bayesian hierarchical models. *The Annals of Statistics*, **36**, 95–117.

Pole, A., West, M. and Harrison, J. (1994) *Applied Bayesian Forecasting and Time Series Analysis*. Chapman & Hall/CRC.

Robert, C. P. (2001) *The Bayesian Choice*. Springer-Verlag: New York.

Ruiz, E. (1994) Quasi-maximum likelihood estimation of stochastic variance models. *Journal of Econometrics*, **63**, 284–306.

Ruppert, D. (2004) *Statistics and Finance*. Springer-Verlag: New York.

Ruppert, R. J. C. D. (1988) *Transformation and Weighting in Regression*. Chapman & Hall: New York.

Sahu, S. K. (2004) Applications of formal model choice to archaeological chronology building. In *Tools for Constructing Chronologies: Cross Disciplinary Boundaries* (eds. C. E. Buck and A. R. Millard), 111–117. Springer-Verlag: London.

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

Silverman, B. W. (1986) *Density Estimation*. Chapman and Hall: London.

Smith, A. F. M. and Roberts, G. (1993) Bayesian Computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, **55**, 3–23.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Linde, A. V. D. (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, **64**, 583–639.

Srivastava, A., Lee, A. B., Simoncelli, E. P. and Zhu, S. (2003) On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, **18**, 17–33.

Tierney, L. (1994) Markov Chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, **22**, 1701–1762.

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*. Springer-Verlag: New York.

West, M. and Harrison, J. (1999) *Bayesian Forecasting and Dynamic Models*. Springer-Verlag: New York.

Yu, J. and Meyer, R. (2006) Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews*, **25**, 361–384.