Statistical Modelling of Equations of State for Carbon Capture, Transport, and Storage



Michael James Thomson

Thesis submitted to The University of Nottingham for the Degree of Doctor of Philosophy

January 5, 2018

Abstract

Capturing CO_2 produced by power plants and storing it in within depleted oil and natural gas reservoirs in the seabed offers a potential means to reduce the emissions that contribute to climate change. This is known as carbon capture and storage (CCS). CO_2 captured from power plants is transported to these reservoirs using pipelines. To determine the parameters of pipeline operation this calls for a need to characterise, via an "equation of state", the physical properties of CO_2 during transport. Doing so is further complicated by the fact that CO_2 captured from power stations is typically a mixture of CO_2 with other fluids. This affects the physical properties of the CO_2 to be transported and so needs to be accounted for.

In this thesis we develop statistical models for equations of state that can account for the physical properties of CO_2 relevant to carbon capture and storage, and which allow us to quantify uncertainty in the predictions from the equation of state. We propose two statistical models for equations of state. Firstly we develop a statistical model which can be applied to any pressure-explicit parametric equation of state. To do so we have developed a novel method by which to rigorously account for uncertainties due to coexistence which is complicated by the fact that it involves perfectly correlated measurements on two fluid phases. We fit this model to pure CO_2 obtaining good agreement with data for most temperatures. We then extended this model and our method for accounting for coexistence to mixtures. We fit this model to real CO_2-H_2 data. Despite mixing well,

the results of this fit do not agree well with the data and equations of state need to be developed further to be able to model mixture data well.

Secondly we develop a non-parametric Gaussian process approach which offers greater flexibility and requires fewer assumptions. This non-parametric model is fit to pure CO_2 reference data for individual subcritical temperatures. We demonstrate how applying a transformation to the covariance function can account for non-stationarity in the data resulting in good agreement between predictions from the fit model and the data.

Acknowledgements

First and foremost I would like to thank my supervisors, Simon Preston, Richard Graham, and Richard Wilkinson, for all of their support, advice and encouragement over the last four years.

I would also like to thank my family, Caroline, Ian, and Zoe, for never failing to lift my spirits when times are tough, and for always believing in me, even when that was foolish.

Lastly I would like to thank the EPSRC for providing the funding that made this thesis possible.

Contents

Figures viii

Tables xiii

1 Carbon Capture and Storage and Equations of State - A Literature Review 1

1.1 Introduction 1

1.2 Carbon Capture and Storage 1

1.3 Properties of CO_2 3

1.3.1 The Critical Point 4

1.3.2 Coexistence of Pure CO_2 6

1.3.3 Limiting Behaviour 10

1.3.4 Coexistence of Mixtures 11

1.4 Equations of State 14

1.4.1 Pressure-explicit and Free-energy-formulations 14

1.4.2 A Simple EoS: The Ideal Gas Law 16

1.5 Review of Previously Proposed Equations of State 16

1.5.1 Peng-Robinson Equation of State 17

1.5.2 Span-Wagner Equation of State 17

1.5.3 Demetriades EoS 19

1.5.4 The Non-Parametric Approach 21

1.6 Summary of Available Data 22

1.6.1 Pure CO_2 Data 22

1.6.2 CO_2-H_2 Mixture Data 27
1.7 Thesis Aims and Plan 29

2 Statistical Literature Review **32**
2.1 Introduction 32
2.2 Likelihood 33
2.2.1 Maximum Likelihood Estimation 34
2.3 Bayesian Parameter Estimation 36
2.4 Errors-in-Variables and Missing Data 37
2.4.1 Bayesian Errors-in-Variables 37
2.4.2 Bayesian Missing Data 39
2.5 Gaussian Processes 41
2.5.1 Bayesian Estimation of Hyperparameters 45
2.6 Markov Chain Monte Carlo 45
2.6.1 Toy Examples 48
2.7 Metropolis-Hastings 51
2.7.1 Toy Examples 53
2.8 Hamiltonian Monte Carlo 56
2.8.1 No-U-Turn Sampler 61
2.8.2 Adaptation of the Stepsize 66
2.8.3 Stan 68
2.8.4 Toy Examples 68
2.9 Parallel Tempering 70
2.9.1 Toy Example 74

3 Statistical Modelling of Parametric EoS **76**
3.1 Introduction 76
3.2 Modelling Homogeneous Data 77
3.3 Modelling Coexistence 78
3.3.1 The Hard Constraint Case 80

3.3.2 The Soft Constraint Case 84
3.4 Modelling Uncertainty in the Coexistence Volumes 90
3.5 Modelling Multiple Observations of Coexistence Pressures and Volumes 91

4 Fitting Parametric EoS for Pure CO_2 94

4.1 Introduction 94
4.2 Single Temperature Homogeneous Synthetic Data 95
4.2.1 Maximum Likelihood Estimation 96
4.2.2 Metropolis-Hastings 100
4.2.3 Hamiltonian Monte Carlo 103
4.2.4 Parallel Tempering with HMC 108
4.2.5 Application of PT-HMC 111
4.3 Multiple Temperature Homogeneous Synthetic Data 115
4.3.1 Application of PT-HMC 117
4.4 Single Temperature with Coexistence Data 120
4.4.1 Hard Constraint 123
4.4.2 Soft Constraint 131
4.5 Multiple Temperatures Synthetic Data with Coexistence Data 136
4.5.1 Application of PT-HMC 139
4.6 Application of the PT-HMC Sampler To Real Pure CO_2 Data 143
4.6.1 Fitting Individual Temperatures 143
4.6.2 Quadratic Temperature Rule 147
4.6.3 Fitting Multiple Temperatures 152

5 Parametric EoS for Mixtures 163

5.1 Introduction 163
5.2 Modelling Homogeneous Mixture Data 164
5.3 Modelling Coexistence Mixture Data 165

5.4 Modelling Missing Coexistence Volumes 173
5.5 Demetriades EoS for Mixtures 175
5.6 Fitting the Demetriades EoS for Mixtures 181

6 Non-Parametric EoS **186**
6.1 Introduction 186
6.2 Modelling Gaussian Processes for EoS 187
6.2.1 Modelling Pressure Observations 189
6.2.2 Modelling Coexistence Data 190
6.3 Fitting the GP EoS Model 195
6.3.1 Fitting an Individual Subcritical Temperature 196
6.3.2 Modelling Non-stationarity 201

7 Conclusions **205**
7.1 Thesis Summary and Results 205
7.2 Future Work 207

Bibliography **210**

List of Figures

1.1 Pressure versus volume plot of NIST data for CO_2 at the critical temperature, 304.1282K, in the region of the critical point. 4

1.2 Pressure versus volume plot demonstrating coexistence for pure CO_2 at 290K. 6

1.3 Pressure versus volume plot of NIST data for CO_2 at the critical temperature, 304.1282K and at large volume. 10

1.4 Coexistence phase envelope for CO_2-H_2 mixture at 295.65K 11

2.1 Contour plots of the densities for the toy examples 49

2.2 Traceplots and marginals densities for an application of Metropolis-Hastings applied to the first toy example. 54

2.3 Traceplots for an application of Metropolis-Hastings applied to the second and third toy examples. 55

2.4 Development of the binary tree constructed in the no-U-turn sampler 64

2.5 Traceplots for an application of the no-U-turn sampler applied to the second and third toy examples. 69

2.6 Traceplot of hottest replica applying MH and parallel tempering to the third toy example. 73

2.7 Traceplots and marginal densities of the coldest replica applying MH and parallel tempering to the third toy example. 74

4.1 Scatterplot of MLE results for homogeneous synthetic data at a single temperature 96
4.2 Histogram of MLE results for homogeneous synthetic data at a single temperature 98
4.3 Traceplot of a sample for MH applied to single temperature synthetic data without coexistence 100
4.4 Traceplot of HMC applied to single temperature synthetic data with no coexistence 104
4.5 Histograms of samples from HMC applied to single temperature synthetic data with no coexistence 105
4.6 Log posterior plot for HMC applied to single temperature synthetic data without coexistence 106
4.7 HMC run at high tempering level for single temperature synthetic data with no coexistence 112
4.8 Traceplots of sampled parameters for PT-HMC applied to single temperature synthetic data without coexistence and after burn-in 114
4.9 Traceplots of PT-HMC applied to the homogeneous multiple temperature synthetic data without burn-in 119
4.10 Traceplots of the initial iterations applying PT-HMC to synthetic data with coexistence at a single temperature for the hard constraint model. 128
4.11 Traceplots of the sampled parameters applying PT-HMC to synthetic data with coexistence at a single temperature for the hard constraint model after burn-in is discarded. 129
4.12 Marginal densities of the sampled parameters applying PT-HMC to synthetic data with coexistence at a single temperature for the hard constraint model. 130

4.13 Traceplots of the initial iterations applying PT-HMC to synthetic data with coexistence at a single temperature for the soft constraint model. 133
4.14 Traceplots of the sampled parameters applying PT-HMC to synthetic data with coexistence at a single temperature for the soft constraint model after burn-in. 134
4.15 Marginal densities of the sampled parameters applying PT-HMC to synthetic data with coexistence at a single temperature for the hard constraint model. 135
4.16 Traceplots of the sampled parameters applying PT-HMC to synthetic data with coexistence at multiple temperatures. . . 140
4.17 Marginal densities of the sampled DEoS parameters applying PT-HMC to synthetic data with coexistence at multiple temperatures. 141
4.18 Marginal densities of the sampled coexistence volume parameters applying PT-HMC to synthetic data with coexistence at multiple temperatures. 142
4.19 Plots of the posterior marginal modes of the model parameters for the DEoS fitted to real data at individual temperatures. 146
4.20 Traceplots of the sampled model parameters applying PT-HMC to fitting the DEoS to real pure CO_2 data at multiple temperatures. 152
4.21 Traceplots of the sampled coexistence volume parameters applying PT-HMC to fitting the DEoS to real pure CO_2 data at multiple temperatures. 154
4.22 Traceplots of the sampled model parameters applying PT-HMC to fitting the DEoS to real pure CO_2 data at multiple temperatures after discarding burn-in. 156

4.23 Traceplots of the sampled coexistence volume parameters applying PT-HMC to fitting the DEoS to real pure CO_2 data at multiple temperatures after discarding burn-in. . . . 157
4.24 Marginal density plots of the sampled model parameters applying PT-HMC to real pure CO_2 data at multiple temperatures. 158
4.25 Marginal density plots of the sampled coexistence volume parameters applying PT-HMC to real pure CO_2 data at multiple temperatures. 159
4.26 Posterior predictive realisations at 296.994K of sample from applying PT-HMC to DEoS to real pure CO_2 data for multiple temperatures. 160
4.27 Posterior predictive realisations at 304.1472K of sample from applying PT-HMC to DEoS to real pure CO_2 data for multiple temperatures. 161

5.1 Traceplots of the sampled model parameters applying PT-HMC to fitting the DEoS to CO_2-H_2 mixture data at 288.15K.177
5.2 Traceplots of the sampled bubblepoint coexistence volume parameters applying PT-HMC to fitting the DEoS to CO_2-H_2 mixture data at 288.15K. 178
5.3 Traceplots of the sampled dewpoint coexistence volume parameters applying PT-HMC to fitting the DEoS to CO_2-H_2 mixture data at 288.15K. 180
5.4 Marginal density plots of the sampled model parameters applying PT-HMC to CO_2-H_2 mixture data at 288.15K. . . . 181
5.5 Marginal density plots of the sampled coexistence volume parameters applying PT-HMC to CO_2-H_2 mixture data at 288.15K. 182

5.6 Posterior realisations at 288.15K and H_2 concentration of 9.97% of sample from applying PT-HMC to DEoS to real CO_2-H_2 mixture data at 288.15K 183
5.7 Posterior sample of coexistence points at 288.15K and H_2 concentration of 9.97% of sample from applying PT-HMC to DEoS to real CO_2-H_2 mixture data at 288.15K 184

6.1 Traceplots and marginal density plots from using HMC to fit the GP EoS to 294K after discarding burn-in. 198
6.2 Plots of 500 posterior realisations of the pressure-volume curve from fitting the GP EoS model to 294K. 199
6.3 Traceplots and marginal density plots from using HMC to fit the GP EoS to 294K after discarding burn-in with non-stationary covariance function. 202
6.4 Plots of 500 posterior realisations of the pressure-volume curve from fitting the GP EoS model to 294K with non-stationary covariance function. 203

List of Tables

1.1 Summary of pure CO_2 data from Dusschek et al. [1990b] . . 23

1.2 Summary of pure CO_2 data from Gilgen et al. [1992] 24

1.3 Summary of pure CO_2 data for coexistence from Dusschek et al. [1990a]. 26

1.4 Summary of homogeneous pure CO_2 data from Dusschek et al. [1990a] 27

1.5 Summary of homogeneous CO_2-H_2 mixture data from Sanchez-Vicente et al. [2013]. 28

1.6 Summary of coexistence CO_2-H_2 mixture data from Fandiño et al. [2015]. 29

Chapter 1

Carbon Capture and Storage and Equations of State - A Literature Review

1.1 Introduction

Chapter 1 is a literature review on the subject of equations of state (EoS) for carbon capture and storage (CCS). An overview of CCS and the need for modelling of the properties of carbon dioxide (CO_2) is provided. This is followed by a description of the properties of CO_2 which EoS need to model. We then present a review of previously proposed EoS, and describe the data on physical properties of CO_2 relevant to CCS which we use in this thesis.

1.2 Carbon Capture and Storage

The presence of large quantities of greenhouse gases such as carbon dioxide (CO_2) in the atmosphere is one of the main contributors to climate change [Stern, 2006]. Fossil fuel-based power plants are one of the main producers of such gases [Seevam et al., 2007]. Carbon capture and storage (CCS) of-

fers a means by which to reduce such emissions and so reduce the effects of the electricity network on climate change [Boot-Handford et al., 2014]. CO_2 produced by power plants is captured and then transported via pipelines to depleted oil and natural gas reservoirs within the seabed where it can be stored. The CO_2 is thus removed from the atmosphere and as such its effect upon the climate mitigated. The use of pipelines for transportation gives rise to many key questions: what should be the standards for the construction and operation; what pressure should be maintained; and how should pressure be maintained. Furthermore CO_2 produced by power plants is not pure CO_2 but a mixture of pure CO_2 and small quantities of other fluids [Seevam et al., 2007]. These impurities affect the behaviour of the CO_2 and thus impact upon how it is transported. This calls for a need to characterise the physical properties of CO_2 (and of mixtures of CO_2 and other fluids) during transport. Such a model for doing so is called an equation of state (EoS). Deriving, and particularly fitting EoS to somewhat noisy data is the main focus of this thesis.

In this thesis we focus on fitting EoS under the conditions that would be expected in standard operation of pipelines for the transportation of CO_2 for carbon storage. The *CCS transport window* is defined by the range of temperatures and pressures a pipeline could be expected to operate at. The temperature range at which a CCS pipeline operates will be affected by the ground temperature of the soil around the pipeline [Mohitpour et al., 2012]. In the United Kingdom it is reasonable to expect a temperature range of between 270 and 313 K for standard pipeline operation.

Typically the CO_2 in a pipeline will be compressed at source and so the pressure of the fluid will decrease along the length of the pipeline. It is most efficient to transport a fluid when it is *supercritical* [Seevam et al., 2007], when it has properties of both a liquid and a gas. This necessitates maintaining the CO_2 in the pipeline at a pressure greater than the critical

pressure for the entire length of the pipeline. Should the pressure drop too much along the pipeline the fluid will enter *coexistence* and form into separate liquid and gas states. For the CO_2 to remain supercritical a pipeline could be expected to operate at pressures of up to 15 MPa [de Visser et al., 2008]. The CO_2 in CCS pipelines will likely not be pure but rather be a mixture with one or more impurities. Important impurities that might be expected in the CCS process include N_2, O_2 and H_2 at expected concentrations of up to 4% [Porter et al., 2015, Seevam et al., 2007, Tenorio et al., 2015, de Visser et al., 2008]. The presence of these impurities alters the conditions under which the CO_2 is supercritical. Thus for efficient pipeline operation it is vital that an EoS can account for the effect of impurities on the behaviour of the fluid to be transported.

We will now describe the properties of CO_2 such as coexistence which will be encountered in the CCS transport window and which need to accounted for in any EoS to be used for CCS.

1.3 Properties of CO_2

There are several properties of fluids which need to be considered for CCS [Orbey and Sandler, 1998, Seevam et al., 2007]. These properties have been precisely characterised empirically and any EoS to be applied for CCS purposes will need to account for these properties. These properties are expressed in terms of pressure, P, the molar volume, v, and the temperature, T, where the molar volume is the volume per mole of the fluid. Henceforth we use the term "volume" to mean "molar volume" unless stated otherwise.

Alternatively rather than define these properties in terms of P, sometimes the Helmholtz free energy, A, is used instead. From Kunz et al. [2007] the relationship between the Helmholtz free energy, A, and pressure of a

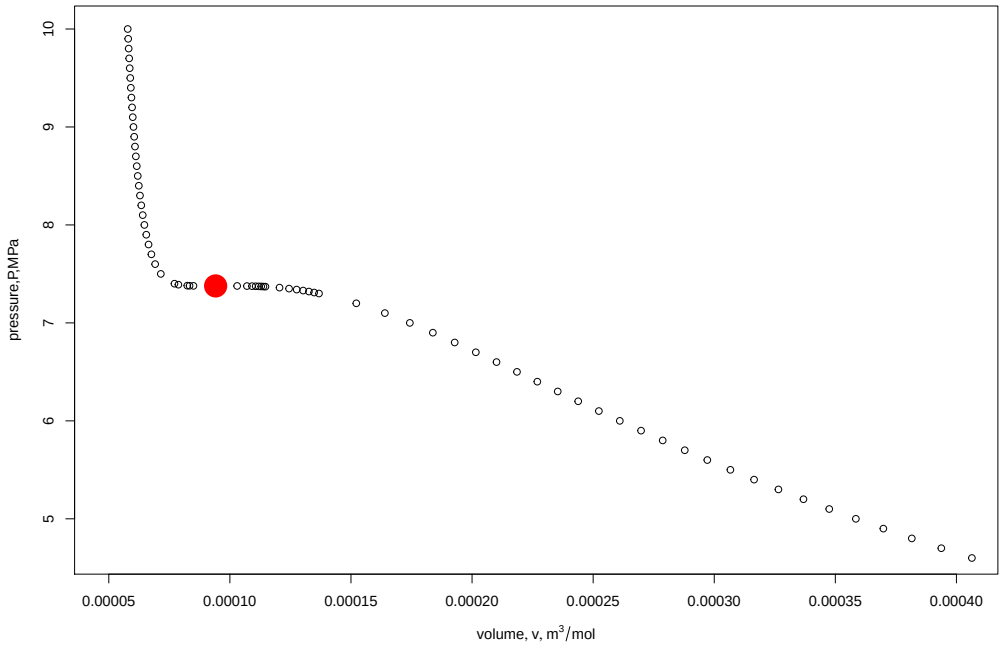


Figure 1.1: Pressure versus volume plot of NIST data for CO_2 at the critical temperature, 304.1282K, in the region of the critical point, $v = v_c, P = P_c$. The black circles are the NIST data points. The red point marks the location of the critical point. From this plot it can be seen that the critical point is a saddle point.

fluid is as follows,

$$\frac{\partial A}{\partial v} = -P. \tag{1.1}$$

So the Helmholtz free energy is the negative integral of pressure over volume up to some arbitrary constant.

In §1.2 we described how for best operation of a CCS pipeline it is desired to keep the fluid in the supercritical state. Supercriticality is defined in relation to the *critical point* of the fluid which we describe now.

1.3.1 The Critical Point

The first property of a fluid of interest to CCS applications is that it has a critical point. This critical point consists of a critical temperature, T_c, pressure, P_c, and volume, v_c. For CO_2 reference values for these are usually

taken to be [Span and Wagner, 1996],

$$T_c = 304.1282\,\mathrm{K}, \tag{1.2}$$

$$P_c = 7.3773\,\mathrm{MPa}, \tag{1.3}$$

$$v_c = 0.000094118\,\mathrm{m^3\,mol^{-1}}. \tag{1.4}$$

These values are not measurements but are instead an extrapolation of a fit to coexistence (see §1.3.2) data from Dusschek et al. [1990a].

When T and P exceed these critical values ($T > T_c, P > P_c$) the fluid is supercritical. When the fluid is supercritical it has properties of both a liquid and a gas and it is in this form that CO_2 is most efficiently transported [Seevam et al., 2007]. The critical point is also a saddle point in volume, as can be seen in Figure 1.1. That is, evaluated at the critical point, the first and second derivatives of pressure with respect to volume are equal to 0,

$$\left.\frac{\partial P}{\partial v}\right|_{v=v_c,T=T_c} = 0, \tag{1.5}$$

$$\left.\frac{\partial^2 P}{\partial v^2}\right|_{v=v_c,T=T_c} = 0. \tag{1.6}$$

We can define this saddle point in terms of A equivalently as,

$$-\left.\frac{\partial^2 A}{\partial v^2}\right|_{v=v_c,T=T_c} = 0, \tag{1.7}$$

$$-\left.\frac{\partial^3 A}{\partial v^3}\right|_{v=v_c,T=T_c} = 0. \tag{1.8}$$

Figure 1.1 plots National Institute of Standards and Technology (NIST) CO_2 data [National Institute of Standards and Technology, 2015] fixed at the critical temperature, T_c, in the region around the critical point which is marked in red. This reference data is from the Span-Wagner EoS (SWEoS) [Span and Wagner, 1996] and we use it here for plotting the properties of CO_2. For actual fitting of EoS we will use the data discussed in §1.6. From

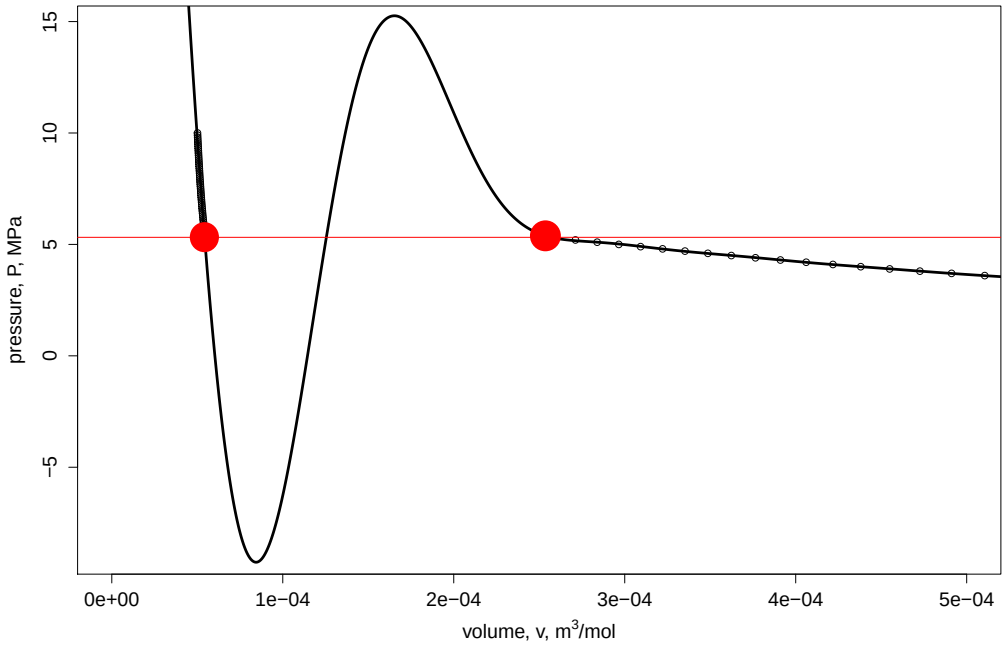


Figure 1.2: Pressure volume plot demonstrating coexistence for pure CO_2 at 290K. Black circles NIST data points. Solid red circles coexistence points. The black line is an example of a function which satisfies the fugacity constraint (equation (1.14)) while passing through the data. The horizontal red line marks out the coexistence pressure P_s.

this plot it can be clearly seen that the critical point is a saddle point.

In §1.2 we also introduced coexistence, another property of CO_2 that needs to be considered for CCS applications and we now provide a full definition and description of coexistence.

1.3.2 Coexistence of Pure CO_2

For pure fluid at temperatures less than T_c (subcritical) distinct gas and liquid phases can co-exist. Depending on the values of volume and pressure the fluid will be either a liquid or a gas or a mixture of the two when in coexistence. When volume is small and pressure high the fluid is a liquid. For large volume and low pressure the fluid is a gas. The issue of *coexistence* arises when the fluid switches state, from liquid to gas, or vice versa.

For a given subcritical temperature, $T < T_c$, there exists a coexistence

pressure, P_s, such that the fluid phase switches from liquid to gas (or vice versa) at this pressure. Suppose we have a fluid in a pipeline, the pressure of which is greater than P_s. Then the fluid is dense, ie. it is a liquid. If the pressure were then to decrease to P_s, bubbles of gas will appear in the liquid as the fluid begins switching states. Were the pressure to start below P_s the fluid would initially be a gas. As pressure increases condensation would appear in the pipeline as the fluid undergoes the opposite phase switch, from gas to liquid. This would lead to the fluid being in two states within the pipeline and so coexistence is also known as the *two-phase region*. When the fluid is in a single state (either liquid or gas) we describe the fluid as *homogeneous*.

Both of the points where a phase switch occurs have an associated volume. We call the volume at which the fluid phase switches from liquid to gas v_l, and the volume at which the fluid switches from gas to liquid v_g. For a liquid subjected to decreasing pressure, the volume increases to v_l, at which point the phase switch from liquid to gas occurs. For a gas subjected to increasing pressure, the volume decreases to v_g, at which point the phase switch from gas to liquid occurs. We refer to the data consisting of a given subcritical temperature, $T < T_c$, the coexistence pressure, P_s, and a coexistence volume, v_l or v_g, as *coexistence points*. For each subcritical temperature there are two coexistence points. Figure 1.2 shows volume and pressure data for pure CO_2 with $T = 285K$ (black circles). As T is subcritical it can be seen that between the coexistence points, marked in red, there is no data. The coexistence pressure is marked by a horizontal red line.

Given a subcritical temperature, for $v \in (v_l, v_g)$ the fluid cannot physically exist. Were a fluid imposed to have such volume the fluid would separate into a liquid and a gas each of which would have volume as at the corresponding coexistence point.

At the coexistence points there are three equilibrium conditions, three equalities between the coexistence points. The first of these is mechanical equilibrium: the pressure at each coexistence point is equal, what we have here called P_s,

$$P_l = P_g = P_s, \tag{1.9}$$

where P_l is the pressure corresponding to v_l and P_g is the pressure corresponding to v_g. The second is thermal equilibrium: the temperature is equal across coexistence points,

$$T_l = T_g = T, \tag{1.10}$$

where T_l is the temperature corresponding to v_l and T_g is the temperature corresponding to v_g. Finally the third condition is equilibrium of the chemical potential. This final equilibrium condition determines the behaviour an EoS will need to model for volumes between v_l and v_g for a given temperature. Typically this third equilibrium condition is defined equivalently in terms of some other property.

In terms of P, the chemical potential equilibrium condition is defined in terms of the *fugacity*, f, which for a pure fluid is [Orbey and Sandler, 1998],

$$f = P\phi, \tag{1.11}$$

where ϕ is the *fugacity coefficient* defined as [Orbey and Sandler, 1998],

$$\phi = \exp\left(\frac{1}{RT}\int_{v'=\infty}^{v}\left(\frac{RT}{v'} - P\right)\,\mathrm{d}v' - \ln\left(\frac{Pv}{RT}\right) + \frac{Pv}{RT} - 1\right), \tag{1.12}$$

where R is the ideal gas constant,

$$R = 8.3144621\,\mathrm{m^3\,Pa\,K^{-1}\,mol^{-1}}. \tag{1.13}$$

The chemical potential equilibrium condition is satisfied if the fugacity evaluated at the coexistence point with volume v_l, which we denote f_l, is equal to the fugacity as evaluated at the coexistence point with volume v_g, which we call f_g [Orbey and Sandler, 1998],

$$f_l = f_g, \tag{1.14}$$

where we call (1.14) the *fugacity constraint*. This constraint is a physical law with a basis in statistical mechanics, which a fluid must obey [Orbey and Sandler, 1998], and thus which an EoS must satisfy. The black line in Figure 1.2 is an example of a function satisfying this constraint.

In terms of A the same constraint can be expressed in terms of the *Gibbs free energy*, G, [Kunz et al., 2007],

$$G = A + Pv. \tag{1.15}$$

The constraint now becomes that the Gibbs free energy must be equal at the coexistence points. So for $T < T_c$ the constraint is

$$G_l = G_g, \tag{1.16}$$

where G_l and G_g correspond to v_l and v_g respectively. We refer to (1.16) as the *Gibbs free energy constraint*. As with the fugacity constraint the Gibbs free energy constraint holding means that the chemical potential is equal at the two coexistence points.

Having described coexistence for pure CO_2 there is one final property of pure CO_2 which an EoS will need to describe, the limiting behaviour for large v.

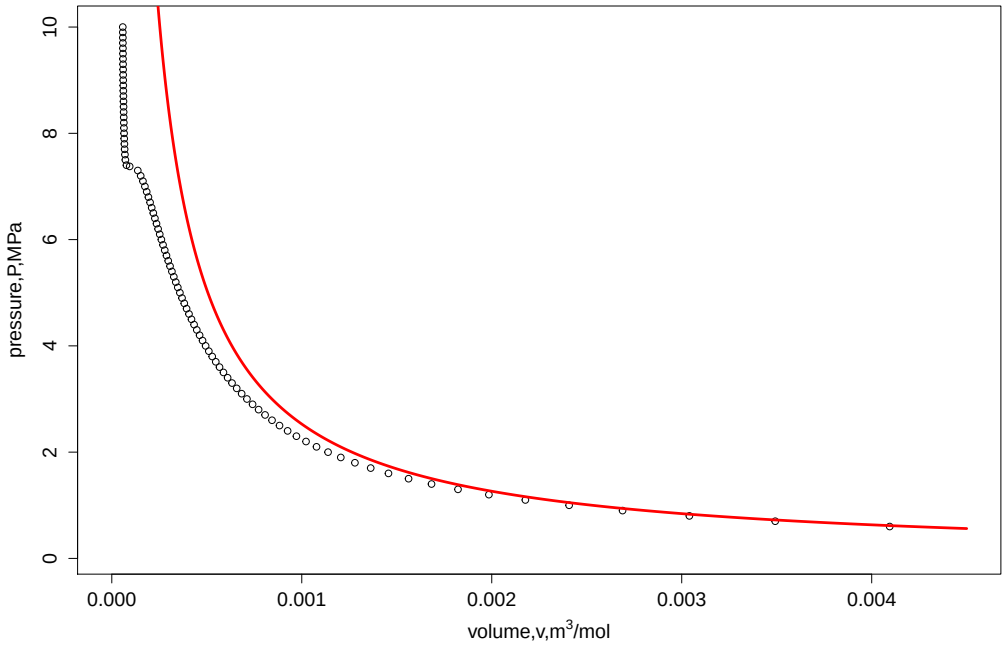


Figure 1.3: Pressure versus volume plot of NIST data for CO_2 at the critical temperature, 304.1282K. The black circles are the NIST data points. The red line plots the ideal gas law (equation (1.17)) at the critical temperature. Here it can be seen the ideal gas law does not describe the data well except as $v \to \infty$.

1.3.3 Limiting Behaviour

It is known that for large v a fluid behaves as an ideal gas [Cengel and Boles, 2001]. An ideal gas is described by the ideal gas law which is

$$P = \frac{RT}{v}. \tag{1.17}$$

As v gets large the behaviour of pressure tends to that of the ideal gas law (equation (1.17)),

$$P \to \frac{RT}{v}, \quad v \to \infty. \tag{1.18}$$

Figure 1.3 demonstrates this at the critical temperature, 304.1282K. In red is the ideal gas law. We can see from figure 1.3 that for small v the ideal gas law does not describe the data at all, however as v increases it becomes more accurate until eventually it describes the data well.

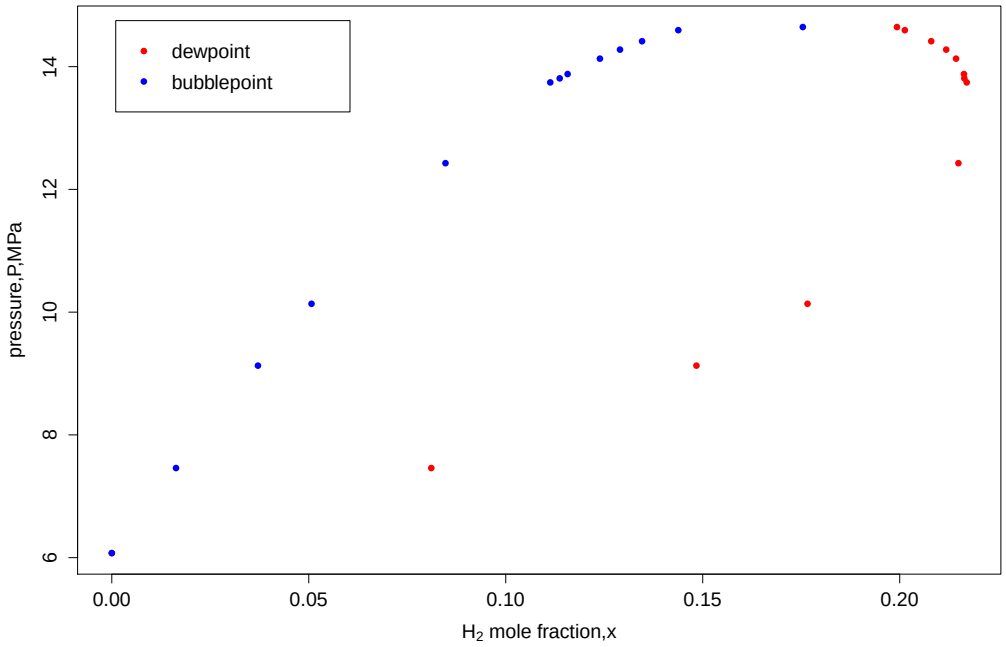


Figure 1.4: Plot of points on the phase envelope for CO_2-H_2 mixture at 295.65K. Data from Fandiño et al. [2015]. Dewpoints are in red and bubblepoints in blue. From this figure we can see how the dewpoint and bubblepoint curves meet at the critical point for this temperature.

We have to this point described the physical properties of pure CO_2 that an EoS must be able to model. It is expected that CO_2 in the CCS system will be a mixture of CO_2 with one or more impurities [Porter et al., 2015, Seevam et al., 2007, Tenorio et al., 2015, de Visser et al., 2008]. When a fluid is a mixture of two or more components the coexistence behaviour becomes more complex. In the next section we describe coexistence for a binary mixture.

1.3.4 Coexistence of Mixtures

Suppose we have a binary mixture of CO_2 and some other fluid which we refer to from here as the *impurity*. Let x be the mole fraction of the impurity, then the mole fraction of CO_2 is $1-x$. In the case of mixtures for a given temperature T at which coexistence can occur there is no longer a single point at which coexistence occurs. Instead for every temperature at

which coexistence can occur there is a *coexistence phase envelope*. Figure 1.4 plots points from the coexistence phase envelope data for a CO_2-H_2 mixture at 295.65K from Fandiño et al. [2015]. The coexistence phase envelope consists of a dewpoint curve and a bubblepoint curve, points from which are in red and blue respectively in figure 1.4. The dewpoint curve marks where condensation to a liquid in the gas phase of the mixture begins to occur and the bubblepoint curve marks where bubbles of gas begin to appear in the liquid phase of the mixture. Where the dewpoint and bubblepoint curves meet is the critical point for this temperature and this mixture. For each point on the dewpoint curve, a dewpoint, there is a corresponding point on the bubblepoint curve, a bubblepoint, with equal pressure. For a pair of bubblepoint and dewpoint with matched pressure, we denote the impurity mole fraction at these points by x_l and x_g respectively, and their volumes by v_l and v_g respectively. When the pressure and impurity mole fraction of the binary system as a whole fall within the boundary defined by the phase envelope, the mixture will phase separate into a liquid with the properties of the bubblepoint (impurity mole fraction x_l and volume v_l), and a gas with the properties of the dewpoint (impurity mole fraction x_g and volume v_g), at that pressure.

At each pair of bubblepoint and dewpoint four equilibrium conditions will hold, mechanical equilibrium,

$$P_l = P_g, \tag{1.19}$$

where P_l and P_g are the pressures corresponding to the bubble and dewpoints respectively, thermal equilibrium,

$$T_l = T_g, \tag{1.20}$$

where T_l and T_g are the temperatures corresponding to the bubble and

dewpoints respectively, and equilibrium of the chemical potential of each component of the mixture, CO_2 and the impurity, at the bubble and dewpoints. As in the pure case equilibrium of the chemical potential at the dew and bubblepoints is equivalent to equilibrium of the fugacity at the dew and bubble points. The fugacity of component i of a mixture is [Orbey and Sandler, 1998],

$$\bar{f}_i = P x_i \bar{\phi}_i, \tag{1.21}$$

where x_i is the mole fraction of component i and $\bar{\phi}_i$ is the fugacity coefficient of component i of the mixture, defined as [Orbey and Sandler, 1998],

$$\bar{\phi}_i = \exp\left(\frac{1}{RT}\int_{V'=\infty}^{V} \frac{RT}{V'} - \left(\frac{\partial P}{\partial N_i}\right)_{T,V,N_{j\neq i}} \mathrm{d}v' - \ln\left(\frac{Pv}{RT}\right)\right), \tag{1.22}$$

where N_i is the number of moles of component i, N is the total number of moles, and $V = Nv$. Note that for a binary mixture of CO_2 with an impurity, $x_i = x$ if i denotes the impurity and $x_i = 1 - x$ if i denotes CO_2. Then for a binary mixture the two chemical equilibrium conditions are

$$\bar{f}_{\mathrm{CO_2},l} = \bar{f}_{\mathrm{CO_2},g}, \tag{1.23}$$

$$\bar{f}_{\mathrm{impurity},l} = \bar{f}_{\mathrm{impurity},g}. \tag{1.24}$$

where $\bar{f}_{i,l}$ is the fugacity of component i of the mixture evaluated at the bubblepoint and $\bar{f}_{i,g}$ is the fugacity of component i of the mixture evaluated at the dewpoint. We once again call these the fugacity constraints.

The presence of impurities in the CO_2 transported by pipelines can have a significant impact upon the coexistence behaviour as described here. This affects the pressure and temperature required to ensure the fluid remains supercritical and so can be most efficiently transported. This calls for a need for an EoS which can fit mixture data well taking into account the altered coexistence behaviour.

The difficulty in defining an EoS comes from satisfying all of these physical properties, the behaviour at the critical point, coexistence behaviour and the limiting behaviour. In the next section we introduce the two commonly used EoS formulations and briefly review several previously proposed equations of state.

1.4 Equations of State

1.4.1 Pressure-explicit and Free-energy-formulations

Many EoS models for CO_2 have been proposed [Demetriades et al., 2013, Kunz et al., 2007, Peng and Robinson, 1976, Span and Wagner, 1996]. These EoS commonly take one of two forms, which we define here for the case of pure CO_2 (see chapter 5 for a discussion of the extension of EoS to mixtures). The first are *pressure-explicit*, meaning they model the pressure P, as a function of molar volume, v, and temperature, T. So a pressure-explicit EoS considers P to be a function of v and T,

$$P = P(v, T). \tag{1.25}$$

We choose to model P, rather than T or v, as the response variable in our model due to the property of *coexistence* (a physical description, and definition, of which is given in §1.3.2). A consequence of this is that the function $P(v, T)$ must satisfy a constraint involving the integral of P with respect to v. As such it is helpful to define an EoS so that it is pressure-explicit rather than having to derive an implicit form for P. Henceforth all references to P presume it to be a function of v and T unless stated otherwise.

Sometimes EoS are defined as functions of density, ρ, rather than v,

where

$$\rho = \frac{M}{v}, \tag{1.26}$$

where M is the molar mass of the fluid (the mass per mole of fluid). As M has known value, when an EoS is defined as a function of ρ, we can substitute this expression for ρ into the EoS to derive an equivalent expression for the EoS as a function of v.

An alternative formulation for EoS is to model the *Helmholtz free energy*, A, of the system as a function of v and T rather than the pressure. So we consider,

$$A = A(v, T). \tag{1.27}$$

Suppose we have a pressure-explicit EoS (1.25) where P has known form, and we know the integral of P with respect to v, P_{int}, up to some constant,

$$P_{int} = \int^{v} P(v', T)\ \mathrm{d}v'. \tag{1.28}$$

Then the pressure-explicit EoS defined by equation (1.25) is equivalent to the free-energy-formulation EoS defined as,

$$A(v, T) = -P_{int}(v, T), \tag{1.29}$$

due to the relationship between P and A (equation (1.1)). As the two formulations are equivalent there is no scientific reason to favour one over the other. Instead the choice of formulation is left to each individual modeller. The pressure-explicit formulation is more intuitive for modelling, pressure-volume data can be plotted and so it is perhaps simpler to propose and understand models. However in the pressure-explicit formulation the fugacity constraints due to coexistence are expressed as integrals, thus we need an EoS which is integrable with respect to volume. In the free-energy-formulation the constraint takes a simpler form, requiring the difference

between two points to be equal to a constant. In chapters 3-5 we will use a parametric pressure-explicit EoS while in chapter 6 we describe a non–parametric free-energy-formulation EoS.

1.4.2 A Simple EoS: The Ideal Gas Law

The simplest example of an EoS is the ideal gas law (1.17),

$$P(v,T) = \frac{RT}{v}. \tag{1.30}$$

This is a pressure-explicit EoS. We saw previously from figure 1.3 how this simple EoS performs well for large v but poorly elsewhere. Thus there is a need for a more complex EoS able to more fully capture the behaviour of CO_2. In §1.5 several examples of proposed EoS (both pressure explicit and free-energy-formulation) are presented and discussed.

1.5 Review of Previously Proposed Equations of State

Previously proposed EoS [Demetriades et al., 2013, Kunz et al., 2007, Peng and Robinson, 1976, Span and Wagner, 1996] are parametric models, meaning they depend upon a set of p parameters which we collect in a p-dimensional vector $\boldsymbol{\theta}$. As discussed in §1.4.1 these are defined in one of two ways, pressure-explicit parametric EoS which have form,

$$P = P(v,T) = h(v,T,\boldsymbol{\theta}), \tag{1.31}$$

or free-energy-formulation parametric EoS,

$$A = A(v,T) = h(v,T,\boldsymbol{\theta}), \tag{1.32}$$

where h is some function.

1.5.1 Peng-Robinson Equation of State

A simple and widely used EoS is the Peng-Robinson EoS (PREoS) proposed by Peng and Robinson [1976],

$$P = \frac{RT}{v-b} - \frac{a(T)}{v(v+b)+b(v-b)}, \tag{1.33}$$

where a and b are the two model parameters of the PREoS. As there are only two model parameters this is a simple model hence its popularity. Due to its simplicity the PREoS is unable to accurately model the behaviour of the fluid around the critical point [Demetriades et al., 2013]. There are not enough terms in the model to account for the change from coexistence to the saddlepoint at the critical point. This suggests a need for a model with a greater number of terms and parameters than the PREoS.

1.5.2 Span-Wagner Equation of State

Examples of free energy defined equations of state are the Span-Wagner EoS (SWEoS) proposed by Span and Wagner [1996] and the GERG EoS [Kunz et al., 2007]. Compared with the PREoS these offer greater accuracy at the cost of higher functional complexity [Demetriades et al., 2013].

Span and Wagner [1996] define their EoS in terms of density, ρ, rather than v. As discussed in §1.4.1 the two methods are equivalent, and the following expressions for the SWEoS could be written in terms of v by substituting in the expression for density in terms of volume given by (1.26).

The SWEoS is,

$$A = A^o + \sum_{i=1}^{7} n_i \left(\frac{\rho}{\rho_c}\right)^{d_i} \left(\frac{T_c}{T}\right)^{t_i} + \sum_{i=8}^{34} n_i \left(\frac{\rho}{\rho_c}\right)^{d_i} \left(\frac{T_c}{T}\right)^{t_i} \exp\left(-\left(\frac{\rho}{\rho_c}\right)^{c_i}\right) + \sum_{i=35}^{39} n_i \left(\frac{\rho}{\rho_c}\right)^{d_i} \left(\frac{T_c}{T}\right)^{t_i} \exp\left(\alpha_i \left(\frac{\rho}{\rho_c} - \epsilon_i\right)^2 - \beta_i \left(\frac{T_c}{T} - \gamma_i\right)^2\right) + \sum_{i=40}^{42} n_i \Delta_i^{b_i} \left(\frac{\rho}{\rho_c}\right) \exp\left(-C_i \left(\frac{\rho}{\rho_c} - 1\right)^2 - D_i \left(\frac{T_c}{T} - 1\right)^2\right), \quad (1.34)$$

where

$$A^o = RT\left(\log\left(\frac{\rho}{\rho_c}\right) + a_1^o + a_1^o \frac{T_c}{T} + a_1^o \log\left(\frac{T_c}{T}\right) + \sum_{i=4}^{8} a_i^o \log\left(1 - \exp\left(-\frac{T_c}{T}\theta_i^o\right)\right)\right) \quad (1.35)$$

is the Helmholtz free energy due to the ideal gas law and

$$\Delta_i = \left(\left(1 - \frac{T_c}{T}\right) + A_i \left(\left(\frac{\rho}{\rho_c} - 1\right)^2\right)^{\frac{1}{2\beta_i}}\right)^2 + B_i \left(\frac{\rho}{\rho_c} - 1\right)^{2a_i}, \quad (1.36)$$

$A_i, B_i, C_i, D_i, a_i, b_i, c_i, d_i, n_i, t_i, \alpha_i, \beta_i, \gamma_i, \epsilon_i, \theta_i$, are the model parameters and ρ_c is the critical density. The SWEoS is clearly very complex, having a large number of parameters. Although more accurate than the PREoS this large number of parameters can prove difficult to fit and the model is computationally expensive [Demetriades et al., 2013]. It is also not straight forward to extend to mixtures. The GERG EoS [Kunz et al., 2007] is similarly complex.

There is thus a need for an EoS model which has more parameters than the PREoS and so is better able to model the behaviour of a fluid in all regions, but which is simpler and so more computationally efficient than the SWEoS. It also needs to be easily generalisable to mixtures. Demetriades et al. [2013] have proposed one such model, the Demetriades EoS (DEoS),

a pressure-explicit EoS which we discuss here.

1.5.3 Demetriades EoS

In proposing their EoS Demetriades et al. [2013] use the following non-dimensionalisation for v, T and P,

$$P = P_c P^*, \tag{1.37}$$

$$T = T_c T^*, \tag{1.38}$$

$$v = \frac{RT_c}{P_c} v^*, \tag{1.39}$$

where P^* is the non-dimensionalised pressure and so on. The reference values for the critical temperature, T_c (1.2), and pressure, P_c (1.3), from Dusschek et al. [1990a] are used in the non-dimensionalisation. From here onwards we shall denote the non-dimensionalised units without *, i.e. all quantities will be non-dimensionalised unless stated otherwise.

The DEoS is

$$h_D(v, T, \boldsymbol{\theta}) = \frac{T}{v + a(T)} + \frac{b(T)}{v^2 + c(T)} + \frac{d(T)}{v^3 + e(T)} + \left(\frac{f(T)}{v - g(T)} \right)^6. \tag{1.40}$$

If we are modelling a single temperature then we set

$$a(T) = a, \tag{1.41}$$

and likewise for $b, c, \ldots, g$. Then the single temperature DEoS has parameter vector,

$$\boldsymbol{\theta} = (a, b, c, d, e, f, g)^T, \tag{1.42}$$

For considering multiple temperatures we would need to define a functional form for $a(T), \ldots, g(T)$. These are known as temperature rules. To demonstrate we describe the linear temperature rule, the most simple temperature

rule, and the first we will implement, here.

Let T_l and T_h be the minimum and maximum temperatures we wish to model. Let a_l and a_h be the value of $a(T)$ at T_l and T_h respectively were they considered as single temperature models. Then for temperatures between T_l and T_h we linearly interpolate the value of a between a_l and a_h. We thus set

$$a(T) = a_l + \frac{(T - T_l)}{(T_h - T_l)}(a_h - a_l), \tag{1.43}$$

and do likewise for $b, \ldots, g$. Thus in the case of using linear temperature rules for all of $a, \ldots, g$, the complete set of model parameters is now a 14-dimensional vector,

$$\boldsymbol{\theta} = (a_l, \ldots, g_l, a_h, \ldots, g_h)^T. \tag{1.44}$$

For CCS we would define the minimum and maximum temperatures of interest, T_l and T_h, to be the minimum and maximum temperatures we would expect to operate a pipeline at. We are not concerned with temperatures outside of this range ($T < T_l$ or $T > T_h$) for our purposes.

To evaluate the fugacity coefficient for a pure fluid (1.12) we need to be able to evaluate the following integral,

$$h_{int}(v, T, \boldsymbol{\theta}) = \int_{v'=\infty}^{v} \frac{T}{v'} - h_D(v', T, \boldsymbol{\theta})\ \mathrm{d}v'. \tag{1.45}$$

In the case of the DEoS the integration to ∞ here converges and we have,

$$\begin{aligned} h_{int}(v, T, \boldsymbol{\theta}) = T \ln\left(\frac{v}{v + a(T)}\right) + \frac{\pi b(T)}{2\sqrt{c(T)}} - \frac{b(T)}{\sqrt{c(T)}} \arctan\left(\frac{v}{\sqrt{c(T)}}\right) \\ + \frac{\pi d(T)}{2\sqrt{3}e(T)^{\frac{2}{3}}} + \frac{d(T)}{3e(T)^{\frac{2}{3}}}\left(\ln\left(\frac{\sqrt{e(T)^{\frac{2}{3}} + ve(T)^{\frac{1}{3}} + v^2}}{e(T)^{\frac{1}{3}} + v}\right)\right) \\ + \frac{d(T)}{\sqrt{3}e(T)^{\frac{2}{3}}} \arctan\left(\frac{e(T)^{\frac{1}{3}} - 2v}{\sqrt{3}e(T)^{\frac{1}{3}}}\right) + \frac{f(T)^6}{5(v - g(T))^5}, \end{aligned} \tag{1.46}$$

if $c > 0$, and,

$$h_{int}(v, T, \boldsymbol{\theta}) = T \ln\left(\frac{v}{v + a(T)}\right) + \frac{\pi d(T)}{2\sqrt{3}e(T)^{\frac{2}{3}}} + \frac{d(T)}{3e(T)^{\frac{2}{3}}}\left(\ln\left(\frac{\sqrt{e(T)^{\frac{2}{3}} + ve(T)^{\frac{1}{3}} + v^2}}{e(T)^{\frac{1}{3}} + v}\right)\right) + \frac{d(T)}{\sqrt{3}e(T)^{\frac{2}{3}}}\arctan\left(\frac{e(T)^{\frac{1}{3}} - 2v}{\sqrt{3}e(T)^{\frac{1}{3}}}\right) + \frac{f(T)^6}{5(v - g(T))^5} + \frac{b(T)\mathrm{i}}{2\sqrt{c(T)}}\left(\ln\left(1 - \frac{\mathrm{i}\sqrt{c(T)}}{v}\right) - \ln\left(1 + \frac{\mathrm{i}\sqrt{c(T)}}{v}\right)\right), \quad (1.47)$$

if $c < 0$, where $\mathrm{i} = \sqrt{-1}$.

Given an EoS model such as the DEoS has significantly more parameters than the PREoS, the question arises of how to estimate these model parameters, and how to quantify our uncertainty about these estimated parameter values and predictions that result from the fitted model. Owing to the dimension of the parameter vector and the non-linearity of the model, parameter estimation is very challenging. We discuss this in detail in chapter 4, along with our proposed solution.

1.5.4 The Non-Parametric Approach

An alternative to building a parametric EoS is to use a *non-parametric* model. Parametric models like we have introduced so far have a fixed number of parameters which determine model behaviour. Inference, based on a given data set, is then made about these parameters, using either Bayesian or frequentist methods. The size of the data set does not affect the dimension of the parameter vector in the model.

Non-parametric models, in contrast, have dimensionality which grows with the size of the data set. The data itself becomes the parameters of the model. Hence such models are sometimes called infinite-parameter models. The number of parameters is constrained only by the size of our data set.

Non-parametric models require fewer assumptions, have the potential for greater flexibility, and offer a natural way to characterise uncertainty in predictions.

One possible non-parametric model for EoS is Gaussian process regression [Rasmussen and Williams, 2006]. We discuss modelling EoS using Gaussian process regression (GPR) in Chapter 6.

1.6 Summary of Available Data

The statistical models in this thesis are all motivated by previous experimental datasets which we describe in this section. We begin by describing the pure CO_2 datasets used in this thesis.

1.6.1 Pure CO_2 Data

Span and Wagner [1996] give an extensive summary of the available data for pure CO_2, including both homogeneous and coexistence data. They provide their judgement on the accuracy of each data set and detail the data sets they used in fitting the SWEoS. The data sets they recommend as most accurate, and primarily used for fitting the SWEoS, are Dusschek et al. [1990b] and Gilgen et al. [1992] for the homogeneous region, and Dusschek et al. [1990a] for both homogeneous and coexistence data. We follow Span and Wagner [1996] in using this pure CO_2 data for fitting EoS.

Dusschek et al. [1990a,b], Gilgen et al. [1992] include data on temperature (measured in K), pressure (measured in MPa) and density (measured in $\mathrm{kg\,m^{-3}}$). The temperatures are measure according to the International Practical Temperature Scale of 1968 (IPTS-68) [Comité International des Poids et Mesures, 1969, 1976]. To be consistent with [Span and Wagner, 1996] we convert these to the most recent International Temperature Scale of 1990 (ITS-90) [Preston-Thomas, 1990] using the formula in Preston-

Temperature (K)	Data Points	Temperature (K)	Data Points
217.0105	2	296.9940	15
220.0101	18	299.9933	42
230.0086	2	302.9925	37
233.1581	1	304.1272	10
240.0069	19	304.1472	13
250.0051	3	304.9919	22
260.0030	18	306.9914	23
270.0007	2	309.9906	25
270.8545	1	312.9898	20
279.9984	29	319.9879	23
289.9959	10	339.9826	22

Table 1.1: Summary of pure CO_2 data from Dusschek et al. [1990b]

Thomas et al. [1990], Rusby [1991],

$$T_{90} = T_{68} + \sum_{i=1}^{8} b_i \left(\frac{T_{68} - 273.15}{630} \right)^i , \tag{1.48}$$

where T_{68} and T_{90} are the temperatures in K according to the IPTS-68 and ITS-90 respectively and,

$$b_1 = -0.148759, \tag{1.49}$$

$$b_2 = -0.267408, \tag{1.50}$$

$$b_3 = 1.080760, \tag{1.51}$$

$$b_4 = 1.269056, \tag{1.52}$$

$$b_5 = -4.089591, \tag{1.53}$$

$$b_6 = -1.871251, \tag{1.54}$$

$$b_7 = 7.438081, \tag{1.55}$$

$$b_8 = -3.536296. \tag{1.56}$$

In this chapter we report the converted temperatures. We convert the densities into volumes using equation (1.26) and the known fact that the

Temperature (K)	Data Points	Temperature (K)	Data Points
220.0101	7	309.9906	14
240.0069	7	312.9898	13
260.0030	7	317.9885	30
279.9984	8	319.9879	15
289.9959	7	323.1371	44
296.9940	6	329.9853	20
299.9933	10	339.9826	10
302.9925	10	349.9800	21
306.9914	14	359.9775	21

Table 1.2: Summary of pure CO_2 data from Gilgen et al. [1992]

molar mass of CO_2 is,

$$M_{CO_2} = 44.0098\,\mathrm{g\,mol^{-1}}. \tag{1.57}$$

Dusschek et al. [1990b] has homogeneous data for temperatures ranging from 217.0105 to 339.9826 K. At these temperatures they recorded data for pressures of up to 9 MPa. In total there are 362 data points at 22 temperatures. Dusschek et al. [1990b] cite estimated uncertainties in pressure of less than 0.006% or 30 Pa, whichever is greater. Table 1.1 shows how many data points there are for each temperature in the Dusschek et al. [1990b] dataset.

Gilgen et al. [1992] expands the temperature and pressure ranges over which we have data from those in Dusschek et al. [1990b]. Gilgen et al. [1992] measure the properties of CO_2 at temperatures ranging from 220.0101 to 359.9775K, and they measure data for pressures up to 13MPa. For those temperatures at which Dusschek et al. [1990b] took measurements Gilgen et al. [1992] take measurements for pressures between 9 to 13 MPa. For temperatures not recorded in Dusschek et al. [1990b] they took measurements for pressures from 1 to 13 MPa. In total there are 264 data points at 18 temperatures. Estimated uncertainties in pressure are again 0.006% or 30 Pa, whichever is greater. Table 1.2 summarises how many data points

there are for each temperature in the Gilgen et al. [1992] dataset.

Dusschek et al. [1990a] records coexistence data for pure CO_2 at 31 temperatures ranging from 216.5905 to 304.0922 K. However rather than have a single measure of P_s, v_l and v_g at every temperature for which data was measured, varying numbers of measurements of each of P_s, v_l and v_g were taken. At each of the 31 temperatures for which data was recorded there is at least 1 measurement of P_s and for 29 of the temperatures there are at least 2 measurements of P_s. For some temperatures the total number of measurements of P_s can be as high as 13. For 30 of the temperatures for which data were recorded there are between 1 and 6 measurements of v_l at that temperature. However at 216.5905 K no measurement of v_l was taken. For 27 of the temperatures for which data were recorded there are between 1 and 6 measurements of v_G at that temperature. However at 216.5905 K, 303.8922 K, 303.9922 K and 304.0922 K (the lowest and three highest temperatures in the dataset) no measurement of v_g was taken. For temperatures near the critical temperature Dusschek et al. [1990a] also recorded homogeneous data in the vicinity of the coexistence points. Table 1.3 lists how many observations there are of P_s, v_l and v_g for each temperature in the Dusschek et al. [1990a] dataset. Table 1.4 lists how many homogeneous data points there are for temperatures in the Dusschek et al. [1990a] dataset where homogeneous measurements were taken. Dusschek et al. [1990a] reports uncertainty in pressure of between $0.012 - 0.016\%$ depending on temperature. Uncertainty in the coexistence volumes of less than 0.015% for v_l and less than 0.025% for v_g is reported for temperatures below 294.9926 K. As the critical temperature is approached the uncertainty in v_l and v_g increases to 0.2% in v_l and 0.25% in v_g.

Dusschek et al. [1990a] reports the critical values given in (1.2)-(1.4) along with associated uncertainties. These values are the results of an extrapolation, using a function for the relationship defined by Dusschek

Temperature (K)	P_s	v_l	v_g
216.5905	2	0	0
217.0105	5	2	1
220.0101	12	3	6
225.0094	2	1	1
230.0086	3	1	2
235.0078	2	1	1
240.0069	6	3	3
245.0060	2	1	1
250.0051	3	1	1
255.0040	2	1	1
260.0030	8	3	3
265.0019	2	1	1
270.0070	3	1	2
274.9996	2	1	1
279.9984	13	6	4
284.9971	2	1	1
289.9959	4	2	1
294.9946	5	4	1
296.9940	2	1	1
298.9935	2	1	1
299.9933	5	3	1
300.9930	2	1	1
301.9927	3	2	1
302.4926	3	1	2
302.9925	3	1	1
303.3924	3	2	1
303.5923	2	1	1
303.7923	2	1	1
303.8922	1	1	0
303.9922	2	1	0
304.0922	1	1	0

Table 1.3: Summary of pure CO_2 data for coexistence from Dusschek et al. [1990a]. P_s, v_l and v_g record how many observations of each respective coexistence identity are included for individual temperatures.

Temperature (K)	Data Points	Temperature (K)	Data Points
294.9946	5	302.9925	9
296.9940	4	303.3924	10
298.9935	5	303.5923	10
299.9933	9	303.7923	9
300.9930	6	303.8922	3
301.9927	5	303.9922	4
302.4926	8		

Table 1.4: Summary of homogeneous pure CO_2 data from Dusschek et al. [1990a].

et al. [1990a], from the measured coexistence values for temperatures near the critical temperature. Thus while we use these estimated critical values in the non-dimensionalisation to put our data on roughly the unit scale as described in §1.5.3 we do not explicitly include the information on the critical point in our statistical model for a parametric EoS (see Chapter 3). This is because the functional form of a parametric EoS such as the DEoS (1.40) is defined such that there should be a temperature at which a point with the properties of the critical point described in §1.3.1 exists, and we can estimate the critical values, and uncertainty about them, from our fit to the data in Dusschek et al. [1990a]. This is essentially the same approach as taken by Dusschek et al. [1990a].

As well as fitting our proposed model to pure CO_2 we also consider the case where the CO_2 in a pipeline is contaminated by an impurity. In this thesis we restrict our attention to a single impurity, H_2, available data for which we now summarise.

1.6.2 CO_2-H_2 Mixture Data

For the CO_2-H_2 mixture we use homogeneous data from Sanchez-Vicente et al. [2013], and coexistence data from Fandiño et al. [2015]. Sanchez-Vicente et al. [2013] include data on temperature (measured in K), pressure (measured in MPa), density (measured in $\mathrm{kg\,m^{-3}}$) and H_2 mole fraction.

Temperature (K)	mole fraction	data points
288.15	0.0997	7
288.15	0.0748	10
288.15	0.0197	14
293.15	0.1000	10
293.15	0.0751	13
293.15	0.0201	15
303.15	0.1000	39
303.15	0.0754	37
303.15	0.0203	29
313.15	0.0755	36
313.15	0.0203	30
323.15	0.1000	35
323.15	0.0754	36
323.15	0.0201	36
333.15	0.1000	36
333.15	0.0754	37
333.15	0.0198	30

Table 1.5: Summary of homogeneous CO_2-H_2 mixture data from Sanchez-Vicente et al. [2013]

We convert the densities into volumes using the known molar masses of CO_2 (1.57) and H_2,

$$M_{H_2} = 2.01588\,\mathrm{g\,mol^{-1}}, \tag{1.58}$$

and the fact that the molar mass of a mixture is the sum of the molar masses of each component of the mixture multiplied by the mole fraction of each component,

$$M_{CO_2\text{-}H_2}(x) = (1-x)M_{CO_2} + xM_{H_2}, \tag{1.59}$$

where $M_{CO_2\text{-}H_2}(x)$ is the molar mass of the mixture of CO_2-H_2 with H_2 mole fraction x, and then applying (1.26). Sanchez-Vicente et al. [2013] report data for 6 temperatures. At each temperature there is data for mixtures of CO_2-H_2 with 3 different concentrations of H_2. The data is for pressures of up to 23MPa. The number of data points for each temperature and each mole fraction are summarised in table 1.5.

Temperature (K)	data points
218.15	11
233.15	11
243.15	10
258.15	9
273.15	9
280.65	8
288.15	8
295.65	12
303.15	8

Table 1.6: Summary of coexistence CO_2-H_2 mixture data from Fandiño et al. [2015]

Fandiño et al. [2015] include data on temperature (measured in K), pressure (measured in MPa) and the coexistence H_2 mole fraction for each bubble and dewpoint pair (x_l and x_g). They did not however record any coexistence volumes. There are varying numbers of measurements for 9 temperatures between 218.15 and 303.15K at pressures of up to 15MPa. Table 1.6 records the numbers of data points at each temperature. Fandiño et al. [2015] report uncertainties in their pressure measurements of 0.003MPa.

1.7 Thesis Aims and Plan

The aim of this thesis is to develop statistical methods to fit EoS. In the case of parametric EoS we seek to quantify our uncertainty in the values of the parameters, to then quantify uncertainty in predictions. This is particularly necessary in the case of mixtures (CO_2 mixed with other gases) for which experimental data are noisy and sparse. We also seek to account for measurement error in coexistence volumes, v_l and v_g, or account for when these volumes are completely unknown. We seek to define a general approach for a pressure-explicit parametric EoS which we demonstrate using the example of the DEoS. We also investigate and compare the alternative approach of using a non-parametric model for the EoS which offers an al-

ternative, more flexible approach by which to quantify uncertainty in the physical properties of CO_2 and mixtures of CO_2 with impurities.

Chapter 2 is a literature review of the statistical methods applied in this thesis. We describe two methods of parameter estimation, maximum likelihood estimation, and the Bayesian approach. Bayesian parameter estimation requires the ability to sample from a probability distribution that need be defined only up to a normalising constant. Markov chain Monte Carlo (MCMC) methods have been developed to achieve this and chapter 2 goes on to describe the MCMC methods, including Hamiltonian Monte Carlo and parallel tempering, used in this thesis. We present toy examples to demonstrate the strengths and weakness of each MCMC method. We also review Bayesian models for missing data and errors-in-variables, and Gaussian processes in this chapter.

Chapter 3 constructs a statistical model for a parametric EoS for pure CO_2 which is able to model all of the pure CO_2 data in §1.6. We first describe our model for homogeneous data. We then show that coexistence can be modelled as a pair of constraints, and give two alternative approaches ("hard" and "soft", depending on whether these constraints are made to hold exactly or approximately) to including these constraints in the model. We define how measurement error in the coexistence volumes can be included in the model, and finally how multiple observations of coexistence pressures or volumes can be accounted for.

In chapter 4 we fit our proposed statistical model for a parametric EoS for pure CO_2 in the specific case of the Demetriades EoS (DEoS). We perform a Bayesian analysis and demonstrate the difficulty in constructing a sampler that can efficiently sample from the posterior distribution of the model parameters through application to synthetic data simulated from the DEoS. We show how using a sampler which combines parallel tempering and Hamiltonian Monte Carlo, what we call the PT-HMC sampler, we

can reliably and consistently fit our statistical model to the DEoS using synthetic data. We then show that our PT-HMC sampler can be used to reliably fit to the real pure CO_2 data as described in §1.6.

Chapter 5 is on the topic of mixtures of CO_2 with impurities. We propose an extension of our model to cover binary mixtures. Mixture coexistence volumes are often not observed, see §1.6 and so we further propose a model for this missing data. We then show that we are able to fit this model to CO_2-H_2 mixture data described §1.6 using the PT-HMC sampler developed in chapter 4.

Chapter 6 considers a non-parametric approach to constructing a statistical model for EoS, Gaussian processes [Rasmussen and Williams, 2006]. We describe how to construct a Gaussian process model of an EoS, showing how a Gaussian process model can ensure that physical constraints, such as coexistence, can be included in the model and made to hold exactly. Fitting to NIST reference data we demonstrate how to define a non-stationary covariance function that gives results with good agreement with the data.

Chapter 7 summarises the results of this thesis and presents possible directions for future research.

Chapter 2

Statistical Literature Review

2.1 Introduction

In this chapter we review statistical models and methods which are applied in this thesis. Firstly we introduce the concept of the likelihood, which forms the core of a statistical model. The likelihood relates data to model parameters which we wish to estimate.

We then describe two alternative methods for estimating parameters. Firstly the classical optimisation method of *maximum likelihood estimation* which obtains a point estimate for the parameter values. Large sample asymptotic approximations are then used to quantify uncertainty in this estimate. Secondly we describe *Bayesian* parameter estimation which accounts for uncertainty in model parameters by treating them as unknown random variables with some associated probability distribution. Application of Bayes rule then allows for evaluation of the probability distribution of these parameters conditioned on observed data, the *posterior* parameter distribution.

We then discuss how a Bayesian model can account for measurement error in, and missing values of, explanatory variables and introduce Gaussian process regression, a non-parametric approach to defining a statistical

model.

In a Bayesian analysis direct evaluation of the posterior distribution is often not possible due to an inability to analytically evaluate a normalising constant. As such a family of methods, *Markov chain Monte Carlo* (MCMC), have been developed to enable sampling from a probability distribution which is not known explicitly. In this chapter we describe several MCMC algorithms which are applied in this thesis, and discuss how different MCMC algorithms are better able to sample from probability distributions with features such as multi-modality and complex correlations between parameters. We use several toy examples to show the strengths and weaknesses of each MCMC method introduced here.

2.2 Likelihood

Suppose we have a data set consisting of n data points, where each data point consists of observations of $s+1$ variables. We collect these variables in the n-dimensional vectors $\boldsymbol{y}, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_s$, where,

$$\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^T, \tag{2.1}$$

$$\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in})^T, \tag{2.2}$$

and each vector consists of the corresponding variable across all observations. The j-th data point consists of $y_j, x_{1j}, \ldots, x_{sj}$. We seek to model $\boldsymbol{y}$ as a function of the $\boldsymbol{x}_i$. We call $\boldsymbol{y}$ our *response variable* and the $\boldsymbol{x}_i$ the *explanatory variables*. We collect the $\boldsymbol{x}_i$ in the n x s dimensional *design matrix*,

$$\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s), \tag{2.3}$$

so the j-th row of $\boldsymbol{X}$ is the explanatory variables corresponding to y_j. A model for this data set then defines a functional relationship between $\boldsymbol{y}$ and

$\boldsymbol{X}$. This relationship also typically depends on a set of p model parameters which we collect in the parameter vector,

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T. \tag{2.4}$$

In a statistical model the relationship between the response and explanatory variables takes the form of a probability distribution for $\boldsymbol{y}$ conditional on $\boldsymbol{X}$ and $\boldsymbol{\theta}$, which has probability density function (p.d.f.), $p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X})$. This probability density function then fully defines the statistical model. The parameter estimation problem is then to determine which values of $\boldsymbol{\theta}$ best fit observed $\boldsymbol{y}$ and $\boldsymbol{X}$. As such the p.d.f. is often defined as a function of $\boldsymbol{\theta}$ known as the *likelihood* and written,

$$L(\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}). \tag{2.5}$$

2.2.1 Maximum Likelihood Estimation

The simplest method of estimating model parameters is via an optimisation method. Optimisation methods treat model parameters as having some true fixed value and seek to estimate these parameters by maximising (or minimising) some objective function. This obtains a point estimate for the assumed true values of the parameters. For maximum likelihood estimation the objective function is the likelihood $L(\boldsymbol{\theta})$. To perform maximum likelihood estimation we wish to find the parameter values $\hat{\boldsymbol{\theta}}$, which maximise $L(\boldsymbol{\theta})$,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}). \tag{2.6}$$

If $\hat{\boldsymbol{\theta}}$ maximises $L(\boldsymbol{\theta})$ then it similarly maximises the log-likelihood,

$$l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}). \tag{2.7}$$

So we wish to calculate

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} l(\boldsymbol{\theta}). \tag{2.8}$$

We call $\hat{\boldsymbol{\theta}}$ the maximum likelihood estimate (MLE). The log-likelihood, $l(\boldsymbol{\theta})$, can be maximised, and the MLE obtained, using standard optimisation methods such as Nelder-Mead [Nelder and Mead, 1965] and simulated annealing [Kirkpatrick et al., 1983].

The MLE, $\hat{\boldsymbol{\theta}}$, has a *sampling distribution*. This is the distribution of MLEs that would arise were we to take repeated samples of fixed size from $p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X})$ and calculate the corresponding MLE for each sample. This distribution can be estimated based on large-sample asymptotic approximations. As our number of data points n increases we have that,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \rightarrow^{d} \mathcal{N}(0, \boldsymbol{I}^{-1}), \tag{2.9}$$

where $\rightarrow^{d}$ indicates convergence in probability distribution and $\boldsymbol{I}$ is the *Fisher information matrix*, the p x p matrix which has elements,

$$I_{ij} = -\mathbb{E}\left(\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right), \tag{2.10}$$

where $\mathbb{E}$ denotes expectation. Therefore the MLE is asymptotically normal and we can use (2.9) to quantify the uncertainty in $\hat{\boldsymbol{\theta}}$.

Given an MLE, $\hat{\boldsymbol{\theta}}$, the predicted value of new observations, $\boldsymbol{y}_{\text{new}}$ with s explanatory variables given by the design matrix, $\boldsymbol{X}_{\text{new}}$, can be obtained by evaluating,

$$p(y_{\text{new}}|\hat{\boldsymbol{\theta}}, \boldsymbol{X}_{\text{new}}). \tag{2.11}$$

The normal approximation to the distribution of the MLE (2.9) is only true asymptotically. Thus for finite n this approximation may not accurately describe the true distribution of the MLE and so our quantification of the uncertainty in the MLE could be inaccurate. We also need to be able

to find the MLE reliably. Depending on the form of the likelihood surface this can be difficult to achieve (see §4.2.1 for the difficulty in achieving this with the models considered in this thesis). This motivates us to consider an alternative *Bayesian* approach to parameter estimation.

2.3 Bayesian Parameter Estimation

An alternative approach to using MLE for parameter estimation is to undertake a *Bayesian* analysis. Bayesian parameter estimation treats the unknown $\boldsymbol{\theta}$, as random variables directly, making inferences about them based on the probability distribution of these parameters conditioned on our data, called the *posterior distribution* of the parameters. This offers a natural way to quantify the uncertainty in our parameter estimates.

Bayesian analysis entails representing our beliefs about the values of parameters as a probability distribution on these parameters, called the *prior*. Such beliefs before seeing the data are represented by the prior, defined through the prior p.d.f., $p(\boldsymbol{\theta})$. The posterior distribution of $\boldsymbol{\theta}$, which represents our updated beliefs about the parameters once information from the data is taking into account, can then be defined through application of Bayes rule,

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X})p(\boldsymbol{\theta})}{p(\boldsymbol{y})}, \tag{2.12}$$

where $p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X})$ is the posterior p.d.f. of the parameters, $p(\boldsymbol{\theta})$ and $p(\boldsymbol{y})$ is the marginal p.d.f. of $\boldsymbol{y}$ defined as,

$$p(\boldsymbol{y}) = \int p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X})p(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}, \tag{2.13}$$

and takes the role of the normalising constant, ensuring that the posterior p.d.f. integrates to 1. Thus the Bayesian approach to parameter estimation takes our initial beliefs about the parameters made before we observe any

data, the prior, and updates these beliefs based on the model we have defined for the data.

Evaluating the marginal distribution $p(\boldsymbol{y})$ is often difficult meaning that the posterior cannot be computed directly. Instead we can *sample* from the posterior distribution, by methods that require only that the posterior be known up to a constant of proportionality,

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}) \propto p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X})p(\boldsymbol{\theta}). \tag{2.14}$$

To sample from the posterior distribution defined by (2.14) we can use methods of Markov chain Monte Carlo (MCMC) which we describe later in this chapter.

It is sometimes the case that some or all of the elements of $\boldsymbol{X}$ either have some associated measurement error or are missing. The case of measurement error is commonly known as *errors-in-variables*. The Bayesian approach offers a simple way to account for the added uncertainty resulting from errors-in-variables or missing data by considering the true variables as additional parameters [Stephens and Dellaportas, 1992, Dellaportas and Stephens, 1995, Gelman et al., 2014]. We describe the resulting models in the next section.

2.4 Errors-in-Variables and Missing Data

2.4.1 Bayesian Errors-in-Variables

We here follow the approach introduced in Stephens and Dellaportas [1992] and Dellaportas and Stephens [1995] to modelling errors-in-variables in a Bayesian context. In the errors-in-variables case we have an observed design matrix $\boldsymbol{X}_O$, and a true design matrix $\boldsymbol{X}_T$, where X_{Tij} is the true value corresponding to the observed X_{Oij}. The Bayesian errors-in-variables

approach is to treat the elements of $\boldsymbol{X}_T$ as additional parameters. We assume that $\boldsymbol{y}$ does not depend on $\boldsymbol{X}_O$ given $\boldsymbol{X}_T$ and so the likelihood is, $p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}_T)$.

To define the joint posterior of $\boldsymbol{\theta}$ and $\boldsymbol{X}_T$ we require a probability model for our parameters $\boldsymbol{\theta}$ and $\boldsymbol{X}_T$ conditional on $\boldsymbol{X}_O$, $p(\boldsymbol{\theta}, \boldsymbol{X}_T|\boldsymbol{X}_O)$. We make the prior assumption that $\boldsymbol{\theta}$ is independent of $\boldsymbol{X}_T$ and $\boldsymbol{X}_O$. Then

$$p(\boldsymbol{\theta}, \boldsymbol{X}_T|\boldsymbol{X}_O) = p(\boldsymbol{\theta})p(\boldsymbol{X}_T|\boldsymbol{X}_O). \tag{2.15}$$

If we can define $p(\boldsymbol{X}_T|\boldsymbol{X}_O)$ directly this is the *Berkson* errors-in-variables case. Alternatively the *classical* approach, applied in this thesis, is to model $\boldsymbol{X}_O$ conditional on $\boldsymbol{X}_T$ with p.d.f. $p(\boldsymbol{X}_T|\boldsymbol{X}_O)$, and define a prior on $\boldsymbol{X}_T$, $p(\boldsymbol{X}_T)$. By application of Bayes rule the p.d.f of $\boldsymbol{X}_T$ conditional on $\boldsymbol{X}_O$ is then,

$$p(\boldsymbol{X}_T|\boldsymbol{X}_O) = \frac{p(\boldsymbol{X}_O|\boldsymbol{X}_T)p(\boldsymbol{X}_T)}{p(\boldsymbol{X}_O)}. \tag{2.16}$$

In the classical approach it is often the case that $\boldsymbol{X}_O$ will depend on some parameter(s) as well as $\boldsymbol{X}_T$. For example if $\boldsymbol{X}_O$ is normally distributed with mean $\boldsymbol{X}_T$ and covariance matrix $\boldsymbol{\Sigma}$ then the entries of $\boldsymbol{\Sigma}$ are additional parameters. Let $\boldsymbol{\phi}$ be the vector of additional parameters. Then by application of Bayes rule,

$$p(\boldsymbol{X}_T, \boldsymbol{\phi}|\boldsymbol{X}_O) = \frac{p(\boldsymbol{X}_O|\boldsymbol{X}_T, \boldsymbol{\phi})p(\boldsymbol{X}_T)p(\boldsymbol{\phi})}{p(\boldsymbol{X}_O)}, \tag{2.17}$$

assuming prior independence of $\boldsymbol{X}_T$ and $\boldsymbol{\phi}$.

We can then define the joint posterior density of $(\boldsymbol{\theta}, \boldsymbol{X}_T, \boldsymbol{\phi})$ as,

$$p(\boldsymbol{\theta}, \boldsymbol{X}_T, \boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{X}_O) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}_T)p(\boldsymbol{X}_O|\boldsymbol{X}_T, \boldsymbol{\phi})p(\boldsymbol{\theta})p(\boldsymbol{X}_T)p(\boldsymbol{\phi})}{p(\boldsymbol{y}, \boldsymbol{X}_O, \boldsymbol{\phi})}, \tag{2.18}$$

where

$$p(\boldsymbol{y}, \boldsymbol{X}_O, \boldsymbol{\phi}) = \int\int\int p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}_T)p(\boldsymbol{X}_O|\boldsymbol{X}_T, \boldsymbol{\phi})p(\boldsymbol{\theta})p(\boldsymbol{X}_T)p(\boldsymbol{\phi})\,\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{X}_O\mathrm{d}\boldsymbol{\phi}. \quad (2.19)$$

As before in the case that we can not derive an analytic expression for the integral in (2.18) the posterior density can be defined up to proportionality,

$$p(\boldsymbol{\theta}, \boldsymbol{X}_T, \boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{X}_O) \propto p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}_T)p(\boldsymbol{X}_O|\boldsymbol{X}_T, \boldsymbol{\phi})p(\boldsymbol{\theta})p(\boldsymbol{X}_T)p(\boldsymbol{\phi}). \quad (2.20)$$

We may then sample from the joint posterior distribution using MCMC methods.

We have here described how to model uncertainty in elements of $\boldsymbol{X}$. We now consider the case when elements of $\boldsymbol{X}$ are not just uncertain but missing entirely.

2.4.2 Bayesian Missing Data

Let us set $\boldsymbol{X}_{\text{mis}}$ denote the missing elements of $\boldsymbol{X}$, and let $\boldsymbol{X}_{\text{obs}}$ denote the observed elements of $\boldsymbol{X}$. Then the full Bayesian approach to missing data is to model the elements of $\boldsymbol{X}_{\text{mis}}$ as parameters [Gelman et al., 2014]. This is akin to how in the errors-in-variables case we modelled the true unknown values of the elements of $\boldsymbol{X}$ which were observed with error as parameters.

Then we need to consider the missingness indicator matrix $\boldsymbol{M}$. The missingness indicator is a a matrix with same dimension as $\boldsymbol{X}$ with $M_{ij} = 1$ when the corresponding element of $\boldsymbol{X}$ is observed and 0 otherwise. We now need to consider the distribution of $\boldsymbol{M}$ conditional on $\boldsymbol{X}$, which has density $p(\boldsymbol{M}|\boldsymbol{X}, \boldsymbol{\phi})$, where $\boldsymbol{\phi}$ is some vector of unknown parameters. The distribution of $\boldsymbol{M}$ is described as taking one of 3 forms [Rubin, 1976]. If

the distribution of the missingness indicator does not depend on $\boldsymbol{X}$, so,

$$p(\boldsymbol{M}|\boldsymbol{X}, \boldsymbol{\phi}) = P(\boldsymbol{M}|\boldsymbol{\phi}), \tag{2.21}$$

the data is missing completely at random (MCAR). In the MCAR case the distribution of the missingness indicator depends only upon the parameters denoted by $\boldsymbol{\phi}$.

If the distribution of $\boldsymbol{M}$ depends upon $\boldsymbol{X}$ through the observed elements $\boldsymbol{X}_{\text{obs}}$ only, and so,

$$p(\boldsymbol{M}|\boldsymbol{X}, \boldsymbol{\phi}) = p(\boldsymbol{M}|\boldsymbol{X}_{\text{obs}}, \boldsymbol{\phi}), \tag{2.22}$$

then the data is defined to be missing at random (MAR).

Finally if the distribution of $\boldsymbol{M}$ depends upon both the observed and missing elements of $\boldsymbol{X}$ and so,

$$p(\boldsymbol{M}|\boldsymbol{X}, \boldsymbol{\phi}) = p(\boldsymbol{M}|\boldsymbol{X}_{\text{mis}}, \boldsymbol{X}_{\text{obs}}, \boldsymbol{\phi}), \tag{2.23}$$

then the data is missing not at random (MNAR). The distribution of $\boldsymbol{M}$ is said to describe the missing-data mechanism.

If the data is MCAR or MAR the missing-data mechanism is ignorable, and we do not need to account for it in our posterior distribution.

We seek to calculate the joint posterior distribution of $\boldsymbol{\theta}$, and $\boldsymbol{X}_{\text{mis}}$. We have likelihood, $p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}_{\text{mis}}, \boldsymbol{X}_{\text{obs}})$. Then we need to define the prior, $p(\boldsymbol{\theta}, \boldsymbol{X}_{\text{mis}}|\boldsymbol{X}_{\text{obs}})$. If we assume that $\boldsymbol{\theta}$ is independent of $\boldsymbol{X}_{\text{mis}}$ and $\boldsymbol{X}_{\text{obs}}$, and allow the distribution of $\boldsymbol{X}_{\text{mis}}$ dependent on $\boldsymbol{X}_{\text{obs}}$ to depend upon some parameters $\boldsymbol{\psi}$ then we have,

$$p(\boldsymbol{\theta}, \boldsymbol{X}_{\text{mis}}|\boldsymbol{X}_{\text{obs}}) = p(\boldsymbol{\theta})p(\boldsymbol{X}_{\text{mis}}|\boldsymbol{X}_{\text{obs}}, \boldsymbol{\psi}). \tag{2.24}$$

The full posterior distribution is then as follows [Gelman et al., 2014],

$$p(\boldsymbol{\theta}, \boldsymbol{X}_{\text{mis}}, \boldsymbol{\psi}|\boldsymbol{y}, \boldsymbol{X}_{\text{obs}}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}_{\text{mis}}, \boldsymbol{X}_{\text{obs}})p(\boldsymbol{X}_{\text{mis}}|\boldsymbol{X}_{\text{obs}}, \boldsymbol{\psi})p(\boldsymbol{\theta})p(\boldsymbol{\psi})}{p(\boldsymbol{y}, \boldsymbol{X}_{\text{obs}})}, \tag{2.25}$$

where

$$\begin{aligned} &p(\boldsymbol{y}, \boldsymbol{X}_{\text{obs}}) \\ &\quad = \int\int\int p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}_{\text{mis}}, \boldsymbol{X}_{\text{obs}})p(\boldsymbol{X}_{\text{mis}}|\boldsymbol{X}_{\text{obs}}, \boldsymbol{\psi})p(\boldsymbol{\theta})p(\boldsymbol{\psi})\, \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{X}_{\text{mis}}\mathrm{d}\boldsymbol{\psi}. \end{aligned} \tag{2.26}$$

As previously it is often impossible to derive an analytic expression for (2.26). In that case the posterior can be defined up to proportionality as,

$$p(\boldsymbol{\theta}, \boldsymbol{X}_{\text{mis}}, \boldsymbol{\psi}|\boldsymbol{y}, \boldsymbol{X}_{\text{obs}}) \propto p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}_{\text{mis}}, \boldsymbol{X}_{\text{obs}})p(\boldsymbol{X}_{\text{mis}}|\boldsymbol{X}_{\text{obs}}, \boldsymbol{\psi})p(\boldsymbol{\theta})p(\boldsymbol{\psi}). \tag{2.27}$$

We could then sample from this posterior using the MCMC methods introduced later in this chapter. Before introducing MCMC methods we consider an alternative approach to parametric modelling, the non-parametric Gaussian process approach.

2.5 Gaussian Processes

Suppose we wish to model some function f as a function of s explanatory variables, $x_1, \ldots, x_s$. Then rather than define some fixed parametric form for f we can attempt to learn the form of f from data by placing a Gaussian process (GP) prior on f. Rasmussen and Williams [2006] define a Gaussian process as follows:

Definition A Gaussian process is a collection of random variables, any finite number of which has a joint Gaussian distribution.

A GP is a non-parametric model, it has dimensionality which grows with

the size of the data set and so learns its form from the data. GPs offer several advantages requiring fewer assumptions and having more flexibility than parametric models when fitted to large datasets. However for small data sets GPs use data inefficiently. Given a small data set that comes from a parametric model the parametric model will fit the data better than a GP. Another advantage offered by GPs is that the linear transformation of a GP is itself a GP [Murray-Smith and Pearlmutter, 2005, Rasmussen and Williams, 2006]. This is an extremely useful property for modelling EoS and we discuss this in more detail in chapter 6.

Given a function $f(\boldsymbol{x})$, a Gaussian process prior treats each realisation of this function for an individual input $\boldsymbol{x} = (x_1, \ldots, x_s)$ as a random variable. We then assume that these realisations are jointly normally distributed. To fully define a GP we need to define a *mean function*,

$$m(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})], \tag{2.28}$$

and *covariance function*,

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}[(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))]. \tag{2.29}$$

where the covariance function defines the covariance between f evaluated at inputs $\boldsymbol{x}$ and $\boldsymbol{x}'$. There is no restriction on the form that the mean function may take, however covariance functions must be positive semidefinite. We write that,

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}, k(\boldsymbol{x}, \boldsymbol{x}')). \tag{2.30}$$

Suppose we know n values of f, which we call $\boldsymbol{f} = (f_1, \ldots, f_n)$ We collect the inputs corresponding to these values of f in the n x s design matrix, $\boldsymbol{X}$, where $\boldsymbol{x}_{.i}$ is the i-th column of $\boldsymbol{X}$ and is the n-dimensional vector of all observations of explanatory variable x_i, and $\boldsymbol{x}_{j.}$ is the j-th row

of $\boldsymbol{X}$ and is the s-dimensional vector consisting of the input corresponding to f_i. Now suppose we wish to evaluate f for m new inputs which we collect in the m x s matrix $\boldsymbol{X}^*$, where as with $\boldsymbol{X}$ the rows of $\boldsymbol{X}^*$, $\boldsymbol{x}^*_{j\cdot}$, correspond to the j-th s-dimensional input at which we wish to evaluate our function f. Let

$$\boldsymbol{f}^* = (f(\boldsymbol{x}^*_{1\cdot}), \ldots, f(\boldsymbol{x}^*_{m\cdot}))^T = (f^*_1, \ldots, f^*_m)^T, \tag{2.31}$$

be the unknown values of f we wish to evaluate. Then from the definition of the GP we have that $\boldsymbol{f}$ and $\boldsymbol{f}^*$ are jointly normally distributed. Thus we can define the distribution of $\boldsymbol{f}^*$ conditional on $\boldsymbol{f}$, as follows [Rasmussen and Williams, 2006],

$$\begin{aligned}\boldsymbol{f}^*|\boldsymbol{f}, \boldsymbol{X}, \boldsymbol{X}^* \sim \mathcal{N}(&m(\boldsymbol{X}^*) + K(\boldsymbol{X}^*, \boldsymbol{X})K(\boldsymbol{X}, \boldsymbol{X})^{-1}(\boldsymbol{f} - m(\boldsymbol{X})),\\ &K(\boldsymbol{X}^*, \boldsymbol{X}^*) - K(\boldsymbol{X}^*, \boldsymbol{X})K(\boldsymbol{X}, \boldsymbol{X})^{-1}K(\boldsymbol{X}, \boldsymbol{X}^*)),\end{aligned} \tag{2.32}$$

where

$$m(\boldsymbol{X}) = (m(\boldsymbol{x}_{1\cdot}), \ldots, m(\boldsymbol{x}_{n\cdot}))^T, \tag{2.33}$$

and $K(\boldsymbol{X}, \boldsymbol{X}')$ is a matrix with elements,

$$K_{ij}(\boldsymbol{X}, \boldsymbol{X}') = k(\boldsymbol{x}_{i\cdot}, \boldsymbol{x}'_{j\cdot}). \tag{2.34}$$

So we can condition the unknown $\boldsymbol{f}^*$ on the observed $\boldsymbol{f}$ and quantify our predictive uncertainty about the $\boldsymbol{f}^*$ through the probability distribution in (2.32), and sample from it to obtain predictions for $\boldsymbol{f}^*$. If f^*_i were the function evaluated at the same input as an element of $\boldsymbol{f}$ we have that the variance in (2.32) would be 0. As such our predictions interpolate the data used to train the GP [Rasmussen and Williams, 2006].

We here assumed that $\boldsymbol{f}$ was observed exactly. Suppose instead that we observe

$$y_i = f_i + \epsilon_i, \tag{2.35}$$

where the ϵ_i have joint normal distribution with mean 0 and covariance matrix $\boldsymbol{\Sigma}$. So we observe $\boldsymbol{f}$ with some error. We assume the ϵ_i are independent of $\boldsymbol{f}$. Then due to additivity of independent Gaussian distributions we have

$$\boldsymbol{y} \sim \mathcal{N}(m(\boldsymbol{X}), K(\boldsymbol{X}, \boldsymbol{X}) + \boldsymbol{\Sigma}). \tag{2.36}$$

Then $\boldsymbol{y}$ and $\boldsymbol{f}^*$ are again jointly normally distributed and we have [Rasmussen and Williams, 2006],

$$\begin{aligned}\boldsymbol{f}^*|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{X}^* \sim \mathcal{N}(m(\boldsymbol{X}^*) + K(\boldsymbol{X}^*, \boldsymbol{X})(K(\boldsymbol{X}, \boldsymbol{X}) + \boldsymbol{\Sigma})^{-1}(\boldsymbol{f} - m(\boldsymbol{X})),\\ K(\boldsymbol{X}^*, \boldsymbol{X}^*) - K(\boldsymbol{X}^*, \boldsymbol{X})(K(\boldsymbol{X}, \boldsymbol{X}) + \boldsymbol{\Sigma})^{-1}K(\boldsymbol{X}, \boldsymbol{X}^*)).\end{aligned} \tag{2.37}$$

Due to the noise on the observations predictions sampled from this distribution will no longer interpolate the data used to train the GP [Rasmussen and Williams, 2006].

Mean and covariance functions can in principle depend upon some set of p parameters which we collect in the p-dimensional parameter vector, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$, so

$$m(\boldsymbol{x}) = m(\boldsymbol{x}, \boldsymbol{\phi}), \tag{2.38}$$

$$k(\boldsymbol{x}, \boldsymbol{x}') = k(\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{\phi}), \tag{2.39}$$

where we call $\boldsymbol{\phi}$ the hyperparameters of the GP. We now describe how we could estimate uncertainty in these hyperparameters using Bayesian methods.

2.5.1 Bayesian Estimation of Hyperparameters

From (2.36) we can write down the likelihood of data $\boldsymbol{y}$,

$$p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\phi}) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{y}-m(\boldsymbol{X}))^T(K(\boldsymbol{X},\boldsymbol{X})+\boldsymbol{\Sigma})^{-1}(\boldsymbol{y}-m(\boldsymbol{X}))\right)}{\sqrt{(2\pi)^n|K(\boldsymbol{X},\boldsymbol{X})+\boldsymbol{\Sigma}|}}. \tag{2.40}$$

Then if we define a prior $p(\boldsymbol{\phi})$ for the hyperparameters we have posterior distribution,

$$p(\boldsymbol{\phi}|\boldsymbol{y},\boldsymbol{X}) = \frac{p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\phi})p(\boldsymbol{\phi})}{p(\boldsymbol{y})}, \tag{2.41}$$

where $p(\boldsymbol{y})$ is defined as in (2.13). Again it is often impossible to exactly evaluate $p(\boldsymbol{y})$ and so we can define the posterior up to proportionality and sample from it using MCMC methods, which we introduce here.

2.6 Markov Chain Monte Carlo

It is often not possible to derive an analytic expression for the normalising constant, $p(\boldsymbol{y})$ (2.13). Then it is not possible to derive an analytic expression for the posterior p.d.f. $p(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{X})$ (2.12). This has led to the development of Markov chain Monte Carlo (MCMC) methods [Robert and Casella, 2004, Gelman et al., 2014]. MCMC methods offer a means of sampling from a probability distribution where the distribution is only known up to proportionality (2.14).

These methods work by constructing a Markov chain which has the probability distribution of interest as its stationary distribution. Once the Markov chain converges to its stationary distribution this is equivalent to sampling from the posterior distribution of interest. The Markov chain is then simulated for a desired number of iterations and these then form the sample from the probability distribution of interest.

To define an MCMC sampler we need to define the *transition density* $W(\boldsymbol{\theta}^{t+1}|\boldsymbol{\theta}^t)$. The transition density defines how at time t the Markov chain

moves from the current parameter values $\boldsymbol{\theta}^t$ to the parameter values at time $t+1$, $\boldsymbol{\theta}^{t+1}$. Then an MCMC sampler consists of a transition density W and a set of initial parameter values, $\boldsymbol{\theta}^0$.

We require the Markov chain to have stationary distribution equal to our distribution of interest and need to choose a transition density accordingly. *Detailed balance* is a sufficient condition for property that the stationary distribution of the Markov chain is the posterior distribution we wish to sample from [Robert and Casella, 2004]. Detailed balance is defined as,

$$p(\boldsymbol{\theta}^t|\boldsymbol{y},\boldsymbol{X})W(\boldsymbol{\theta}^{t+1}|\boldsymbol{\theta}^t) = p(\boldsymbol{\theta}^{t+1}|\boldsymbol{y},\boldsymbol{X})W(\boldsymbol{\theta}^t|\boldsymbol{\theta}^{t+1}). \tag{2.42}$$

If a Markov chain has stationary distribution equal to our distribution of interest then were that Markov chain run for an infinite amount of time we would have guaranteed convergence to the posterior distribution of interest [Robert and Casella, 2004]. However we can obviously only run an MCMC sampler for a finite amount of time. This calls for a need to use MCMC methods that are able to converge to the posterior distribution of interest in a reasonable finite amount of time.

Suppose we have two transition densities, W_1 and W_2, for which Markov chains which transition according to these distributions have stationary distribution which is our distribution of interest. Then a Markov chain which alternates transitions between W_1 and W_2, will also have stationary distribution equal to our distribution of interest so long as we use the same fixed sequence of steps (for example transition twice according to W_1 and then once according to W_2) called a *scan* [Geyer, 1991]. This argument can be extended to any arbitrary number of transition densities.

It is typical for the first N iterations of an MCMC sampler to be discarded, called *burn-in*. This is because these iterations reflect the chosen starting point rather than the distribution of interest [Gelman et al., 2014]. The Markov chain does not sample from the distribution of interest until

it has converged to its stationary distribution and so burn-in discards the early iterations of the Markov chain before it has converged.

To determine the value of N we can consider the *traceplot* of the sampler. The traceplot plots each iteration of the sampler in order. From this we can judge after which iteration the sampler has converged to its stationary distribution, ie. is no longer affected by the choice of $\boldsymbol{\theta}^0$, and discard as burn-in all iterations up to the point at which we judge convergence to have taken place.

From the traceplot we can also judge how well our sampler is *mixing*. We say a sampler is mixing well if $\boldsymbol{\theta}^{t+1}$ is independent of $\boldsymbol{\theta}^t$. If our sampler is mixing poorly $\boldsymbol{\theta}^{t+1}$ will depend greatly on $\boldsymbol{\theta}^t$. As such each iteration of the sampler will only be a small way removed from the previous iteration. The sampler will thus explore the parameter space slowly. To correct a poorly mixing sampler it is common to *thin* the iterations of the sampler, to keep only every k-th iteration Gelman et al. [2014]. So say we have run our sampler for R iterations, our sample for $\boldsymbol{\theta}$ is, $(\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^R)$. Then if we were to thin this run of our sampler, keeping only every k-th iteration, where R is divisible by k, our thinned sample would be

$$\boldsymbol{\theta}^{rk}, \quad r = 1, 2, \ldots, \frac{R}{k}. \tag{2.43}$$

From traceplots we can further judge if a sampler fails to converge to a stationary distribution. It is common to make multiple runs of the sampler using different starting values $\boldsymbol{\theta}^0$. We call these *chains* here. The running of multiple chains of the sampler allows us to determine if the sampler is consistently converging to the same distribution. A sampler which works well will consistently converge to the same distribution.

Use of multiple chains also allows us to determine if we are attempting to sample from a multi-modal distribution. If a distribution is multi-modal while it may appear that all of our chains have converged, comparison of

chains can demonstrate that they have converged in different locations. These locations are modes of the distribution, each is the location of a local maximum of the posterior density. If the sampler is able to transition between modes in reasonable time our sampler is working well, it is exploring a multi-modal distribution with multiple modes of significant posterior probability. However if chains get trapped in a local mode of much lower posterior probability than the global mode, we have a sampler that is performing poorly. Were we able to run the sampler for an infinite amount of time we would find a swap to the global mode happens eventually due to the guaranteed convergence of a Markov chain to its stationary distribution [Robert and Casella, 2004]. However to achieve convergence in reasonable finite time we need to consider MCMC algorithms that are able to swap between modes within a reasonable amount of time.

Often MCMC algorithms depend on some tuning parameters. We may choose to fix these parameters when we initiate the runs of our sampler or adapt these tuning parameters with each iteration of the sampler [Andrieu and Thoms, 2008].

In the remainder of this chapter we describe three different MCMC algorithms applied in this thesis; Metropolis-Hastings, a simple and commonly used MCMC algorithm; Hamiltonian Monte Carlo, an MCMC algorithm which seeks to avoid inefficient random walk behaviour (explained in more detail later); and parallel tempering, a method for dealing with multimodality. To demonstrate the strengths and weaknesses of these algorithms we apply them to some toy examples which we introduce here.

2.6.1 Toy Examples

Each of the toy examples used here is two-dimensional. We set,

$$\boldsymbol{\theta} = (\theta_1, \theta_2)^T, \tag{2.44}$$

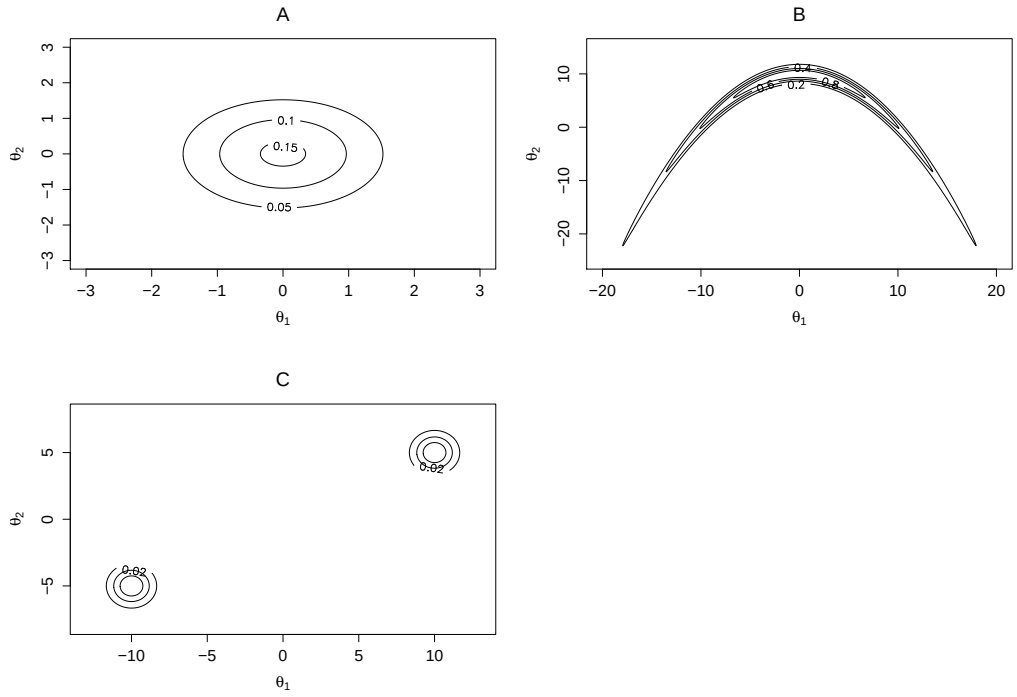


Figure 2.1: Contour plots of the densities used in toy examples in this chapter. Plot A are the contours of the multivariate normal density p_1 (2.45). Plot B are the contours for the "banana-shaped" density p_2 (2.46). Plot C are the contours of the multi-modal mixture of Gaussians density p_3 (2.49).

and define three probability distributions for $\boldsymbol{\theta}$ which we wish to sample from. The first toy example is the multivariate normal distribution with zero mean and the identity matrix as the covariance matrix. This has density,

$$p_1(\boldsymbol{\theta}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(\theta_1^2 + \theta_2^2\right)\right). \tag{2.45}$$

Plot A in figure 2.1 is a contour plot of this density. From this we can see that as we would expect p_1 is uni-modal and there is no correlation between the parameters, as such we would expect this to be a relatively simple distribution to sample from.

For the second toy example we use a "banana-shaped" distribution first introduced by Haario et al. [1999, 2001]. Defined up to proportionality this

distribution has density,

$$p_2(\boldsymbol{\theta}) \propto \exp\left(-\frac{\theta_1^2}{100} - \frac{1}{2}(\theta_2 + 0.1\theta_1^2 - 10)^2\right). \tag{2.46}$$

Plot B in figure 2.1 is a contour plot of this un-normalised density. From this plot we can see that there is a complex correlation between θ_1 and θ_2, hence the distribution's description as"banana-shaped". We would expect this relationship between θ_1 and θ_2 to make the distribution harder to sample from. As such since its introduction (2.46) has been a common example used to demonstrate the performance of MCMC algorithms [Betancourt et al., 2017, Bornkamp, 2011, Mbalawata et al., 2015, Roberts and Rosenthal, 2009].

The final distribution we consider is a mixture of two multivariate Gaussians, with respective means,

$$\boldsymbol{\mu}_1 = (10, 5)^T, \tag{2.47}$$

$$\boldsymbol{\mu}_2 = (-10, -5)^T, \tag{2.48}$$

and where each has covariance matrix equal to the identity matrix. This has density,

$$p_3(\boldsymbol{\theta}) = \frac{1}{4\pi}\exp\left(-\frac{1}{2}\left((\theta_1 - 10)^2 + (\theta_2 - 5)^2\right)\right) + \frac{1}{4\pi}\exp\left(-\frac{1}{2}\left((\theta_1 + 10)^2 + (\theta_2 + 5)^2\right)\right). \tag{2.49}$$

Plot C in figure 2.1 is a contour plot of this density. As we can see there are two distinct modes at $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. Thus we use p_3 to demonstrate how well various MCMC algorithms are able to handle multi-modality.

We now introduce the MCMC algorithms used in this thesis and which we will demonstrate through application to the toy examples here. The first MCMC algorithm we introduce is Metropolis-Hastings, a simple and

widely used MCMC algorithm.

2.7 Metropolis-Hastings

We here describe the Metropolis-Hastings algorithm [Hastings, 1970] and provide toy examples of its implementation.

To sample from $p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X})$ using the Metropolis-Hastings (MH) algorithm, a *proposal density*, q, needs to be defined. Given an iteration of the sampler, $\boldsymbol{\theta}^t$, then $q(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)$ is a probability density function which defines a probability distribution from which we are able to sample directly to obtain a proposal for the next iteration in the Markov chain, $\boldsymbol{\theta}'$. A decision is then taken as to whether to accept this proposal. If the proposal is accepted we set $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}'$. If the proposal is rejected we set $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t$. We derive as follows the probability that we accept the proposed move, $A(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)$, from the requirement that the detailed balance condition (2.42) is satisfied.

For the MH algorithm we can define the transition density as,

$$W(\boldsymbol{\theta}'|\boldsymbol{\theta}^t) = q(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)A(\boldsymbol{\theta}'|\boldsymbol{\theta}^t), \tag{2.50}$$

the probability of transitioning from $\boldsymbol{\theta}^t$ to $\boldsymbol{\theta}'$, is the probability of proposing $\boldsymbol{\theta}'$ given $\boldsymbol{\theta}^t$, $q(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)$, multiplied by the probability of accepting the move from $\boldsymbol{\theta}^t$ to $\boldsymbol{\theta}'$, $A(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)$. Then the detailed balance condition (2.42) for the MH algorithm is,

$$p(\boldsymbol{\theta}^t|\boldsymbol{y}, \boldsymbol{X})q(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)A(\boldsymbol{\theta}'|\boldsymbol{\theta}^t) = p(\boldsymbol{\theta}'|\boldsymbol{y}, \boldsymbol{X})q(\boldsymbol{\theta}^t|\boldsymbol{\theta}')A(\boldsymbol{\theta}^t|\boldsymbol{\theta}'). \tag{2.51}$$

We can rearrange this as,

$$\frac{A(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)}{A(\boldsymbol{\theta}^t|\boldsymbol{\theta}')} = \frac{p(\boldsymbol{\theta}'|\boldsymbol{y}, \boldsymbol{X})q(\boldsymbol{\theta}^t|\boldsymbol{\theta}')}{p(\boldsymbol{\theta}^t|\boldsymbol{y}, \boldsymbol{X})q(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)} = \frac{p(\boldsymbol{y}|\boldsymbol{\theta}', \boldsymbol{X})p(\boldsymbol{\theta}')q(\boldsymbol{\theta}^t|\boldsymbol{\theta}')}{p(\boldsymbol{y}|\boldsymbol{\theta}^t, \boldsymbol{X})p(\boldsymbol{\theta}^t)q(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)}. \tag{2.52}$$

Then detailed balance holds if we set [Hastings, 1970],

$$A(\boldsymbol{\theta}'|\boldsymbol{\theta}^t) = \min(1, \alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)), \tag{2.53}$$

where

$$\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}^t) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta}', \boldsymbol{X})p(\boldsymbol{\theta}')q(\boldsymbol{\theta}^t|\boldsymbol{\theta}')}{p(\boldsymbol{y}|\boldsymbol{\theta}^t, \boldsymbol{X})p(\boldsymbol{\theta}^t)q(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)}. \tag{2.54}$$

Then the Metropolis-Hastings algorithm is,

1. Given current state $\boldsymbol{\theta}^t$ propose new state $\boldsymbol{\theta}' \sim q(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)$.
2. Calculate $\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}^t) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta}', \boldsymbol{X})p(\boldsymbol{\theta}')q(\boldsymbol{\theta}^t|\boldsymbol{\theta}')}{p(\boldsymbol{y}|\boldsymbol{\theta}^t, \boldsymbol{X})p(\boldsymbol{\theta}^t)q(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)}$.
3. With probability $A(\boldsymbol{\theta}'|\boldsymbol{\theta}^t) = \min(1, \alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}^t))$, set $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}'$, else set $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t$.

In order to run the MH algorithm, we need to choose the proposal density $q(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)$. This often depends on some tuning parameters. For example, say a multivariate normal distribution with mean equal to $\boldsymbol{\theta}^t$ and covariance matrix, $\boldsymbol{\Sigma}$, was used, so

$$\boldsymbol{\theta}' \sim \mathcal{N}(\boldsymbol{\theta}^t, \boldsymbol{\Sigma}). \tag{2.55}$$

Such a choice leads to what is called a *Gaussian random walk* MH algorithm. Some choice for the covariance matrix, $\boldsymbol{\Sigma}$, needs to be made. A MH Gaussian random walk is sensitive to choice of $\boldsymbol{\Sigma}$. If the proposal variance is too large then few proposals will be accepted and the sampler will remain in a state for long periods of time and so it will take longer to explore the parameter space, while if the proposal variance is too small then proposals will be accepted more often but will not move very far and so its possible that only a small region of the parameter space about the initial position will be explored even if the sampler is run for a significant amount of time [Gelman et al., 2014].

Often the choice for the elements of $\boldsymbol{\Sigma}$ is made based on the percentage of proposals which are accepted. Roberts et al. [1997] suggest for a 1-parameter model that 45% of proposals should be accepted, this criterion decreasing to 23% as the number of parameters increases. The σ_i^2 are then chosen such that this target percentage of proposals are accepted.

Furthermore if we use a Gaussian proposal density we have that

$$q(\boldsymbol{\theta}'|\boldsymbol{\theta}^t) = \frac{1}{\sqrt{(2\pi)^p|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta}' - \boldsymbol{\theta}^t)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}' - \boldsymbol{\theta}^t)\right) = q(\boldsymbol{\theta}^t|\boldsymbol{\theta}'), \tag{2.56}$$

So the proposal density, $q(\cdot|\cdot)$ is *symmetric* in its arguments. Proposal densities that are symmetric in their arguments are called symmetric proposal densities.

Symmetric proposal densities allow us to simplify the acceptance probability as follows,

$$\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}^t) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta}', \boldsymbol{X})p(\boldsymbol{\theta}')}{p(\boldsymbol{y}|\boldsymbol{\theta}^t, \boldsymbol{X})p(\boldsymbol{\theta}^t)}. \tag{2.57}$$

This acceptance probability holds for all symmetric proposal densities, not just the Gaussian proposal density with diagonal covariance matrix.

Having introduced the MH algorithm we now apply it to the three toy examples introduced in §2.6.1, demonstrating where it performs well and what distributions it is unable to efficiently sample from.

2.7.1 Toy Examples

We apply Metropolis-Hastings to sample from the three toy examples detailed in §2.6.1. To fit the toy examples we need to choose a proposal density $q(\boldsymbol{\theta}'|\boldsymbol{\theta}^t)$. We here use the Gaussian random walk proposal density

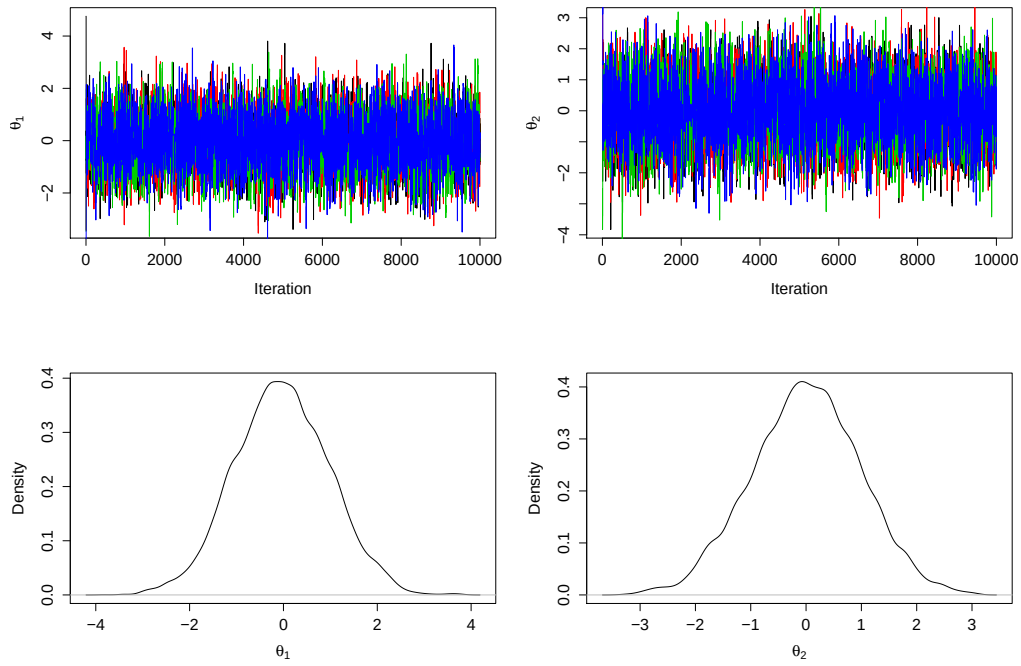


Figure 2.2: Traceplots and marginal density plots from an application of Metropolis-Hastings to the first toy example (2.45). The top row plots traceplots for both parameters and for four runs of 10000 iterations of the sampler from different starting locations (in different colours). We can see clearly that the samplers mix well and have converged to the same location. The second row plots the marginal densities for the two parameters from our sample. We can see that both are approximately $\mathcal{N}(0, 1)$ and so we have recovered the correct marginal distributions.

(2.55) with diagonal covariance matrix,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}. \tag{2.58}$$

Then to sample from each model we need to choose σ_1^2 and σ_2^2. We follow the suggestion of Roberts et al. [1997] and choose σ_1^2 and σ_2^2 to be such that approximately 33% of proposals are accepted.

To sample from p_1 (2.45) we set $\sigma^2 = \sigma_1^2 = \sigma_2^2$ as we know the density is symmetrical. We find that $\sigma^2 = 1.8$ leads to approximately 33% of proposals being accepted. The top row of figure 2.2 plots the traces for 4 runs of this sampler for 10000 iterations from 4 different starting locations.

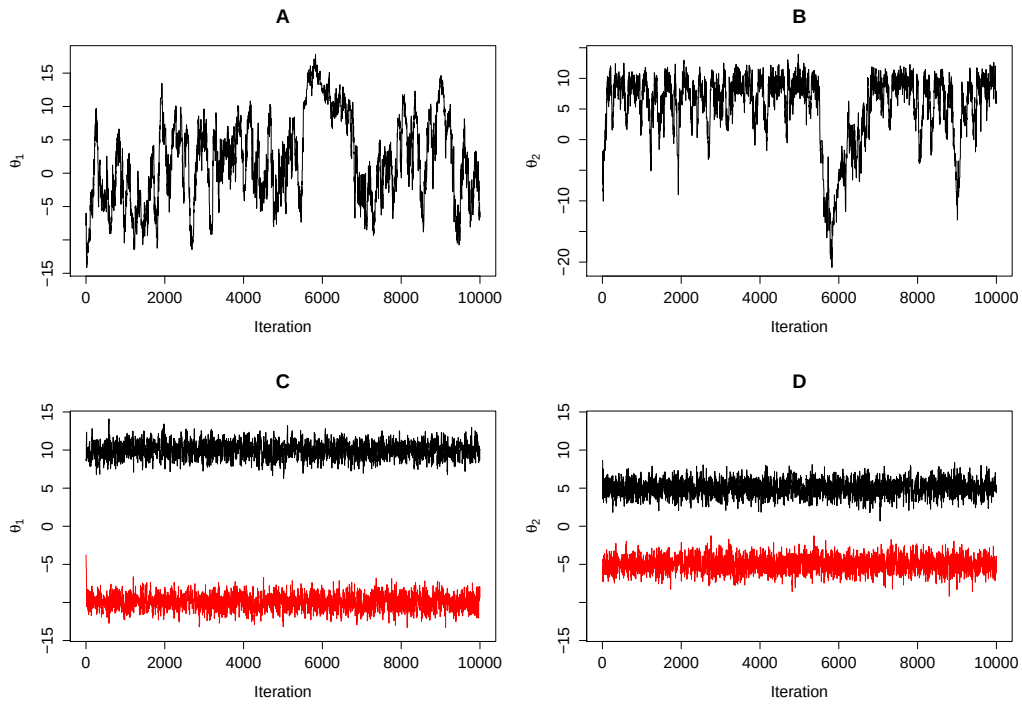


Figure 2.3: Traceplots from an application of Metropolis-Hastings to the second (2.46) and third toy examples (2.49). A and B plot the traces of the parameters for a 10000 iteration run of Metropolis-Hastings applied to p_2. We can see that in comparison to the application of the sampler to p_1 here we are not mixing as well due to the correlation between θ_1 and θ_2. C and D plot traces for two runs of 10000 iterations of Metropolis-Hastings applied to p_3. We see that each run has found its way into one of the two modes of the distribution and then become trapped there.

We can see that these 4 traces have all converged to the same location and all mix well within the region of high density. Thus, as we would expect for such a simple distribution, the sampler is performing well. The second row in 2.3 plots the marginal densities derived from our samples. We can see that these are approximately $\mathcal{N}(0, 1)$ which are the true marginal densities. Thus we can see that our sample accurately reflects the true distribution of the parameters.

To sample from p_2 (2.46) we try various values for σ_1^2 and σ_2^2. Plots A and B in figure 2.3 plot traces for θ_1 and θ_2 for a 10000 iteration long run of the Metropolis-Hastings sampler with $\sigma_1^2 = 1.4$, $\sigma_2^2 = 1.8$, which sees approximately 33% of proposals accepted. We can see that in comparison to our sample from p_1 we do not have very good mixing. Iterations of the

sampler are highly correlated and so we explore the parameter space slower. This is due to the correlation between θ_1 and θ_2. We try various other choices for σ_1^2 and σ_2^2 which give similar acceptance percentages. Traces from such runs are similar to those in figure 2.3. Random walk Metropolis-Hastings is inefficient when applied to distributions of highly correlated variables.

To sample from p_3 (2.49) we again set $\sigma_1^2 = \sigma_2^2 = \sigma^2$ as at each mode the distribution is symmetric. We find that $\sigma^2 = 1.8$ leads to approximately 33% of proposals being accepted. Plots C and D in figure 2.3 are the traces for two runs of the Metropolis-Hastings sampler with different initial positions. We can see that each run has found one of the two modes of p_3, and has then remained trapped there. Were we to run the sampler for long enough the guarantee of convergence of MCMC methods as $t \to \infty$ would see us eventually propose a move between modes. However after 10000 iterations here no such move has been proposed. Were we to plot the marginal densities of either of these runs we would incorrectly observe a uni-modal density. Metropolis-Hastings is not efficiently able to deal with multi-modal target distributions.

2.8 Hamiltonian Monte Carlo

We saw from the toy examples in the previous section that MH struggled to sample from p_2 in comparison with p_1. Iterations of the MH sampler were more strongly correlated and so the sampler explored the parameter space more slowly. Thus the MH sampler was less efficient when applied to p_2 and we would need to run the sampler for longer to obtain an accurate sample from our target distribution of interest. This is because the variables over which p_2 is defined, θ_1 and θ_2, are strongly correlated.

Hamiltonian (also known as hybrid) Monte Carlo (HMC) [Duane et al.,

1987, Neal, 1994] is an MCMC method which makes use of Hamiltonian dynamics to better explore the parameter space. HMC avoids some of the unsatisfactory behaviour of a MH random walk [Neal, 2011], HMC allows proposals which are a large distance from the current state, to be accepted with much higher probability than with a MH random walk. It does so by using *Hamiltonian dynamics* to make proposals instead of some proposal density q. This allows HMC to better explore the parameter space than MH, particularly if the probability distribution has correlation between parameters, resulting in a likelihood surface with complex shape, which standard MCMC methods such as MH are unable to accurately sample from. However HMC, like MH, can have issues with sampling from a strongly multi-modal distribution. Here we describe the HMC algorithm for sampling from the posterior distribution. We then illustrate the strengths of HMC over MH through a toy example.

For each parameter we wish to sample we introduce a corresponding momentum parameter. So for each θ_i in $\boldsymbol{\theta}$ we introduce the corresponding momentum parameter ϕ_i, so that we have the p-dimensional vector of momentum parameters,

$$\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p). \tag{2.59}$$

In the Bayesian framework we treat the elements of $\boldsymbol{\phi}$ as a random variable, just as with the model parameters. We assume that $\boldsymbol{\phi}$ is independent of $\boldsymbol{y}, \boldsymbol{X}$ and $\boldsymbol{\theta}$. Then $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ have joint posterior distribution

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{X}) = p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X})p(\boldsymbol{\phi}), \tag{2.60}$$

where $p(\boldsymbol{\phi})$ is the probability distribution of the momentum. Sampling from this joint distribution is equivalent to sampling from our posterior distribution of interest (2.12) for $\boldsymbol{\theta}$ as the marginal distribution of $p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{X})$

is

$$\int p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{X})\,\mathrm{d}\boldsymbol{\phi} = p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}) \int p(\boldsymbol{\phi})\,\mathrm{d}\boldsymbol{\phi} = p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}). \tag{2.61}$$

We define the Hamiltonian on the system defined by (2.60) as [Neal, 2011],

$$\begin{aligned} H(\boldsymbol{\theta}, \boldsymbol{\phi}) &= -\log\left(p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X})p(\boldsymbol{\theta})p(\boldsymbol{\phi})\right), && (2.62)\\ &= -\log\left(p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X})p(\boldsymbol{\theta})\right) - \log p(\boldsymbol{\phi}), && (2.63)\\ &= U(\boldsymbol{\theta}) + K(\boldsymbol{\phi}), && (2.64) \end{aligned}$$

where $U(\boldsymbol{\theta})$ is the potential energy and $K(\boldsymbol{\phi})$ is the kinetic energy. Note we once again only need to know the posterior distribution up to proportionality. The Hamiltonian of a system describes the energy of a position-momentum system. Here we have position, $\boldsymbol{\theta}$, and momentum, $\boldsymbol{\phi}$. Then Hamilton's equations [Neal, 2011] are,

$$\begin{aligned} \frac{d\theta_i}{dt} &= \frac{\partial H}{\partial \phi_i} = \frac{\partial K}{\partial \phi_i}, && (2.65)\\ \frac{d\phi_i}{dt} &= -\frac{\partial H}{\partial \theta_i} = -\frac{\partial U}{\partial \theta_i}. && (2.66) \end{aligned}$$

So Hamilton's equations form a system of $2p$ differential equations. Hamilton's equations, (2.65)-(2.66), describe Hamiltonian dynamics. Hamiltonian dynamics describe the evolution of the system consisting of position, $\boldsymbol{\theta}$, and momentum, $\boldsymbol{\phi}$, through time.

HMC uses Hamiltonian dynamics to propose the value for the next iteration in the Markov chain. In this way these equations correspond to the proposal density q discussed in §2.7 for the MH algorithm. Suppose we are at iteration $\boldsymbol{\theta}^t$ of our sampler. We propose value, $\boldsymbol{\phi}'$, for the next iteration of the sampler, $\boldsymbol{\phi}^{t+1}$, by evolving the system described by Hamilton's equations, (2.65)-(2.66), over some period of time s, where s is a tuning

parameter of the HMC algorithm.

If we could evolve the system described by (2.65)-(2.66) exactly then the value of H would be unchanged, due to the conservation property of the Hamiltonian [Neal, 2011]. Thus if we have small potential, U, corresponding to being in a state with low probability with respect to our distribution of interest, we have high kinetic energy. The evolution of Hamilton's equations through time will then see the proposal trajectory explore the parameter space, the high momentum will ensure a significant change in position. Should a region with greater potential be reached, i.e. potential energy has increased, the conservation property means kinetic energy must have decreased accordingly. Thus there is lower momentum and so we are unlikely to leave the region of high probability, instead being able to explore it thoroughly.

We can not typically solve Hamilton's equations exactly; instead, a numerical integrator is required. We use the *leapfrog* integrator [Neal, 2011]. This requires the time interval of interest, s, to be discretised into L steps of size ϵ, $s = L\epsilon$. We call ϵ the stepsize. The leapfrog integrator then iterates through the following steps,

$$\phi_i\left(t + \frac{\epsilon}{2}\right) = \phi_i(t) - \left(\frac{\epsilon}{2}\right)\frac{\partial U}{\partial \theta_i}(\boldsymbol{\theta}(t)), \tag{2.67}$$

$$\theta_i(t+\epsilon) = \theta_i(t) + \epsilon\frac{\partial K}{\phi_i}\left(\boldsymbol{\phi}\left(t + \frac{\epsilon}{2}\right)\right), \tag{2.68}$$

$$\phi_i(t+\epsilon) = \phi_i\left(t + \frac{\epsilon}{2}\right) - \left(\frac{\epsilon}{2}\right)\frac{\partial U}{\partial \theta_i}(\boldsymbol{\theta}(t+\epsilon)). \tag{2.69}$$

Given current iteration $(\boldsymbol{\theta}^t, \boldsymbol{\phi}^t)$ the HMC algorithm for determining the $(t+1)$th iteration, $(\boldsymbol{\theta}^{t+1}, \boldsymbol{\phi}^{t+1})$, proceeds as follows:

1. Sample $\boldsymbol{\phi}^0$ from its independent distribution $p(\boldsymbol{\phi}^0)$.

2. Simulate Hamiltonian dynamics using the leapfrog integrator for L steps of size ϵ from initial state $(\boldsymbol{\theta}^t, \boldsymbol{\phi}^0)$.

3. To make the proposal symmetric we negate the momentum (set $\boldsymbol{\phi'} = -\boldsymbol{\phi'}$).

4. With probability $\min[1, \exp(-H(\boldsymbol{\theta'}, \boldsymbol{\phi'}) + H(\boldsymbol{\theta}^t, \boldsymbol{\phi}^0)]$ set $(\boldsymbol{\theta}^{t+1}, \boldsymbol{\phi}^{t+1}) = (\boldsymbol{\theta'}, \boldsymbol{\phi'})$, else set $(\boldsymbol{\theta}^{t+1}, \boldsymbol{\phi}^{t+1}) = (\boldsymbol{\theta}^t, \boldsymbol{\phi}^0)$.

The final step is a Metropolis acceptance step required to correct for the numerical errors introduced by the leapfrog integrator. If we could evaluate the Hamiltonian dynamics exactly, the probability of acceptance would be 1.

HMC requires a distribution over the momentum parameters. Typically this is taken to be a zero mean multivariate normal distribution,

$$\boldsymbol{\phi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{M}), \tag{2.70}$$

where $\boldsymbol{M}$ is the covariance matrix of this distribution called the *mass matrix* or *metric* [Betancourt and Stein, 2011]. Typically $\boldsymbol{M}$ is presumed to be diagonal [Neal, 2011].

HMC has three tuning parameters for which we need to determine values, ϵ, L and $\boldsymbol{M}$. Neal [2011] recommends a trial and error approach to determining these parameters, running the HMC sampler several times for different values of the tuning parameters. However this is an inefficient way to select the tuning parameters. Furthermore it is possible that in different regions of the parameter space different values of the tuning parameters may be more suitable, rather than applying one universal value for all tuning parameters. One possible solution to this is to adapt these tuning parameters.

Hoffman and Gelman [2014] present a method which uses dual averaging [Nesterov, 2009] to adapt ϵ, and they present a modification of HMC, the no-U-turn sampler. The no-U-turn sampler sets L adaptively. Being able to adaptively set the tuning parameters rather than have to perform expensive

trial and error runs of the HMC sampler in the hope of finding suitable tuning parameter values reduces the complexity and difficulty of the task of fitting our model parameters using the HMC sampler.

In this work we fix $\boldsymbol{M}$ to be the identity matrix, $\boldsymbol{I}$. Compared to estimating $\boldsymbol{M}$, we would expect fixing $\boldsymbol{M}$ to be a poorer choice of metric. In the case of a poorly defined metric, i.e. a metric that does not reflect the correlations between the elements of the distribution we seek to sample from, the no-U-turn sampler takes a larger number of steps at smaller stepsize to compensate [Stan Development Team, 2016b]. We find this to be the case and our results to be unaffected and so throughout the thesis we use the unit metric. We now present a description of the no-U-turn sampler.

2.8.1 No-U-Turn Sampler

The no-U-turn sampler [Hoffman and Gelman, 2014] is an extension of HMC which eliminates the need to hand tune the number of steps to take, L. The key idea behind the no-U-turn sampler is that we should no longer take any more steps, i.e. we should no longer evaluate the leapfrog integrator, (2.67)-(2.69), once the trajectory our proposed parameter values, $\boldsymbol{\theta'}$ takes away from our starting parameters, $\boldsymbol{\theta}^t$, begins to double back on itself, i.e. makes a U-turn.

A criterion for determining when a U-turn has taken place is [Hoffman and Gelman, 2014],

$$\frac{d}{dt}\frac{(\boldsymbol{\theta'}-\boldsymbol{\theta}^t)\cdot(\boldsymbol{\theta'}-\boldsymbol{\theta}^t)}{2} = (\boldsymbol{\theta'}-\boldsymbol{\theta}^t)\cdot\frac{d}{dt}(\boldsymbol{\theta'}-\boldsymbol{\theta}^t) = (\boldsymbol{\theta'}-\boldsymbol{\theta}^t)\cdot\boldsymbol{\phi'}. \quad (2.71)$$

When the expression in (2.71) becomes negative this suggest the trajectory of our proposed parameter values is reversing. We cannot however just take steps of size ϵ until the quantity in (2.71) becomes negative as this

does not guarantee time reversibility, a condition which needs to hold for detailed balance (2.42) to hold. So if we cannot guarantee time reversibility we cannot guarantee that our sampler will converge to the distribution of interest.

As such the no-U-turn sampler proposed by Hoffman and Gelman [2014] preserves time-reversibility by making leapfrog steps both forwards and backwards in time. The first step in the no-U-turn sampler is the introduction of u, which has conditional distribution,

$$u|\boldsymbol{\theta}, \boldsymbol{\phi} \sim \mathcal{U}(0, \exp(-U(\boldsymbol{\theta}) - K(\boldsymbol{\phi}))), \tag{2.72}$$

where u is known as the slice variable, and $\mathcal{U}(a, b)$ is the uniform distribution on the interval $[a, b]$. The conditional distribution of our HMC parameters conditioned on u is

$$\boldsymbol{\theta}, \boldsymbol{\phi}|u \sim \mathcal{U}(\{\boldsymbol{\theta}, \boldsymbol{\phi}| \exp(-U(\boldsymbol{\theta}) - K(\boldsymbol{\phi})) \geq u\}), \tag{2.73}$$

where $\mathcal{U}(A)$, where A is a set, is the uniform distribution defined on the set A. So here $A = \{\boldsymbol{\theta}, \boldsymbol{\phi}| \exp(-U(\boldsymbol{\theta}) - K(\boldsymbol{\phi})) \geq u\}$. The marginal distribution of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ is as before. The no-U-turn sampler then functions as follows. The slice variable u is resampled, and the leapfrog integrator, (2.67)-(2.69), is evolved either forwards or backwards in time. This is done so that a *balanced binary tree* is constructed which has as nodes proposal values for $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. This balanced binary tree is created by first evaluating the leapfrog integrator forwards or backwards 1 step, then evaluating the leapfrog integrator forwards or backwards 2 steps, and so on, doubling the number of steps over which we evaluate the leapfrog integrator each time.

We determine whether to move forwards or backwards in time randomly. Suppose we are at the step where we make 2^j steps of the leapfrog integrator. Let $(\boldsymbol{\theta}^+, \boldsymbol{\phi}^+)$ be the node of our current tree which is furthest

forward in time and $(\boldsymbol{\theta}^-, \boldsymbol{\phi}^-)$ be the node which is furthest back in time. Then we sample $v_j \sim \mathcal{U}(\{-1, 1\})$. If $v_j = 1$ we evaluate the leapfrog integrator forward in time for 2^j steps from $(\boldsymbol{\theta}^+, \boldsymbol{\phi}^+)$. If $v_j = -1$ we evaluate the leapfrog integrator backwards in time for 2^j steps from $(\boldsymbol{\theta}^-, \boldsymbol{\phi}^-)$.

We stop expanding the tree when one of two stopping conditions occurs. The first is when the trajectory for a balanced subtree of the overall larger tree turns back on itself, i.e. makes a U-turn as discussed above [Hoffman and Gelman, 2014]. Say the full binary tree has i balanced subtrees. Then this first stopping condition is

$$(\boldsymbol{\theta}_i^+ - \boldsymbol{\theta}_i^-) \cdot \boldsymbol{\phi}_i^- < 0, \text{ or } (\boldsymbol{\theta}_i^+ - \boldsymbol{\theta}_i^-) \cdot \boldsymbol{\phi}_i^+ < 0, \tag{2.74}$$

for some i, where $(\boldsymbol{\theta}_i^+, \boldsymbol{\phi}_i^+)$ is the most forward node in time for subtree i and $(\boldsymbol{\theta}_i^-, \boldsymbol{\phi}_i^-)$ is the node furthest back in time for subtree i.

The second stopping condition is if the error in our simulation of Hamiltonian dynamics, using the leapfrog integrator, gets too large. This second stopping condition can be defined as [Hoffman and Gelman, 2014]

$$-U(\boldsymbol{\theta}) - K(\boldsymbol{\phi}) - \log u < -\Delta_{max}, \tag{2.75}$$

for some large positive Δ_{max}.

$\mathcal{B}$ is set to be all nodes of the balanced binary tree (each node corresponding to a value for $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$) and $\mathcal{C} \subseteq \mathcal{B}$ to be the nodes of the binary tree we can move to and have detailed balance hold. $\mathcal{C}$ is constructed deterministically from $\mathcal{B}$ and includes,

1. The starting value for this iteration of the sampler, $(\boldsymbol{\theta}^t, \boldsymbol{\phi}_0)$.

2. Any nodes, $(\boldsymbol{\theta}, \boldsymbol{\phi})$, of $\mathcal{B}$ such that $\exp(-U(\boldsymbol{\theta}) - K(\boldsymbol{\phi})) < u$.

3. Exclude from $\mathcal{C}$ any state which could not have generated $\mathcal{B}$ (this corresponds to doubling having been stopped due to one of the two

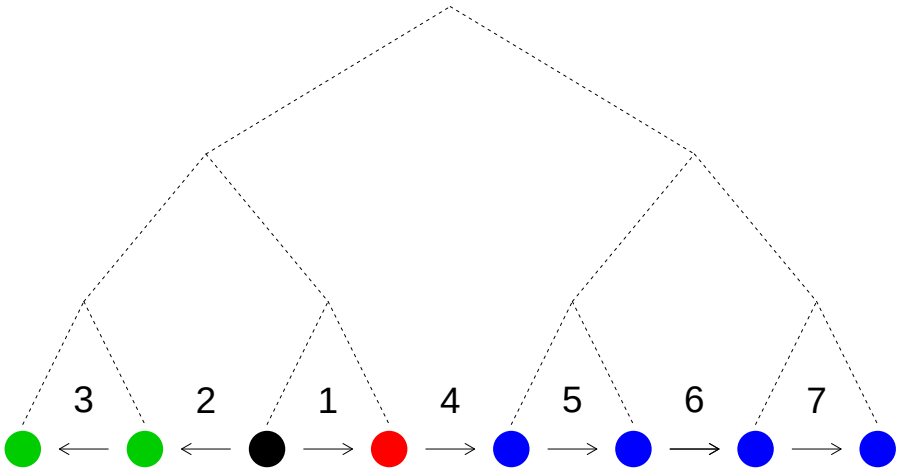


Figure 2.4: This figure details the development of the binary tree for a single iteration of the no-U-turn sampler. The black circle represents the initial location of the parameters. The binary tree is then developed by repeated doubling. A direction forwards or backwards in time is sampled uniformly at random. Here the initial direction sampled is forwards and so the leapfrog integrator is evaluated forwards in time from the black point giving the parameter values denoted by the red point. The arrow indicates from which state the leapfrog integrator is evaluated and the number above the arrow indicates that this is the first evaluation of the leapfrog integrator. The next doubling occurs backwards in time giving the green points. The arrows indicate that this is done from the point furthest back in time, here the black point, and the numbers indicate that these are the second and third evaluations of the leapfrog integrator here. The final doubling is forwards in time and gives the blue points which arise via evaluation of the leapfrog integrator for 4 steps beginning at the point foremost in time which is here the red point. The arrows and numbers again denote the direction and order in which the evaluations of the leapfrog indicator occur. The dashed lines give the structure of the binary tree, with the points where two dashed lines meet indicating a subtree. If during the final doubling, given by the blue points one of the stopping conditions, (2.74) and (2.75), are violated then the blue points would be discarded and we would sample uniformly from the black, red and green points to give the next iteration of the no-U-turn sampler.

stopping conditions being met).

This construction of $\mathcal{C}$ ensures detailed balance holds [Hoffman and Gelman, 2014]. As all elements of $\mathcal{C}$ satisfy detailed balance, we sample the next iteration of the sampler uniformly from $\mathcal{C}$.

Then the no-U-turn sampler algorithm for determining iteration $\boldsymbol{\theta}^{t+1}$ given iteration $\boldsymbol{\theta}^t$ is as follows, with initial value of $j = 0$,

1. Sample $\boldsymbol{\phi}^0$ from its independent distribution $p(\boldsymbol{\phi}^0)$.
2. Sample u from $p(u|\boldsymbol{\theta}, \boldsymbol{\phi})$ (2.72).
3. Sample $v_j \sim \mathcal{U}(\{-1, 1\})$.
4. Set $(\boldsymbol{\theta}^+, \boldsymbol{\phi}^+)$ to be the node of our current tree which is furthest forward in time and $(\boldsymbol{\theta}^-, \boldsymbol{\phi}^-)$ to be the node which is furthest back in time.
5. If $v_j = 1$ evaluate the leapfrog integrator forward in time for 2^j steps from $(\boldsymbol{\theta}^+, \boldsymbol{\phi}^+)$. If $v_j = -1$ evaluate the leapfrog integrator backwards in time for 2^j steps from $(\boldsymbol{\theta}^-, \boldsymbol{\phi}^-)$.
6. Stop doubling tree size if either of the stopping conditions, (2.74)-(2.75), holds and proceed to step 8.
7. If neither of these stopping conditions are met set $j = j + 1$ and return to step 3.
8. Construct $\mathcal{C}$ as described above.
9. Sample $(\boldsymbol{\theta}^{t+1}, \boldsymbol{\phi}^{t+1})$ uniformly from $\mathcal{C}$.

Figure 2.4 presents the development of the binary tree for a single interation of the no-U-turn sampler as detailed in this algorithm. Starting from the black point the tree is doubled forwards (the red point), backwards (the green points), and then forwards (the blue points) in time, where the

direction of doubling is sampled uniformly at random. The arrows indicate the direction of the evaluations of the leapfrog integrator from one point to another, with the numbers above the arrows being the order in which evaluations of the leapfrog integrator are performed. The dashed lines give the structure of the binary tree formed. Where two dashed lines meet any nodes of the tree beneath this meeting point form a subtree. If during the final doubling either of the stopping conditions, (2.74) and (2.75), is met, the doubling stops, the subtree consisting of the blue points is discarded and the next iteration of the sampler is uniformly sampled from the black, red and green points.

The ability of the no-U-turn sampler to adaptively select the number of steps L removes one obstacle to using HMC, determining the tuning parameter. Furthermore the no-U-turn sampler allows for the use of different L for each iteration of the sampler. Different values of L may be needed in different regions of the posterior and the no-U-turn sampler is able to do this. We now discuss how to adapt another of the tuning parameters, the stepsize ϵ.

2.8.2 Adaptation of the Stepsize

Hoffman and Gelman [2014] also propose a method by which to adapt the stepsize, ϵ, using dual averaging [Nesterov, 2009]. We here describe the use of dual averaging to adapt step-size for HMC.

It is important to get a good value for the stepsize, ϵ, as it has a similar impact on the sampler as the proposal variance, σ_g^2, in the Metropolis-Hastings Gaussian random walk algorithm we used in §2.7. If ϵ is too small the sampler will propose values for the next iteration which are close to our current state and so the sampler will mix slowly. While if ϵ is too large we will rarely accept proposals, and so our sampler will be stationary for a large number of iterations.

The dual averaging algorithm adapts ϵ so that Markov chain meets some target acceptance probability (the probability of accepting a proposed state), δ. This requires a choice for the targeted value of δ. Beskos et al. [2013] and Neal [2011] suggest that $\delta = 0.65$, is the ideal target acceptance probability while the Stan Development Team [2016b] find $\delta = 0.8$ to work well. The choice of δ is to some extent arbitrary but values in this range are agreed to be sensible by the literature [Beskos et al., 2013, Neal, 2011, Hoffman and Gelman, 2014].

Suppose that we have used stepsize ϵ^t to determine iteration $\boldsymbol{\theta}^{t+1}$ of the sampler from current iteration $\boldsymbol{\theta}^t$. Then after doing this we can adapt the stepsize to its updated value ϵ^{t+1} through the following algorithm [Hoffman and Gelman, 2014];

1. Set $\bar{H}^{t+1} = \left(1 - \frac{1}{(t+1)+t_0}\right)\bar{H}^t + \frac{1}{(t+1)+t_0}(\delta - A)$.
2. Set $\log \epsilon_{t+1} = \mu - \frac{\sqrt{t+1}}{\gamma}\bar{H}^{t+1}$.
3. Set $\log \bar{\epsilon}_{t+1} = (t+1)^{-\kappa}\log \epsilon_{t+1} + (1-(t+1)^{-\kappa})\log \bar{\epsilon}_t$.
4. Set $\epsilon_{t+1} = \bar{\epsilon}_{t+1}$.

Here ϵ_0 is our initial stepsize (we can select this as we wish), $\bar{H}^0 = 0$, $\bar{\epsilon}_0 = 1$, A is the probability of the sampler accepting the proposal to move from $\boldsymbol{\theta}^t$ to $\boldsymbol{\theta}^{t+1}$ and μ, t_0, γ, κ are parameters of the dual averaging algorithm. The parameter μ is a freely chosen point that the iterates $\log(\epsilon_{t+1})$ are shrunk towards, t_0 stabilises the initial iterations of the dual averaging algorithm, γ controls the amount of shrinkage to μ, and $\kappa \in (0.5, 1]$ ensures that the steps in the dual averaging algorithm are such that it converges so that the target acceptance probability is met [Hoffman and Gelman, 2014]. Hoffman and Gelman [2014] find that setting $\mu = \log(10\epsilon_0)$, $t_0 = 10$, $\gamma = 0.05$, and $\kappa = 0.75$, leads to good performance of the dual averaging algorithm in almost all cases and so we follow Hoff-

man and Gelman [2014] and Stan Development Team [2016b], and fix those values here.

Adaptively setting the stepsize and number of steps using dual averaging and the no-U-turn sampler removes the difficulty of attempting to determine efficient values for the HMC tuning parameters, allowing us to more easily implement HMC. Also aiding implementation of HMC is the use of Stan, a probabilistic programming language for the no-U-turn sampler.

2.8.3 Stan

Stan [Stan Development Team, 2016a] is a probabilistic programming language which implements HMC and the no-U-turn sampler. Stan is also able to optimise the stepsize using the dual averaging method described above.

Stan by default applies the adaptation of ϵ only to the burn-in period. If the sampler is working well we would hope that, by the end of the burn-in period, the sampler will have converged and that our adapted tuning parameters will be suitable for the stationary distribution. For adapting ϵ through the dual averaging method described above Stan sets $\delta = 0.8$ [Stan Development Team, 2016b]. We use Stan to implement HMC and the no-U-turn sampler, firstly in toy examples in the next section and then to EoS models in Chapters 4 and 5.

2.8.4 Toy Examples

We use Stan to apply the no-U-turn sampler to the second (2.46) and third (2.49) toy distributions which MH was unable to efficiently sample from. Figure 2.5 plots traces from these runs. All runs were again for 10000 iterations. Plots A and B in figure 2.5 plot the traces for θ_1 and θ_2 of an application of the no-U-turn sampler to p_2 (2.46). We see that in compar-

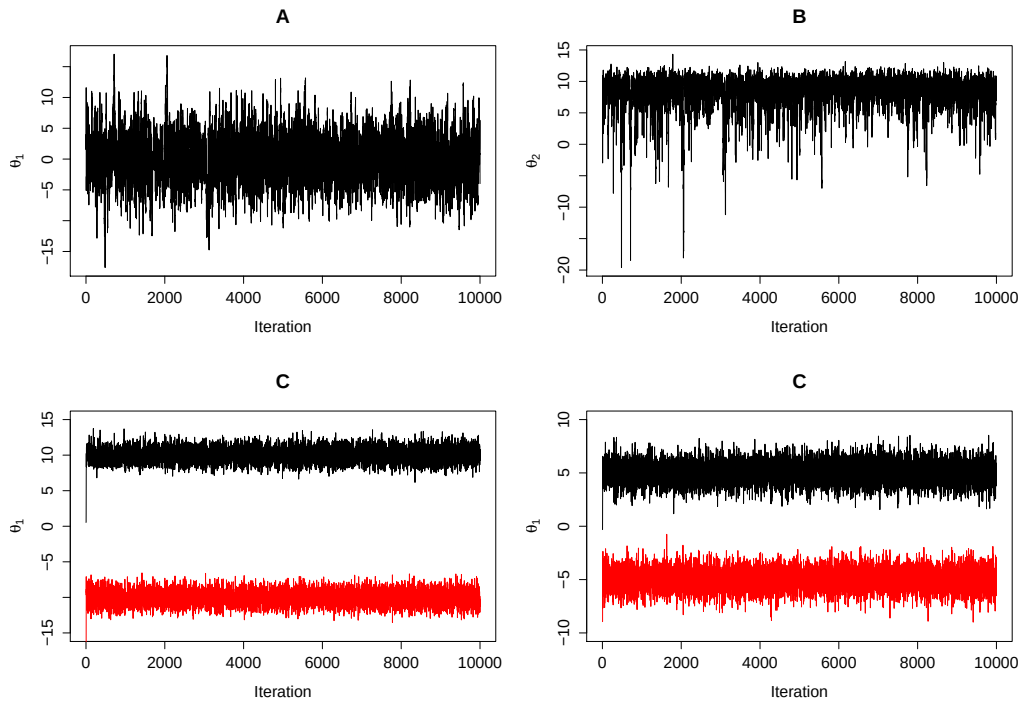


Figure 2.5: Traceplots from an application of the no-U-turn sampler to the second (2.46) and third toy examples (2.49). A and B plot the traces of the parameters for a 10000 iteration run of the no-U-turn sampler applied to p_2. We can see that in comparison to the MH sampler the parameters mix much better due to the no-U-turn samplers ability to avoid random walk behaviour and better cope with the correlation between θ_1 and θ_2. C and D plot traces for two runs of 10000 iterations of Metropolis-Hastings applied to p_3. We see that once again each run has found its way into one of the two modes of the distribution and then become trapped there.

ison to the MH case we have much better mixing of the parameters due to the no-U-turn samplers ability to use Hamiltonian dynamics to make proposals and so better deal with correlations between variables. Other runs of the sampler from different initial positions return similar results. Plots C and D in figure 2.5 plot the traces for θ_1 and θ_2 for runs of the no-U-turn sampler applied to p_3. We see that once again each run has become trapped in one of the modes of the distribution. This is because while HMC and the no-U-turn sampler are able to more efficiently sample from distributions with highly correlated variables they do not make proposals that allow us to move between isolated modes. Once again were we to plot the marginal densities of either of these runs we would inaccurately

see a uni-modal density. Again were we to run the sampler for long enough the sampler would eventually move between the modes. However to efficiently sample from the distribution we need an MCMC sampler which can more readily move propose moves between modes. This has motivated the development of several MCMC samplers which can efficiently sample from a multi-modal distribution such as p_3, including *parallel tempering* [Swendsen and Wang, 1986, Geyer, 1991].

2.9 Parallel Tempering

We have seen from the third toy example (2.49) how multi-modal distributions can lead to issues for MCMC samplers such as MH or HMC. Runs of the sampler find one of the modes and then become trapped there for significant amounts of time and so fail to efficiently explore the parameter space, leading to inaccurate samples not being from the whole distribution of interest. Gelman et al. [2014] describe several possible methods for dealing with multi-modality in an MCMC sampler, including simulated tempering [Geyer and Thompson, 1992], parallel tempering [Swendsen and Wang, 1986, Geyer, 1991] and particle filters [Gordon et al., 1993]. These methods allow the sampler to switch between modes. We take a parallel tempering approach, also sometimes called replica-exchange Monte Carlo (REMC).

We take the parallel tempering approach firstly because it can be combined easily with MCMC methods such as Metropolis-Hastings and HMC, which often perform well within individual modes but cannot move easily between them, by performing MH or HMC steps and then parallel tempering steps. We can thus combine the methods to develop a sampler which works well in all respects, using MH or HMC to sample within modes and parallel tempering to switch between them. Secondly with a simulated

tempering sampler run for a set number of iterations only a subset of iterations will be from our distribution of interest. The size of this subset is unknown when running the sampler and so to obtain a sample of fixed size we must run our sampler for an indeterminate number of iterations. Using parallel tempering has the advantage that a run for a fixed number of iterations gives a sample of that size from our distribution of interest. Finally it is relatively simple to implement parallel tempering in parallel on multiple CPUs allowing us to make more efficient use of computational resources including the University of Nottingham's high performance computing facility.

The idea behind parallel tempering is to sample independently from k probability distributions, called here replicas, which differ from each other by raising either some or all of the posterior to differing powers. So given a probability distribution p this is tempered by raising to the power β^{-1}, $p^{\frac{1}{\beta}}$, where β is called the tempering level.

When considering a posterior distribution, replicas at higher tempering levels become dominated by the prior distribution allowing greater exploration of the parameter space as the distribution is flattened by the tempering, removing the multi-modality. Swaps of parameter values between replicas are then proposed allowing higher tempering level replicas to more greatly explore the parameter space. The coldest replica, the replica with tempering level $\beta = 1$, corresponds to a sample from the untempered distribution of interest, here this will be the posterior distribution of $\boldsymbol{\theta}$ defined by equation (2.12) which we wish to sample from.

Parallel tempering as described by Earl and Deem [2005] tempers the entire posterior and Fukunishi et al. [2002] extend this to the HMC case. An alternative approach, as used by Paquet [2007], is to temper only the likelihood. In chapter 4 we extend the likelihood tempering case to HMC.

In the case that we temper the entire posterior we have that replica k

is defined as,

$$p_k(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}) = p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X})^{\frac{1}{\beta_k}}, \tag{2.76}$$

where β_k is the tempering level of replica k. Parallel tempering would then sample from each individual replica k independently for a given number of iterations using some MCMC sampler, such as MH. Swaps would then be proposed between neighbouring replicas. Let $\boldsymbol{\theta}_k$ be the current parameters at replica k. Let $\boldsymbol{\theta}_j$ be the current parameters at replica j. Then we would accept a swap of parameter values between replicas k and j with probability [Earl and Deem, 2005],

$$\alpha = \min\left(1, \exp\left(\left(\frac{1}{\beta_k} - \frac{1}{\beta_j}\right)\left(\log p(\boldsymbol{\theta}_j|\boldsymbol{y}, \boldsymbol{X}) - \log p(\boldsymbol{\theta}_k|\boldsymbol{y}\boldsymbol{X})\right)\right)\right). \tag{2.77}$$

If we temper only the likelihood then we define replica k as ,

$$p_k(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}) \propto p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\theta})^{\frac{1}{\beta_k}} p(\boldsymbol{\theta}), \tag{2.78}$$

and swaps between replicas k and j are accepted with probability [Paquet, 2007],

$$\alpha = \min\left(1, \exp\left(\left(\frac{1}{\beta_k} - \frac{1}{\beta_j}\right)\left(\log p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\theta}) - \log p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\theta})\right)\right)\right). \tag{2.79}$$

The following property of tempered normal distributions provides some motivation for tempering only the likelihood. Let y be normally distributed with mean μ and standard deviation σ. Suppose σ is fixed but we wish to estimate μ. Then y has density

$$p(y|\mu, \sigma) \propto \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right). \tag{2.80}$$

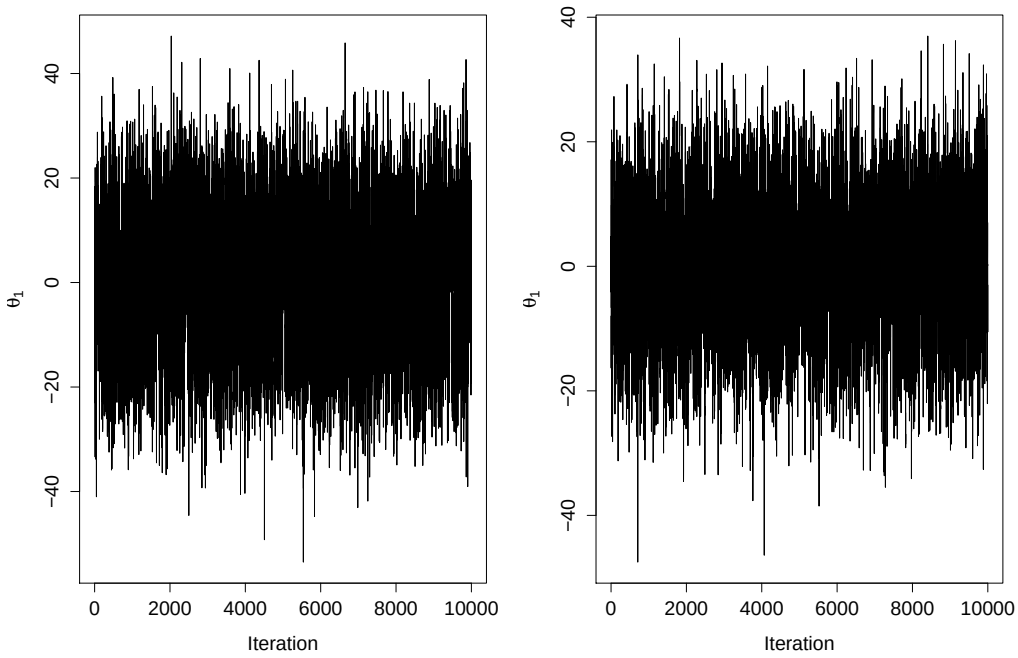


Figure 2.6: Traceplots of the replica with highest tempering level from applying MH and parallel tempering to the third toy example (2.49). We can see that the sampler at this replica explores the entire parameter space covering both modes of the distribution of interest and is itself uni-modal.

Let $p^{\beta}(y|\mu,\sigma)$ be this density to the power $\frac{1}{\beta}$. Then

$$L(\mu) = (p(y|\mu,\sigma))^{\frac{1}{\beta}} \propto \exp\left(-\frac{(y-\mu)^2}{2\beta\sigma^2}\right), \tag{2.81}$$

where $L(\mu)$ is the likelihood of μ. So a tempered normal likelihood for the mean μ with known standard deviation σ, is itself a normal likelihood, with mean μ and standard deviation $\sqrt{\beta}\sigma$. This argument can be extended to the case where μ is a function of a vector of model parameters, $\boldsymbol{\theta}$. Then if we have a normal likelihood with known standard deviation our tempered likelihoods are all proportional to normal likelihoods with a known standard deviation. This aids implementation of the PT method. We now apply the PT method to the third toy example (2.49).

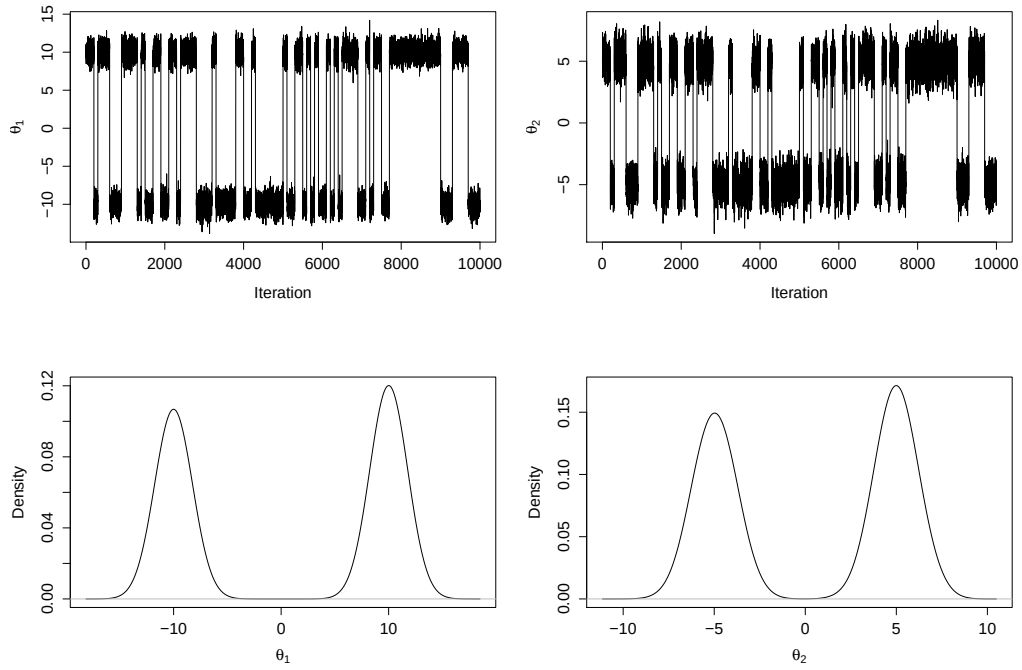


Figure 2.7: Traceplots and marginal density plots for the replica with lowest tempering level, i.e. our distribution of interest, from applying MH and parallel tempering to the third toy example (2.49). We can see from the traceplots that swaps occur between modes and from the density plots that our sample accurately reflects the multi-modality in the model.

2.9.1 Toy Example

We here apply parallel tempering and MH to sampling from p_3 (2.49). At each replica we run 100 iterations of MH followed by proposing swaps between neighbouring replicas beginning at the replica with highest tempering level. Here we temper the entire distribution. At every replica we use the same Gaussian random walk proposal density with covariance matrix,

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1.8 & 0 \\ 0 & 1.8 \end{pmatrix}, \tag{2.82}$$

as was used when applying MH to this model previously, as although there was no movement between modes within modes the sampler was mixing well. We then need to choose a set of tempering levels. We find that setting the maximum tempering level to $\beta = 100$ the corresponding replica

mixes throughout the whole parameter space covering both modes. Figure 2.6 is a traceplot for a run of MH demonstrating this good mixing over the whole parameter space. We then use a total of 5 tempering levels, $\boldsymbol{\beta} = (1, \sqrt{10}, 10, 10\sqrt{10}, 100)^T$, which we find leads to a good rate of swaps between replicas. So $\beta_1 = 1$ corresponds to our target distribution, p_3.

We run our sampler for 10000 iterations. Figure 2.7 plots traces and marginal densities for the coldest replicas which samples from our distribution of interest, p_3. We can see from the traces that swaps are occurring between modes and from the marginal densities that our sample accurately reflects the multi-modality in the model. So we have demonstrated here how parallel tempering allows us to sample from multi-modal distributions.

Having reviewed all of the statistical methods applied in this thesis in the next chapter we propose a statistical model which can be applied to pressure-explicit parametric EoS.

Chapter 3

Statistical Modelling of Parametric EoS

3.1 Introduction

In this chapter we define a statistical model for a pressure-explicit parametric EoS for pure CO_2. We first describe the modelling of homogeneous pure CO_2. We then consider the case of how to model coexistence. We show that a complete model of coexistence leads to a pair of constraints. We examine two ways of including these constraints which we call the *hard constraint* and *soft constraint* approaches. As discussed in §1.6 it is often the case that measurements of the coexistence volumes, v_l and v_g, have some associated uncertainty. We then show how the Bayesian approach to modelling measurement error allows us to account for this additional uncertainty in our model. Finally we describe how we model the case when there are multiple observations of the coexistence pressures and volumes at a single temperature.

We now describe the modelling of homogeneous data for a general pressure-explicit EoS which models pressure as a function of volume, temperature and a p-dimensional vector of EoS parameters, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$,

as described in §1.5 and equation (1.31). As such in this chapter we use $h(v, T, \boldsymbol{\theta})$ to denote our general pressure-explicit equation of state.

3.2 Modelling Homogeneous Data

Suppose we have n homogeneous data points. Then our response variable is the pressure, P, and we have an n-dimensional vector of observed pressures, $\boldsymbol{P}^{\text{obs}} = (P_1^{\text{obs}}, \ldots, P_n^{\text{obs}})^T$. For each P_i^{obs} there is a corresponding volume, v_i, and temperature T_i. Then if $\boldsymbol{v} = (v_1, \ldots, v_n)^T$, $\boldsymbol{T} = (T_1, \ldots, T_n)^T$, we have design matrix (2.3),

$$\boldsymbol{X} = (\boldsymbol{v},\ \boldsymbol{T}). \tag{3.1}$$

Our statistical model then defines the dependence of $\boldsymbol{P}^{\text{obs}}$ on the elements of $\boldsymbol{X}$ as follows,

$$P_i^{\text{obs}} = h(v_i, T_i, \boldsymbol{\theta}) + \epsilon_i, \quad i = 1, \ldots, n \tag{3.2}$$

where the ϵ_i are independent and identically distributed (i.i.d.) normal random variables with mean 0 and variance σ^2. This relationship includes an additional parameter σ which can either be fixed based on reported experimental uncertainty or treated as a parameter. Then our model for $\boldsymbol{P}^{\text{obs}}$ dependent on $\boldsymbol{\theta}$ and σ is,

$$p(\boldsymbol{P}^{\text{obs}}|\boldsymbol{\theta}, \sigma, \boldsymbol{X}) = \prod_{i=1}^{n} \phi(P_i^{\text{obs}}; h(v_i, T_i, \boldsymbol{\theta}), \sigma^2), \tag{3.3}$$

where $\phi(x; \mu, \sigma^2)$ is the normal probability density function of x with mean μ and variance σ^2. An EoS needs to satisfy coexistence as well as model the homogeneous data well. To do this we first need to define the fugacity constraint (1.14) such that it can be related to experimental measurements. We can then define a statistical model for the coexistence data. Because this amounts to working with perfectly correlated data defining a model is

not trivial. We describe our method for including coexistence behaviour in the model in the next section.

3.3 Modelling Coexistence

Coexistence of a fluid is defined by the constraints that at coexistence the pressure, temperature and fugacity must be equal. Equivalence of temperature is trivial in our model as temperature is an explanatory variable and so the model is defined dependent upon temperature.

The fugacity coefficient was defined in (1.12) and in our non-dimensionalised units, (1.37)-(1.39), is

$$\phi(v,T) = \exp\left(\frac{1}{T}\int_{v'=\infty}^{v}\left(\frac{T}{v'} - P(v',T)\right)\,\mathrm{d}v' - \ln\left(\frac{P(v,T)v}{T}\right) + \frac{P(v,T)v}{T} - 1\right). \quad (3.4)$$

For a given subcritical temperature $T < T_c$, with coexistence volumes v_l and v_g, and coexistence pressure P_s, and substituting in the fact that $P(v_l,T) = P(v_g,T) = P_s$, the fugacity constraint (1.14) is

$$\begin{aligned}
& f(v_l,T) = f(v_g,T),\\
\implies & P(v_l,T)\phi(v_l,T) = P(v_g,T)\phi(v_g,T)\\
\implies & P_s\phi(v_l,T) = P_s\phi(v_g,T),\\
\implies & \phi(v_l,T) = \phi(v_g,T),\\
\implies & \ln\phi(v_l,T) = \ln\phi(v_g,T),\\
\implies & \frac{1}{T}\int_{v'=\infty}^{v_l}\left(\frac{T}{v'} - P(v',T)\right)\,\mathrm{d}v' - \ln\left(\frac{P_s v_l}{T}\right) + \frac{P_s v_l}{RT} - 1\\
= & \frac{1}{T}\int_{v'=\infty}^{v_g}\left(\frac{T}{v'} - P(v',T)\right)\,\mathrm{d}v' - \ln\left(\frac{P_s v_g}{T}\right) + \frac{P_s v_g}{T} - 1, \qquad (3.5)
\end{aligned}$$

We may rearrange equation (3.5) to obtain the following form for the fu-

gacity constraint,

$$\begin{aligned}
&\frac{1}{T}\int_{v'=\infty}^{v_l}\left(\frac{T}{v'}-P(v',T)\right)\,\mathrm{d}v'-\frac{1}{T}\int_{v'=\infty}^{v_g}\left(\frac{T}{v'}-P(v',T)\right)\,\mathrm{d}v'\\
&=\ln\left(\frac{P_s v_l}{T}\right)-\ln\left(\frac{P_s v_g}{T}\right)+\frac{P_s v_g}{T}-\frac{P_s v_l}{T},\\
\implies &\frac{1}{T}\int_{v'=\infty}^{v_l}\left(\frac{T}{v'}-P(v',T)\right)\,\mathrm{d}v'+\frac{1}{T}\int_{v_g}^{v'=\infty}\left(\frac{T}{v'}-P(v',T)\right)\,\mathrm{d}v'\\
&=\ln\left(\frac{v_l}{v_g}\right)+\frac{P_s(v_g-v_l)}{T}.
\end{aligned} \tag{3.6}$$

If we define P to be a function of v and T (equation (1.25)) such that we know the integral to ∞ in the log-fugacity (equation (3.4)) converges, we may rewrite the fugacity constraint (equation (3.6)) as follows,

$$\frac{1}{T}\int_{v_g}^{v_l}\left(\frac{T}{v'}-P(v',T)\right)\,\mathrm{d}v'=\ln\left(\frac{v_l}{v_g}\right)+\frac{P_s(v_g-v_l)}{T}. \tag{3.7}$$

Evaluating the integral in (3.7) term by term gives,

$$\int_{v_g}^{v_l}\frac{1}{v'}\,\mathrm{d}v'+\frac{1}{T}\int_{v_l}^{v_g}P(v',T)\,\mathrm{d}v'=\ln\left(\frac{v_l}{v_g}\right)+\frac{P_s(v_g-v_l)}{T}, \tag{3.8}$$

$$\implies \ln\left(\frac{v_l}{v_g}\right)+\frac{1}{T}\int_{v_l}^{v_g}P(v',T)\,\mathrm{d}v'=\ln\left(\frac{v_l}{v_g}\right)+\frac{P_s(v_g-v_l)}{T}, \tag{3.9}$$

$$\implies \int_{v_l}^{v_g}P(v',T)\,\mathrm{d}v'=P_s(v_g-v_l). \tag{3.10}$$

It is then possible to define the coexistence pressure P_s from the fugacity constraint as,

$$P_s=\frac{\int_{v_l}^{v_g}P(v',T)\,\mathrm{d}v'}{v_g-v_l}. \tag{3.11}$$

Given an observation of P_s we can then use (3.11) to incorporate the fugacity constraint in our model.

However this is not a trivial task due to the pressure constraint at coexistence (1.9). The pressures at each coexistence point must be equal

and are,

$$P_s = P(v_l, T) = P(v_g, T). \tag{3.12}$$

This means we can define P_s in three ways, as the pressure at each coexistence point (3.12), and as the pressure derived from the fugacity constraint (3.11). This leads to choices in how to model coexistence. We now describe two possible choices for how to model this data depending on whether the constraints are imposed in a hard (exact) or soft (approximate) manner depending on how we relate our model to the true pressure.

3.3.1 The Hard Constraint Case

To describe the hard constraint case we shall first consider coexistence data for a single temperature T. Then the data is an observed coexistence pressure, P_s^{obs}, and liquid and gas coexistence volumes, v_l and v_g which here we assume are observed without error.

Let P_l^{true} be the true pressure at the liquid coexistence volume, v_l. Let P_g^{true} be the true pressure at the gas coexistence volume, v_g. Let P_f^{true} be the true pressure derived from the fugacity constraint as given by (3.11). From (3.11) and (3.12) we know that

$$P_l^{\text{true}} = P_g^{\text{true}} = P_f^{\text{true}} = P_s^{\text{true}}. \tag{3.13}$$

To relate P_s^{obs} and P_s^{true} we assume that

$$P_s^{\text{obs}} = P_s^{\text{true}} + \epsilon_s, \tag{3.14}$$

where ϵ_s is normally distributed with zero mean and variance σ_s^2,

$$\implies P_l^{\text{obs}} = P_l^{\text{true}} + \epsilon_s, \tag{3.15}$$

$$\implies P_g^{\text{obs}} = P_g^{\text{true}} + \epsilon_s, \tag{3.16}$$

$$\implies P_f^{\text{obs}} = P_f^{\text{true}} + \epsilon_s. \tag{3.17}$$

Here ϵ_s represents the observation error on P_s^{obs}; P_l^{obs} and P_g^{obs} are used to denote the observed pressure at v_l and v_g respectively; and P_f^{obs} to denote the observed pressure derived from the fugacity constraint. Note that,

$$P_s^{\text{obs}} = P_l^{\text{obs}} = P_g^{\text{obs}} = P_f^{\text{obs}}. \tag{3.18}$$

From hereon we assume that $\sigma_s = \sigma$, so that the observation error on the coexistence pressure is i.i.d. to the observation error on the homogeneous pressures.

As we have just a single observation, P_s^{obs}, this suggests including just one of P_l^{obs}, P_g^{obs} or P_f^{obs} as a data point. However this discards the information that the pressure must be matched at the coexistence volumes and that this pressure must also equal that derived from the fugacity constraint. If we were to include just P_l^{obs} this would place no conditions on what the value of the pressure at v_g must be equal to and vice versa as well as disregarding the fugacity. Thus we must look at ways of including the constraint that the coexistence pressure can be defined in each of these ways in our model.

In the hard constraint case we assume that our model perfectly describes the true pressure. That is that $P(v,T) = h(v,T,\boldsymbol{\theta})$. Let

$$H(v_1, v_2, T, \boldsymbol{\theta}) = \int_{v_1}^{v_2} h(v, T, \boldsymbol{\theta}) \, \mathrm{d}v. \tag{3.19}$$

Then we have

$$P_l^{\text{true}} = h(v_l, T, \boldsymbol{\theta}), \tag{3.20}$$

$$P_g^{\text{true}} = h(v_g, T, \boldsymbol{\theta}), \tag{3.21}$$

$$P_f^{\text{true}} = \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}, \tag{3.22}$$

and so the observed values are defined as,

$$P_l^{\text{obs}} = h(v_l, T, \boldsymbol{\theta}) + \epsilon_s, \tag{3.23}$$

$$P_g^{\text{obs}} = h(v_g, T, \boldsymbol{\theta}) + \epsilon_s, \tag{3.24}$$

$$P_f^{\text{obs}} = \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l} + \epsilon_s. \tag{3.25}$$

Then the three coexistence observations are jointly normally distributed with coexistence matrix,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2 \end{pmatrix}, \tag{3.26}$$

which is clearly singular. Jointly normally distributed random variables with singular covariance matrix are said to be degenerate. To define the probability density function of the degenerate multivariate normal distribution we need the *pseudo-determinant*, $|\boldsymbol{\Sigma}|^*$, of $\boldsymbol{\Sigma}$, which is the product of the non-zero eigenvalues of $\boldsymbol{\Sigma}$, and the *generalised inverse* (or *g-inverse*), $\boldsymbol{\Sigma}^-$, of $\boldsymbol{\Sigma}$, such that $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^-\boldsymbol{\Sigma} = \boldsymbol{\Sigma}$. Here $\boldsymbol{\Sigma}$ has pseudo-determinant

$$|\boldsymbol{\Sigma}|^* = 3\sigma^2, \tag{3.27}$$

and generalised inverse,

$$\mathbf{\Sigma}^{-} = \begin{pmatrix} \Sigma_{11}^{-} & \Sigma_{12}^{-} & \Sigma_{13}^{-} \\ \Sigma_{21}^{-} & \Sigma_{22}^{-} & \Sigma_{23}^{-} \\ \Sigma_{31}^{-} & \Sigma_{32}^{-} & \Sigma_{33}^{-} \end{pmatrix}, \tag{3.28}$$

where

$$\sum_{i=1}^{3}\sum_{j=1}^{3} \Sigma_{ij}^{-} = \frac{1}{\sigma^2}. \tag{3.29}$$

Then the degenerate multivariate normal distribution of $(P_l^{\text{obs}}, P_g^{\text{obs}}, P_f^{\text{obs}})^T$ is in two parts [Rao, 1973], an expression for the density, and an expression for the subspace on which this density is defined. The subspace is defined by two constraints,

$$P_l^{\text{obs}} - h(v_l, T, \boldsymbol{\theta}) = P_g^{\text{obs}} - h(v_g, T, \boldsymbol{\theta}), \tag{3.30}$$

$$P_l^{\text{obs}} - h(v_l, T, \boldsymbol{\theta}) = P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}. \tag{3.31}$$

Then substituting in these constraints and the definition of the generalised inverse (3.28 and 3.29) and the pseudo-determinant (3.27) in this case the density is,

$$p(P_s^{\text{obs}}|\boldsymbol{\theta}, \sigma, v_l, v_g, T) = \frac{1}{\sqrt{6\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(P_s^{\text{obs}} - h(v_l, T, \boldsymbol{\theta}))^2\right\}. \tag{3.32}$$

This density is equal to the density for P_l^{obs} alone multiplied by $\sqrt{3}$. Given some coexistence data the constraints may be rearranged to give expressions for 2 parameters in terms of the other parameters thus removing them from the model and reducing the size of the parameter space by 2. As we assume that this coexistence data is independent of the homogeneous data a joint

likelihood for P_s^{obs} and $\boldsymbol{P}^{\text{obs}}$ can be derived by multiplying (3.3) and (3.32),

$$p(P_s^{\text{obs}}, \boldsymbol{P}^{\text{obs}}|\boldsymbol{\theta}, \sigma, v_l, v_g, \boldsymbol{X}) = \frac{1}{\sqrt{6\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(P_s^{\text{obs}} - h(v_l, T, \boldsymbol{\theta}))^2\right\} \prod_{i=1}^{n} \phi(P_i^{\text{obs}}; h(v_i, T_i, \boldsymbol{\theta}), \sigma^2). \tag{3.33}$$

We now consider the case of coexistence data for multiple temperatures. If we make the reasonable assumption that coexistence data at each temperature is independent then for every additional temperature we introduce to the model we would gain a further two constraints of the type given by (3.30) and (3.31). Thus each additional temperature for which we have coexistence data reduces the size of the parameter space by 2. Thus issues with the hard constraint model arise either if the number of constraints exceeds the number of parameters or if it is not possible to rearrange the constraints so as to define the required number of parameters in terms of the remaining parameters and thus remove them from the model. As such we now propose an alternative model which instead uses a "soft" version of the constraint, in which it is only approximately imposed.

3.3.2 The Soft Constraint Case

An alternative approach is to assume that our model is an imperfect representation of the truth, i.e. that the model does not exactly describe the true behaviour of pressure. This involves assuming there is some model error. The difficulty here lies in defining the model error. Perhaps the simplest representation of model error for coexistence data is to assume the model error is independent and identically normally distributed for each volume with mean 0 and variance σ_e^2. This is equivalent to assuming that the model error is a white noise process with finite variance. There is then no model error on the fugacity observation as this is an observation of an

integral and the integral of a white noise process with finite variance over any interval has variance 0. As such the variance of the model error on P_f^{true} is 0, and the covariance between the model error on P_f^{true} and both P_l^{true} and P_g^{true} is similarly 0. We thus have,

$$P_l^{\text{obs}} = P_l^{\text{true}} + \epsilon_s, \tag{3.34}$$

$$P_g^{\text{obs}} = P_g^{\text{true}} + \epsilon_s, \tag{3.35}$$

$$P_f^{\text{obs}} = P_f^{\text{true}} + \epsilon_s, \tag{3.36}$$

$$P_l^{\text{true}} = h(v_l, T, \boldsymbol{\theta}) + \xi_l, \tag{3.37}$$

$$P_g^{\text{true}} = h(v_g, T, \boldsymbol{\theta}) + \xi_g, \tag{3.38}$$

$$P_f^{\text{true}} = \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}. \tag{3.39}$$

where ξ_l and ξ_g are independent normally distributed random variables with mean 0 and variance σ_e^2 which represent model error, and ϵ_s is as in the hard constraint case (and so has standard deviation σ).

We have little information on what reasonable values of σ_e might be. We assume that our model should describe the data well and so would expect most noise to come from observation error. We thus assume that $\sigma_e \ll \sigma$.

Then $P_l^{\text{obs}}, P_g^{\text{obs}}, P_f^{\text{obs}}$ are joint normally distributed with covariance matrix,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} \sigma^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2 \end{pmatrix} = \begin{pmatrix} \sigma_e^2 + \sigma^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma_e^2 + \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2 \end{pmatrix}. \tag{3.40}$$

Then $\boldsymbol{\Sigma}$ has determinant,

$$|\boldsymbol{\Sigma}| = \sigma_e^4 \sigma^2, \tag{3.41}$$

and inverse

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_e^2\sigma^2}\begin{pmatrix} \sigma^2 & 0 & -\sigma^2 \\ 0 & \sigma^2 & -\sigma^2 \\ -\sigma^2 & -\sigma^2 & \sigma_e^2 + 2\sigma^2 \end{pmatrix}. \tag{3.42}$$

Clearly as long as σ_e^2 and σ^2 are greater than 0 then $\boldsymbol{\Sigma}$ is non-singular. Then the density of the coexistence data is,

$$\begin{aligned} &p(P_l^{\text{obs}}, P_g^{\text{obs}}, P_f^{\text{obs}}|\boldsymbol{\theta}, v_l, v_g, T, \sigma, \sigma_e) = \\ &\frac{1}{\sqrt{(2\pi)^3\sigma_e^4\sigma^2}}\exp\left\{-\frac{1}{2\sigma_e^2\sigma^2}\left(\sigma^2(P_l^{\text{obs}} - h(v_l, T, \boldsymbol{\theta}))^2 + \sigma^2(P_g^{\text{obs}} - h(v_g, T, \boldsymbol{\theta}))^2\right.\right. \\ &\qquad - 2\sigma^2(P_l^{\text{obs}} - h(v_l, T, \boldsymbol{\theta}))\left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right) \\ &\qquad - 2\sigma^2(P_g^{\text{obs}} - h(v_g, T, \boldsymbol{\theta}))\left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right) \\ &\left.\left. + 2\sigma^2\left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right)^2 + \sigma_e^2\left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right)^2\right)\right\}. \end{aligned} \tag{3.43}$$

Then the log-likelihood $l(\boldsymbol{\theta})$ (up to an additive constant and treating σ and σ_e as fixed) is,

$$\begin{aligned} l(\boldsymbol{\theta}) = -\frac{1}{2\sigma_e^2\sigma^2}&\left(\sigma^2(P_l^{\text{obs}} - h(v_l, T, \boldsymbol{\theta}))^2 + \sigma^2(P_g^{\text{obs}} - h(v_g, T, \boldsymbol{\theta}))^2\right. \\ &- 2\sigma^2(P_l^{\text{obs}} - h(v_l, T, \boldsymbol{\theta}))\left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right) \\ &- 2\sigma^2(P_g^{\text{obs}} - h(v_g, T, \boldsymbol{\theta}))\left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right) \\ &\left. + 2\sigma^2\left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right)^2 + \sigma_e^2\left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right)^2\right). \end{aligned} \tag{3.44}$$

We can write

$$l(\boldsymbol{\theta}) = l_1(\boldsymbol{\theta}) + l_2(\boldsymbol{\theta}), \tag{3.45}$$

where,

$$l_1(\boldsymbol{\theta}) = -\frac{1}{2\sigma_e^2}\Bigg((P_l^{\text{obs}} - h(v_l, T, \boldsymbol{\theta}))^2 + (P_g^{\text{obs}} - h(v_g, T, \boldsymbol{\theta}))^2 + 2\left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right)^2 - 2(P_l^{\text{obs}} - h(v_l, T, \boldsymbol{\theta}))\left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right) - 2(P_g^{\text{obs}} - h(v_g, T, \boldsymbol{\theta}))\left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right)\Bigg), \quad (3.46)$$

and

$$l_2(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2}\left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right)^2. \quad (3.47)$$

This term, $l_2(\boldsymbol{\theta})$, is the likelihood for the fugacity observation alone. It gives high likelihood to values of the parameters such that the model co-existence pressure matches the observed pressure within the experimental uncertainty. We can rewrite (3.46) as,

$$l_1(\boldsymbol{\theta}) = -\frac{1}{2\sigma_e^2}\Bigg(\Bigg\{ (P_l^{\text{obs}} - h(v_l, T, \boldsymbol{\theta}))^2 - 2(P_l^{\text{obs}} - h(v_l, T, \boldsymbol{\theta}))\left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right) + \left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right)^2 \Bigg\} + \Bigg\{ (P_g^{\text{obs}} - h(v_g, T, \boldsymbol{\theta}))^2 - 2(P_g^{\text{obs}} - h(v_g, T, \boldsymbol{\theta}))\left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right) + \left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}\right)^2 \Bigg\}\Bigg). \quad (3.48)$$

Then rearranging further and substituting in that $P_l^{\text{obs}} = P_g^{\text{obs}} = P_f^{\text{obs}}$, we

have,

$$
\begin{aligned}
l_1(\boldsymbol{\theta}) = -\frac{1}{2\sigma_e^2}\Bigg(& \left\{ (P_l^{\text{obs}} - h(v_l, T, \boldsymbol{\theta})) - \left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l} \right) \right\}^2 \\
& + \left\{ (P_g^{\text{obs}} - h(v_g, T, \boldsymbol{\theta})) - \left(P_f^{\text{obs}} - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l} \right) \right\}^2 \Bigg) \\
= -\frac{1}{2\sigma_e^2}\Bigg(& \left\{ (P_l^{\text{obs}} - P_f^{\text{obs}}) - \left(h(v_l, T, \boldsymbol{\theta}) - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l} \right) \right\}^2 \\
& + \left\{ (P_g^{\text{obs}} - P_f^{\text{obs}}) - \left(h(v_g, T, \boldsymbol{\theta}) - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l} \right) \right\}^2 \Bigg) \\
= -\frac{1}{2\sigma_e^2}\Bigg(& \left\{ 0 - \left(h(v_l, T, \boldsymbol{\theta}) - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l} \right) \right\}^2 \\
& + \left\{ 0 - \left(h(v_g, T, \boldsymbol{\theta}) - \frac{H(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l} \right) \right\}^2 \Bigg). \quad (3.49)
\end{aligned}
$$

Then the final expression for $l_1(\boldsymbol{\theta})$ in (3.49) corresponds to the two constraints that need to be satisfied. These constraints give high likelihood to values of the parameters such that the pressure at the two coexistence volumes predicted by the model are consistent with the coexistence pressure derived from the fugacity constraint as predicted from the model to within model error. As $\sigma_e^2 \to 0$ the constraints must hold more strongly eventually resulting in the hard constraint case. As discussed previously we assume there is no model error on the integral and thus there is less error on the fugacity observation than on the observations at each coexistence pressure. Therefore the observation we seek to fit is the fugacity constraint and we define the constraints in terms of the fugacity. We could also interpret (3.45) to be a penalised likelihood where we include in the standard likelihood only one observation (the fugacity constraint) (3.47) and then add an additional penalty term for the constraints (3.49).

Suppose we have data at coexistence for n_c temperatures. Then our coexistence data consists of a vector of observed coexistence pressures $\boldsymbol{P}_s^{\text{obs}} = (P_{s1}^{\text{obs}}, \ldots, P_{sn_c}^{\text{obs}})^T$, vectors of coexistence volumes, $\boldsymbol{v}_l = (v_{l1}, \ldots, v_{ln_c})^T$ and

$\boldsymbol{v}_g = (v_{g1}, \ldots, v_{gn_c})^T$ and the vector of temperatures at which we have coexistence data, $\boldsymbol{T}_c = (T_{c1}, \ldots, T_{cn_c})^T$. Then making the reasonable assumption that coexistence data at each temperature is independent and treating σ and σ_e as fixed, the density of the coexistence data can be defined proportionally as,

$$
\begin{aligned}
p(\boldsymbol{P}_s^{\text{obs}} | \boldsymbol{\theta}, \sigma, \sigma_e, \boldsymbol{v}_l, \boldsymbol{v}_g, \boldsymbol{T}_c) \propto & \\
& \prod_{i=1}^{n_c} \phi\left(P_{si}^{\text{obs}}; \frac{H(v_{li}, v_{gi}, T_{ci}, \boldsymbol{\theta})}{v_{gi} - v_{li}}, \sigma^2\right) \\
& \phi\left(0; h(v_{li}, T_{ci}, \boldsymbol{\theta}) - \frac{H(v_{li}, v_{gi}, T_{ci}, \boldsymbol{\theta})}{v_{gi} - v_{li}}, \sigma_e^2\right) \\
& \phi\left(0; h(v_{gi}, T_{ci}, \boldsymbol{\theta}) - \frac{H(v_{li}, v_{gi}, T_{ci}, \boldsymbol{\theta})}{v_{gi} - v_{li}}, \sigma_e^2\right). \quad (3.50)
\end{aligned}
$$

Then as we model homogeneous and coexistence data as independent the joint density for homogeneous and coexistence data defined up to proportionality is,

$$
\begin{aligned}
p(\boldsymbol{P}_s^{\text{obs}}, \boldsymbol{P}^{\text{obs}} | \boldsymbol{\theta}, \sigma, \sigma_e, \boldsymbol{v}_l, \boldsymbol{v}_g, \boldsymbol{T}_c, \boldsymbol{X}) \propto & \\
& \prod_{i=1}^{n_c} \phi\left(P_{si}^{\text{obs}}; \frac{H(v_{li}, v_{gi}, T_{ci}, \boldsymbol{\theta})}{v_{gi} - v_{li}}, \sigma^2\right) \\
& \phi\left(0; h(v_{li}, T_{ci}, \boldsymbol{\theta}) - \frac{H(v_{li}, v_{gi}, T_{ci}, \boldsymbol{\theta})}{v_{gi} - v_{li}}, \sigma_e^2\right) \\
& \phi\left(0; h(v_{gi}, T_{ci}, \boldsymbol{\theta}) - \frac{H(v_{li}, v_{gi}, T_{ci}, \boldsymbol{\theta})}{v_{gi} - v_{li}}, \sigma_e^2\right) \\
& \prod_{j=1}^{n} \phi(P_j^{\text{obs}}; h(v_j, T_j, \boldsymbol{\theta}), \sigma^2). \quad (3.51)
\end{aligned}
$$

We have until now only considered the case that we know the true values for the coexistence volumes. In the next section we will describe how to model the cases where the coexistence volumes are measured with some error.

3.4 Modelling Uncertainty in the Coexistence Volumes

Suppose we have coexistence data for n_c temperatures as described in the previous section, but now instead of knowing the true coexistence volumes, $\boldsymbol{v}_l$ and $\boldsymbol{v}_g$, we instead have two vectors of observed coexistence volumes, $\boldsymbol{v}_l^{\text{obs}} = (v_{l1}^{\text{obs}}, \ldots, v_{ln_c}^{\text{obs}})^T$ and $\boldsymbol{v}_g^{\text{obs}} = (v_{g1}^{\text{obs}}, \ldots, v_{gn_c}^{\text{obs}})^T$ due to measurement error.

In §2.4.1 we described the Dellaportas and Stephens [1995] approach to Bayesian errors-in-variables models. This approach involves treating the true coexistence volumes, $\boldsymbol{v}_l$ and $\boldsymbol{v}_g$ as parameters. Thus the density for the homogeneous and coexistence pressure data (3.51) is unchanged, but now $\boldsymbol{v}_l$ and $\boldsymbol{v}_g$ are unknown and we will seek to sample them along with the model parameters $\boldsymbol{\theta}$.

To complete the errors-in-variables model we need to define the relationship between $\boldsymbol{v}_l^{\text{obs}}$ and $\boldsymbol{v}_l$, and between $\boldsymbol{v}_g^{\text{obs}}$ and $\boldsymbol{v}_g$. A simple, physically reasonable model for the measurement error is to assume that the measurement error on each observation of v_l and v_g is i.i.d. normally distributed. We then define $\boldsymbol{v}_l^{\text{obs}}$ and $\boldsymbol{v}_g^{\text{obs}}$ as follows,

$$v_{li}^{\text{obs}} = v_{li} + \epsilon_{v_l i}, \quad \forall i \in 1, \ldots, n_c, \tag{3.52}$$

$$v_{gi}^{\text{obs}} = v_{gi} + \epsilon_{v_g i}, \quad \forall i \in 1, \ldots, n_c, \tag{3.53}$$

where the $\epsilon_{v_l i}$ and $\epsilon_{v_g i}$ are normally distributed with mean 0 and variance σ_v^2. We then have,

$$p(\boldsymbol{v}_l^{\text{obs}}, \boldsymbol{v}_g^{\text{obs}} | \boldsymbol{v}_l, \boldsymbol{v}_g) = \prod_{i=1}^{n_c} \phi(v_{li}^{\text{obs}}; v_{li}, \sigma_v^2) \phi(v_{gi}^{\text{obs}}; v_{gi}, \sigma_v^2). \tag{3.54}$$

The full model for our data is then (3.54) multiplied by the likelihood for the coexistence and homogeneous pressures in either the hard (3.33) or soft

(3.51) cases. To fully define the Bayesian errors-in-variables model for the coexistence volumes the only further requirement is a prior on $\boldsymbol{v}_l$ and $\boldsymbol{v}_g$ which is discussed in the next chapter.

As was described in §1.6 for a given temperature with coexistence data there are sometimes multiple measurements of one or more of P_s, v_l, and v_g.

3.5 Modelling Multiple Observations of Coexistence Pressures and Volumes

For temperature T_{ci}, at which we have coexistence data, let n_{pi}, n_{li}, and n_{gi}, denote the number of observations of P_s, v_l, and v_g, at that temperature respectively. We show here that if we fix σ and σ_v we can equivalently consider the means of our observed coexistence data.

Suppose we have n_y observations of y, $y_1, \ldots, y_{n_y}$ where y_i is the i-th observation and,

$$y_i = y + \epsilon_{yi}, \tag{3.55}$$

where the ϵ_{yi} are i.i.d. normally distributed with mean 0 and s.d. σ_y. If we model y using a function f of parameters $\boldsymbol{\theta}$ then,

$$y_i = f(\boldsymbol{\theta}) + \epsilon_{yi}. \tag{3.56}$$

Then the log-likelihood of $\boldsymbol{\theta}$ and σ_y is

$$\begin{aligned} \log L(\boldsymbol{\theta}, \sigma_y | y_1, \ldots, y_{n_y}) &= \log \left(\prod_{i=1}^{n_y} (2\pi\sigma_y^2)^{-\frac{1}{2}} \exp\left(-\frac{(y_i - f(\boldsymbol{\theta}))^2)}{2\sigma_y^2} \right) \right) \\ &= -\frac{n_y}{2} \log(2\pi\sigma_y^2) - \sum_{i=1}^{n} \frac{(y_i - f(\boldsymbol{\theta}))^2}{2\sigma_y^2}. \end{aligned} \tag{3.57}$$

Now consider the mean of the observations,

$$\bar{y} = \frac{\sum_{i=1}^{n_y} y_i}{n_y}. \tag{3.58}$$

We know that $\bar{y} \sim \mathcal{N}\left(f(\boldsymbol{\theta}), \frac{\sigma_y^2}{n_y}\right)$. Then the log-likelihood of $\boldsymbol{\theta}$ and σ_y based on $\bar{y}$ is

$$\begin{aligned}
\log L(\boldsymbol{\theta}, \sigma_y|\bar{y}) =& \log\left(\left(\frac{2\pi\sigma_y^2}{n_y}\right)^{-\frac{1}{2}} \exp\left(-n_y \frac{(\bar{y} - f(\boldsymbol{\theta}))^2)}{2\sigma_y^2}\right)\right) \\
=& -\frac{1}{2}\log\left(\frac{2\pi\sigma_y^2}{n_y}\right) - \frac{\left(\frac{(\sum_{i=1}^{n_y} y_i)^2}{n_y} - 2f(\boldsymbol{\theta})\sum_{i=1}^{n_y} y_i + n_y f(\boldsymbol{\theta})^2\right)}{2\sigma_y^2} \\
=& -\frac{1}{2}\log\left(\frac{2\pi\sigma_y^2}{n_y}\right) \\
& - \frac{\left(\frac{(\sum_{i=1}^{n_y} y_i^2}{n_y} - 2f(\boldsymbol{\theta})\sum_{i=1}^{n_y} y_i + n_y f(\boldsymbol{\theta})^2 + \frac{2}{n_y}\sum_{i<j} y_i y_j\right)}{2\sigma_y^2} \\
=& -\frac{1}{2}\log\left(\frac{2\pi\sigma_y^2}{n_y}\right) \\
& - \frac{\left(\sum_{i=1}^{n_y} y_i^2 - 2f(\boldsymbol{\theta})\sum_{i=1}^{n_y} y_i + n_y f(\boldsymbol{\theta})^2\right)}{2\sigma_y^2} \\
& + \frac{\frac{n_y-1}{n_y}\sum_{i=1}^{n_y} y_i^2 + \frac{2}{n_y}\sum_{i<j} y_i y_j}{2\sigma_y^2} \\
=& -\frac{1}{2}\log\left(\frac{2\pi\sigma_y^2}{n_y}\right) - \sum_{i=1}^{n_y}\frac{(y_i - f(\boldsymbol{\theta}))^2}{2\sigma_y^2} \\
& + \frac{\frac{n_y-1}{n_y}\sum_{i=1}^{n_y} y_i^2 + \frac{2}{n_y}\sum_{i<j} y_i y_j}{2\sigma_y^2}.
\end{aligned} \tag{3.59}$$

Then if we treat σ_y as fixed $\log L(\boldsymbol{\theta}|y_1, \ldots, y_{n_y})$ and $\log L(\boldsymbol{\theta}|\bar{y})$ differ only up to an additive constant and so the posterior for $\boldsymbol{\theta}$ is unchanged. This is not the case if σ_y is free and so we can replace individual observations with the mean only when we fix σ_y.

In our case σ_y corresponds to σ for the coexistence pressure and σ_v for the coexistence volume. If we fix both of these we can use the means of the observations of the coexistence volumes and pressures rather than the

individual values. Let $\bar{P}_{si}^{\text{obs}}, \bar{v}_{li}^{\text{obs}}$, and $\bar{v}_{gi}^{\text{obs}}$, be the mean of the observations of P_s, v_l, and v_g, respectively for temperature T_{ci}. Then these means have variances $\frac{\sigma}{n_{si}}, \frac{\sigma_v}{n_{li}}$, and $\frac{\sigma_v}{n_{gi}}$, respectively. Then when we have multiple observations of coexistence volumes or pressures we simply replace $P_{si}^{\text{obs}}, v_{li}^{\text{obs}}$ and v_{gi}^{obs} with $\bar{P}_{si}^{\text{obs}}, \bar{v}_{li}^{\text{obs}}$ and $\bar{v}_{gi}^{\text{obs}}$ respectively in (3.33), (3.51) and (3.54). We also replace σ^2 and σ_v^2 with $\frac{\sigma}{n_{si}}, \frac{\sigma_v}{n_{li}}$, and $\frac{\sigma_v}{n_{gi}}$ in terms which include the means of the coexistence data.

In this chapter we have described the construction of a statistical model for a generic equation of state which accounts for homogeneous data, coexistence, measurement error on coexistence volumes and multiple observations of coexistence pressures and volumes. This model depends upon some parameters, $\boldsymbol{\theta}$ and in the case of measurement error, $\boldsymbol{v}_l$ and $\boldsymbol{v}_g$. In order to quantify uncertainty in the model's predictions we need to quantify uncertainty in these parameters. In the next chapter we will describe a Bayesian fit of this statistical model for synthetic and real data using the Demetriades EoS [Demetriades et al., 2013] described in Chapter 1.

Chapter 4

Fitting Parametric EoS for Pure CO_2

4.1 Introduction

In the previous chapter we proposed a statistical model for a generic EoS. In this chapter we fit this model for the Demetriades EoS [Demetriades et al., 2013] to both synthetic and real data for pure CO_2. We perform these fits using an MCMC sampler to sample from the posterior distribution of the model parameters, the background for which was described in Chapter 2.

We firstly fit synthetic data to test our MCMC sampler in particular to see if it works effectively in situations where we know the true parameter values. We firstly fit to synthetic homogeneous data for a single temperature and multiple temperatures before incorporating coexistence and measurement error on coexistence volumes. We then use the sampler we have developed to fit our model to the real pure CO_2 data from §1.6. We firstly fit to individual subcritical temperatures in order to propose temperature rules and then use these proposed temperature rules to fit multiple temperatures.

4.2 Single Temperature Homogeneous Synthetic Data

We now describe the single temperature synthetic data used. For the single temperature model, we need to fix the temperature at some value. Here we use $T = 0.9371048$. In dimensional units this corresponds to 285K. Our synthetic data set will consist of 100 data points evenly spaced in volume from 0.105 to 2.24, called here $\boldsymbol{v}^{\text{sim}}$. We produce synthetic data using the single-temperature Demetriades EoS,

$$h_D(v, T, \boldsymbol{\theta}) = \frac{T}{v + a} + \frac{b}{v^2 + c} + \frac{d}{v^3 + e} + \left(\frac{f}{v - g}\right)^6. \tag{4.1}$$

with 7-dimensional parameter vector,

$$\boldsymbol{\theta} = (a, b, c, d, e, f, g)^T. \tag{4.2}$$

To evaluate the Demetriades EoS at the selected volumes we need to select some true values for our parameters. We choose these to be the values for the fitted model at $T = 0.9371048$ as found by Demetriades et al. [2013],

$$a_{\text{TRUE}} = 0.3796146, \tag{4.3}$$

$$b_{\text{TRUE}} = -0.03867881, \tag{4.4}$$

$$c_{\text{TRUE}} = 0.0002303163, \tag{4.5}$$

$$d_{\text{TRUE}} = -0.03452484, \tag{4.6}$$

$$e_{\text{TRUE}} = 0.06297804, \tag{4.7}$$

$$f_{\text{TRUE}} = 0.1027511, \tag{4.8}$$

$$g_{\text{TRUE}} = 0.0497095. \tag{4.9}$$

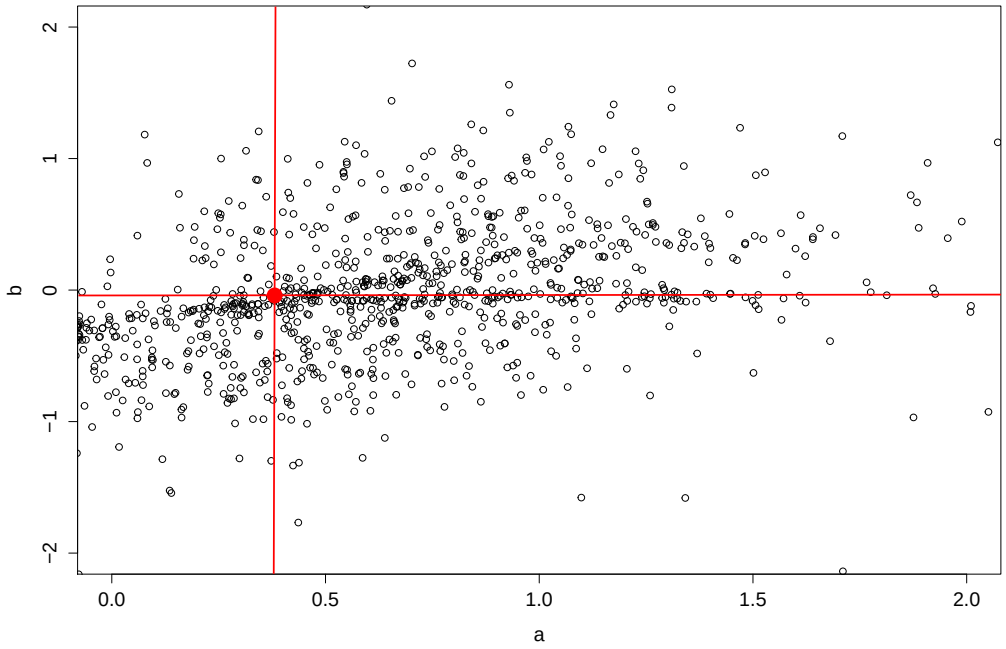


Figure 4.1: Scatter plot of the optimised values for parameters a and b found using MLE for the single temperature Demetriades EoS model without coexistence using synthetic data. The red point at the centre of the crosshairs marks the true values for a and b. Here we can see how despite the Nelder-Mead optimiser reporting convergence the optimised values of these parameters often differ significantly from the true values.

Then

$$\boldsymbol{\theta}_{\text{TRUE}} = (a_{\text{TRUE}}, b_{\text{TRUE}}, \ldots, g_{\text{TRUE}})^T. \tag{4.10}$$

Then our synthetic pressure values are

$$P_i^{\text{sim}} = h_D(v_i^{\text{sim}}, T, \boldsymbol{\theta}_{\text{TRUE}}) + \epsilon_i, \quad i = 1, \ldots, 100, \tag{4.11}$$

where the ϵ_i are iid normal random variables (r.v.) with zero mean and standard deviation $\sigma = 10^{-4}$.

4.2.1 Maximum Likelihood Estimation

We firstly attempt to fit this model using maximum likelihood estimation (MLE) as was described in §2.2.1. From (3.3) the likelihood for our model

using the Demetriades EoS and only homogeneous data is,

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \phi(P_i^{\text{sim}}; h_D(v_i^{\text{sim}}, T, \boldsymbol{\theta}), \sigma^2), \tag{4.12}$$

$$\implies L(\boldsymbol{\theta}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(P_i^{\text{sim}} - h_D(v_i^{\text{sim}}, T, \boldsymbol{\theta}))^2}{2\sigma^2}\right). \tag{4.13}$$

The log-likelihood is then,

$$l(\boldsymbol{\theta}) = -\frac{\sum_{i=1}^{n}(P_i^{\text{sim}} - h_D(v_i^{\text{sim}}, T, \boldsymbol{\theta}))^2}{2\sigma^2} - \frac{n}{2}\log(2\pi\sigma^2). \tag{4.14}$$

This is then the objective function we wish to maximise. Note that here we fix σ to be its true value 10^{-4} rather than estimate it. The optimiser should be better able to estimate a lesser number of parameters and given this is synthetic data and we know the true value of σ this gives us the best chance of finding the true parameter values using MLE. For this synthetic data, we then maximise the log-likelihood, using the Nelder-Mead [Nelder and Mead, 1965] optimisation algorithm.

We run the optimiser 1000 times for 2500 iterations with initial values sampled from the following distributions,

$$a \sim \mathcal{U}(0.1, 1), \tag{4.15}$$

$$b \sim \mathcal{U}(-1, 1), \tag{4.16}$$

$$c \sim \mathcal{U}(-0.01, 1), \tag{4.17}$$

$$d \sim \mathcal{U}(-1, 1), \tag{4.18}$$

$$e \sim \mathcal{U}(0, 1), \tag{4.19}$$

$$f \sim \mathcal{U}(0, 0.15), \tag{4.20}$$

$$g \sim \mathcal{U}(0, 0.1). \tag{4.21}$$

These distributions are chosen such that for parameters in these ranges and values of T and v which we consider, and which might be reasonably

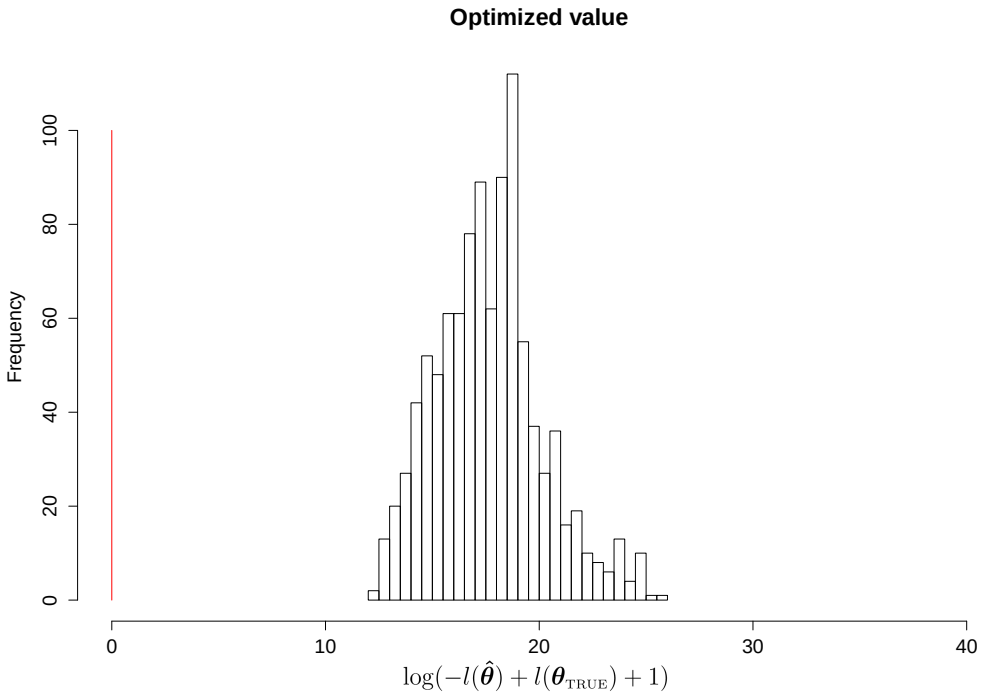


Figure 4.2: Histogram of $\log(-l(\hat{\boldsymbol{\theta}})+l(\boldsymbol{\theta}_{\text{TRUE}})+1)$ where $\hat{\boldsymbol{\theta}}$, are the optimised parameters found using MLE for the single temperature Demetriades EoS model without coexistence using synthetic data. Here we can see how despite the Nelder-Mead optimiser reporting convergence the optimised log-likelihoods are considerably less than that for the true parameter values, marked in red.

encountered in a CCS pipeline, there are no singularities in any of the terms in the Demetriades EoS and none of these terms are constant over the values of v and T we consider. As such these are sensible values for the parameters to take.

The optimiser reports that it has converged for 982 out of the 1000 times it was run. However when we consider the results of these runs of the optimiser we can see that it is performing poorly. Figure 4.1 shows the optimised values of the first two model parameters, a and b. The true parameter values are marked in red. We can see clearly from this that the optimiser is not finding the true values of the parameters. Figure 4.2 is a histogram of

$$\log(-l(\hat{\boldsymbol{\theta}}) + l(\boldsymbol{\theta}_{\text{TRUE}}) + 1), \tag{4.22}$$

for each set of optimised parameters $\hat{\boldsymbol{\theta}}$. We plot this rather than just plotting the log-likelihood in order for the plot to have better scaling. If

$$l(\hat{\boldsymbol{\theta}}) > l(\boldsymbol{\theta}_{\text{TRUE}}), \tag{4.23}$$

then

$$\log(-l(\hat{\boldsymbol{\theta}}) + l(\boldsymbol{\theta}_{\text{TRUE}}) + 1) < 0, \tag{4.24}$$

if

$$l(\hat{\boldsymbol{\theta}}) < l(\boldsymbol{\theta}_{\text{TRUE}}), \tag{4.25}$$

then

$$\log(-l(\hat{\boldsymbol{\theta}}) + l(\boldsymbol{\theta}_{\text{TRUE}}) + 1) > 0, \tag{4.26}$$

and if

$$l(\hat{\boldsymbol{\theta}}) = l(\boldsymbol{\theta}_{\text{TRUE}}), \tag{4.27}$$

then

$$\log(-l(\hat{\boldsymbol{\theta}}) + l(\boldsymbol{\theta}_{\text{TRUE}}) + 1) = 0. \tag{4.28}$$

The red line in figure 4.2 marks out $\log(l(\hat{\boldsymbol{\theta}}) + l(\boldsymbol{\theta}_{\text{TRUE}}) + 1) = 0$ which correspond to $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_{\text{TRUE}}$. As we can see the bars of the histogram are located for values of $\log(l(\hat{\boldsymbol{\theta}}) + l(\boldsymbol{\theta}_{\text{TRUE}}) + 1)$ which are significantly greater than 0. This means the optimised parameters in fact have a log-likelihood much less than the log-likelihood of the true parameter values. This demonstrates the failure of the optimisation routine in this case.

Optimisation methods such as MLE are how previous attempts at estimating model parameters for EoS models have been made. However as we have seen here such methods are not an effective method for parameter estimation for EoS models. The optimisation routine fails to recover the true parameter values in this case. The MLE method also relies upon large sample approximations in order to quantify uncertainty about the estimated

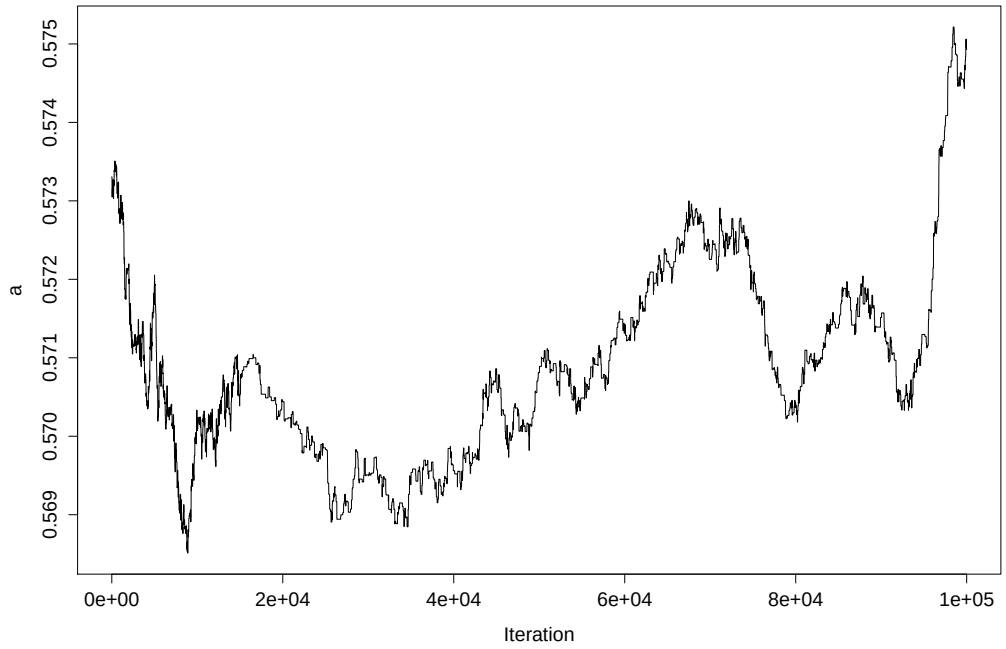


Figure 4.3: Traceplot of sample of parameter a for 100000 iterations of an application of the Metropolis-Hastings sampler with Gaussian random walk proposal density. We can see that iterations are highly correlated and as such we have inefficient exploration of the parameter space.

parameters. If we can not reliably identify the MLE, then this method has two weaknesses, for finite data the approximation to the distribution of the MLE may be inaccurate, and if we can not identify the MLE then we can not apply this approximation at all. This motivates us to consider the alternative Bayesian approach to parameter estimation, where we place a probability distribution upon the unknown parameters directly, condition upon data to obtain estimates and where we can use MCMC methods to reliably sample from the posterior distribution.

4.2.2 Metropolis-Hastings

In §2.3 we described the Bayesian approach to fitting statistical models. This is based around sampling from the posterior distribution for the parameters. As described in §2.3 to sample from this posterior distribution we require a likelihood and a prior. The likelihood is unchanged from the

MLE case discussed in the previous section and so our likelihood for the single temperature synthetic data without coexistence is given by (4.13). We then need to define a prior over our parameters. We set the priors to be the independent uniform distributions defined by (4.15)-(4.21). We use these uniform distributions again for the same reasons as before, for the values of v and T we are working with, these distributions were chosen to ensure that all terms in the Demetriades EoS contribute to the modelling of P. These ranges for the parameters remove from the parameter space regions where, for values of T and v we consider, the individual terms of the Demetriades EoS become constant and so do not contribute to the model and ensure that the terms in the Demetriades EoS can take values over the entire range of P of interest. We once more fix σ to be equal to its true value, 10^{-4}. The resulting posterior distribution,

$$p(\boldsymbol{\theta}|\boldsymbol{P}^{\text{sim}}, \boldsymbol{v}^{\text{sim}}) \propto L(\boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{4.29}$$

can not be expressed analytically. As such we must use MCMC methods as were described in chapter 2 to sample from the posterior.

We firstly apply the Metropolis-Hastings algorithm to the single temperature synthetic data without coexistence. The MH algorithm as described in §2.7 is the simplest MCMC algorithm. For our proposal density we use a Gaussian random walk as was defined in equation (2.55), with diagonal

covariance matrix,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_a^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_b^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_c^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_d^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_f^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_g^2 \end{pmatrix}. \tag{4.30}$$

Then we need to choose values for our proposal standard deviations, $\boldsymbol{\sigma} = (\sigma_a, \ldots, \sigma_g)^T$.

For a 6-parameter model Roberts et al. [1997] suggest an acceptance rate of 25% leads to optimal mixing. As our model has 7 parameters we will aim for our proposals to be accepted at a similar rate. We find that to achieve such an acceptance rate we have to set the elements of $\boldsymbol{\sigma}$ to be very small,

$$\boldsymbol{\sigma} = (0.0001, 0.0001, 0.0005, 0.0001, 0.0002, 0.000035, 0.000025)^T. \tag{4.31}$$

This then leads to very inefficient exploration of the parameter space with iterations being highly correlated.

Figure 4.3 shows the traceplot of the sample of a for a run of the MH sampler for 100000 iterations. From figure 4.3 we can see that even after so many iterations the chain has not explored much of the prior parameter space and has certainly not converged. This is due to the need for small proposals in order to ensure proposals are accept at a reasonable rate. Similar behaviour is observed in the traceplots of the other model parameters. This is most likely due to the model parameters being strongly correlated. In chapter 2 we used the toy example of a "banana" shaped distribution to

demonstrate that MH performs inefficiently at sampling from distributions of variables which are highly correlated.

Within the group at Nottingham Nelson et al. [2015] have applied several more advanced Monte Carlo methods to the EoS discussed here. These include slice sampling [Neal, 2003], parallel tempering with MH [Swendsen and Wang [1986],Geyer [1991]], and adaptive Monte Carlo [Andrieu and Thoms, 2008]. They have found that these methods encounter similar problems sampling the DEoS model parameters to MH. As such we now explore a Hamiltonian Monte Carlo (HMC) approach, which should be better able to sample from a distribution of highly correlated variables.

4.2.3 Hamiltonian Monte Carlo

As described in §2.8 Hamiltonian Monte Carlo, and its extension the no-U-turn sampler, is an MCMC algorithm designed to better avoid undesirable random walk behaviour such as the MH algorithm can be prone to. By using Hamiltonian dynamics to propose moves the sampler is better able to sample from distributions of variables which are highly correlated as appears to be the case here.

Using Stan as was described in §2.8 we are able to use the no-U-turn sampler algorithm to sample from our posterior distribution of interest. In this case to begin we sample σ rather than fix it. Thus our likelihood is now a function of $\boldsymbol{\theta}$ and σ,

$$L(\boldsymbol{\theta}, \sigma) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(P_i^{\text{sim}} - h_D(v_i^{\text{sim}}, T, \boldsymbol{\theta}))^2}{2\sigma^2}\right)). \tag{4.32}$$

We use the same independent prior distributions for the Demetriades EoS model parameters, $(a, \ldots, g)$, as we used in the MH, (4.15)-(4.21). We also now need to place a prior distribution on σ. We choose this prior distribu-

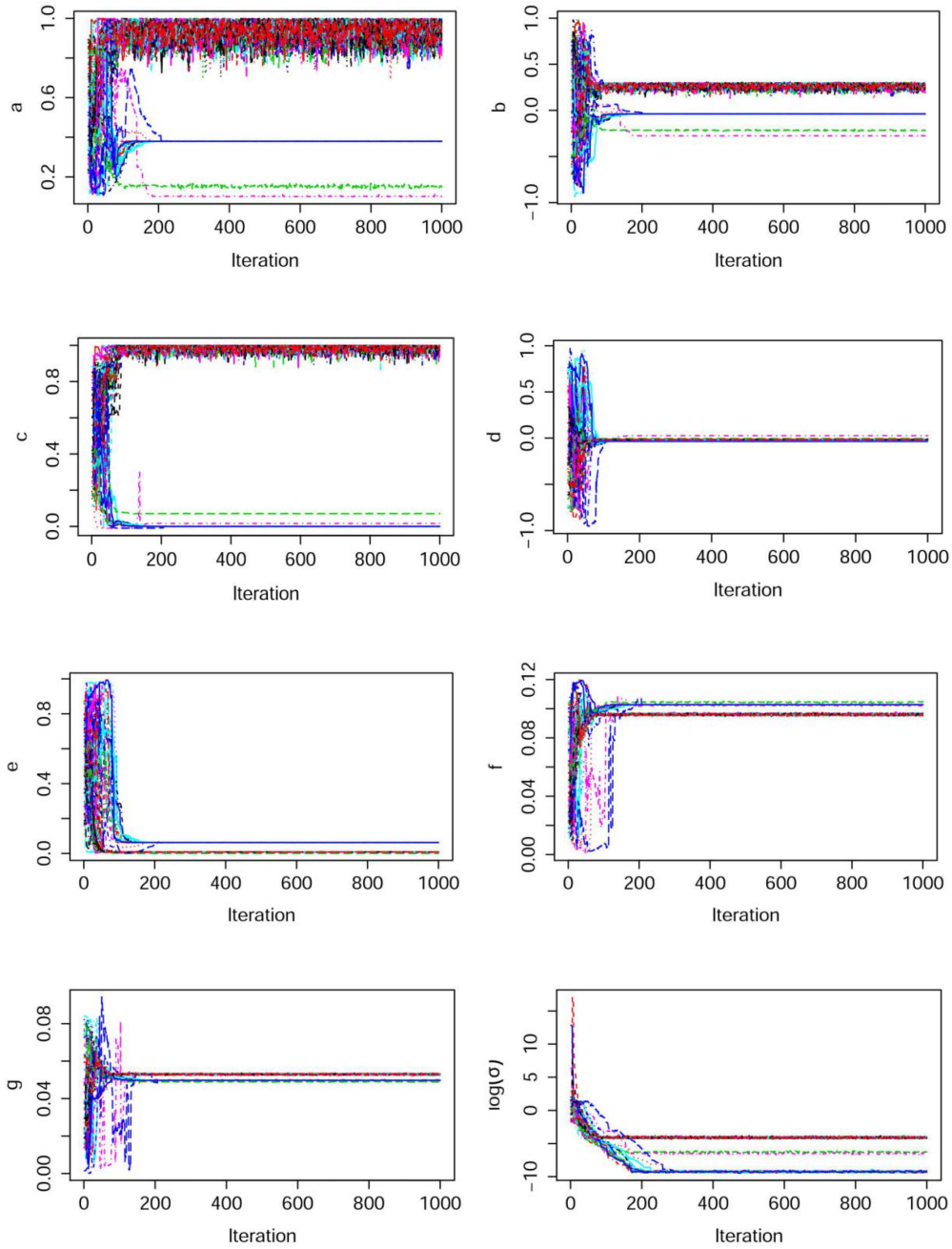


Figure 4.4: Traceplot of the parameter values for 50 chains of the no-U-turn sampler run for 1000 iterations using synthetic data for a single temperature. We have taken the log of σ so that it scales better for plotting. Different chains are marked in different colours. From this plot we can see that we have multi-modality in the parameters, particularly a, b, c and σ. We are also able to note that chains are not switching between modes, instead they are trapped.

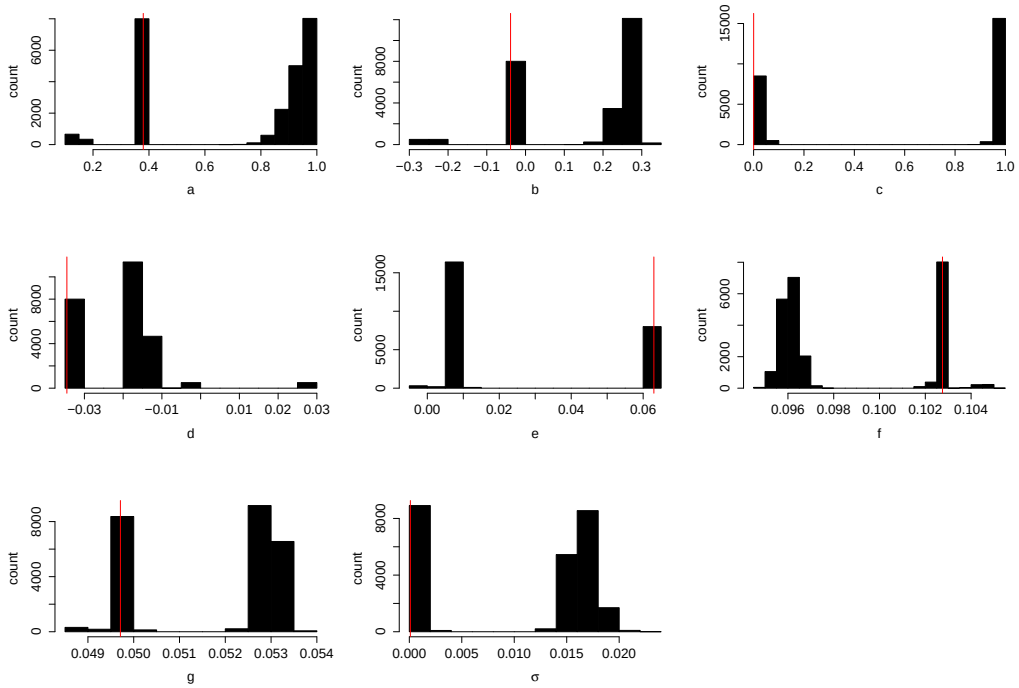


Figure 4.5: Histograms of the samples of the parameters from 50 chains of the no-U-turn sampler run for 1000 iterations using synthetic data for a single temperature. The red vertical lines mark the true parameter values. We can see here that while HMC often finds the mode about the true parameter values, it also finds other modes.

tion to be independent of the prior distributions of the other parameters,

$$p(\boldsymbol{\theta}, \sigma) = p(\boldsymbol{\theta})p(\sigma). \tag{4.33}$$

We initially gave σ the non-informative prior on the interval $[0, \infty)$ (we constrain σ to be positive), which has form

$$p(\sigma) \propto 1,\ \sigma \in [0, \infty). \tag{4.34}$$

This prior essentially states that before being presented with data we assume all positive values of σ are equally likely. This is not necessarily a good assumption as our non-dimensionalisation (1.37)-(1.39) puts the pressure data on roughly the unit scale. As such we would expect any observation error on the pressures to have variance less than 1. However,

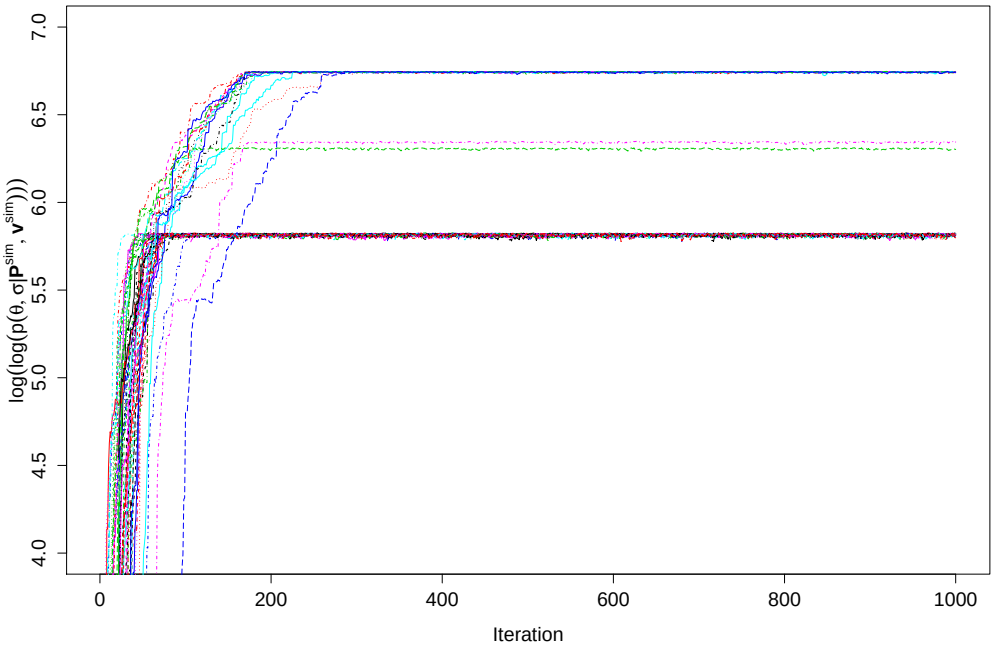


Figure 4.6: Traceplot of $\log\log p(\boldsymbol{\theta}, \sigma|\boldsymbol{P}^{\text{sim}}, \boldsymbol{v}^{\text{sim}})$ (up to a normalising constant) from 50 chains of the no-U-turn sampler run for 1000 iterations using synthetic data for a single temperature. Here we can see that a significant number of chains are not exploring the regions where the posterior probability is greatest. Instead these chains are trapped in local modes.

using a non-informative prior has the advantage of being convenient, and given a sufficient quantity of data this choice should not matter too much; with a greater quantity of data the model will dominate the prior and thus this possibly poor choice of prior will not have too great of an effect on the posterior. We later tried an alternate prior for σ which put more prior density on smaller values of σ, using an exponential distribution with various rate parameters ranging from 1 to 5. These exponential prior distributions place greater prior probability on smaller values of σ. Such values of σ are consistent with our non-dimensionalisation which puts our data on roughly the unit scale. We found that this change of prior had no discernible difference from the results obtained using the non-informative prior which are presented below.

We run 50 chains of the no-U-turn sampler algorithm starting at initial

points sampled from the prior, $p(\boldsymbol{\theta})$. Each chain consists of a 1000 iterations the first 500 of which are discarded as burn-in. Figure 4.4 shows the traces (including burn-in) of these chains for each of the eight parameters, $(a, b, c, d, e, f, g, \sigma)$, where we have taken the logarithm of σ so that it scales better for plotting. As can be seen for each of the parameters, the chains find a mode, become trapped there, and there is no movement between modes. Figure 4.5 shows a histogram of the posterior samples for the eight parameters averaged over all 50 chains. As can be seen, while some of the chains have found the true values of the parameters a significant number are trapped in other modes. Figure 4.6 is a traceplot of the log of the log-posterior (taking logs for scaling purposes), $\log\log p(\boldsymbol{\theta}, \sigma|\boldsymbol{P}^{\text{sim}}, \boldsymbol{v}^{\text{sim}})$ up to a normalising constant for all 50 chains. As we can see for each chain the log-posterior increases to some value before staying roughly constant, it is at these points the chains are becoming trapped in modes. Some chains achieve the highest log-posterior value, as they have found the mode about the true parameter values, $\boldsymbol{\theta}_{\text{TRUE}}$. We can see that the other modes correspond to a significantly lower log-posterior. As such were our sampler functioning well we would expect states at these modes with lower log-posterior to spend only a short time there, eventually moving to the mode about the true parameter values, and perhaps, should we run the sampler for a significant amount of time, to see a chain move briefly from the mode about the true parameter values to another mode.

Running the sampler for longer (10000 iterations), to see if there is eventually movement between modes we find that the chains remain trapped in the first mode they found their way into. Any MCMC sampler that is correctly defined to have the target distribution as its stationary distribution (and so satisfies detailed balance) will given enough time always converge to the distribution of interest. Were we to run our sampler here for a long enough time we would eventually see the chains propose a move into the

global mode and move there. However in order to get convergence to occur in a short enough timescale that we can obtain results we need to consider alternative approaches.

We have made several alterations to this model to see if this reduces the multi-modality in the posterior; trying different priors for σ that place a greater probability density on lower values; fixing σ so that the size of the parameter space is reduced by 1; removing terms from the Demetriades EoS, to simplify the model; and reparameterising the Demetriades EoS as,

$$h_D(v, T, \boldsymbol{\theta}) = \frac{T}{v + a} + \frac{b}{v^2 + c} + \frac{d}{v^3 + e} + \frac{f_r}{(v - g)^6}, \tag{4.35}$$

by setting $f_r = f^6$ where f_r is our reparameterised f. We do this reparameterisation in the hopes that removing the parameter to a significant power (6) we might smooth out the likelihood surface. None of these alterations have any effect; the posterior remains multi-modal in every case, and we find our sampler always gets stuck in modes which it is then not able to escape.

For this reason, we need to investigate different approaches that enable the sampler to move between modes. One such approach which we consider in the following section is parallel tempering [Swendsen and Wang, 1986, Geyer, 1991].

4.2.4 Parallel Tempering with HMC

In §2.9 we described how parallel tempering allows an MCMC sampler to move between local modes by sampling from a system of probability distributions and proposing swaps between them. As we have found our posterior distribution of interest to be multi-modal we add PT to our HMC algorithm. In §2.9 we described two alternative approaches to parallel tempering, tempering the entire distribution of interest [Earl and Deem,

2005, Fukunishi et al., 2002] or tempering only some part of the distribution of interest, generally the likelihood and so called the likelihood-tempering approach [Paquet, 2007].

We firstly derive an acceptance probability for applying the likelihood-tempering method to the case of HMC and the no-U-turn sampler, which we will call PT-HMC. So for replica k we define the probability distribution

$$\begin{aligned} p_k(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{P}^{\text{sim}}, \boldsymbol{v}^{\text{sim}}) &= p_k(\boldsymbol{\theta}|\boldsymbol{P}^{\text{sim}}, \boldsymbol{v}^{\text{sim}})p(\boldsymbol{\phi}), \\ &\propto p(\boldsymbol{P}^{\text{sim}}|\boldsymbol{v}^{\text{sim}}, \boldsymbol{\theta})^{\frac{1}{\beta_k}} p(\boldsymbol{\theta})p(\boldsymbol{\phi}). \end{aligned} \tag{4.36}$$

where β_k is the tempering level of replica k and as in §2.8 $\boldsymbol{\phi}$ are the momentum parameters introduced by the HMC algorithm. The tempering levels are defined on the interval $[1, \infty)$. When $\beta_k = 1$ this is the posterior distribution of interest.

The Hamiltonian for replica k is

$$\begin{aligned} H_k(\boldsymbol{\theta}, \boldsymbol{\phi}) &= -\log(p(\boldsymbol{P}^{\text{sim}}|\boldsymbol{\theta}, \boldsymbol{v}^{\text{sim}})^{\frac{1}{\beta_k}} p(\boldsymbol{\theta})) - \log(p(\boldsymbol{\phi})) \\ &= -\left[\frac{1}{\beta_k}\log p(\boldsymbol{P}^{\text{sim}}|\boldsymbol{\theta}, \boldsymbol{v}^{\text{sim}}) + \log p(\boldsymbol{\theta})\right] - \log p(\boldsymbol{\phi}) \\ &= U_k(\boldsymbol{\theta}) + K_k(\boldsymbol{\phi}), \end{aligned} \tag{4.37}$$

where U_k and K_k are the potential and kinetic energies for replica k respectively. Note for all k, the kinetic energy is the same as we do not temper the distribution of the momentum variables. So

$$K_k(\boldsymbol{\phi}) = K(\boldsymbol{\phi}) = -\log(p(\boldsymbol{\phi})) \quad \forall k, \tag{4.38}$$

Set $C_k = (\boldsymbol{\theta}_k, \boldsymbol{\phi}_k)$ to be the current state of replica k. Then the probability density of replica k is

$$p_k(C_k|\boldsymbol{P}^{\text{sim}}, \boldsymbol{v}^{\text{sim}}) = \frac{1}{Z_k}\exp(-H_k(C_k)), \tag{4.39}$$

where Z_k is the normalising constant for replica k and the joint probability distribution for m replicas is

$$P(C_1, H_1; C_2, H_2; ...; C_m, H_m) = \prod_{i=1}^{m} p_i(C_i | \boldsymbol{P}^{\text{sim}}, \boldsymbol{v}^{\text{sim}}). \tag{4.40}$$

Then we update $\boldsymbol{\theta}_k$ and $\boldsymbol{\phi}_k$ for each replica k independently using HMC or the no-U-turn sampler.

After having sampled from the individual replicas for a given number of iterations we then propose swaps between neighbouring replicas. Define the transition probability that configuration $C = (\boldsymbol{\theta}, \boldsymbol{\phi})$ of the m-th replica is swapped with the configuration $C' = (\boldsymbol{\theta}', \boldsymbol{\phi}')$ of the n-th replica as $W(C, H_m; C', H_n)$.

In order for the untempered density to have stationary distribution equal to the posterior distribution we wish to sample from we require detailed balance (2.42) to hold. If detailed balance holds then

$$\begin{aligned} &P(...; C, H_m; C', H_n; ...) W(C, H_m; C', H_n) \\ &\qquad = P(...; C', H_m; C, H_n; ...) W(C', H_m; C, H_n), \end{aligned} \tag{4.41}$$

$$
\begin{aligned}
\implies \frac{W(C, H_m; C', H_n)}{W(C', H_m; C, H_n)} &= \frac{\frac{1}{Z_m}\exp(-H_m(C'))\frac{1}{Z_n}\exp(-H_n(C))}{\frac{1}{Z_m}\exp(-H_m(C))\frac{1}{Z_n}\exp(-H_n(C'))} \\
&= \exp(-H_m(C') - H_n(C) + H_m(C) + H_n(C')) \\
&= \exp(-U_m(\boldsymbol{\theta}') - U_n(\boldsymbol{\theta}) + U_m(\boldsymbol{\theta}) + U_n(\boldsymbol{\theta}')) \\
&= \exp\Bigg(\frac{1}{\beta_m}\log p(\boldsymbol{P}^{\text{sim}}|\boldsymbol{\theta}', \boldsymbol{v}^{\text{sim}}) + \log p(\boldsymbol{\theta}') \\
&+ \frac{1}{\beta_n}\log p(\boldsymbol{P}^{\text{sim}}|\boldsymbol{\theta}, \boldsymbol{v}^{\text{sim}}) + \log p(\boldsymbol{\theta}) \\
&- \frac{1}{\beta_m}\log p(\boldsymbol{P}^{\text{sim}}|\boldsymbol{\theta}, \boldsymbol{v}^{\text{sim}}) - \log p(\boldsymbol{\theta}) \\
&- \frac{1}{\beta_n}\log p(\boldsymbol{P}^{\text{sim}}|\boldsymbol{\theta}', \boldsymbol{v}^{\text{sim}}) + \log p(\boldsymbol{\theta}')\Bigg) \\
&= \exp\Bigg(\frac{1}{\beta_m}\log p(\boldsymbol{P}^{\text{sim}}|\boldsymbol{\theta}', \boldsymbol{v}^{\text{sim}}) \\
&+ \frac{1}{\beta_n}\log p(\boldsymbol{P}^{\text{sim}}|\boldsymbol{\theta}, \boldsymbol{v}^{\text{sim}}) - \frac{1}{\beta_m}\log p(\boldsymbol{P}^{\text{sim}}|\boldsymbol{\theta}, \boldsymbol{v}^{\text{sim}}) \\
&- \frac{1}{\beta_n}\log p(\boldsymbol{P}^{\text{sim}}|\boldsymbol{\theta}', \boldsymbol{v}^{\text{sim}})\Bigg) \\
&= \exp(-\boldsymbol{\Delta}),
\end{aligned}
\tag{4.42}
$$

where

$$
\boldsymbol{\Delta} = \left(\frac{1}{\beta_n} - \frac{1}{\beta_m}\right)(\log p(\boldsymbol{P}^{\text{sim}}|\boldsymbol{\theta}', \boldsymbol{v}^{\text{sim}}) - \log p(\boldsymbol{P}^{\text{sim}}|\boldsymbol{\theta}, \boldsymbol{v}^{\text{sim}})). \tag{4.43}
$$

So we have a Metropolis acceptance step and accept swaps between replicas with probability $\min(1, \exp(-\boldsymbol{\Delta}))$.

4.2.5 Application of PT-HMC

We now apply our PT-HMC algorithm to the synthetic data for a single temperature without coexistence. We wish to sample from the posterior distribution of the model of this data with fixed σ. For the synthetic data we know $\sigma = 10^{-4}$ and so fix it accordingly. We use the same independent, uniform prior distributions as previously, (4.15)-(4.21). We need to choose the number of replicas and the value of the tempering level at each replica,

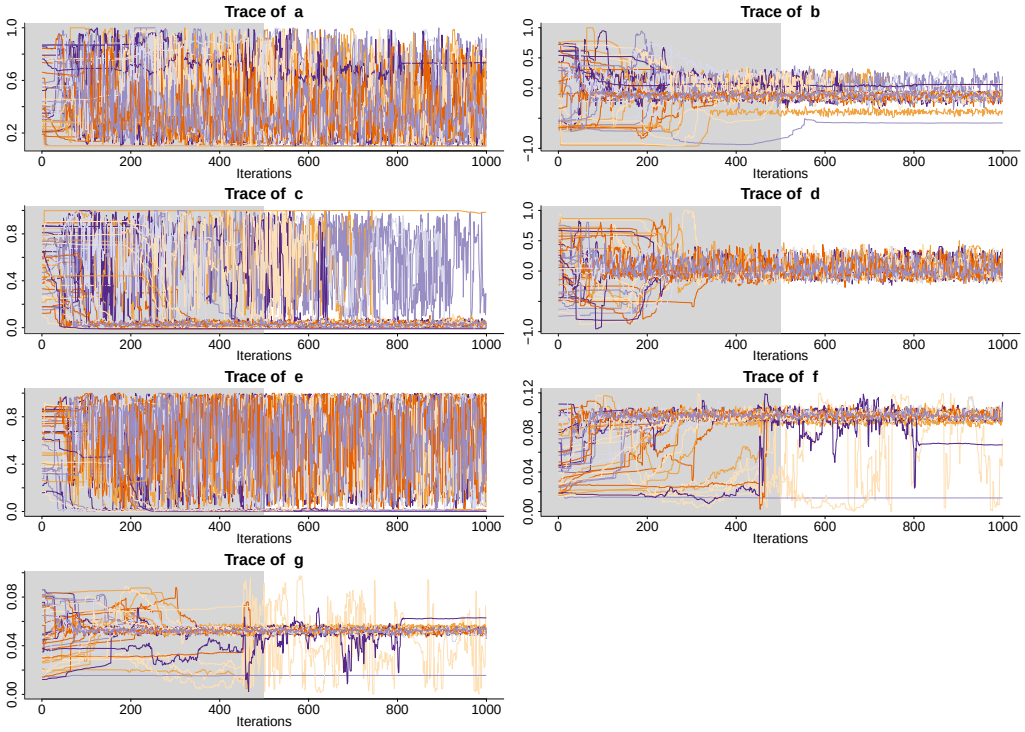


Figure 4.7: Traceplot of iterations of parameters for chains of no-U-turn sampler for single temperature DEoS with tempering level $\beta = 10^6$. The grey background marks the initial burn-in period of 500 iterations where stepsize is adapted. The white background is the period after burn-in where stepsize is fixed. We can see the chains explore either the entire parameter space or one large mode. We thus choose to make this the replica with the highest tempering level.

$(\beta_1, \beta_2, \dots)$. This is somewhat a matter of trial and error, we run replicas at different tempering levels to see how well they explore the space. What is most important is to choose the correct value for the highest level. We need the highest tempering level to be such that the corresponding replica explores the space.

To determine the maximum tempering level, the no-U-turn sampler is applied 50 times, running for 1000 iterations, using different initial values for $\boldsymbol{\theta}$ sampled from the prior distribution $p(\boldsymbol{\theta})$, to sample from the posterior distribution of replicas of our model with increasing β. Considering the traceplots of these allows us to determine when the distribution is no longer dominated by the multi-modality of the likelihood. Figure 4.7 shows such a traceplot for $\beta = 10^6$. As can be seen the chains are either exploring

the entirety of the parameter space or a single wide mode rather than the number of tightly constrained modes we saw with our previous runs of the sampler 4.4. We thus choose our maximum tempering level to be 10^6. This corresponds to a standard deviation of 0.1 for the highest replica as,

$$\frac{\sigma}{\sqrt{10^{-6}}} = \frac{10^{-4}}{10^{-3}} = 0.1. \tag{4.44}$$

Paquet [2007] suggests that the set of tempering levels should be a geometric series, so that the probability of swaps between replicas is constant for all replicas, and so we make an initial choice to have tempering levels increase by a factor of 10. So we have a set of 7 tempering levels

$$\boldsymbol{\beta} = (1, 10^1, 10^2, 10^3, 10^4, 10^5, 10^6)^T. \tag{4.45}$$

So $\beta_1 = 1$ corresponds to our distribution of interest.

A choice needs to be made of how many iterations of the no-U-turn sampler to do before proposing swaps between replicas. We choose to do 10 iterations. Proposing a swap after every iteration of the no-U-turn sampler would be inefficient, since after just 1 iteration there is unlikely to be a significant difference from the values of the previous iteration. As such if our previously proposed swap was rejected, we are unlikely to accept this proposed swap, and if it was accepted then we want our lower level replicas to explore the new region of parameter space and not swap back. 10 iterations allows the possibility of significant movement in the sampler between swaps, while still proposing swaps regularly enough that parameter values, $\boldsymbol{\theta}$, with higher posterior probability should be able to move quickly from the hotter replicas to the coldest. We propose swaps between neighbouring replicas beginning with the highest level replicas and continuing to the

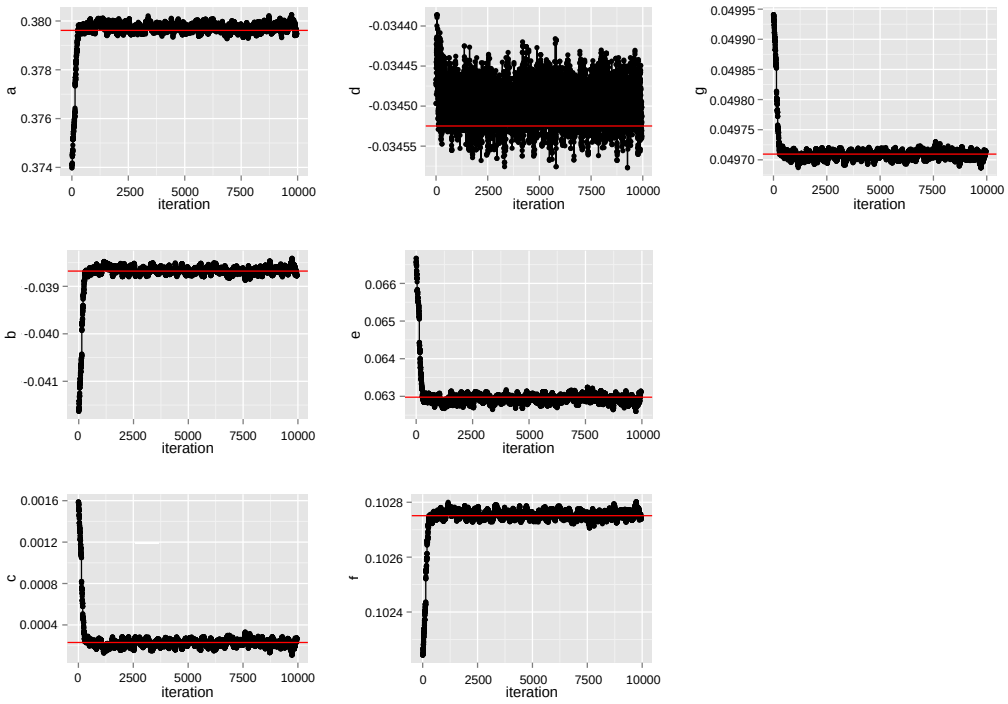


Figure 4.8: Traceplots of the sampled parameters for a single chain of the PT-HMC sampler applied to the single temperature DEoS with synthetic data. The first 1000 iterations of the sampler were discarded as burn-in and are not plotted here. The red lines mark the true parameter values. The sampler has converged and is mixing well. Traceplots for other runs of the PT-HMC sampler from different starting positions show the same convergence to the mode containing the true parameter values. We have developed a method of parameter estimation for the single temperature model which works well.

lowest replicas. So we propose swaps as follows,

$$S(p_k, p_{k-1}), \quad k = 7, \ldots, 1, \tag{4.46}$$

where $S(p_k, p_{k-1})$ indicates we are proposing a swap between replicas k and $k-1$,

We sample each set of 10 iterations of the no-U-sampler in Stan before proposing swaps as described above. As we enter and leave Stan when proposing swaps we must input the current values for our tuning parameters each time we enter Stan. Stan does not allow user-defined metrics, $\boldsymbol{M}$, except for $\boldsymbol{M} = \boldsymbol{I}$, setting the metric to be the identity matrix. Compared

to estimating $\boldsymbol{M}$, we would expect fixing $\boldsymbol{M} = \boldsymbol{I}$ to be a poorer choice of metric. In the case of a poorly defined metric the no-U-turn sampler takes a larger number of steps at smaller stepsize to compensate [Stan Development Team, 2016b]. Running the no-U-turn sampler for our model as described with the unit metric finds that this is indeed the case. This compensation however does not affect the results and so we proceed with the unit metric as an acceptable choice of metric.

A plot of the resulting traceplots for one run of the sampler for 11000 iterations is shown in figure 4.8. We discard the first 1000 iterations as burn-in and they are not plotted in figure 4.8 to better show the mixing about the true parameters. A number of swaps between replicas are typically accepted in the first few proposals allowing the sampler to get near the mode about the true values. The adapted stepsize then allows reasonably quick exploration of this mode and the traceplots show the sampler for the untempered distribution, p_1, mixing well around the true parameter values. Once the sampler is mixing around the true values of the parameters occasional swaps are still proposed. These moves remain within the same mode. Runs of the sampler from different starting positions perform similarly well, converging to the mode about the true parameter values and mixing well once there. As such the PT-HMC sampler seems to be working well for the single temperature DEoS with synthetic data for a single temperature without coexistence.

4.3 Multiple Temperature Homogeneous Synthetic Data

Having developed a sampler which works well for the homogeneous single temperature case we apply our PT-HMC sampler to the multiple temperature case. To do so we need to define temperature rules for the parameters

in the Demetriades EoS. For this synthetic data we use linear temperature rules (1.43) as were described in Chapter 1. To produce synthetic data for the multiple temperature Demetriades EoS, a minimum and maximum temperature of interest are required for the linear temperature rule; we set these to correspond, in dimensional units, to 273K and 304.1282K, the critical temperature, respectively. So in non-dimensionalised units,

$$T_l = 0.8976478, \quad T_h = 1. \tag{4.47}$$

We simulate data at two temperatures $T = 0.9206644$ and 0.9634095 (280K and 293K in dimensional units respectively). Let $\boldsymbol{T}^{\text{sim}}$ be the vector of length 200 which collects our synthetic T. At each temperature we use the same 100 values for volume as were used in the single temperature case. We thus have 200 data points. Our vector of synthetic volumes, $\boldsymbol{v}^{\text{sim}}$, is a 200-dimensional vector. We again take the true values for the parameters at the extremal temperatures from Demetriades et al. [2013]. These are as follows,

$$a_{l,\text{TRUE}} = 0.405015,\ a_{h,\text{TRUE}} = 0.339126, \tag{4.48}$$

$$b_{l,\text{TRUE}} = -0.051129,\ b_{h,\text{TRUE}} = -0.018833, \tag{4.49}$$

$$c_{l,\text{TRUE}} = 0.000265,\ c_{h,\text{TRUE}} = 0.000183, \tag{4.50}$$

$$d_{l,\text{TRUE}} = -0.037169,\ d_{h,\text{TRUE}} = -0.030310, \tag{4.51}$$

$$e_{l,\text{TRUE}} = 0.066872,\ e_{h,\text{TRUE}} = 0.056771, \tag{4.52}$$

$$f_{l,\text{TRUE}} = 0.101726,\ f_{h,\text{TRUE}} = 0.104385, \tag{4.53}$$

$$g_{l,\text{TRUE}} = 0.047521,\ g_{h,\text{TRUE}} = 0.053198. \tag{4.54}$$

These are collected in the parameter vector, $\boldsymbol{\theta}_{\text{TRUE}}$. Then our synthetic pressure values are calculated by using these parameter values in the Demetriades EoS with our synthetic volumes for each of the two temperatures

of interest,

$$P_i^{\text{sim}} = h_D(v_i^{\text{sim}}, T_i^{\text{sim}}, \boldsymbol{\theta}_{\text{TRUE}}) + \epsilon_i, \quad i = 1, \ldots, 200, \tag{4.55}$$

where the ϵ_i are again iid normal r.v. with zero mean and standard deviation $\sigma = 10^{-4}$.

4.3.1 Application of PT-HMC

For the multiple temperature model the likelihood is unchanged from (4.13) except now $\boldsymbol{\theta}$ is the 14-dimensional parameter vector,

$$\boldsymbol{\theta} = (a_l, \ldots, g_l, a_h, \ldots, g_h)^T, \tag{4.56}$$

as defined by the linear temperature rule, and as such temperature is no longer fixed,

$$L(\boldsymbol{\theta}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(P_i^{\text{sim}} - h_D(v_i^{\text{sim}}, T_i^{\text{sim}}, \boldsymbol{\theta}))^2}{2\sigma^2}\right). \tag{4.57}$$

We once more fix, $\sigma = 10^{-4}$. To obtain a posterior to sample from prior distributions are defined on the parameters. We apply the same independent uniform prior distribution to a_l and a_h as was applied to a in our previous samplers (4.15). We do the same for the other parameters so our

priors are,

$$a_l \sim \mathcal{U}(0.1, 1), \tag{4.58}$$

$$b_l \sim \mathcal{U}(-1, 1), \tag{4.59}$$

$$c_l \sim \mathcal{U}(-0.01, 1), \tag{4.60}$$

$$d_l \sim \mathcal{U}(-1, 1), \tag{4.61}$$

$$e_l \sim \mathcal{U}(0, 1), \tag{4.62}$$

$$f_l \sim \mathcal{U}(0, 0.15), \tag{4.63}$$

$$g_l \sim \mathcal{U}(0, 0.1), \tag{4.64}$$

$$a_h \sim \mathcal{U}(0.1, 1), \tag{4.65}$$

$$b_h \sim \mathcal{U}(-1, 1), \tag{4.66}$$

$$c_h \sim \mathcal{U}(-0.01, 1), \tag{4.67}$$

$$d_h \sim \mathcal{U}(-1, 1), \tag{4.68}$$

$$e_h \sim \mathcal{U}(0, 1), \tag{4.69}$$

$$f_h \sim \mathcal{U}(0, 0.15), \tag{4.70}$$

$$g_h \sim \mathcal{U}(0, 0.1). \tag{4.71}$$

Consideration of the traceplots of the no-U-turn sampler applied to the multiple temperature synthetic data with β fixed at various values suggests that once again maximum tempering level of $\beta = 10^6$ (corresponding to the maximum replica having standard deviation, 0.1) is suitable. As increasing tempering levels by a factor of 10 worked in the single temperature case, we again use that as an initial choice here. As such our choice of tempering levels is the same 7 as in the single temperature case. We again propose swaps between neighbouring replicas, beginning at the maximum tempering level after every 10 iterations of the no-U-turn sampler. We also retain the 1000 iteration burn-in and then run the sampler until such time as a sample of size 10000 is completed. We set $\boldsymbol{M} = \boldsymbol{I}$ for the same reasons as before.

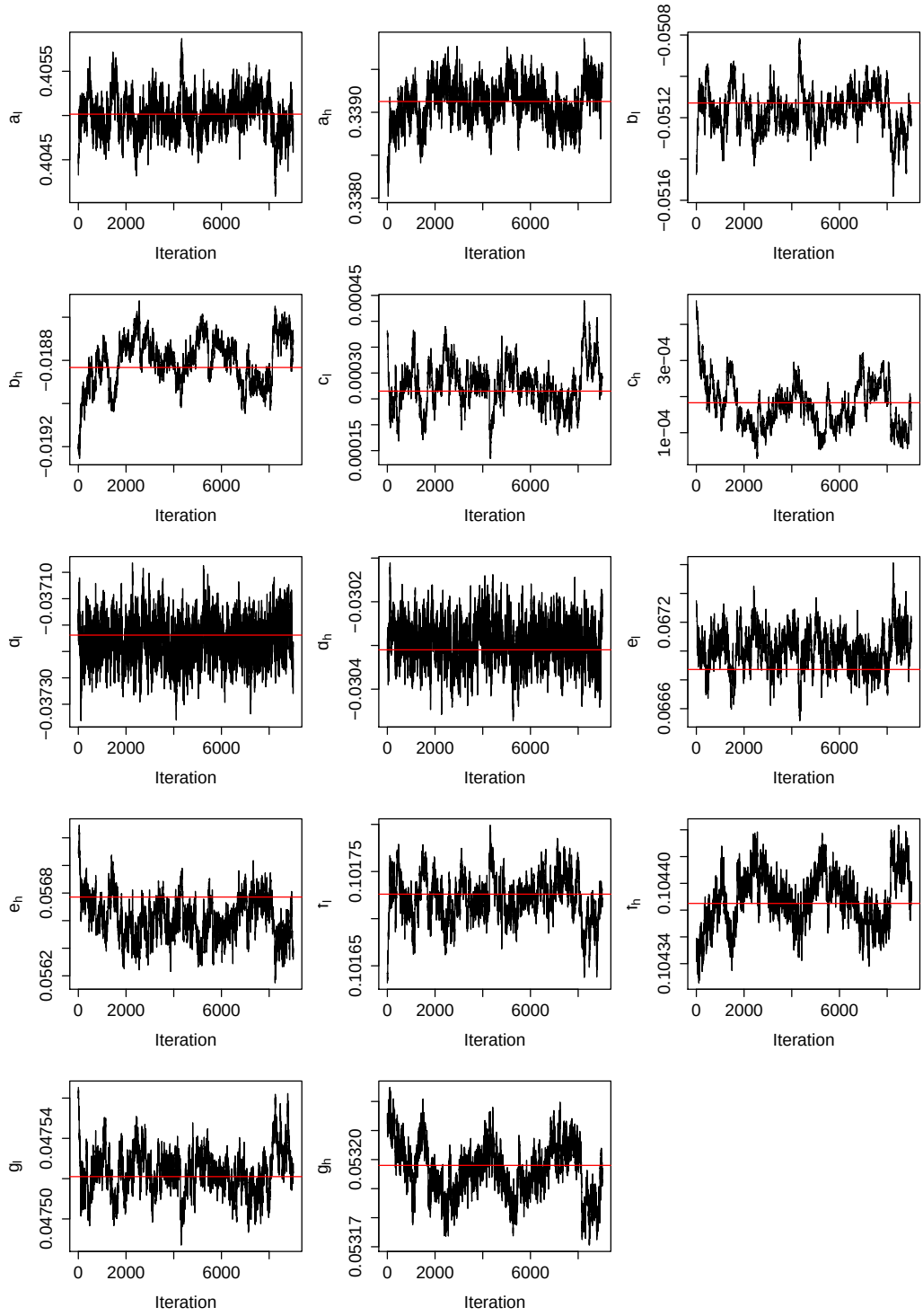


Figure 4.9: Traceplots of the sampled parameters for a single chain of the PT-HMC sampler applied to the multiple temperature DEoS with synthetic data. The first 1000 iterations were discarded as burn-in and are not plotted here. The red lines mark the true parameter values. Although the sampler has converged, for a significant number of the parameters we can see the sampler is mixing poorly. This appears to be due to a small adapted stepsize. This suggests a need to either run the sampler for longer or thin the samples.

Based on the results from the single temperature case we once more adapt the stepsize at every iteration of the no-U-turn sampler.

The above sampler was run using 10 sets of initial values randomly sampled from the prior. As in the single temperature case these all found the true values of the parameters. Figure 4.9 is a traceplot for the parameters for one of these chains. The behaviour seen here is repeated across chains. The sampler has succeeded in finding the mode about the true parameter values. However with the exception of parameters d_l and d_h the sampler is showing signs of poor mixing around these true values, suggesting a poor stepsize. Running the sampler for a greater number of iterations and thinning corrects for the poor mixing.

Having found a sampler which works well in both the single and multiple temperature cases with homogeneous data we now expand the model to include coexistence.

4.4 Single Temperature with Coexistence Data

We first introduce data including coexistence for a single temperature. We fix temperature to be $T = 0.9371048$ in our non-dimensionalised units, as in the previous case of using synthetic data for a single temperature. We use the same true parameter values as in the case without coexistence data (4.3)-(4.9). To fit the model we will require the integral of the Demetriades EoS between two volumes v_1 and v_2 which in the notation of chapter 3 is,

$$H_D(v_1, v_2, T, \boldsymbol{\theta}) = \int_{v_1}^{v_2} h_D(v, T, \boldsymbol{\theta})\, \mathrm{d}v. \tag{4.72}$$

In chapter 1 we defined (1.45),

$$h_{int}(v, T, \boldsymbol{\theta}) = \int_{v'=\infty}^{v} \frac{T}{v'} - h_D(v', T, \boldsymbol{\theta})\ \mathrm{d}v', \tag{4.73}$$

and saw that this integral converged and had analytic form given by (1.46) in the case that c is positive and by (1.47) in the case that c is negative. Then we have,

$$H_D(v_1, v_2, T, \boldsymbol{\theta}) = h_{\mathrm{int}}(v_1, T, \boldsymbol{\theta}) - h_{\mathrm{int}}(v_2, T, \boldsymbol{\theta}) + T \ln\left(\frac{v_g}{v_l}\right). \tag{4.74}$$

Given T and the true parameter values at coexistence the equilibrium conditions define a non-linear system of three equations with three unknowns, P_s, v_l, v_g,

$$P_s = h_D(v_l, T, \boldsymbol{\theta}_{\mathrm{TRUE}}), \tag{4.75}$$

$$P_s = h_D(v_g, T, \boldsymbol{\theta}_{\mathrm{TRUE}}), \tag{4.76}$$

$$P_s = \frac{H_D(v_l, v_g, T, \boldsymbol{\theta}_{\mathrm{TRUE}})}{v_g - v_l}. \tag{4.77}$$

We use the Newton-Raphson root solving method [Ryaben'kii and Tsynkov, 2006] to find these unknowns. Using Newton-Raphson the true coexistence volumes and coexistence pressure are found to be,

$$v_l = 0.1515569, \tag{4.78}$$

$$v_g = 0.9272321, \tag{4.79}$$

$$P_s = 0.6319629. \tag{4.80}$$

We introduce error into the coexistence volumes as follows,

$$v_l^{\mathrm{obs}} = v_l + \epsilon_l, \tag{4.81}$$

$$v_g^{\mathrm{obs}} = v_g + \epsilon_g, \tag{4.82}$$

where ϵ_l and ϵ_g are iid normally distributed r.v. with zero mean and standard deviation $\sigma_v = 10^{-3}$. Similarly error is introduced into the coexistence pressure as follows,

$$P_s^{\text{obs}} = P_s + \epsilon_s, \tag{4.83}$$

where ϵ_s is a normally distributed r.v. with zero mean and standard deviation $\sigma = 10^{-4}$

We now create some synthetic, homogeneous data. Previously we used 0.105 and 2.24 as the bounds for our synthetic volume. We use these as bounds again here. However when considering coexistence it is important to note that no data can physically exist in the interval (v_l, v_g). As such we choose the synthetic volumes in our data set to be whichever of the 100 evenly spaced points from 0.105 to 2.24 which are not contained in the interval (v_l, v_g). The resulting vector of synthetic volumes $\boldsymbol{v}^{\text{sim}}$ is of size 64. Then our synthetic pressure values are,

$$P_i^{\text{sim}} = h_D(v_i^{\text{sim}}, T, \boldsymbol{\theta}_{\text{TRUE}}) + \epsilon_i, \quad i = 1, \ldots, 64, \tag{4.84}$$

where the ϵ_i are iid normal r.v. with zero mean and standard deviation $\sigma = 10^{-4}$. This is the same noise as was used on the coexistence pressure value. As v_l and v_g are treated as parameters we have 7-dimensional EoS vector of parameters,

$$\boldsymbol{\theta} = (a, b, c, d, e, f, g)^T, \tag{4.85}$$

and 9 total parameters collected in the vector,

$$\tilde{\boldsymbol{\theta}} = (a, b, c, d, e, f, g, v_l, v_g)^T. \tag{4.86}$$

We will now fit our statistical model for the specific case of the Demetriades EoS using both the hard constraint and soft constraint methods described in Chapter 3.

4.4.1 Hard Constraint

As described in chapter 3 the likelihood for our statistical model in the hard constraint case is a degenerate multivariate normal distribution consisting of two parts, a density (3.32), and a subspace on which this density is defined, (3.30) and (3.31). This subspace is defined by the two constraints,

$$P_s^{\text{obs}} - h_D(v_l, T, \boldsymbol{\theta}) = P_s^{\text{obs}} - h_D(v_g, T, \boldsymbol{\theta}), \tag{4.87}$$

$$P_s^{\text{obs}} - h_D(v_l, T, \boldsymbol{\theta}) = P_s^{\text{obs}} - \frac{H_D(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}. \tag{4.88}$$

Substituting in the forms of h_D and H_D we can rearrange these two constraints to provide expressions for the two parameters which are linear in the model, b and d, in terms of the other parameters. This removes these two parameters from the model and ensures the constraints always hold. We firstly rearrange (4.87) to give an expression for b in terms of the other parameters,

$$b = \frac{\frac{T}{\log(v_g+a)} - \frac{T}{\log(v_l+a)} + d\left(\frac{1}{v_g^3+e} - \frac{1}{v_l^3+e}\right) + \frac{f^6}{(v_g-g)^6} - \frac{f^6}{(v_l-g)^6}}{\frac{1}{v_l^2+c} - \frac{1}{v_g^2+c}}. \tag{4.89}$$

We can then substitute (4.89) into (4.88) and rearrange so that it is defined in terms of d as,

$$d^* = \frac{d_1}{d_2}, \tag{4.90}$$

where,

$$\begin{aligned} d_1 = {} & \frac{T}{v_l + a} + \left(\frac{\frac{T}{v_g+a} - \frac{T}{v_l+a} + \frac{f^6}{(v_g-g)^6} - \frac{f^6}{(v_l-g)^6}}{(v_l^2 + c)\left(\frac{1}{v_l^2+c} - \frac{1}{v_g^2+c}\right)} \right) + \frac{f^6}{(v_l - g)^6} \\ & - \frac{1}{v_g - v_l}\left(T\log(v_g + a) - T\log(v_l + a) - \frac{f^6}{5(v_g - g)^5} + \frac{f^6}{5(v_l - g)^5} \right) \\ & - \left(\frac{\frac{T}{v_g+a} - \frac{T}{v_l+a} + \frac{f^6}{(v_g-g)^6} - \frac{f^6}{(v_l-g)^6}}{(v_g - v_l)\sqrt{c}\left(\frac{1}{v_l^2+c} - \frac{1}{v_g^2+c}\right)} \right) \left(\arctan\left(\frac{v_g}{\sqrt{c}}\right) - \arctan\left(\frac{v_l}{\sqrt{c}}\right) \right), \end{aligned} \tag{4.91}$$

and,

$$
\begin{aligned}
d_2 = & \frac{1}{v_g - v_l}\left(\frac{\frac{1}{v_g^3+e} - \frac{1}{v_l^3+e}}{\sqrt{c}\left(\frac{1}{v_l^2+c} - \frac{1}{v_g^2+c}\right)}\right)\left(\arctan\left(\frac{v_g}{\sqrt{c}}\right) - \arctan\left(\frac{v_l}{\sqrt{c}}\right)\right) \\
& - \frac{1}{6e^{\frac{2}{3}}(v_g - v_l)}\left(2\sqrt{3}\left(\arctan\left(\frac{1 - \frac{2v_g}{e^{\frac{1}{3}}}}{\sqrt{3}}\right) - \arctan\left(\frac{1 - \frac{2v_l}{e^{\frac{1}{3}}}}{\sqrt{3}}\right)\right)\right) \\
& - \frac{1}{6e^{\frac{2}{3}}(v_g - v_l)}\left(\log\left(\frac{e^{\frac{2}{3}} - e^{\frac{1}{3}}v_g + v_g^2}{e^{\frac{2}{3}} - e^{\frac{1}{3}}v_l + v_l^2}\right) + 2\log\left(\frac{v_l + e^{\frac{1}{3}}}{v_g + e^{\frac{1}{3}}}\right)\right) \\
& - \left(\frac{\frac{1}{v_g^3+e} - \frac{1}{v_l^3+e}}{(v_l^2 + c)\left(\frac{1}{v_l^2+c} - \frac{1}{v_g^2+c}\right)} + \frac{1}{v_l^3 + e}\right).
\end{aligned}
\tag{4.92}
$$

We can then substitute d^* back into (4.89) giving,

$$
b^* = \frac{\frac{T}{\log(v_g+a)} - \frac{T}{\log(v_l+a)} + d^*\left(\frac{1}{v_g^3+e} - \frac{1}{v_l^3+e}\right) + \frac{f^6}{(v_g-g)^6} - \frac{f^6}{(v_l-g)^6}}{\frac{1}{v_l^2+c} - \frac{1}{v_g^2+c}}. \tag{4.93}
$$

We then replace b and d in the Demetriades EoS with b^* and d^* which are defined entirely in terms of the other parameters. As such our vector of EoS parameters is reduced to,

$$
\boldsymbol{\theta} = (a, c, e, f, g)^T, \tag{4.94}
$$

and the full vector of parameters is reduced to

$$
\tilde{\boldsymbol{\theta}} = (a, c, e, f, g, v_l, v_g)^T. \tag{4.95}
$$

Then from (3.33) the likelihood of our statistical model in the hard constraint case with the synthetic data described above is,

$$L(\tilde{\boldsymbol{\theta}}) = \frac{1}{\sqrt{6\pi\sigma_s^2}} \exp\left\{-\frac{1}{2\sigma^2}(P_s^{\text{obs}} - h_D(v_l, T, \boldsymbol{\theta}))^2\right\} \prod_{i=1}^{n} \phi(P_i^{\text{sim}}; h_D(v_i^{\text{sim}}, T, \boldsymbol{\theta}), \sigma^2). \quad (4.96)$$

To completely define the posterior we need to define $p(v_l^{\text{obs}}, v_g^{\text{obs}}|v_l, v_g)$ which from (3.54) is,

$$p(v_l^{\text{obs}}, v_g^{\text{obs}}|v_l, v_g) = \phi(v_l^{\text{obs}}; v_l, \sigma_v^2)\phi(v_g^{\text{obs}}; v_g, \sigma_v^2), \quad (4.97)$$

and the prior distributions on the parameters. We use independent prior distributions for each parameter. We use the same prior distributions on the five EoS parameters as were previously,

$$a \sim \mathcal{U}(0.1, 1), \quad (4.98)$$

$$c \sim \mathcal{U}(-0.01, 1), \quad (4.99)$$

$$e \sim \mathcal{U}(0, 1), \quad (4.100)$$

$$f \sim \mathcal{U}(0, 0.15), \quad (4.101)$$

$$g \sim \mathcal{U}(0, 0.1). \quad (4.102)$$

For the prior distributions on the coexistence volumes we use

$$v_l \sim \mathcal{U}(0.1, v_c), \quad (4.103)$$

$$v_g \sim \mathcal{U}(v_c, 10), \quad (4.104)$$

where v_c is the non-dimensionalised critical volume given by Dusschek et al. [1990a]. We use v_c to partition the prior to ensure that v_l is less than v_g and prevent singularities from arising from the expression of the pressure

derived from fugacity. Similarly the lower bound on v_l ensures that there is no singularity arising from the term involving parameter g in the model. The upper bound on v_g was chosen to be much larger than any reasonable possible values for a coexistence volume for temperatures in our window of interest. Other choices for upper bound were explored and found to have no effect on the results which follow. Another benefit of the prior is that it reduces the number of points which trivially satisfy the coexistence constraints terms in the log-likelihood (3.49). These constraints are trivially satisfied when $v_l = v_g$. By partitioning the prior at v_c we ensure that there is only one point at which this condition is satisfied. Otherwise allowing v_l and v_g to be equal for any physically reasonable value leads to a ridge in the likelihood surface which we have found it is difficult for the sampler to escape from. Given priors the posterior up to proportionality is

$$p(\tilde{\boldsymbol{\theta}}|P_s^{\text{obs}}, \boldsymbol{P}^{\text{sim}}, v_l^{\text{obs}}, v_g^{\text{obs}}) \propto L(\tilde{\boldsymbol{\theta}})p(v_l^{\text{obs}}, v_g^{\text{obs}}|v_l, v_g)p(\tilde{\boldsymbol{\theta}}). \tag{4.105}$$

Given this posterior distribution defined up to proportionality we now construct an MCMC sampler which can sample from this distribution.

Application of PT-HMC to the Hard Constraint Case

We once again use parallel tempering with Hamiltonian Monte Carlo (PT-HMC) to sample from this posterior. To do so we firstly need to define a system of replicas from which we are sampling. We choose to temper $L(\tilde{\boldsymbol{\theta}})$ and $p(v_l^{\text{obs}}, v_g^{\text{obs}}|v_l, v_g)$ in (4.105) as these are the terms in the posterior which we expect to introduce any multi-modality. Then replica k has probability distribution up to proportionality,

$$p_k(\tilde{\boldsymbol{\theta}}|P_s^{\text{obs}}, \boldsymbol{P}^{\text{sim}}, v_l^{\text{obs}}, v_g^{\text{obs}}) \propto \Big(L(\tilde{\boldsymbol{\theta}})p(v_l^{\text{obs}}, v_g^{\text{obs}}|v_l, v_g)\Big)^{\frac{1}{\beta_k}} p(\tilde{\boldsymbol{\theta}}), \tag{4.106}$$

where β_k is the tempering level for replica k. Then the replicas with highest tempering level will sample from the prior,

$$p(\tilde{\boldsymbol{\theta}}) = p(a)p(c)p(e)p(f)p(g)p(v_l)p(v_g), \tag{4.107}$$

and so explore the whole parameter space allowing the sampler to escape local modes. Then the acceptance probability for swaps between replicas m and n is $\min(1, \exp(-\Delta))$, where

$$\Delta = \left(\frac{1}{\beta_n} - \frac{1}{\beta_m}\right)\Bigg(\log\left(L(\tilde{\boldsymbol{\theta}}_n)p(v_l^{\text{obs}}, v_g^{\text{obs}}|v_{ln}, v_{gn})\right) - \log\left(L(\tilde{\boldsymbol{\theta}}_m)p(v_l^{\text{obs}}, v_g^{\text{obs}}|v_{lm}, v_{gm})\right)\Bigg), \tag{4.108}$$

where $\tilde{\boldsymbol{\theta}}_m$ are the current parameters for replica m and v_{lm} and v_{gm} are the current coexistence volume parameters for replica m. $\tilde{\boldsymbol{\theta}}_n$, v_{ln} and v_{gn} are defined likewise for replica n.

To choose the highest tempering level we once more use the no-U-turn sampler applied 50 times, running for 1000 iterations, using different initial values for the parameters sampled from the prior distribution, to sample from the posterior distribution of replicas of our model with increasing β. We find that a maximum tempering level of $\beta = 2^{30}$ ensures the replicas with the highest tempering levels are sampling from distributions close to the prior and thus exploring the whole parameter space. We once again use a geometric series of tempering levels with $\beta_1 = 1$, which corresponds to the posterior distribution of interest, and

$$\beta_k = 2\beta_{k-1}, \quad k = 2, \ldots, 31, \tag{4.109}$$

so that each tempering level is twice the previous tempering level until we have reached our maximum tempering level. We thus have 31 total replicas to sample from. The tempering levels are closer together than previously

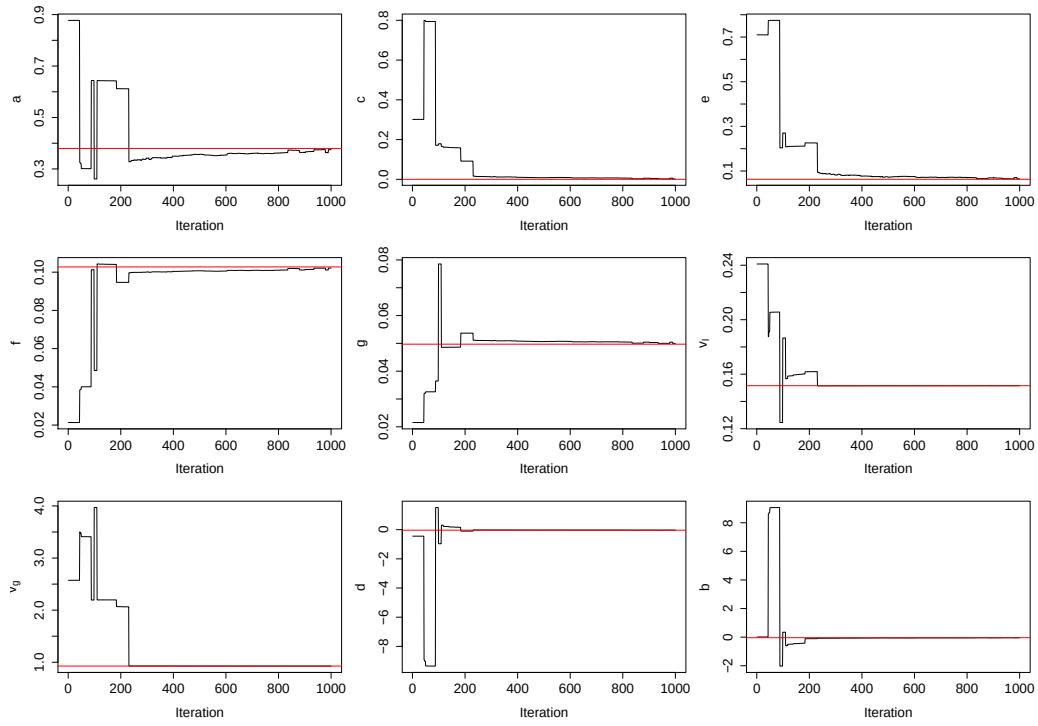


Figure 4.10: Traceplot of the initial 1000 iterations for a single run of the PT-HMC sampler applied to the single temperature Demetriades EoS with synthetic data including coexistence and the hard constraint. Red lines mark the true parameter values. Here d and b are derived from the other parameters, not sampled directly. Traceplots for other runs are similar. The sampler consistently finds it way from its initial position to the vicinity of the true parameter values.

to increase the number of swaps. Applying the previous case where each tempering level was 10 times the previous tempering level led to too few swaps occurring as the distributions of the replicas were too far apart. We again propose swaps between neighbouring replicas after 10 iterations of the no-U-turn sampler to allow each replica to make more significant moves before swaps are proposed again.

We have then run this PT-HMC sampler from 10 different starting locations sampled from the prior for 110000 iterations. This number was chosen as initial runs of 10000 iterations found that while the sampler reached the mode containing the true parameter values convergence within the mode was somewhat poor. These initial runs had not fully explored the mode containing the true parameter values to the extent that the true param-

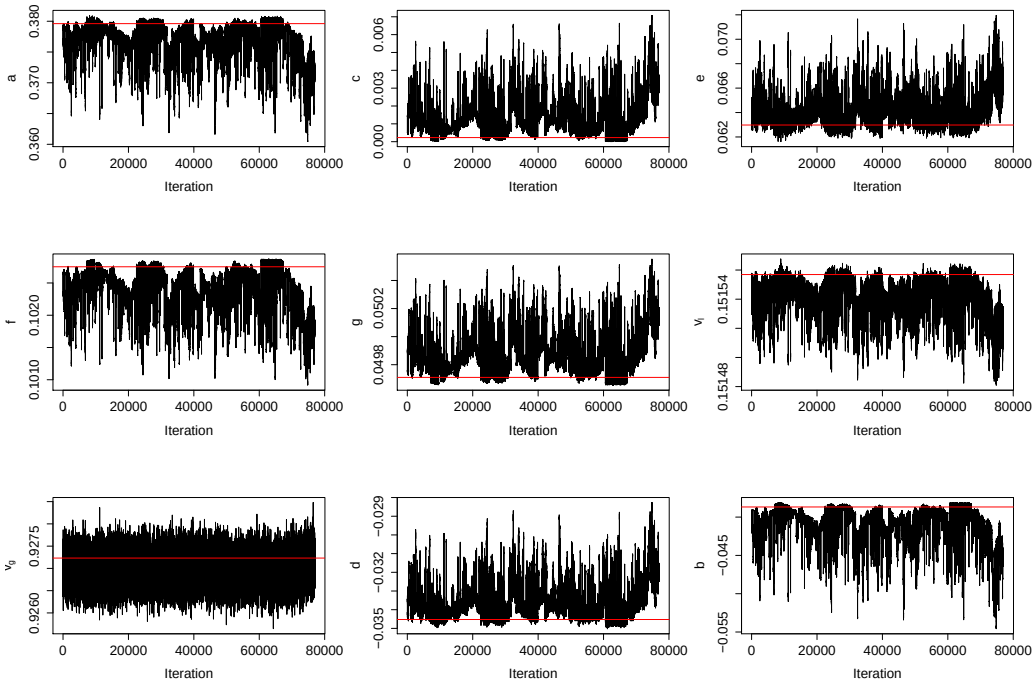


Figure 4.11: Traceplot of the sampled parameters for a single run of the PT-HMC sampler applied to the single temperature Demetriades EoS with synthetic data including coexistence and the hard constraint with first 33000 iterations discarded as burn-in. Here d and b are derived from the other parameters, not sampled directly. Red lines mark the true parameter values. Traceplots for other runs are similar. We see that while neighbouring iterations are still somewhat correlated our sampler is mixing reasonably well about the mode containing the true parameter values. Given this and the fact that our sampler finds this mode consistently we can say that it has converged to the correct posterior distribution.

eter values were not within range covered by the samples after reaching the mode. As such the sampler was run for an additional 100000 iterations, starting from the final parameter values for the initial runs of 10000 iterations, to ensure that enough samples were obtained to accurately characterise the posterior distributions of the parameters.

Each run of the sampler converged to the same mode which contained the true parameter values. Figure 4.10 plots the traces for initial 1000 iterations of the bottom replica ($\beta = 1$), which is our posterior distribution of interest, showing how the sampler moves into the region around the true parameter values. Included in the plot are traces for d and b which are

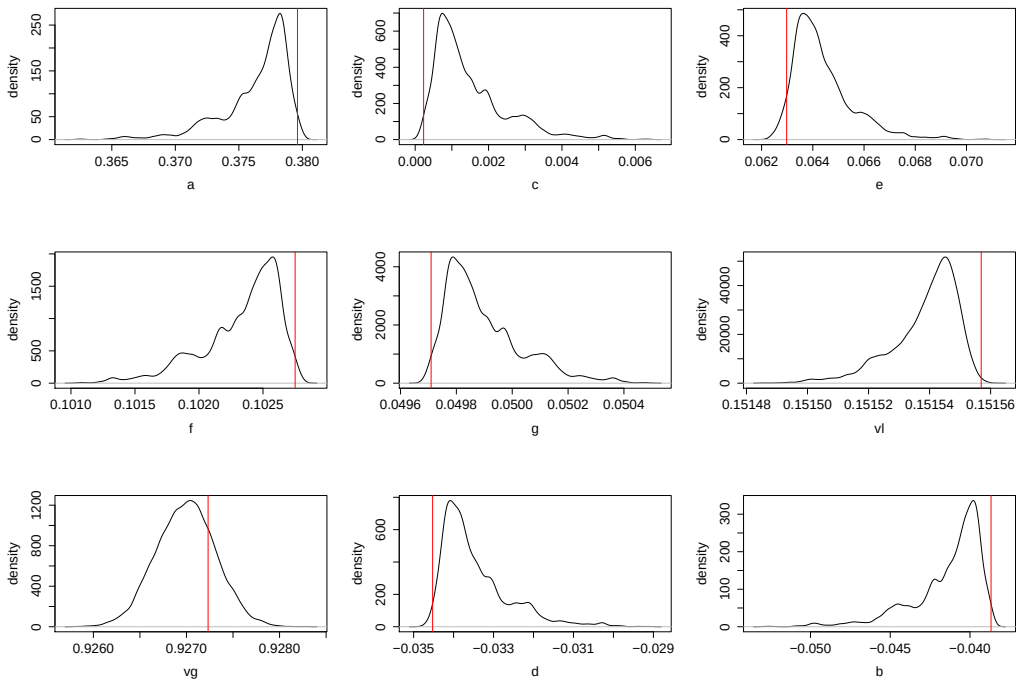


Figure 4.12: Marginal Densities of the sampled parameters for a single run of the PT-HMC sampler applied to the single temperature Demetriades EoS with synthetic data including coexistence and the hard constraint. Here d and b are derived from the other parameters, not sampled directly. Red lines mark the true parameter values.

calculated from the other parameters in order to compare them to the true parameter values used to simulate the data. Although the sampler quickly moves to the region around the true parameter values it takes some time for the sampler to fully explore the entirety of the mode. Figure 4.11 plots the traces after discarding the first 33000 iterations (first 30% of the iterations) as burn-in to remove the effect of the initial starting location and the time taken to reach equilibrium within the mode. From this we can see that our sampler is mixing reasonably well within the mode. The fact that our sampler is mixing and finds the same mode, the mode which contains the true parameter values, when started from different initial conditions allows us to conclude our sampler is converging correctly to the distribution of interest.

Figure 4.12 plots the marginal densities of the parameters from this

sample after burn-in. As can be seen in figure 4.12 for each of these parameters the region of high posterior density is much smaller than the prior range and contains the true parameter values used to simulate the synthetic data, even if they are found in the tails of the region of high posterior density. We have here a sampler which can recover the true parameter values for the single temperature Demetriades EoS model with hard constraint when applied to synthetic data.

However we have seen how the hard constraint removes two parameters from our model. With every additional temperature for which we have coexistence data applying the hard constraint would remove a further two parameters. As with every additional temperature for which we have coexistence data we would add a further two parameters for the coexistence volume at that temperature we do not face the issue of running out of parameters. However it is not possible to derive simple forms for other parameters as was done for b and d in this case. As such we now explore fitting the alternative soft constraint model to our synthetic data as that does not lead to a reduction of the parameter space.

4.4.2 Soft Constraint

The soft constraint model was described in chapter 3. We assume that there is some normally distributed i.i.d. model error with standard deviation σ_e^2 on our data, this leads to the constraints being included as additional terms in the likelihood rather than reducing the size of the parameter space. As such our EoS parameter vector is once more 7-dimensional,

$$\boldsymbol{\theta} = (a, b, c, d, e, f, g)^T, \tag{4.110}$$

and the full parameter vector including the true coexistence volumes is,

$$\tilde{\boldsymbol{\theta}} = (a, b, c, d, e, f, g, v_l, v_g)^T. \tag{4.111}$$

So there are 9 parameters in total to sample. From (3.51) with our synthetic data the resulting likelihood for the soft constraint Demetriades EoS model is,

$$\begin{aligned} L(\tilde{\boldsymbol{\theta}}) = \phi\left(P_s^{\text{obs}}; \frac{H_D(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}, \sigma^2\right) \\ \phi\left(0; h_D(v_l, T, \boldsymbol{\theta}) - \frac{H_D(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}, \sigma_e^2\right) \\ \phi\left(0; h_D(v_g, T, \boldsymbol{\theta}) - \frac{H_D(v_l, v_g, T, \boldsymbol{\theta})}{v_g - v_l}, \sigma_e^2\right) \\ \prod_{i=1}^{n} \phi(P_i^{\text{sim}}; h_D(v_i^{\text{sim}}, T, \boldsymbol{\theta}), \sigma^2). \end{aligned} \tag{4.112}$$

To completely define the posterior up to a normalising constant we need $p(v_l^{\text{obs}}, v_g^{\text{obs}} | v_l, v_g)$ which is defined as before (4.97) and priors over the parameters. We use the same independent priors on the parameters as in the hard constraint case (4.98)-(4.104), with the additional following priors on b and d,

$$b \sim \mathcal{U}(-1, 1), \tag{4.113}$$

$$d \sim \mathcal{U}(-1, 1). \tag{4.114}$$

Then the posterior up to a normalising constant is,

$$p(\tilde{\boldsymbol{\theta}} | P_s^{\text{obs}}, \boldsymbol{P}^{\text{sim}}, v_l^{\text{obs}}, v_g^{\text{obs}}) \propto L(\tilde{\boldsymbol{\theta}}) p(v_l^{\text{obs}}, v_g^{\text{obs}} | v_l, v_g) p(\tilde{\boldsymbol{\theta}}). \tag{4.115}$$

The only additional modelling choice to be made is the value at which to fit σ_e. Here we do this by running the no-U-turn sampler starting at the true parameter values for both the hard constraint model and the soft

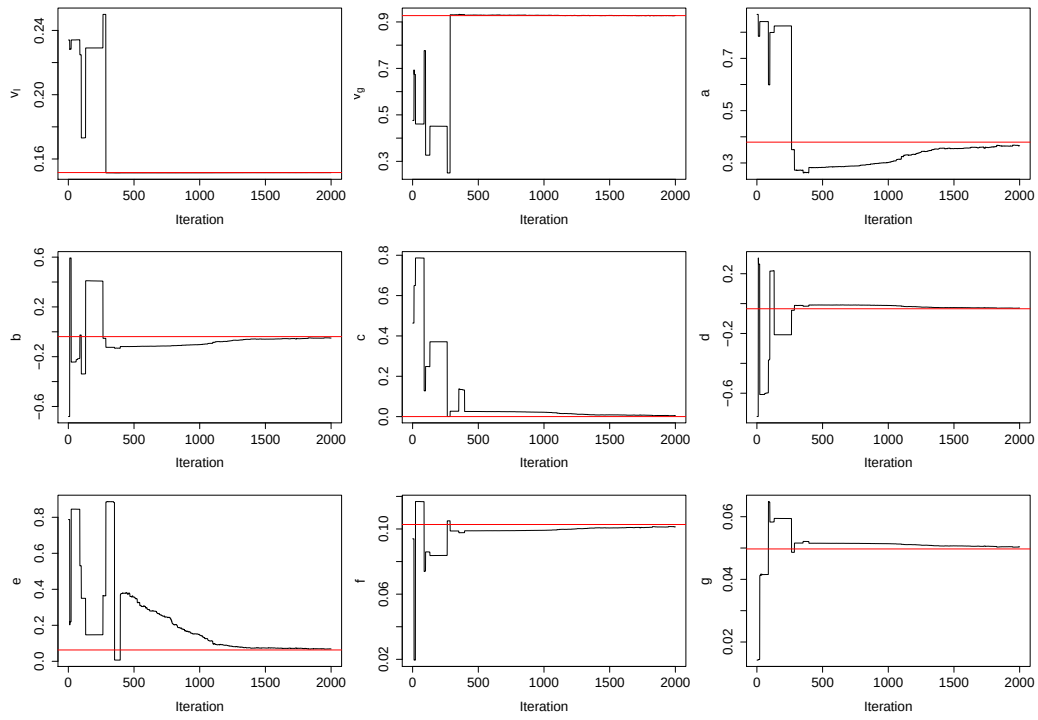


Figure 4.13: Traceplot of the initial 2000 iterations for a single run of the PT-HMC sampler applied to the single temperature Demetriades EoS with synthetic data including coexistence and the soft constraint. Red lines mark the true parameter values. Traceplots for other runs are similar. The sampler has found its way into the vicinity of the true parameters.

constraint model with various values of σ_e to determine what value for σ_e leads to the soft constraint case being a good approximation to the hard constraint case. We find that $\sigma_e = 10^{-5}$ so 10 times smaller than σ leads to the soft constraint sampler run starting at the true parameters being within sampling variation of the hard constraint sampler run starting at the true values. As such we fix $\sigma_e = 10^{-5}$. This σ_e being smaller than σ or σ_v means that we prioritise parameter values for which the constraints hold. Having defined our soft constraint model for the synthetic data fully we now construct an MCMC sampler to sample from this model.

Application of PT-HMC to the Soft Constraint Case

We again use a PT-HMC sampler in order to overcome multi-modality in the model. We temper as in the hard constraint case, tempering the likeli-

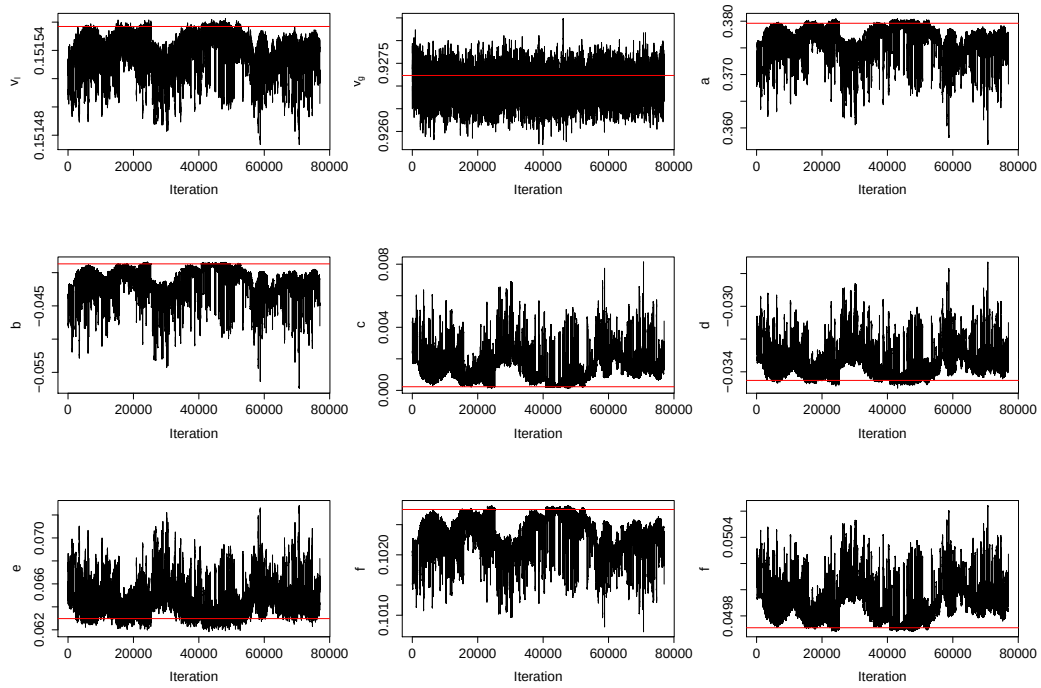


Figure 4.14: Traceplot of the sampled parameters for a single run of the PT-HMC sampler applied to the single temperature Demetriades EoS with synthetic data including coexistence and the soft constraint after discarding first 33000 iterations as burn-in. Red lines mark the true parameter values. Traceplots for other runs are similar. The sampler has found its way into the mode about the true parameters and is mixing about the mode.

hood and the errors-in-variables term on the coexistence volumes. We also use the same tempering levels as in the hard constraint case finding that the top replicas once again sample from the prior and a good rate of swaps occurs. We run the sampler for 110000 iterations, for consistency with the hard constraint case, from 10 initial starting values sampled randomly from the prior and find that the individual runs of the sampler consistently converge to the same mode which includes the true parameter values. Figure 4.13 shows the traceplots for the first 2000 iterations of one of these runs showing that the sampler has found its way from the initial position to the region around the true parameters. We discard the first 33000 iterations as burn-in to remove the effect of the initial starting location and time taken for the sampler to reach equilibrium. Figure 4.14 plots the traces after burn-in. From this we can see that our sampler is mixing if some-

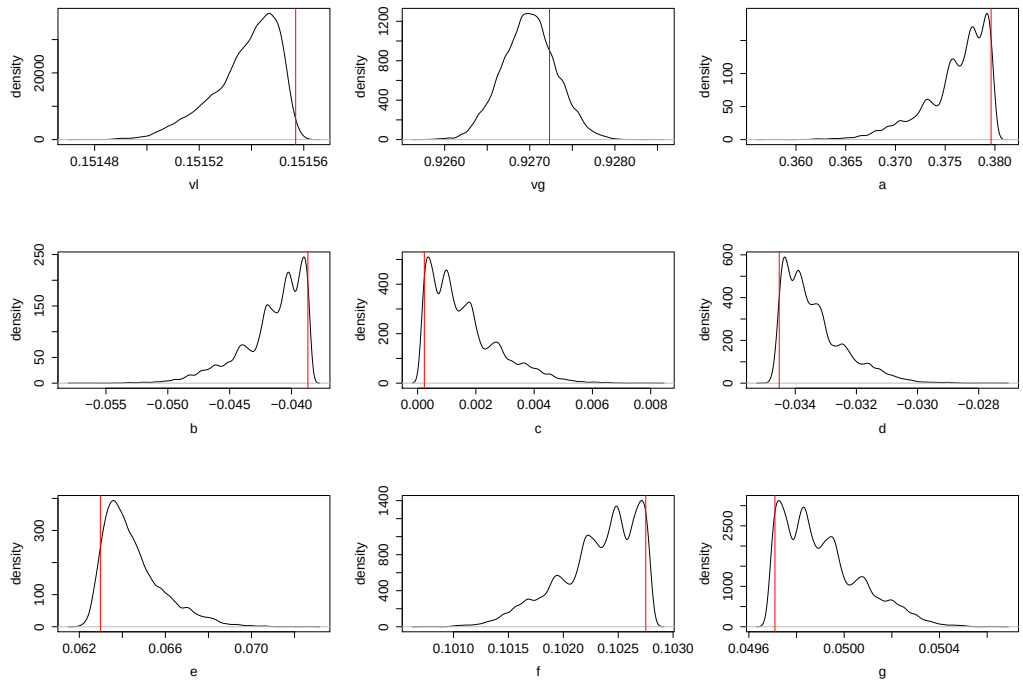


Figure 4.15: Marginal Densities of the sampled parameters for a single run of the PT-HMC sampler applied to the single temperature Demetriades EoS with synthetic data including coexistence and the soft constraint. Red lines mark the true parameter values.

what poorly. The fact that our sampler is mixing and consistently finds the mode about the true parameter values when started from different initial conditions leads us to conclude our sampler is converging correctly to the distribution of interest. Figure 4.15 plots the marginal densities of the parameters from this sample after burn-in. Once again as can be seen in figure 4.15 each of these parameters region of high posterior density is much smaller than the prior range and contains the true parameter values used to simulate the synthetic data. It is also noted that there a number of peaks in the densities for some of the parameters. These are most likely artefacts as a result of the finite run of the sampler and would be smoothed out in a longer run. We have found here that our PT-HMC sampler which can recover the true parameter values for the single temperature Demetriades EoS model with soft and hard constraints when applied to synthetic data. We now apply this sampler to synthetic data for multiple temperatures

including coexistence to test how the sampler copes with the expanded parameter space and temperature dependence.

4.5 Multiple Temperatures Synthetic Data with Coexistence Data

We first introduce data including coexistence for multiple temperatures. We once more simulate data at two temperatures, $T = 0.9206644$ and $T = 0.9634095$ in our non-dimensionalised units, as in the previous case of using synthetic data for multiple temperatures. We then set $\boldsymbol{T}_c = (0.9206644, 0.9634095)^T$. We again use linear temperature rules (1.43) for the EoS parameters with the minimum and maximum temperature of interest set as before,

$$T_l = 0.8976478, \quad T_h = 1. \tag{4.116}$$

We use the same true parameter values as previously (4.48)-(4.54). To find the values of the coexistence points for the DEoS at the two temperatures at which we are simulating data with these parameter values we use a root solving method approach to solve the system of three non-linear equations in three unknowns (4.75)-(4.77) at both temperatures for which we are simulating data. We again use the Newton-Raphson method [Ryaben'kii and Tsynkov, 2006]. Using the Newton-Raphson method the true coexistence

volumes and coexistence pressure are found to be,

$$v_{l1} = 0.1452966, \tag{4.117}$$

$$v_{g1} = 1.0848770, \tag{4.118}$$

$$P_{s1} = 0.5603271, \tag{4.119}$$

$$v_{l2} = 0.1656542, \tag{4.120}$$

$$v_{g2} = 0.6979179, \tag{4.121}$$

$$P_{s2} = 0.7642022, \tag{4.122}$$

where v_{l1}, v_{g1}, P_{s1} are the coexistence data for $T = 0.9206644$ and v_{l2}, v_{g2}, P_{s2} are similarly for $T = 0.9634095$. We introduce error into the coexistence volumes as follows,

$$v_{li}^{\text{obs}} = v_{li} + \epsilon_{li}, \quad i = 1, 2, \tag{4.123}$$

$$v_{gi}^{\text{obs}} = v_{gi} + \epsilon_{gi}, \quad i = 1, 2, \tag{4.124}$$

where the ϵ_{li} and ϵ_{gi} are iid normally distributed r.v. with zero mean and standard deviation $\sigma_v = 10^{-3}$. Similarly error is introduced into the coexistence pressure as follows,

$$P_{si}^{\text{obs}} = P_{si} + \epsilon_{si}, \quad i = 1, 2, \tag{4.125}$$

where the ϵ_{si} are normally distributed r.v. with zero mean and standard deviation $\sigma = 10^{-4}$

We now create some synthetic, homogeneous data. Previously we used 0.105 and 2.24 as the bounds for our synthetic volumes at each temperature. We use these as bounds again here. However when considering coexistence it is once more important to note that no data can physically exist in the interval (v_l, v_g). As such we choose the synthetic volumes in our data

set to be whichever of the 100 evenly spaced points from 0.105 to 2.24 which are not contained in the interval (v_l, v_g). The resulting vector of synthetic volumes $\boldsymbol{v}^{\text{sim}}$ is of size 131 with 56 homogeneous data points at $T = 0.2206644$ and 75 homogeneous data points at $T = 0.9634085$. We collect the temperatures which correspond to these volumes in the vector $\boldsymbol{T}^{\text{sim}}$. Then our synthetic pressure values are,

$$P_i^{\text{sim}} = h_D(v_i^{\text{sim}}, T_i^{\text{sim}}, \boldsymbol{\theta}_{\text{TRUE}}) + \epsilon_i, \quad i = 1, \ldots, 131, \tag{4.126}$$

where the ϵ_i are iid normal r.v. with zero mean and standard deviation $\sigma = 10^{-4}$. This is the same noise as was used on the coexistence pressure value. We have 14 EoS parameters,

$$\boldsymbol{\theta} = (a_l, a_h, b_l, b_h, c_l, c_h, d_l, d_h, e_l, e_h, f_l, f_h, g_l, g_h)^T, \tag{4.127}$$

and 18 total parameters including the coexistence volumes which we collect in the vector,

$$\tilde{\boldsymbol{\theta}} = (a_l, a_h, b_l, b_h, c_l, c_h, d_l, d_h, e_l, e_h, f_l, f_h, g_l, g_h, v_{l1}, v_{g1}, v_{l2}, v_{g2})^T. \tag{4.128}$$

As was described in the previous section it is difficult to apply the hard constraint model to multiple temperatures due to difficulty eliminating parameters using the constraints. As such we apply only the soft constraint

model here. From (3.51) the likelihood is then,

$$L(\tilde{\boldsymbol{\theta}}) = \prod_{i=1}^{2} \phi\left(P_{si}^{\text{obs}}; \frac{H_D(v_{li}, v_{gi}, T_{ci}, \boldsymbol{\theta})}{v_{gi} - v_{li}}, \sigma^2\right) \phi\left(0; h_D(v_{li}, T_{ci}, \boldsymbol{\theta}) - \frac{H_D(v_{li}, v_{gi}, T_{ci}, \boldsymbol{\theta})}{v_{gi} - v_{li}}, \sigma_e^2\right) \phi\left(0; h_D(v_{gi}, T_{ci}, \boldsymbol{\theta}) - \frac{H_D(v_{li}, v_{gi}, T_{ci}, \boldsymbol{\theta})}{v_{gi} - v_{li}}, \sigma_e^2\right) \prod_{j=1}^{n} \phi(P_j^{\text{sim}}; h(v_j^{\text{sim}}, T_j^{\text{sim}}, \boldsymbol{\theta}), \sigma^2). \quad (4.129)$$

To fully define the posterior we then require the errors-in-variables term for the coexistence volumes (3.54) which here is,

$$p(\boldsymbol{v}_l^{\text{obs}}, \boldsymbol{v}_g^{\text{obs}} | \boldsymbol{v}_l, \boldsymbol{v}_g) = \prod_{i=1}^{2} \phi(v_{li}^{\text{obs}}; v_{li}, \sigma_v^2) \phi(v_{gi}^{\text{obs}}; v_{gi}, \sigma_v^2), \quad (4.130)$$

and priors on all the parameters. We use the same independent uniform priors for all parameters as were used before and use $p(\tilde{\boldsymbol{\theta}})$ to denote the resulting joint distribution. Then the posterior we wish to sample from is

$$p(\tilde{\boldsymbol{\theta}} | \boldsymbol{P}_s^{\text{obs}}, \boldsymbol{P}^{\text{sim}}, \boldsymbol{v}_l^{\text{obs}}, \boldsymbol{v}_g^{\text{obs}}) \propto L(\tilde{\boldsymbol{\theta}}) p(\boldsymbol{v}_l^{\text{obs}}, \boldsymbol{v}_g^{\text{obs}} | \boldsymbol{v}_l, \boldsymbol{v}_g) p(\tilde{\boldsymbol{\theta}}). \quad (4.131)$$

up to a normalising constant. Having defined the posterior we now describe the MCMC sampler applied to it.

4.5.1 Application of PT-HMC

We once more use PT-HMC as our MCMC sampler of choice due to its successes overcoming multi-modality in the homogeneous case with multiple and single temperatures and the coexistence case with a single temperature. As before we temper the likelihood (4.129) and errors-in-variables (4.130) terms in the posterior so that the highest replicas sample from a distribution that is close to the prior, $\tilde{\boldsymbol{\theta}}$ and so explore the whole param-

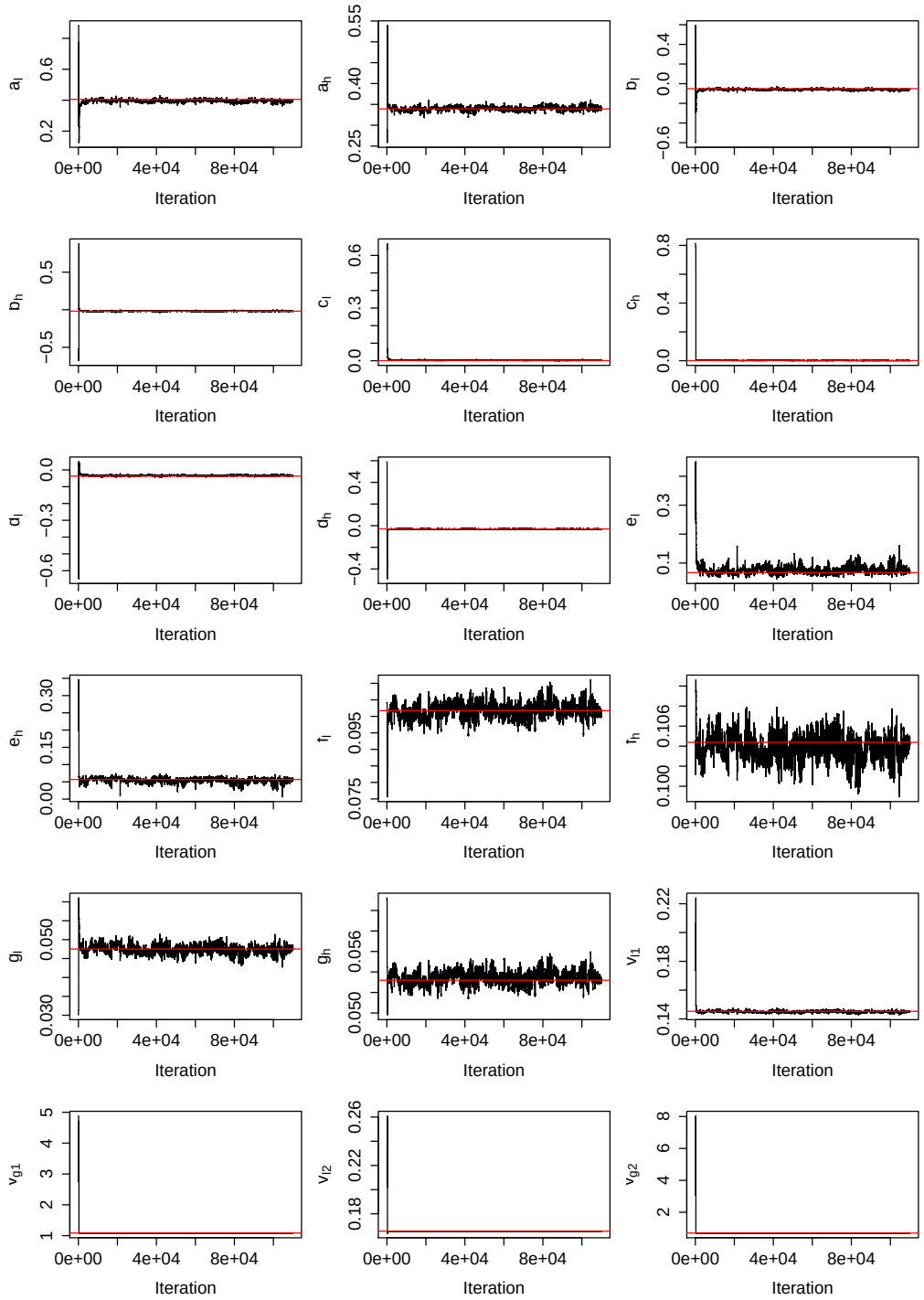


Figure 4.16: Traceplot of the sampled parameters for a single run of the PT-HMC sampler applied to the multiple temperature Demetriades EoS with linear temperature rules and synthetic data including coexistence and the soft constraint. Red lines mark the true parameter values. Traceplots for other runs are similar. The sampler consistently finds its way into the mode about the true parameter values and so appears to have converged.

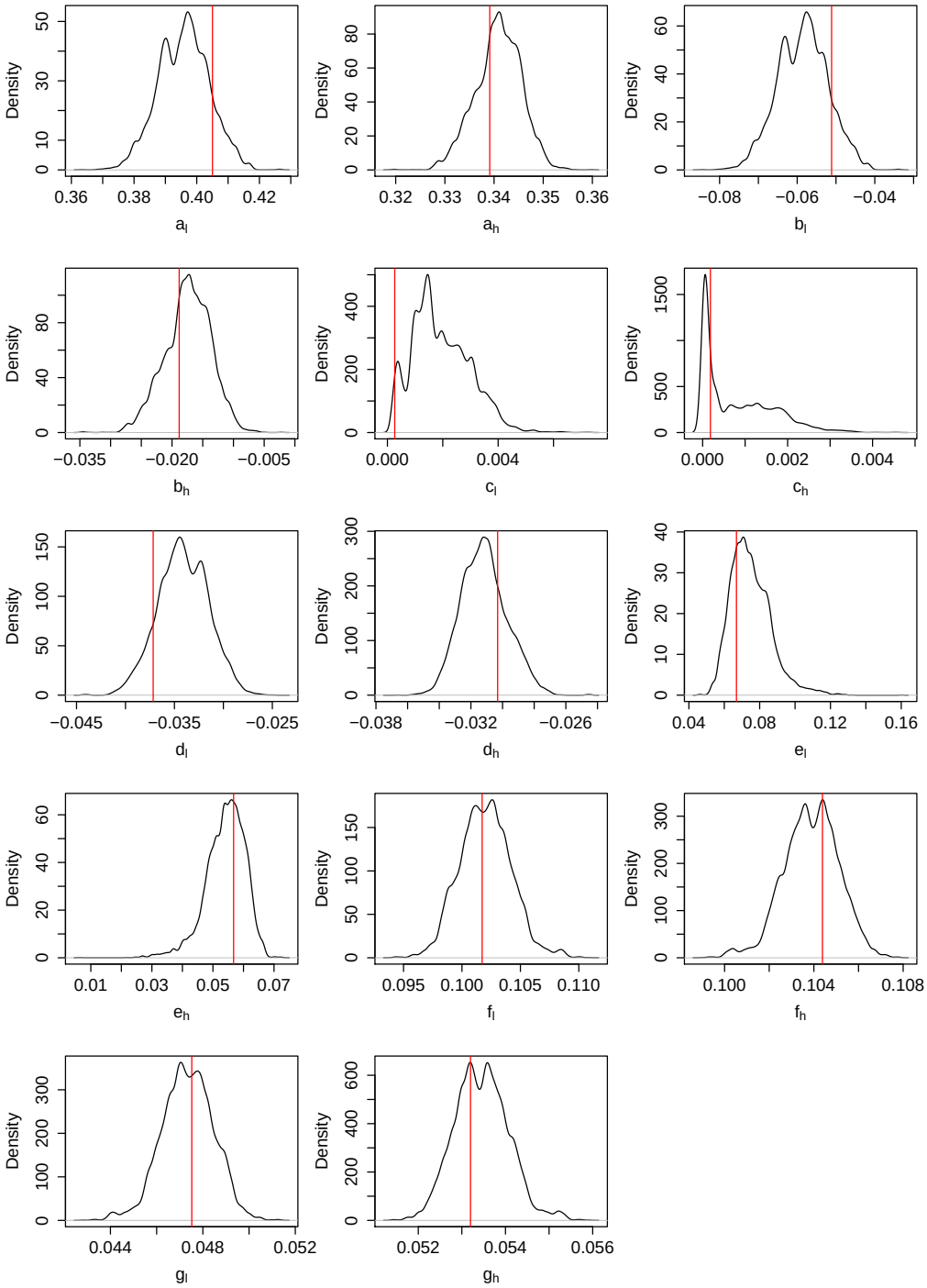


Figure 4.17: Marginal Densities of the sampled DEoS parameters for a single run of the PT-HMC sampler applied to the multiple temperature Demetriades EoS linear temperature rules and synthetic data including coexistence and the soft constraint. Red lines mark the true parameter values.

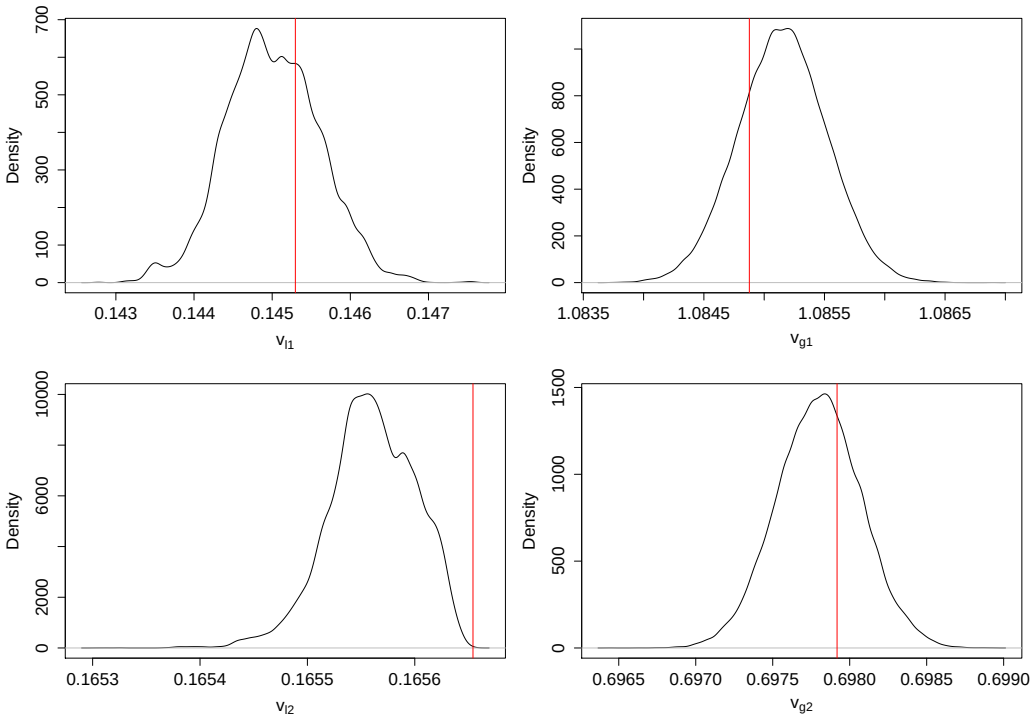


Figure 4.18: Marginal Densities of the sampled coexistence volume parameters for a single run of the PT-HMC sampler applied to the multiple temperature Demetriades EoS linear temperature rules and synthetic data including coexistence and the soft constraint. Red lines mark the true parameter values.

eter space so as best to escape local modes. We find that the same set of 31 tempering levels as was used with the single temperature synthetic data with coexistence works well here. The highest replicas are close to the prior and explore the whole of the parameter space and there are a good rate of swaps between neighbouring replicas. The sampler was run for 110000 iterations (for consistency with the single temperature runs) from 10 starting locations randomly sampled from the prior. Running in parallel on 16 CPUs on the University of Nottingham's high performance computing facility this takes approximately 2.5 days. Each run of the sampler converged to the same mode which includes the true parameter values. Figure 4.16 plots the traces for each parameter for one such run of the sampler. Due to the scale we can see that the f and g parameters are mixing well. We have examined the traceplots after discarding burn-in and from them

find that the other parameters mix similarly well. Figures 4.17 and 4.18 plot the marginal densities, of the DEoS parameters and the coexistence volume parameters respectively, from this run of the sampler after the first 33000 iterations were discarded as burn-in. As can be seen the high posterior density region is once again much smaller than the prior range and each of the true parameter values is in the high posterior density region. We note that the densities for c_l and c_h are least smooth. This is because the high posterior density region hits the lower bound on these parameters and this causes the no-U-turn sampler to struggle to adapt and so it mixes somewhat slowly for these parameters. This lower bound is necessary to avoid singularities in the model and so cannot be extended downwards but longer runs of the MCMC sampler and thinning result in a more accurate prediction of the posterior marginal distributions.

In summary we have applied the PT-HMC sampler to synthetic data for both single and multiple temperatures and found that we are able to consistently recover the true parameter values. Having confirmed that our sampler functions well with synthetic data we now apply to the case of real pure CO_2 data.

4.6 Application of the PT-HMC Sampler To Real Pure CO_2 Data

4.6.1 Fitting Individual Temperatures

We now apply the PT-HMC sampler to the pure CO_2 data that was described in §1.6. To fit the DEoS model to this data fully we need to propose temperature rules for the model parameters. It is not intuitively obvious what rules to propose. Thus we firstly fit several temperatures individually. This will enable us to see how the parameters change over these

temperatures which we can then use to propose temperature rules for the parameters.

In §1.2 we described the CCS transport window which determines the range of temperature values which we would need to model for CCS applications. There are 6 temperatures in our range of interest for which there are both homogeneous and coexistence data. For one of these temperatures, 270.0007K (in dimensional units), which marks the lowest temperature in the CCS transport window, there are only 2 homogeneous data points, not enough to make any inference about the parameter values. Thus we extend the temperature range which we consider down to 260K, where there is both homogeneous and coexistence data, in order to avoid having to extrapolate our fit to lower temperatures. Thus we fit the following six temperatures individually, 260.003K, 279.9984K, 289.9959K, 296.994K, 299.9933K and 302.9925K. We will denote these temperatures in non-dimensionalised units as T_1 through T_6 respectively. We do not fit any supercritical temperatures independently as we expect that for temperatures above the critical point the data has less features and so there will be a larger region of the parameter space which can fit the data and so we will not be able to learn as much about the parameters. Based on the stated experimental uncertainty in Dusschek et al. [1990a], Dusschek et al. [1990b] and Gilgen et al. [1992], we fix $\sigma = 0.0002$, and $\sigma_v = 0.001$. Based on our results in the synthetic data case we set $\sigma_e = \frac{\sigma}{10} = 0.00002$.

As was described in §1.6 for a given temperature with coexistence data there are sometimes multiple measurements of the coexistence data P_s, v_l, and v_g. For temperature T_i let n_{pi}, n_{li} and n_{gi}, denote the number of observations of P_s, v_l, and v_g at that temperature respectively. Then in §3.5 we showed how assuming we fix σ and σ_v we can equivalently consider the means of these observations rather than the individual observations themselves. Let $\bar{P}_{si}^{\text{obs}}, \bar{v}_{li}^{\text{obs}}$, and $\bar{v}_{gi}^{\text{obs}}$, be the mean of the observations of

P_s, v_l, and v_g, respectively for temperature T_i. Then these means have variances $\frac{\sigma}{n_{si}}$, $\frac{\sigma_v}{n_{li}}$, and $\frac{\sigma_v}{n_{gi}}$, respectively. Let $\boldsymbol{P}_i$ and $\boldsymbol{v}_i$ be the vectors of the homogeneous pressure and volume observations for temperature T_i where,

$$\boldsymbol{P}_i = (P_{i1}, P_{i2}, \ldots, P_{in_i})^T, \tag{4.132}$$

$$\boldsymbol{v}_i = (v_{i1}, v_{i2}, \ldots, v_{in_i})^T, \tag{4.133}$$

and n_i is the number of homogeneous observations for temperature T_i. Then we can write the likelihood for temperature T_i as follows,

$$\begin{aligned} L(\tilde{\boldsymbol{\theta}}) = \phi\left(\bar{P}_{si}; \frac{H_D(v_l, v_g, T_i, \boldsymbol{\theta})}{v_g - v_l}, \frac{\sigma^2}{n_{si}}\right) \\ \phi\left(0; h_D(v_l, T_i, \boldsymbol{\theta}) - \frac{H_D(v_l, v_g, T_i, \boldsymbol{\theta})}{v_g - v_l}, \sigma_e^2\right) \\ \phi\left(0; h_D(v_g, T_i, \boldsymbol{\theta}) - \frac{H_D(v_l, v_g, T_i, \boldsymbol{\theta})}{v_g - v_l}, \sigma_e^2\right) \\ \prod_{j=1}^{n_i} \phi(P_{ij}; h_D(v_{ij}, T_i, \boldsymbol{\theta}), \sigma^2), \end{aligned} \tag{4.134}$$

where in this expression v_l and v_g are parameters and

$$\tilde{\boldsymbol{\theta}} = (a, b, c, d, e, f, g, v_l, v_g)^T. \tag{4.135}$$

The means of the observed coexistence volumes have density,

$$p(\bar{v}_{li}, \bar{v}_{gi} | v_l, v_g) = \phi\left(\bar{v}_{li}; v_l, \frac{\sigma_v}{n_{li}}\right) \phi\left(\bar{v}_{gi}; v_g, \frac{\sigma_v}{n_{gi}}\right). \tag{4.136}$$

Then the posterior up to a normalising constant is,

$$p(\tilde{\boldsymbol{\theta}} | \bar{P}_{si}, \boldsymbol{P}_i, \bar{v}_{li}, \bar{v}_{gi}) \propto L(\tilde{\boldsymbol{\theta}}) p(\bar{v}_{li}, \bar{v}_{gi} | v_l, v_g) p(\tilde{\boldsymbol{\theta}}), \tag{4.137}$$

where $p(\tilde{\boldsymbol{\theta}})$ is the prior of $\tilde{\boldsymbol{\theta}}$.

We use the same independent uniform priors on the parameters as were

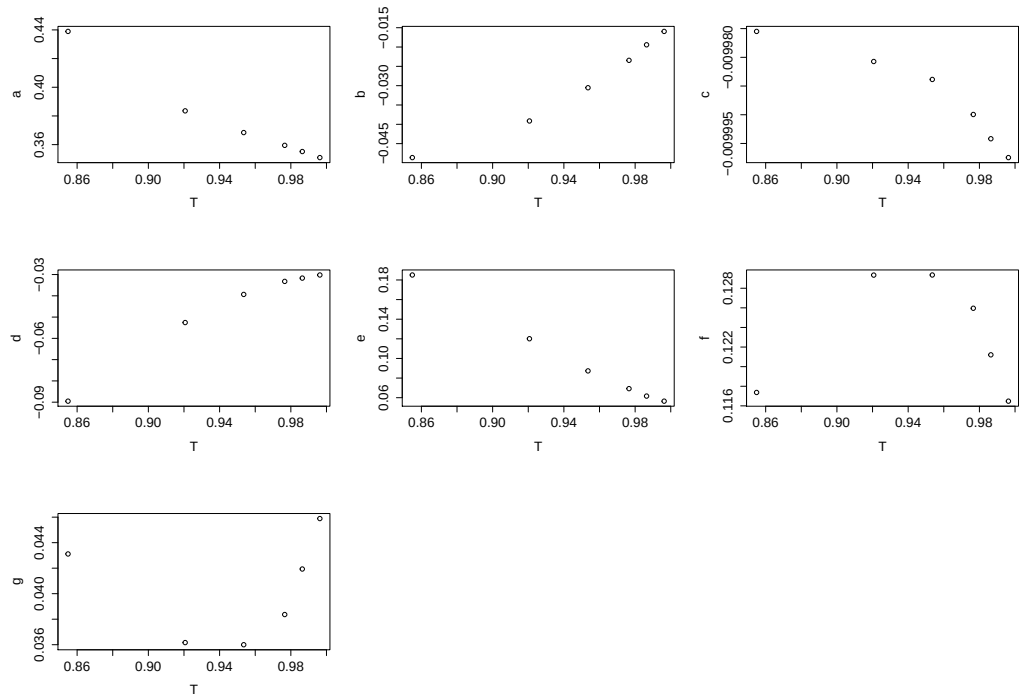


Figure 4.19: Plots of the modes of the posterior marginal distribution of each model parameter for the DEoS fitted to real data at individual temperatures.

used in the synthetic data case (4.15)-(4.21). We also use the same independent uniform priors on the coexistence volumes as were previously used in the synthetic case (4.103)-(4.104). Also as in the synthetic data case we used 31 tempering levels with each tempering level being twice the previous level and a maximum tempering level of 2^{30}. For each T_i we ran the PT-HMC algorithm from 5 different sets of initial parameter values for 11000 iterations and found that for a given T_i each run converged to the same parameter values.

Figure 4.19 plots the mode of the posterior marginal distributions of the seven model parameters $a, \ldots, g$, for each T_i. From figure 4.19 a through e seem approximately linear for higher temperatures but there appears to be some curvature for lower temperatures. However we find that the high posterior density regions for the parameters increase in size as temperature decreases. This means that while a linear temperature rule may not describe exactly the marginal modes of these parameters at each temperature

it can pass through the high posterior density region at each temperature. As such it is possible that linear temperature rules could be a reasonable approximation for $a, b, c, d,$ and e. From figure 4.19 we can see that the behaviour of f and g as a function of temperature appears to be approximately quadratic. This suggests that to fit multiple temperatures for real CO_2 we need to define a form of quadratic temperature rule and priors on such a rule such that within our temperature range of interest it can only take values within the bounds used by our previous priors.

4.6.2 Quadratic Temperature Rule

Suppose we have some parameter θ which is a function of temperature, T, so $\theta = \theta(T)$. We have some temperature range of interest such that,

$$T_{\min} \leq T \leq T_{\max}, \tag{4.138}$$

where $T_{\min}$ and $T_{\max}$ are the minimum and maximum temperatures of interest respectively. For temperatures in this range we have bounds on the value θ can take for any given temperature. So

$$\theta_{\min} \leq \theta(T) \leq \theta_{\max}, \quad \forall T \in [T_{\min}, T_{\max}], \tag{4.139}$$

where $\theta_{\min}$ and $\theta_{\max}$ are the minimum and maximum values we allow θ to take on the temperature range of interest respectively.

Suppose we wish to use a quadratic temperature rule for the functional form of θ. We can define such a rule as follows,

$$\theta(T) = \theta_a(T - \theta_b)^2 + \theta_c. \tag{4.140}$$

Our temperature rule consists of three parameters, $\theta_a, \theta_b, \theta_c$, on which we need to place priors such that (4.139) holds. We see that θ has a stationary

point at $T = \theta_b$ and that it's nature (maximum or minimum) is determined by the sign of θ_a. Substituting $T = \theta_b$ into (4.140) we find that the stationary point has value θ_c. This then allows us to place the following uniform priors on θ_b and θ_c,

$$\theta_b \sim \mathcal{U}(T_{\min}, T_{\max}), \tag{4.141}$$

$$\theta_c \sim \mathcal{U}(\theta_{\min}, \theta_{\max}). \tag{4.142}$$

The prior on θ_b is derived from the fact that we want the stationary point to fall within our temperature range of interest, while the prior on θ_c comes from the fact that it is a value the temperature rule takes over the temperature range of interest and so (4.139) must hold. We now note that to ensure that the temperature rule only takes values as defined by (4.139) it is sufficient that the values of the function at the stationary point and the edges of the temperature region of interest can only take values in this range of interest. Our prior on θ_c ensures this is true for the stationary point. Then we need that

$$\theta_{\min} \leq \theta(T_{\min}) = \theta_a(T_{\min} - \theta_b)^2 + \theta_c \leq \theta_{\max}, \tag{4.143}$$

$$\theta_{\min} \leq \theta(T_{\max}) = \theta_a(T_{\max} - \theta_b)^2 + \theta_c \leq \theta_{\max}. \tag{4.144}$$

We firstly consider (4.143). There are 3 cases to consider, $\theta_a < 0, \theta_a = 0, \theta_a > 0$, which determines if the stationary point is a maximum, minimum or if $\theta(T)$ is constant. If $\theta_a = 0$ then $\theta(T) = \theta_c$ and so (4.143) holds from our prior on θ_c. We now consider $\theta_a > 0$. In this case we have,

$$\theta_a(T_{\min} - \theta_b)^2 \geq 0, \tag{4.145}$$

$$\implies \theta_a(T_{\min} - \theta_b)^2 + \theta_c \geq \theta_c \geq \theta_{\min}, \tag{4.146}$$

again from the prior on θ_c and so the lower bound in (4.143) holds for all

$\theta_a > 0$. Considering the upper bound in (4.143) we have,

$$\theta_a(T_{\min} - \theta_b)^2 \leq \theta_a(T_{\min} - T_{\max})^2 = \theta_a {T_{\text{diff}}}^2, \tag{4.147}$$

where $T_{\text{diff}} = T_{\max} - T_{\min}$. Then we have

$$\theta_a(T_{\min} - \theta_b)^2 + \theta_c \leq \theta_a {T_{\text{diff}}}^2 + \theta_c. \tag{4.148}$$

So the upper bound in (4.143) holds so long as,

$$\theta_a {T_{\text{diff}}}^2 + \theta_c \leq \theta_{\max}, \tag{4.149}$$

$$\implies \theta_a \leq \frac{\theta_{\max} - \theta_c}{{T_{\text{diff}}}^2}. \tag{4.150}$$

As such we now have an upper bound for θ_a which depends on θ_c. Finally we consider the case that $\theta_a < 0$. In this case the stationary point is a maximum and so we have

$$\theta_a(T_{\min} - \theta_b)^2 + \theta_c \leq \theta_c \leq \theta_{\max}, \tag{4.151}$$

trivially. Then to determine the range of values of θ_a for which the lower bound in (4.143) we use the fact that for $\theta_a < 0$ we have,

$$\theta_a(T_{\min} - \theta_b)^2 \geq \theta_a {T_{\text{diff}}}^2, \tag{4.152}$$

and so

$$\theta_a(T_{\min} - \theta_b)^2 + \theta_c \geq \theta_a {T_{\text{diff}}}^2 + \theta_c. \tag{4.153}$$

Then the lower bound in (4.143) holds so long as,

$$\theta_a {T_{\text{diff}}}^2 + \theta_c \geq \theta_{\min}, \tag{4.154}$$

$$\implies \theta_a \geq \frac{\theta_{\min} - \theta_c}{{T_{\text{diff}}}^2}. \tag{4.155}$$

We now have the range of values of θ_a (dependent on θ_c) for which (4.143) holds,

$$\frac{\theta_{\min} - \theta_c}{{T_{\text{diff}}}^2} \leq \theta_a \leq \frac{\theta_{\max} - \theta_c}{{T_{\text{diff}}}^2}. \tag{4.156}$$

We can similarly derive the same bounds from (4.144). Defining a prior on θ_a is not as simple as defining the priors on θ_b and θ_c as it depends on θ_c. As such we define a joint prior on θ_a and θ_c using the identity,

$$\pi(\theta_a, \theta_c) = \pi(\theta_a|\theta_c)\pi(\theta_c). \tag{4.157}$$

Now we have

$$\pi(\theta_c) = \frac{1}{\theta_{\max} - \theta_{\min}}, \quad \theta_{\min} \leq \theta_c \leq \theta_{\max}. \tag{4.158}$$

Then if we define the conditional distribution of $\theta_a|\theta_c$ to be uniform,

$$\theta_a|\theta_c \sim \mathcal{U}\left(\frac{\theta_{\min} - \theta_c}{{T_{\text{diff}}}^2}, \frac{\theta_{\max} - \theta_c}{{T_{\text{diff}}}^2}\right), \tag{4.159}$$

$$\implies \pi(\theta_a|\theta_c) = \frac{{T_{\text{diff}}}^2}{\theta_{\max} - \theta_{\min}}, \quad \frac{\theta_{\min} - \theta_c}{{T_{\text{diff}}}^2} \leq \theta_a \leq \frac{\theta_{\max} - \theta_c}{{T_{\text{diff}}}^2}. \tag{4.160}$$

Then the joint prior distribution of θ_a and θ_c has density,

$$\pi(\theta_a, \theta_c) = \frac{{T_{\text{diff}}}^2}{(\theta_{\max} - \theta_{\min})^2}, \quad \theta_{\min} \leq \theta_c \leq \theta_{\max}, \frac{\theta_{\min} - \theta_c}{{T_{\text{diff}}}^2} \leq \theta_a \leq \frac{\theta_{\max} - \theta_c}{{T_{\text{diff}}}^2}. \tag{4.161}$$

The resulting marginal distribution for θ_a has density

$$\pi(\theta_a) = \frac{{T_{\text{diff}}}^2}{\theta_{\max} - \theta_{\min}}\left(1 - \frac{{T_{\text{diff}}}^2|\theta_a|}{\theta_{\max} - \theta_{\min}}\right), -\frac{\theta_{\max} - \theta_{\min}}{{T_{\text{diff}}}^2} \leq \theta_a \leq \frac{\theta_{\max} - \theta_{\min}}{{T_{\text{diff}}}^2}. \tag{4.162}$$

We can now use this quadratic temperature rule and corresponding priors for the f and g parameters in our model, as was suggested to be necessary by the fits for individual temperatures. In our case $T_{\max}$ and $T_{\min}$ are here $T_h = 1.029171$ and $T_l = 0.8549026$ respectively. In dimensional

units these correspond to 313 and 260K respectively. $f_{\max}, f_{\min}, g_{\max}$ and $g_{\min}$ are given by the bounds of the uniform priors we have previously used on f and g. So we have,

$$f_{\max} = 0.15, \tag{4.163}$$

$$f_{\min} = 0, \tag{4.164}$$

$$g_{\max} = 0.1, \tag{4.165}$$

$$g_{\min} = 0. \tag{4.166}$$

Then the quadratic temperature rule for f is,

$$f(T) = f_a(T - f_b)^2 + f_c, \tag{4.167}$$

with the following priors on the parameters,

$$\pi(f_b) = \frac{1}{(T_h - T_l)^2}, \quad T_l \leq f_b \leq T_h, \tag{4.168}$$

$$\pi(f_a, f_c) = \frac{(T_h - T_l)^2}{0.15^2}, \quad 0 \leq f_c \leq 0.15, \frac{-f_c}{(T_h - T_l)^2} \leq f_a \leq \frac{0.15 - f_c}{(T_h - T_l)^2}. \tag{4.169}$$

Similarly the quadratic temperature rule for g is,

$$g(T) = g_a(T - g_b)^2 + g_c, \tag{4.170}$$

with the following priors on the parameters,

$$\pi(g_b) = \frac{1}{(T_h - T_l)^2}, \quad T_l \leq g_b \leq T_h, \tag{4.171}$$

$$\pi(g_a, g_c) = \frac{(T_h - T_l)^2}{0.1^2}, \quad 0 \leq g_c \leq 0.1, \frac{-g_c}{(T_h - T_l)^2} \leq g_a \leq \frac{0.1 - g_c}{(T_h - T_l)^2}. \tag{4.172}$$

So with quadratic temperature rules on f and g, and linear temperature

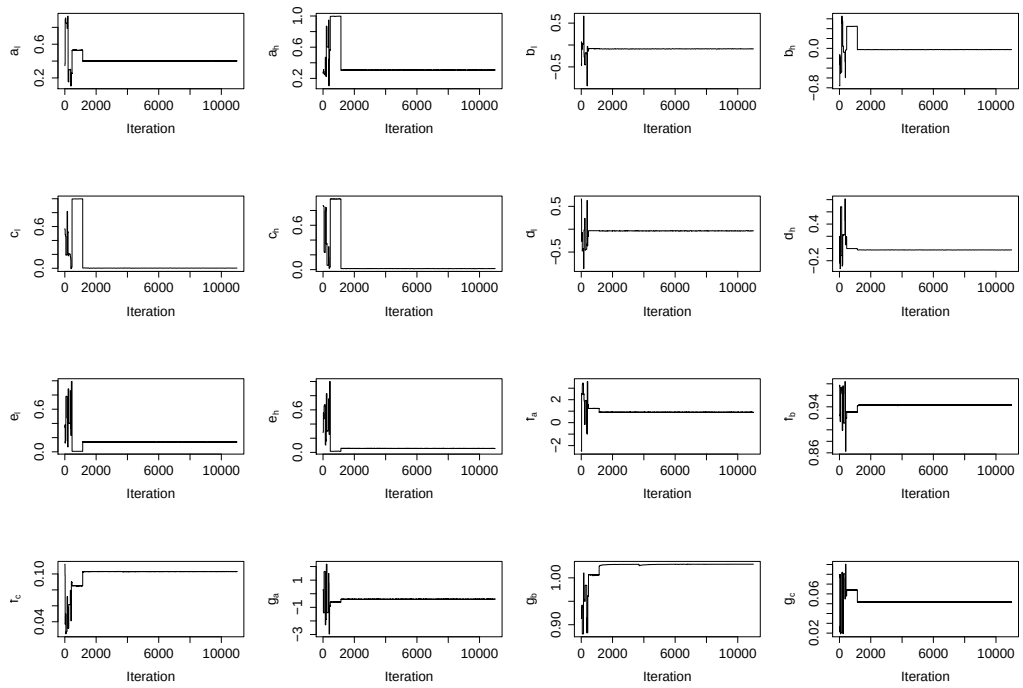


Figure 4.20: Traceplot of the sampled model parameters for a single run of the PT-HMC sampler applied to the multiple temperature Demetriades EoS with real pure CO_2 data and a combination of linear and quadratic temperature rules. Traceplots for other runs are similar. We can see here that the sampler remains in some small subset of the parameter space after the first 2000 iterations.

rules on a through e we have 16 total model parameters collected in the parameter vector $\boldsymbol{\theta}$ as follows,

$$\boldsymbol{\theta} = (a_l, a_h, b_l, b_h, c_l, c_h, d_l, d_h, e_l, e_h, f_a, f_b, f_c, g_a, g_b, g_c)^T. \tag{4.173}$$

In the next section we describe the fitting of the model using this combination of linear and quadratic temperature rules to real pure CO_2 data.

4.6.3 Fitting Multiple Temperatures

We fit the DEoS model with linear temperature rules for a through e and quadratic temperature rules on f and g to real pure CO_2 data for 11 temperatures. These 11 temperatures are the 6 subcritical temperatures which we fitted individually as described in §4.6.1 and the 5 supercritical temper-

atures, within our temperature range of interest, for which we have data, 304.1472K, 304.9919K, 306.9914K, 309.9906K and 312.9898K. This means we will be fitting our model to data over our whole temperature range of interest 260-313K. As we used individual fits of subcritical temperatures to propose temperature rules we withhold the homogeneous and coexistence data we have for the other subcritical temperatures. We can then use this withheld data as test data once posterior samples of the parameters are found to evaluate how well our model describes data and temperatures which were not used in the fitting. In total at the 11 temperatures used to fit here there are 365 homogeneous data points. We collect the pressures, volumes, and temperatures at each of these data points in the vectors,

$$\boldsymbol{P} = (P_1, P_2, \ldots, P_{365})^T, \tag{4.174}$$

$$\boldsymbol{v} = (v_1, v_2, \ldots, v_{365})^T, \tag{4.175}$$

$$\boldsymbol{T} = (T_1, T_2, \ldots, T_{365})^T, \tag{4.176}$$

respectively, where the i-th data points has pressure P_i, volume v_i and temperature T_i. We have coexistence data for the 6 subcritical temperatures. At each of these 6 temperatures, which we collect in the vector $\boldsymbol{T}_x$, there are one or more measurements of v_l, v_g, and P_s. As in §4.6.1 we are fixing σ and σ_v, to 0.0002 and 0.001 respectively. Then rather than using the individual measurements of v_l, v_g, and P_s we can use their means. We once again set σ_e to be a tenth of σ, 0.00002. Let T_{xi} be the i-th element of $\boldsymbol{T}_x$. Let n_{li} be the number of v_l measurements at this temperature and let n_{gi} and n_{si} be similarly for v_g and P_s.

Let $\bar{v}_{li}$ be the mean of the v_l measurements at T_{xi} and let $\bar{v}_{gi}$ and $\bar{P}_{si}$

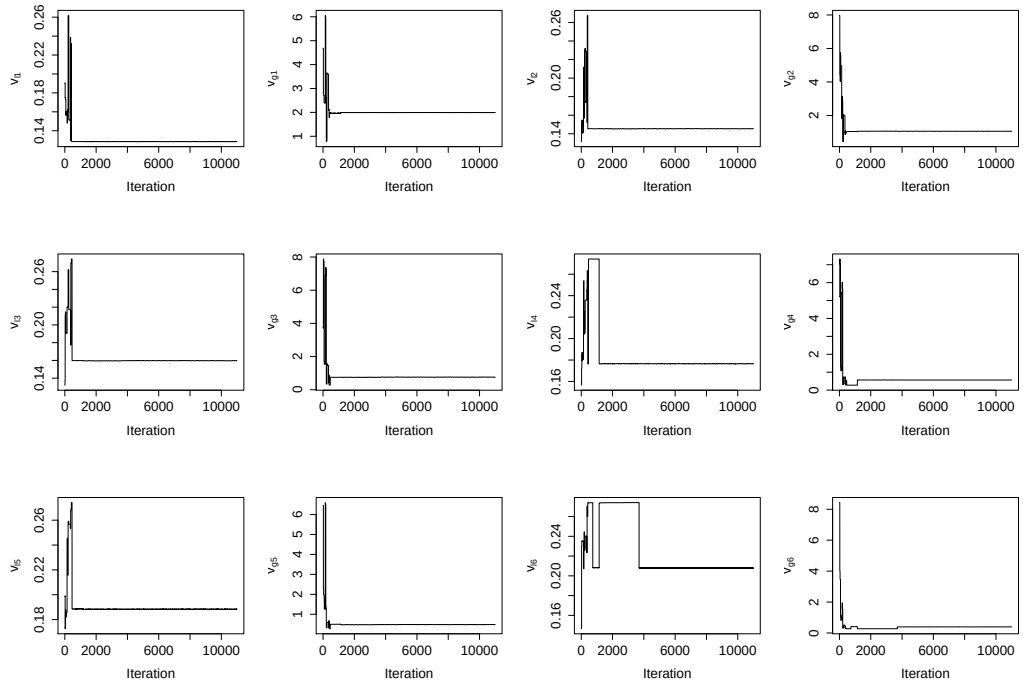


Figure 4.21: Traceplot of the sampled coexistence volume parameters for a single run of the PT-HMC sampler applied to the multiple temperature Demetriades EoS with real pure CO_2 data and a combination of linear and quadratic temperature rules. Traceplots for other runs are similar. We can see here that the sampler remains in some small subset of the parameter space after the first 2000 iterations.

be similarly for v_g and P_s. Then we collect these in the vectors,

$$\boldsymbol{n}_l = (n_{l1}, n_{l2}, \ldots, n_{l6})^T, \tag{4.177}$$

$$\boldsymbol{n}_g = (n_{g1}, n_{g2}, \ldots, n_{g6})^T, \tag{4.178}$$

$$\boldsymbol{n}_s = (n_{s1}, n_{s2}, \ldots, n_{s6})^T, \tag{4.179}$$

$$\bar{\boldsymbol{v}}_l = (\bar{v}_{l1}, \bar{v}_{l2}, \ldots, \bar{v}_{l6})^T, \tag{4.180}$$

$$\bar{\boldsymbol{v}}_g = (\bar{v}_{g1}, \bar{v}_{g2}, \ldots, \bar{v}_{g6})^T, \tag{4.181}$$

$$\bar{\boldsymbol{P}}_s = (\bar{P}_{s1}, \bar{P}_{s2}, \ldots, \bar{P}_{s6})^T. \tag{4.182}$$

There are then 12 coexistence volume parameters,

$$\boldsymbol{v}_\theta = (v_{l1}, v_{l2}, \ldots, v_{l6}, v_{g1}, v_{g2}, \ldots, v_{g6})^T, \tag{4.183}$$

where v_{li} is the errors-in-variables parameter for the liquid coexistence volume at temperature T_{xi} and v_{gi} is similarly the gas coexistence volume at temperature T_{xi}. Then with the 16 model parameters we have 28 parameters total which we collect in the vector,

$$\tilde{\boldsymbol{\theta}} = \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{v}_\theta. \end{pmatrix}. \tag{4.184}$$

The likelihood is then,

$$\begin{aligned} L(\tilde{\boldsymbol{\theta}}) = \prod_{i=1}^{6} \phi\left(\bar{P}_{si}; \frac{H_D(v_{li}, v_{gi}, T_{xi}, \boldsymbol{\theta})}{v_{gi} - v_{li}}, \frac{\sigma^2}{n_{si}}\right) \\ \phi\left(0; h_D(v_{li}, T_{xi}, \boldsymbol{\theta}) - \frac{H_D(v_{li}, v_{gi}, T_{xi}, \boldsymbol{\theta})}{v_{gi} - v_{li}}, \sigma_e^2\right). \\ \phi\left(0; h_D(v_{gi}, T_{xi}, \boldsymbol{\theta}) - \frac{H_D(v_{li}, v_{gi}, T_{xi}, \boldsymbol{\theta})}{v_{gi} - v_{li}}, \sigma_e^2\right) \\ \prod_{j=1}^{365} \phi(P_j; h(v_j, T_j, \boldsymbol{\theta}), \sigma^2). \end{aligned} \tag{4.185}$$

To fully define the posterior we then require the errors-in-variables term for the coexistence volumes which here is,

$$p(\bar{\boldsymbol{v}}_l, \bar{\boldsymbol{v}}_g | \boldsymbol{v}_\theta) = \prod_{i=1}^{6} \phi\left(\bar{v}_{li}; v_{li}, \frac{\sigma_v^2}{n_{li}}\right) \phi\left(\bar{v}_{gi}; v_{gi}, \frac{\sigma_v^2}{n_{gi}}\right), \tag{4.186}$$

and priors on all the parameters. We use the same independent uniform priors on the parameters in the linear temperature rules as previously, (4.58)-(4.62) and (4.65)-(4.69), for the parameters in the quadratic temperature rules we use the priors described in §4.6.2 and we use the same independent uniform priors on the volume parameters as before.

We then sample from the posterior distribution described by these terms using the PT-HMC sampler. To fit the model as before we temper the likelihood and errors-in-variables term so that the hottest replicas are sampling from the prior. We find that in order for a sufficient number of swaps to

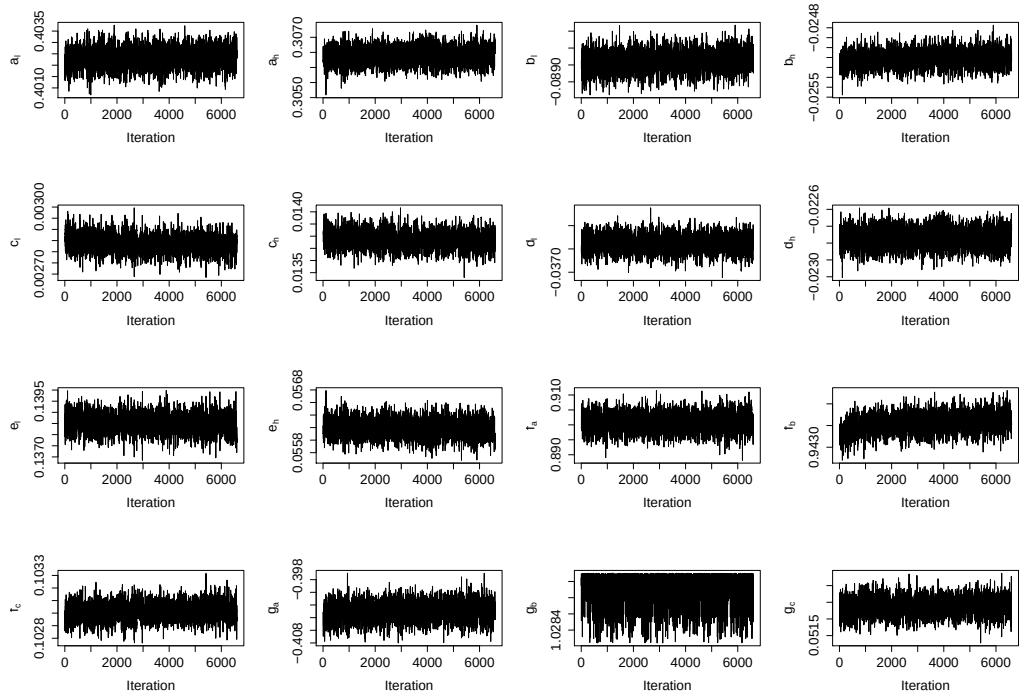


Figure 4.22: Traceplot of the sampled model parameters for a single run of the PT-HMC sampler applied to the multiple temperature Demetriades EoS with real pure CO_2 data and a combination of linear and quadratic temperature rules with burn-in discarded. Traceplots for other runs are similar. We can see here that the sampler mixes well. Combined with our sampler consistently finding the same mode when started with different initial parameter values this suggests the sampler has converged to the high posterior probability region.

occur between replicas for the sampler to converge more quickly we need to decrease the distance between neighbouring tempering levels. Thus we now set each tempering level to be 1.5 times the previous tempering level. We find that if we set the maximum tempering level to be $\beta = \left(\frac{3}{2}\right)^{52}$ the sampler at that level is able to explore the entire prior parameter space. As such we now use 53 replicas with β for each tempering level as follows,

$$\beta_i = \left(\frac{3}{2}\right)^{i-1}, \quad i = 1, \ldots, 53. \tag{4.187}$$

We run the sampler from various initial parameter values for 11000 iterations and find that in each case the sampler converges to the same location. The choice of how many iterations to run is somewhat arbitrary. The num-

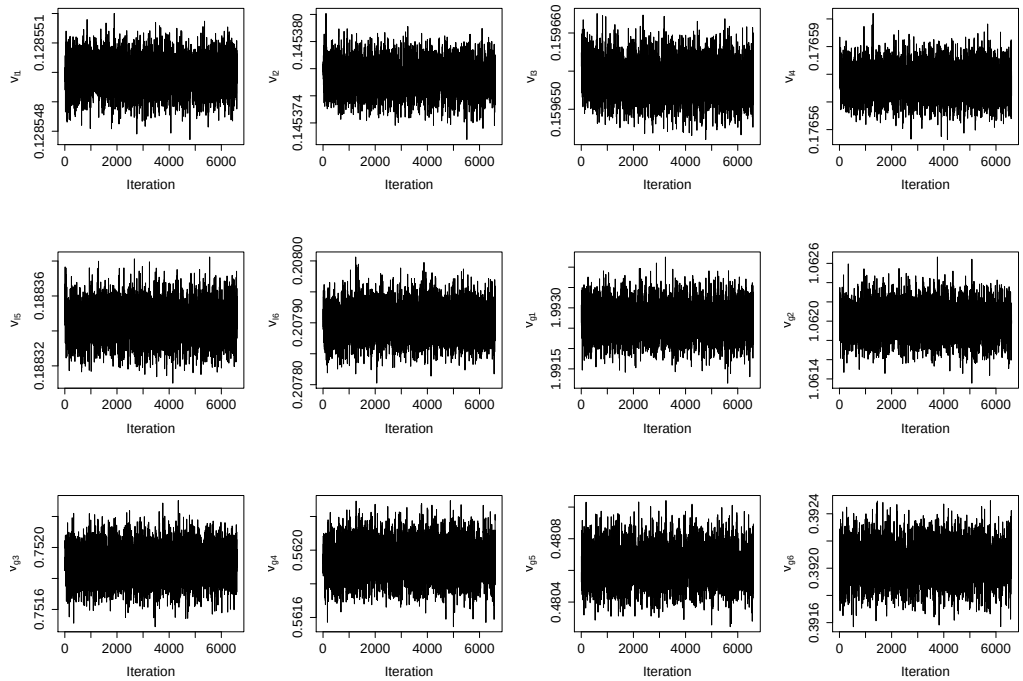


Figure 4.23: Traceplot of the sampled coexistence volume parameters for a single run of the PT-HMC sampler applied to the multiple temperature Demetriades EoS with real pure CO_2 data and a combination of linear and quadratic temperature rules with burn-in discarded. Traceplots for other runs are similar. We can see here that the sampler mixes well. Combined with our sampler consistently finding the same mode when started with different initial parameter values this suggests the sampler has converged to the high posterior probability region.

ber of iterations needs to be such that the sampler has time to converge and that post convergence it mixes well. We found that runs of 11000 iterations ensured these two conditions were met and completed in a reasonable time. Performing one of these runs takes approximately 2.5 days run in parallel on 16 CPUs the University of Nottingham's high performance computing facilities. Figures 4.20 and 4.21 plot the traces for one of these runs of the sampler for the model and coexistence volume parameters respectively demonstrating that after several thousand iterations the sampler remains in a small region of the parameter space. Figures 4.22 and 4.23 are the traceplots after discarding the first 4400 iterations as burn-in. From these we can see that our sampler appears to be mixing well. Thus our sampler consistently finds the same region of high posterior probability when

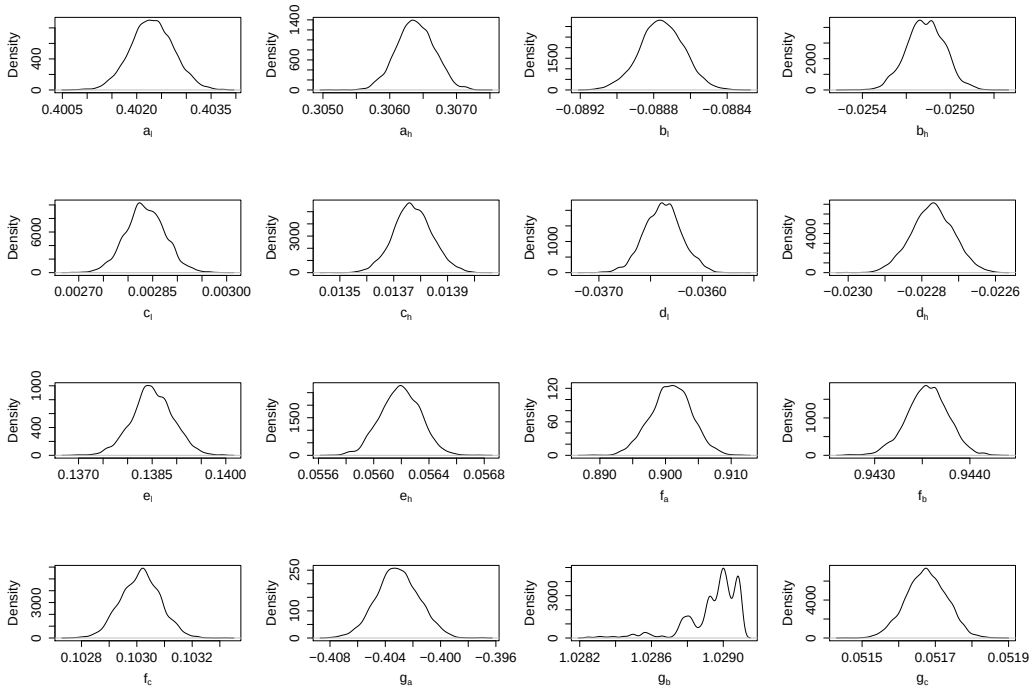


Figure 4.24: Marginal density plots of the sampled model parameters for a single run of the PT-HMC sampler applied to the multiple temperature Demetriades EoS with real pure CO_2 data and a combination of linear and quadratic temperature rules. Only the high density region is shown.

started from different initial parameter values, and mixes well within this region. This all suggests that our sampler has converged and is sampling from the posterior distribution of interest.

We plot the marginal densities for the model and coexistence volume parameters from our samples after discarding burn-in in figures 4.24 and 4.25. These figures display only the high density regions, the priors are all much wider than the the x-axis on each density plot. To evaluate the fit of the model we take a sample of 50 sets of parameters from our posterior sample and plot the resulting realisations of the model against the data used in the fitting as a posterior predictive check of the models fit [Gelman et al., 2014]. In figure 4.26 we plot 50 posterior coexistence points, the magenta crosses, and 50 posterior pressure-volume curves against the experimental data at 296.994K, (homogeneous data are the hollow black circles, coexistence data are the solid black circles). We can see that the

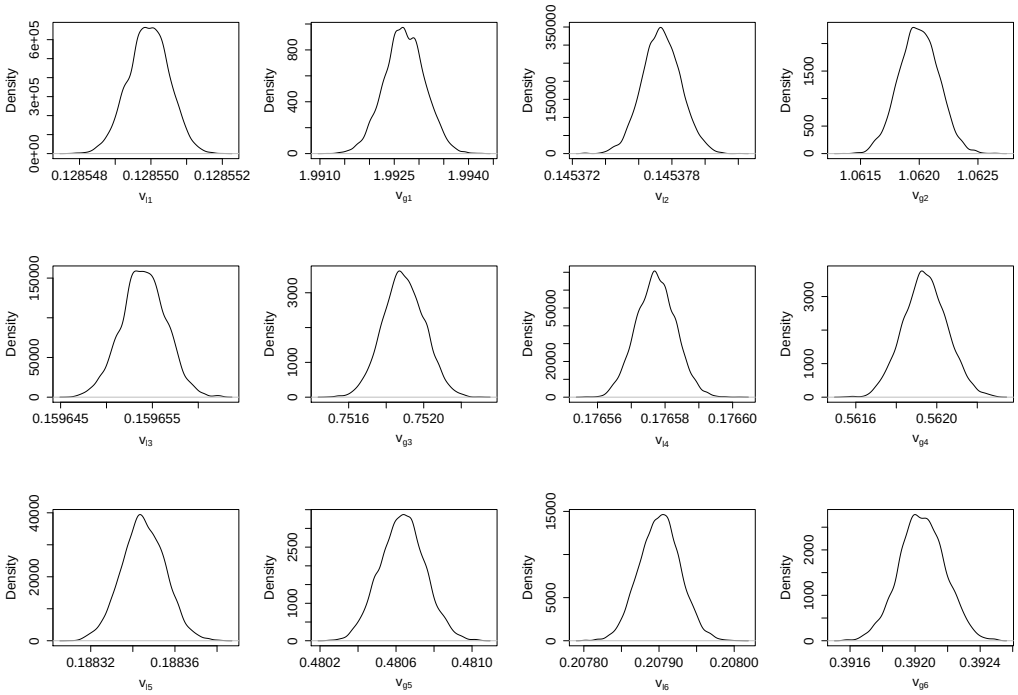


Figure 4.25: Marginal density plots of the sampled coexistence volume parameters for a single run of the PT-HMC sampler applied to the multiple temperature Demetriades EoS with real pure CO_2 data and a combination of linear and quadratic temperature rules. Only the high density region is shown.

posterior realisations go through the data points describing the homogeneous data well and that the sampled coexistence points are very similar to the experimental data. Plots of realisations for other subcritical temperatures and the highest 3 supercritical temperatures used to fit the model show similarly good fit. However at the two lowest subcritical temperatures, 304.1472K and 304.9919K, we find that the posterior realisations do not align so well with the data.

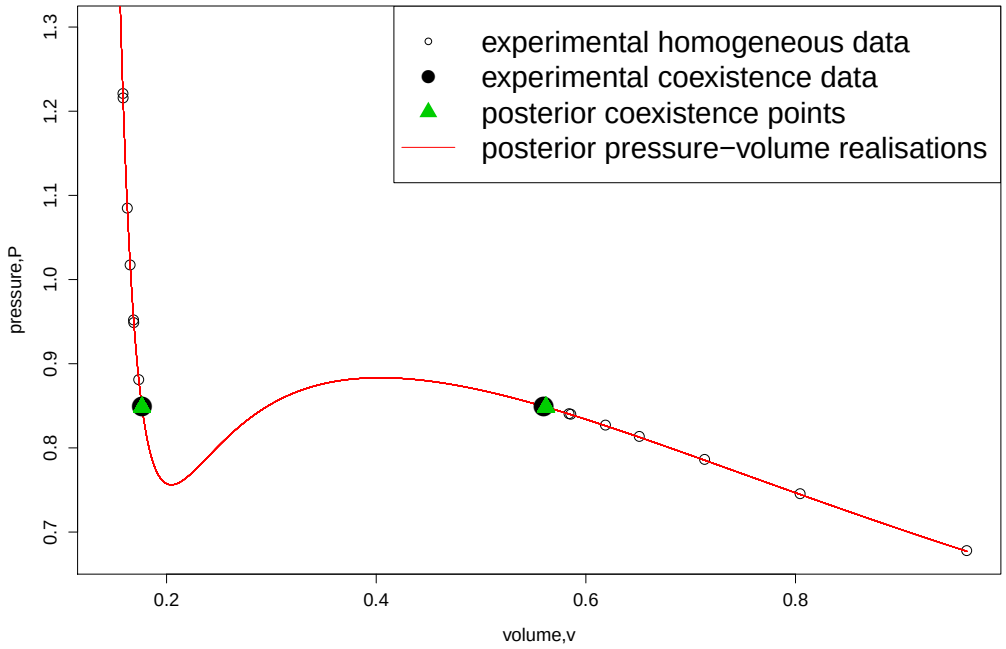


Figure 4.26: Plot of 50 posterior realisations at 296.994K using sampled parameters from a run of the PT-HMC sampler applied to the multiple temperature Demetriades EoS with real pure CO_2 data and a combination of linear and quadratic temperature rules. Hollow black circles are homogeneous data at 296.994K. Solid black circles are coexistence data for 296.994K. Green triangles are 50 posterior coexistence points at 296.994K. Red lines are 50 posterior realisations of the pressure-volume curve.

Figure 4.27 plots 50 posterior pressure-volume curves at 304.1472K, red lines, against the experimental data, black circles. As can be seen from figure 4.27 the data has a near saddlepoint due to being at a temperature slightly above the critical point. The realisations from the model however are still showing behaviour characteristic of coexistence. This would appear to be because the temperature rules used do not have the flexibility to capture the change of behaviour in the data around the critical point. Considering posterior realisations at 307K and above we find that the posterior realisations and the data are in accordance. So the model works well for most of the supercritical region of interest. This poor fit is a result of slight inflexibility of the DEoS model fitted here, it not being flexible enough to capture the behaviour of the data in the vicinity of the critical

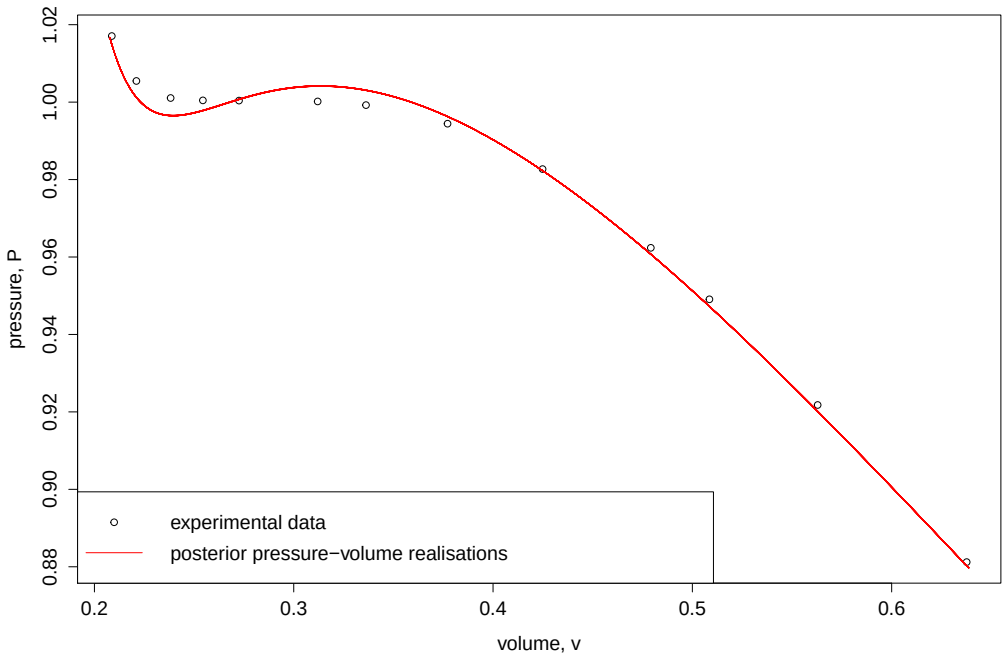


Figure 4.27: Plot of 50 posterior realisations at 304.1472K using sampled parameters from a run of the PT-HMC sampler applied to the multiple temperature Demetriades EoS with real pure CO_2 data and a combination of linear and quadratic temperature rules. Hollow black circles are homogeneous data at 304.1472K. Red lines are 50 posterior realisations of the pressure-volume curve.

point.

However our primary aim here is not to develop the most accurate model. Instead we sought to show how we could fit the statistical model for an EoS we proposed in chapter 3. In this chapter we have successfully, in terms of sampling from the correct posterior distribution, fitted our statistical model applied to a form of the DEoS. We have, despite the difficulties that entails, developed an MCMC sampler which can consistently sample from the posterior distribution of the model parameters, overcoming multimodality and checked this sampler against synthetic data simulated from the model. We then fitted our model to pure CO_2 data and can use the results here to propose future improvements to the DEoS. We also demonstrated how posterior realisations can be used to check model fit. We now look to extend our statistical model to mixtures of CO_2 with impurities

and demonstrate a fit of the DEoS using this model and our sampler to the CO_2-H_2 system.

Chapter 5

Parametric EoS for Mixtures

5.1 Introduction

In the previous chapter we fitted our statistical model for pure CO_2 from chapter 3 to real data. The CO_2 in CCS applications will typically not be pure but a mixture of CO_2 with other fluids (see §1.2 for expected impurities and levels). Phase behaviour for mixtures can differ quite significantly from that for pure CO_2 (see §1.3.4). As such we need to be able to extend our EoS for pure CO_2 to mixtures of CO_2 with other fluids for relevance to CCS applications. In this chapter we propose an extension of our statistical model for a pressure-explicit parametric EoS to a binary mixture of CO_2 with an impurity and then fit this model to the CO_2-H_2 data that was described in §1.6. We begin by describing the model for homogeneous mixture data for a general pressure-explicit EoS. We then propose a way to incorporate coexistence for mixtures in our model based on how coexistence was included for pure CO_2 in chapter 3. It is often the case that coexistence volumes are not recorded for mixtures of CO_2 with an impurity. We thus describe how we can account for this missing data in our model. We then define our statistical model for the DEoS and fit to a single temperature from the CO_2-H_2 data that was described in §1.6.

5.2 Modelling Homogeneous Mixture Data

For mixtures we have an additional input to our general pressure-explicit EoS compared to the pure CO_2 case. Now there is an additional dependence upon x, the mole fraction of the impurity in the mixture. As such, we now use $h_{\text{mix}}(v, T, x, \boldsymbol{\theta})$ to denote our general pressure-explicit EoS for mixtures, where $\boldsymbol{\theta}$ is once again a p-dimensional vector of parameters.

Suppose we have n_m homogeneous mixture data points. The response variable is the pressure, P, and we collect the n observed pressures in the vector, $\boldsymbol{P}_{\text{mix}}^{\text{obs}} = (P_{\text{mix},1}^{\text{obs}}, \ldots, P_{\text{mix},n_m}^{\text{obs}})^T$. For each $P_{\text{mix},i}^{\text{obs}}$ there is a corresponding volume, $v_{\text{mix},i}$, temperature, $T_{\text{mix},i}$ and impurity mole fraction $x_{\text{mix},i}$. We collect these in the vectors, $\boldsymbol{v}_{\text{mix}}, \boldsymbol{T}_{\text{mix}}$ and $\boldsymbol{x}_{\text{mix}}$ respectively. We then have design matrix,

$$\boldsymbol{X}_{\text{mix}} = (\boldsymbol{v}_{\text{mix}}, \boldsymbol{T}_{\text{mix}}, \boldsymbol{x}_{\text{mix}}). \tag{5.1}$$

We then define

$$P_{\text{mix},i}^{\text{obs}} = h_{\text{mix}}(v_{\text{mix},i}, T_{\text{mix},i}, x_{\text{mix},i}, \boldsymbol{\theta}) + \epsilon_{\text{mix},i}, \quad i = 1, \ldots, n_m, \tag{5.2}$$

where the $\epsilon_{\text{mix},i}$ are i.i.d. normal random variables with mean 0 and variance σ_{mix}^2. As previously we can either treat σ_{mix}^2 as an additional parameter or fix it based on reported experimental uncertainty. In this chapter we will only consider fixed σ_{mix}^2. Our model for $\boldsymbol{P}_{\text{mix}}^{\text{obs}}$ then has probability density function,

$$p(\boldsymbol{P}_{\text{mix}}^{\text{obs}} | \boldsymbol{\theta}, \boldsymbol{X}_{\text{mix}}, \sigma_{\text{mix}}^2) = \prod_{i=1}^{n_m} \phi(P_{\text{mix},i}^{\text{obs}}; h_{\text{mix}}(v_{\text{mix},i}, T_{\text{mix},i}, x_{\text{mix},i}, \boldsymbol{\theta}), \sigma_{\text{mix}}^2), \tag{5.3}$$

where $\phi(\cdot; \mu, \sigma^2)$ is the normal p.d.f. of x with mean μ and variance σ^2. This then fully defines the likelihood for the homogeneous mixture data. As with pure CO_2 we must now consider how to incorporate coexistence for a mixture with 2 components into our model. We propose a model based on

that which we derived for pure CO_2 in chapter 3 which we describe now.

5.3 Modelling Coexistence Mixture Data

For a mixture of 2 components, in our case CO_2 and an impurity, there are four constraints that must be satisfied for coexistence to hold at a pair of points. Firstly the fugacity of CO_2 must be equal at each coexistence point. Secondly the fugacity of the impurity must be equal at each coexistence point. Thirdly the pressure must be equal at the coexistence points. Finally the temperature must be equal at the coexistence points. As was the case with pure CO_2 the temperature constraint is trivially fulfilled as our model is dependent on temperature as an explanatory variable. When considering the fugacity constraints we can equivalently consider the log-fugacities. From (1.22) the log-fugacity coefficient of the i-th component in the mixture is

$$\ln \bar{\phi}_i(v,T,x) = \frac{1}{RT}\int_{V'=\infty}^{V} \frac{RT}{V'} - \left(\frac{\partial P(v,T,x)}{\partial N_i}\right)_{T,V,N_{j\neq i}} \mathrm{d}v' - \ln\left(\frac{P(v,T,x)v}{RT}\right), \quad (5.4)$$

where N_i is the number of moles of component i, N is the total number of moles, and $V = Nv$. Here and henceforth V, T and P are dimensional unless stated otherwise. Demetriades and Graham [2016] used application of the chain rule and integration by parts in order to remove dependence of $\ln \bar{\phi}_i$ on the numbers of moles of the component i, N_i, and to rewrite the mixture fugacity integral of the pure fluid fugacity integral, thus ensuring that if the pure integral can be done analytically then so can the mixture integral. We here follow a similar process.

From Demetriades and Graham [2016] through application of the chain

rule we can write,

$$\left(\frac{\partial P(v,T,x)}{\partial N_i}\right)_{T,V,N_{j\neq i}} = \frac{\partial P(v,T,x)}{\partial v}\left(-\frac{V}{N^2}\right) + \frac{\partial P(v,T,x)}{\partial x}\frac{\partial x}{\partial N_i}, \quad (5.5)$$

$$= -\frac{\partial P(v,T,x)}{\partial V}\frac{V}{N} + \frac{\partial P(v,T,x)}{\partial x}\frac{\partial x}{\partial N_i}. \quad (5.6)$$

Then substituting this into (5.4) gives,

$$\ln \bar{\phi}_i(v,T,x) = \frac{1}{RT}\int_{V'=\infty}^{V} \frac{RT}{V'} + \frac{\partial P(v,T,x)}{\partial V'}\frac{V'}{N}\,\mathrm{d}v' - \ln\left(\frac{P(v,T,x)v}{RT}\right) - \frac{1}{RT}\int_{V'=\infty}^{V}\frac{\partial P(v,T,x)}{\partial x}\frac{\partial x}{\partial N_i}\,\mathrm{d}v'. \quad (5.7)$$

We now again follow Demetriades and Graham [2016] and integrate by parts giving,

$$\frac{1}{RT}\int_{V'=\infty}^{V}\frac{RT}{V'} + \frac{\partial P(v,T,x)}{\partial V'}\frac{V'}{N}\,\mathrm{d}v' - \ln\left(\frac{P(v,T,x)v}{RT}\right) = \ln\phi(v,T,x), \quad (5.8)$$

where ϕ is the dimensional fugacity coefficient of a pure fluid. Considering the second integral in (5.7) we have

$$-\frac{1}{RT}\int_{V'=\infty}^{V}\frac{\partial P(v,T,x)}{\partial x}\frac{\partial x}{\partial N_i}\,\mathrm{d}v' = \frac{1}{RT}\int_{V'=\infty}^{V} -\frac{\partial P(v,T,x)}{\partial x}\,\mathrm{d}v'\frac{\partial x}{\partial N_i}. \quad (5.9)$$

We follow Demetriades and Graham [2016] and set

$$F(v,T,x) = \ln\phi(v,T,x) + \ln\left(\frac{P(v,T,x)v}{RT}\right) - \left(\frac{P(v,T,x)v}{RT}\right) + 1 \quad (5.10)$$

$$= \frac{1}{RT}\int_{V'=\infty}^{V}\frac{RT}{V'} - \frac{P(v',T,x)}{N}\,\mathrm{d}v'. \quad (5.11)$$

Then differentiating F with respect to x (we here depart from Demetriades and Graham [2016] who at this point differentiated with respect to the

model parameters) gives,

$$N\frac{\partial F(v,T,x)}{\partial x} = \frac{1}{RT}\int_{V'=\infty}^{V} -\frac{\partial P}{\partial x}\,\mathrm{d}v'. \tag{5.12}$$

Substituting this into (5.9) gives

$$-\frac{1}{RT}\int_{V'=\infty}^{V} \frac{\partial P(v,T,x)}{\partial x}\frac{\partial x}{\partial N_i}\,\mathrm{d}v' = N\frac{\partial F(v,T,x)}{\partial x}\frac{\partial x}{\partial N_i}, \tag{5.13}$$

and so we have

$$\ln\bar{\phi}_i(v,T,x) = \ln\phi(v,T,x) + N\frac{\partial F(v,T,x)}{\partial x}\frac{\partial x}{\partial N_i}. \tag{5.14}$$

If we apply our non-dimensionalisation (1.37)-(1.39), then we have

$$\begin{aligned}\ln\phi(v,T,x) = \frac{1}{T}\int_{v'=\infty}^{v} \frac{T}{v'} - P(v',T,x)\,\mathrm{d}v' \\ - \ln\left(\frac{P(v,T,x)v}{T}\right) + \frac{P(v,T,x)v}{T} - 1,\end{aligned} \tag{5.15}$$

and

$$F(v,T,x) = \frac{1}{T}\int_{v'=\infty}^{v} \frac{T}{v'} - P(v',T,x)\,\mathrm{d}v', \tag{5.16}$$

in non-dimensional units. From here v, T, P, ϕ_i and F are non-dimensional unless stated otherwise. So for a given temperature T with corresponding coexistence points with respective coexistence volumes, v_l and v_g, and respective coexistence impurity mole fractions, x_l and x_g, the two fugacity constraints are

$$\bar{f}_{\mathrm{CO_2}}(v_l,T,x_l) = \bar{f}_{\mathrm{CO_2}}(v_g,T,x_g), \tag{5.17}$$

$$\bar{f}_{\mathrm{impurity}}(v_l,T,x_l) = \bar{f}_{\mathrm{impurity}}(v_g,T,x_g). \tag{5.18}$$

From (1.21) we can write these constraints as,

$$P(v_l, T, x_l)(1 - x_l)\bar{\phi}_{\mathrm{CO_2}}(v_l, T, x_l) = P(v_g, T, x_g)(1 - x_g)\bar{\phi}_{\mathrm{CO_2}}(v_g, T, x_g), \tag{5.19}$$

$$P(v_l, T, x_l)x_l\bar{\phi}_{\mathrm{impurity}}(v_l, T, x_l) = P(v_g, T, x_g)x_g\bar{\phi}_{\mathrm{impurity}}(v_g, T, x_g), \tag{5.20}$$

As with the pure CO_2 case we have that $P(v_l, T, x_l) = P(v_g, T, x_g) = P_s$, where P_s is the coexistence pressure, substituting this in and taking logs gives,

$$\ln(1 - x_l) + \ln\bar{\phi}_{\mathrm{CO_2}}(v_l, T, x_l) = \ln(1 - x_g) + \ln\bar{\phi}_{\mathrm{CO_2}}(v_g, T, x_g), \tag{5.21}$$

$$\ln(x_l) + \ln\bar{\phi}_{\mathrm{impurity}}(v_l, T, x_l) = \ln(x_g) + \ln\bar{\phi}_{\mathrm{impurity}}(v_g, T, x_g). \tag{5.22}$$

We also have that

$$\frac{\partial x}{\partial N_{\mathrm{CO_2}}} = -\frac{x}{N}, \tag{5.23}$$

$$\frac{\partial x}{\partial N_{\mathrm{impurity}}} = \frac{1 - x}{N}. \tag{5.24}$$

Then substituting $P(v_l, T, x_l) = P(v_g, T, x_g) = P_s$ into the $\bar{\phi}_i$ and rearranging the constraints so that they are defined in terms of P_s (as we did in the pure CO_2 case in chapter 3), the CO_2 component fugacity constraint becomes,

$$\begin{aligned} P_s = \frac{1}{v_g - v_l}\Bigg(&\int_{v_g}^{\infty} \frac{T}{v'} - P(v', T, x_g)\,\mathrm{d}v' + \int_{\infty}^{v_l} \frac{T}{v'} - P(v', T, x_l)\,\mathrm{d}v' \\ &+ T\ln\left(\frac{v_g}{v_l}\right) + T\ln\left(\frac{1 - x_l}{1 - x_g}\right) - x_g\frac{\partial}{\partial x}\int_{v_g}^{\infty} \frac{T}{v'} - P(v', T, x_g)\,\mathrm{d}v' \\ &- x_l\int_{\infty}^{v_l} \frac{T}{v'} - P(v', T, x_l)\,\mathrm{d}v'\Bigg). \end{aligned} \tag{5.25}$$

Similarly the impurity fugacity constraint can be written in terms of P_s as

$$P_s = \frac{1}{v_g - v_l}\Bigg(\int_{v_g}^{\infty} \frac{T}{v'} - P(v', T, x_g)\,\mathrm{d}v' + \int_{\infty}^{v_l} \frac{T}{v'} - P(v', T, x_l)\,\mathrm{d}v' + T\ln\left(\frac{v_g}{v_l}\right) + T\ln\left(\frac{x_l}{x_g}\right) + (1 - x_g)\frac{\partial}{\partial x}\int_{v_g}^{\infty} \frac{T}{v'} - P(v', T, x_g)\,\mathrm{d}v' + (1 - x_l)\int_{\infty}^{v_l} \frac{T}{v'} - P(v', T, x_l)\,\mathrm{d}v' \Bigg). \quad (5.26)$$

Then along with the constraint that the pressures must also equal to P_s at each coexistence point,

$$P(v_l, T, x_l) = P_s, \quad (5.27)$$

$$P(v_g, T, x_g) = P_s, \quad (5.28)$$

we have four constraints for each pair of coexistence points. In the pure CO_2 case we had three constraints (see §3.3), there being only one fugacity constraint as there was only one species in the fluid.

In the pure CO_2 case the soft constraint model we derived was akin to a penalised likelihood model (3.50) [Cole et al., 2014, Green, 2006]. We fit one of the observations of P_s directly and then matched the other constraints to the observation of P_s that was directly fitted within a bound that was defined by a model error parameter, σ_e^2. We will follow the same approach for mixtures but with one more constraint to be matched to the fitted P_s observation. For consistency with the CO_2 approach we choose to fit the observation of P_s derived from the CO_2 component fugacity directly. Then

we have,

$$P_s = \frac{1}{v_g - v_l}\Bigg(\int_{v_g}^{\infty} \frac{T}{v'} - h_{\text{mix}}(v', T, x_g, \boldsymbol{\theta})\,\mathrm{d}v' + \int_{\infty}^{v_l} \frac{T}{v'} - h_{\text{mix}}(v', T, x_g, \boldsymbol{\theta})\,\mathrm{d}v' + T\ln\left(\frac{v_g}{v_l}\right) + T\ln\left(\frac{1-x_l}{1-x_g}\right) - x_g \frac{\partial}{\partial x}\int_{v_g}^{\infty} \frac{T}{v'} - h_{\text{mix}}(v', T, x_g, \boldsymbol{\theta})\,\mathrm{d}v' - x_l \int_{\infty}^{v_l} \frac{T}{v'} - h_{\text{mix}}(v', T, x_g, \boldsymbol{\theta})\,\mathrm{d}v' \Bigg) + \epsilon_{\text{mix}}, \quad (5.29)$$

where ϵ_{mix} has a normal distribution with mean 0 and variance σ^2_{mix}. We set

$$\begin{aligned} &h_{\text{mix,CO}_2}(v_l, v_g, T, x_l, x_g, \boldsymbol{\theta}) \\ &= \frac{1}{v_g - v_l}\Bigg(\int_{v_g}^{\infty} \frac{T}{v'} - h_{\text{mix}}(v', T, x_g, \boldsymbol{\theta})\,\mathrm{d}v' + \int_{\infty}^{v_l} \frac{T}{v'} - h_{\text{mix}}(v', T, x_g, \boldsymbol{\theta})\,\mathrm{d}v' \\ &\quad + T\ln\left(\frac{v_g}{v_l}\right) + T\ln\left(\frac{1-x_l}{1-x_g}\right) - x_g \frac{\partial}{\partial x}\int_{v_g}^{\infty} \frac{T}{v'} - h_{\text{mix}}(v', T, x_g, \boldsymbol{\theta})\,\mathrm{d}v' \\ &\quad - x_l \int_{\infty}^{v_l} \frac{T}{v'} - h_{\text{mix}}(v', T, x_g, \boldsymbol{\theta})\,\mathrm{d}v' \Bigg). \end{aligned} \quad (5.30)$$

Similarly we set

$$\begin{aligned} &h_{\text{mix,impurity}}(v_l, v_g, T, x_l, x_g, \boldsymbol{\theta}) \\ &= \frac{1}{v_g - v_l}\Bigg(\int_{v_g}^{\infty} \frac{T}{v'} - h_{\text{mix}}(v', T, x_g, \boldsymbol{\theta})\,\mathrm{d}v' + \int_{\infty}^{v_l} \frac{T}{v'} - h_{\text{mix}}(v', T, x_g, \boldsymbol{\theta})\,\mathrm{d}v' \\ &\quad + T\ln\left(\frac{v_g}{v_l}\right) + T\ln\left(\frac{x_l}{x_g}\right) + (1-x_g)\frac{\partial}{\partial x}\int_{v_g}^{\infty} \frac{T}{v'} - h_{\text{mix}}(v', T, x_g, \boldsymbol{\theta})\,\mathrm{d}v' \\ &\quad + (1-x_l)\int_{\infty}^{v_l} \frac{T}{v'} - h_{\text{mix}}(v', T, x_g, \boldsymbol{\theta})\,\mathrm{d}v' \Bigg). \end{aligned} \quad (5.31)$$

Then the three functions to be matched to $h_{\text{mix,CO}_2}(v_l, v_g, T, x_l, x_g, \boldsymbol{\theta})$ are $h_{\text{mix,impurity}}(v_l, v_g, T, x_l, x_g, \boldsymbol{\theta})$, $h_{\text{mix}}(v_l, T, x_l, \boldsymbol{\theta})$ and $h_{\text{mix}}(v_g, T, x_g, \boldsymbol{\theta})$. We set $\sigma_{\text{mix},e}$ to be the model error standard deviation for our mixtures. As in the pure case we will fix $\sigma_{\text{mix},e}$ to be some small value. Then our proposed

model for a pair of coexistence points of a binary mixture has pdf,

$$\begin{gathered}
p(P_s|\boldsymbol{\theta}, v_l, v_g, x_l, x_g, \sigma_{\text{mix}}, \sigma_{\text{mix},e}) \propto \phi\left(P_s; h_{\text{mix,CO}_2}(v_l, v_g, T, x_l, x_g, \boldsymbol{\theta}), \sigma^2_{\text{mix}}\right) \\
\phi(0; h_{\text{mix}}(v_g, T, x_g, \boldsymbol{\theta}) - h_{\text{mix,CO}_2}(v_l, v_g, T, x_l, x_g, \boldsymbol{\theta}), \sigma^2_{\text{mix},e}) \\
\phi(0; h_{\text{mix}}(v_l, T, x_l, \boldsymbol{\theta}) - h_{\text{mix,CO}_2}(v_l, v_g, T, x_l, x_g, \boldsymbol{\theta}), \sigma^2_{\text{mix},e}) \\
\phi(0; h_{\text{mix,impurity}}(v_l, v_g, T, x_l, x_g, \boldsymbol{\theta}) - h_{\text{mix,CO}_2}(v_l, v_g, T, x_l, x_g, \boldsymbol{\theta}), \sigma^2_{\text{mix},e}).
\end{gathered} \tag{5.32}$$

The first term in (5.32) is the direct observation of P_s due to the fugacity of the CO_2 component in the mixture. This corresponds to the term in the pure CO_2 likelihood which derived from direct fitting of the fugacity constraint. The other three terms in (5.32) come from requiring the pressure due to the impurity component fugacity constraint, and the pressure at each coexistence point to be equal to that derived from the fugacity of the CO_2 component. These correspond to the two terms in the pure CO_2 model which required the pressure at each coexistence point to match the pressure due to the fugacity constraint. Our model here is defined for just one set of coexistence points.

Suppose we have n_{mc} sets of coexistence points then we collect our data in the vectors,

$$\boldsymbol{T}_x = (T_{x1}, \ldots, T_{xn_{mc}})^T, \tag{5.33}$$

$$\boldsymbol{x}_l = (x_{l1}, \ldots, x_{ln_{mc}})^T, \tag{5.34}$$

$$\boldsymbol{x}_g = (x_{g1}, \ldots, x_{gn_{mc}})^T, \tag{5.35}$$

$$\boldsymbol{P}_s = (P_{s1}, \ldots, P_{sn_{mc}})^T, \tag{5.36}$$

$$\boldsymbol{v}_l = (v_{l1}, \ldots, v_{ln_{mc}})^T, \tag{5.37}$$

$$\boldsymbol{v}_g = (v_{g1}, \ldots, v_{gn_{mc}})^T, \tag{5.38}$$

where $\boldsymbol{P}_s$, is our vector of observed coexistence pressures, $\boldsymbol{x}_l$ and $\boldsymbol{x}_g$ are

the vectors of observed impurity concentrations, $\boldsymbol{v}_l$ and $\boldsymbol{v}_g$ are the vectors of observed coexistence volumes, and $\boldsymbol{T}_x$ is the vector of corresponding temperatures at which we have observed coexistence. Note that P_{si}, x_{li}, x_{gi}, v_{li}, v_{gi} and T_{xi} correspond to the same observation of coexistence for index i.

Then, assuming each set of coexistence points is independent, from (5.32) the full density for n_{mc} sets of coexistence points is given by,

$$\begin{aligned} p(\boldsymbol{P}_s|\boldsymbol{\theta}, \boldsymbol{v}_l, \boldsymbol{v}_g, \boldsymbol{x}_l, \boldsymbol{x}_g, \boldsymbol{T}_x, \sigma_{\text{mix}}, \sigma_{\text{mix},e}) \propto \\ \prod_{i=1}^{n_{mc}} \phi\left(P_{si}; h_{\text{mix,CO}_2}(v_{li}, v_{gi}, T_{xi}, x_{li}, x_{gi}, \boldsymbol{\theta}), \sigma^2_{\text{mix}}\right) \\ \phi(0; h_{\text{mix}}(v_{gi}, T_{xi}, x_{gi}, \boldsymbol{\theta}) - h_{\text{mix,CO}_2}(v_{li}, v_{gi}, T_{xi}, x_{li}, x_{gi}, \boldsymbol{\theta}), \sigma^2_{\text{mix},e}) \\ \phi(0; h_{\text{mix}}(v_{li}, T_{xi}, x_{li}, \boldsymbol{\theta}) - h_{\text{mix,CO}_2}(v_{li}, v_{gi}, T_{xi}, x_{li}, x_{gi}, \boldsymbol{\theta}), \sigma^2_{\text{mix},e}) \\ \phi(0; h_{\text{mix,impurity}}(v_{li}, v_{gi}, T_{xi}, x_{li}, x_{gi}, \boldsymbol{\theta}) \\ - h_{\text{mix,CO}_2}(v_{li}, v_{gi}, T_{xi}, x_{li}, x_{gi}, \boldsymbol{\theta}), \sigma^2_{\text{mix},e}). \end{aligned} \quad (5.39)$$

Then assuming independence between homogeneous and coexistence data our full model has density given by,

$$\begin{aligned} p(\boldsymbol{P}^{\text{obs}}_{\text{mix}}, \boldsymbol{P}_s|\boldsymbol{\theta}, \boldsymbol{X}_{\text{mix}}, \boldsymbol{v}_l, \boldsymbol{v}_g, \boldsymbol{x}_l, \boldsymbol{x}_g, \boldsymbol{T}_x, \sigma^2_{\text{mix}}, \sigma_{\text{mix},e}) \propto \\ p(\boldsymbol{P}_s|\boldsymbol{\theta}, \boldsymbol{v}_l, \boldsymbol{v}_g, \boldsymbol{x}_l, \boldsymbol{x}_g, \boldsymbol{T}_x, \sigma_{\text{mix}}, \sigma_{\text{mix},e}) p(\boldsymbol{P}^{\text{obs}}_{\text{mix}}|\boldsymbol{\theta}, \boldsymbol{X}_{\text{mix}}, \sigma^2_{\text{mix}}). \end{aligned} \quad (5.40)$$

There is one final part required for the model to be complete. In §1.6 we described how for the CO_2-H_2 coexistence data no coexistence volumes were recorded. It is not uncommon for coexistence volumes to be missing from experimental data recorded for mixtures of CO_2 with impurities. Thus we need to include in our model a means of accounting for these missing values.

5.4 Modelling Missing Coexistence Volumes

In §2.4.2 we described how to model missing data in the Bayesian case by treating each missing data value, here the v_{li} and v_{gi}, as a parameter to estimated from the data. Where v_{li} and v_{gi} appear in (5.39) then they are now to be considered as parameters rather than data. To define our full statistical model including the missing data we need to consider the distribution of the missingness indicator, M, where $M = 1$ indicates the data is missing and $M = 0$ indicates the data is observed. In this case we have that the coexistence volumes were not recorded at all. We thus have density for M as follows,

$$p(M) = 1, \quad M = 1. \tag{5.41}$$

As the coexistence volumes were chosen not to be recorded then we have that M must be equal to 1 deterministically. Clearly (5.41) does not depend on either the missing or observed data. Our missing data is thus missing completely at random (MCAR) and so we can ignore the missing-data mechanism when defining the model. We seek to estimate the vectors of missing coexistence volumes, $\boldsymbol{v}_l$ and $\boldsymbol{v}_g$ from the data. Let $\boldsymbol{X}_{\text{obs}}$ be all other fully observed explanatory variables ($\boldsymbol{T}_x$, $\boldsymbol{x}_l$, $\boldsymbol{x}_g$). Then from §2.4.2 we need to define the density, $p(\boldsymbol{v}_l, \boldsymbol{v}_g|\boldsymbol{X}_{\text{obs}}, \boldsymbol{\psi})$, where $\boldsymbol{\psi}$ is some vector of parameters which this density could depend upon. We here make the assumptions that the distribution of $\boldsymbol{v}_l$ and $\boldsymbol{v}_g$ does not depend on $\boldsymbol{X}_{\text{obs}}$, and that the model for each pair of coexistence volumes, v_{li} and v_{gi} is independent. Then we have

$$p(\boldsymbol{v}_l, \boldsymbol{v}_g|\boldsymbol{X}_{\text{obs}}, \boldsymbol{\psi}) = \prod_{i=1}^{n_{mc}} p(v_{li}, v_{gi}|\boldsymbol{\psi}). \tag{5.42}$$

To define this model we adapt slightly our approach from the pure CO_2 case. In the pure CO_2 case we used the critical volume to partition the parameter spaces of the v_g and v_l ensuring that they maintained the correct ordering, $v_l \leq v_g$. In the mixture case there we have that for every temperature at which coexistence can occur there is a critical volume, v_c, such that $v_l \leq v_c \leq v_g$ for all v_l and v_g at that temperature. For a pair of coexistence volumes, v_{li} and v_{gi} let v_{ci} be the corresponding critical volume (so the critical volume at T_{xi}), then our model for v_{li} and v_{gi} is,

$$v_{li} \sim \mathcal{U}(0.1, v_{ci}), \tag{5.43}$$

$$v_{gi} \sim \mathcal{U}(v_{ci}, 10), \tag{5.44}$$

The bounds are chosen as in the pure case, the lower bound 0.1 assures no singularities in the model arise in the data space, while the upper bound 10, is much larger than we would reasonably expect any coexistence volume to be. These define $p(v_{li}|v_{ci})$ and $p(v_{gi}|v_{ci})$ respectively so then we have,

$$p(\boldsymbol{v}_l, \boldsymbol{v}_g|\boldsymbol{v}_c) = \prod_{i=1}^{n_{mc}} p(v_{li}, v_{gi}|v_{ci}) = \prod_{i=1}^{n_{mc}} p(v_{li}|v_{ci})p(v_{gi}|v_{ci}), \tag{5.45}$$

and $\boldsymbol{v}_c$ corresponds to $\boldsymbol{\psi}$ in (5.42). Note that we can reduce the dimension of $\boldsymbol{v}_c$ by noting that when $T_{xi} = T_{xj}$ we have that $v_{ci} = v_{cj}$, and we only need as many v_c parameters as there are unique temperatures for which we are fitting coexistence data. For a full Bayesian analysis we need to define a prior, $p(\boldsymbol{v}_c)$. We here choose our prior for each element of $\boldsymbol{v}_c$, v_{ci} to be independent, and set,

$$v_{ci} \sim \mathcal{U}(0.1, 10), \tag{5.46}$$

where the bounds have been chosen as previously. Then the posterior distribution of our full model including missing coexistence volumes up to

proportionality is,

$$p(\boldsymbol{\theta}, \boldsymbol{v}_l, \boldsymbol{v}_g, \boldsymbol{v}_c | \boldsymbol{P}_{\text{mix}}^{\text{obs}}, \boldsymbol{P}_s, \boldsymbol{X}_{\text{mix}}, \boldsymbol{x}_l, \boldsymbol{x}_g, \boldsymbol{T}_x, \sigma_{\text{mix}}^2, \sigma_{\text{mix},e}) \propto$$
$$p(\boldsymbol{P}_{\text{mix}}^{\text{obs}}, \boldsymbol{P}_s | \boldsymbol{\theta}, \boldsymbol{X}_{\text{mix}}, \boldsymbol{v}_l, \boldsymbol{v}_g, \boldsymbol{x}_l, \boldsymbol{x}_g, \boldsymbol{T}_x, \sigma_{\text{mix}}^2, \sigma_{\text{mix},e}) p(\boldsymbol{v}_l, \boldsymbol{v}_g | \boldsymbol{v}_c) p(\boldsymbol{\theta}) p(\boldsymbol{v}_c). \tag{5.47}$$

We have now fully defined a statistical model for a general EoS which models the mixture of CO_2 with an impurity. We now present an extension of the DEoS to modelling such a mixture and a fit of the DEoS to CO_2-H_2 data at a single temperature using the PT-HMC sampler developed in chapter 4 and our proposed statistical model.

5.5 Demetriades EoS for Mixtures

Demetriades and Graham [2016] extended the DEoS to binary mixtures by introducing the mole fraction of the impurity to the temperature rules for the parameters. They set,

$$a(T, x) = (1 - x) a_{\text{CO}_2}(T) + x a_{\text{impurity}}(T), \tag{5.48}$$

where $a_{\text{CO}_2}(T)$ is the temperature rule for pure CO_2 and a_{impurity} is the similar temperature rule for the impurity. The other parameters, $b, \ldots, g$ are defined similarly. We can see that if $x = 0$, i.e. there is no impurity, then we obtain back the DEoS for pure CO_2 which we fitted in chapter 4. We will here use this same temperature rule. Let $h_{D,\text{mix}}(v, T, \boldsymbol{\theta})$ be the DEoS with this impurity concentration dependent temperature rule. Note that $h_{D,\text{mix}}$ depends upon x only through the model parameters $\boldsymbol{\theta}$. We will use the same temperature rules for the CO_2 parameters as in chapter 4: a_{CO_2} through e_{CO_2} use linear temperature rules, see (1.43) while f_{CO_2} and g_{CO_2} have quadratic temperature rules, see equations (4.167) and (4.170). We

could when fitting mixture data refit all of these parameters, along with the new parameters that will be derived from the likes of $a_{\text{impurity}}(T)$. However due to time constraints and preliminary runs showing that convergence is not achieved quickly we instead here choose to use the samples of the pure CO_2 parameters derived from our fit in chapter 4. We take the set of sampled parameter values which maximise the posterior and fix the pure CO_2 parameters to these values. This discards the quantification of the uncertainty in these parameter values but as the posterior distributions are very narrow we do not lose too much information.

For our initial fit of the mixture data we choose to fit only a single temperature as we did in the pure CO_2 case to assess convergence of our MCMC sampler and the fit of the model. The temperature we choose to fit is 288.15K, in dimensional units. This is because this is one of two temperatures for which we have both coexistence and homogeneous mixture data; see §1.6. We choose 288.15K over the other temperature for which there is mixture data, 303.15K, as that is nearer the critical temperature which we know from chapter 4 our current temperature rules for pure CO_2 struggle to fit. At 288.15K there are 31 homogeneous measurements (see table 1.5), for three different concentrations of H_2. There are 8 sets of coexistence points (see table 1.6). As we are fitting a single temperature we set,

$$a_{\text{H}_2}(T) = a_{\text{H}_2}, \tag{5.49}$$

and similarly for the other model parameters giving us a total of 7 model parameters, $\boldsymbol{\theta} = (a_{\text{H}_2}, b_{\text{H}_2}, \ldots, g_{\text{H}_2})^T$. We then need to define the likelihood of our data. Note that in both $h_{D,\text{mix},\text{CO}_2}(v_l, v_g, T, x_l, x_g, \boldsymbol{\theta})$ and $h_{D,\text{mix},\text{H}_2}(v_l, v_g, T, x_l, x_g, \boldsymbol{\theta})$ (as we are now working with a defined impurity, we have replaced the impurity in the subscript of $h_{\text{mix,impurity}}$ with H_2) there are terms of the form $\frac{\partial}{\partial x}\int_v^\infty \frac{T}{v'} - h_{D,\text{mix}}(v', T, x, \boldsymbol{\theta})\,\mathrm{d}v'$. The integral to ∞ does converge with the DEoS, see equations (1.46)-(1.47), and so we

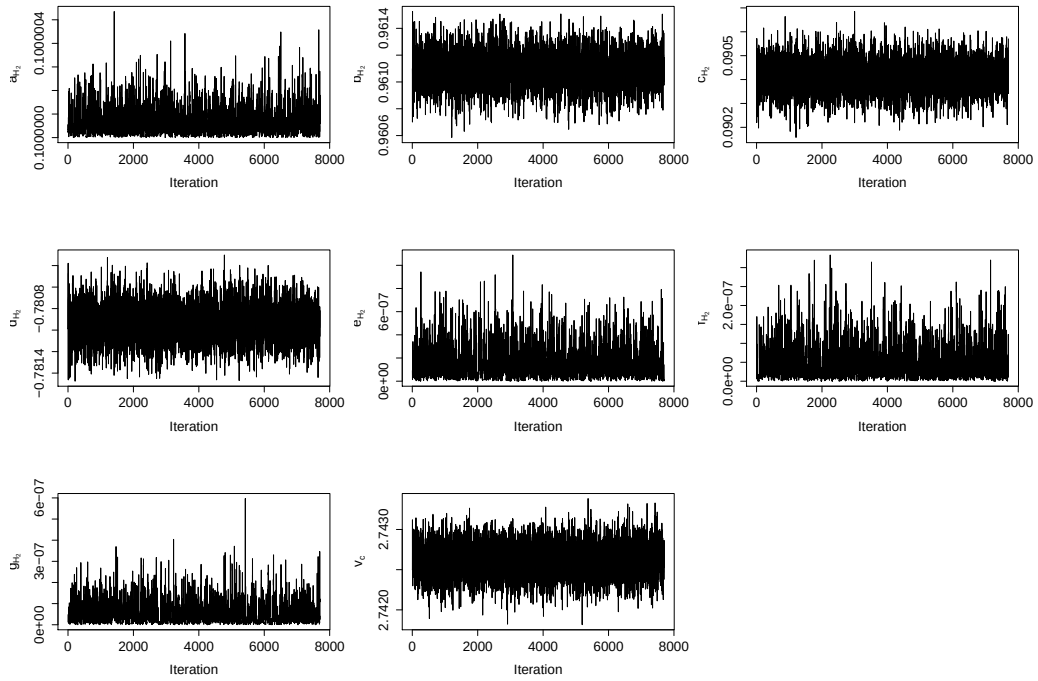


Figure 5.1: Traceplot of the sampled model parameters and v_c for a single run of the PT-HMC sampler applied to the Demetriades EoS with real CO_2-H_2 mixture data at 288.15K after discarding burn-in. Traceplots for other runs are similar. We can see here that our sampler is mixing well. As repeated runs of the sampler from different starting locations find the same mode it appears we have consistent convergence.

can evaluate all of these terms. Furthermore through application of the chain rule we have,

$$\frac{\partial}{\partial x}\int_{v_g}^{\infty}\frac{T}{v'}-h_{D,\mathrm{mix}}(v',T,x,\boldsymbol{\theta})\,\mathrm{d}v'=\sum_{i=1}^{7}\frac{\partial\theta_i}{\partial x}\frac{\partial}{\partial\theta_i}\int_{v_g}^{\infty}\frac{T}{v'}-h_{D,\mathrm{mix}}(v',T,\boldsymbol{\theta})\,\mathrm{d}v', \tag{5.50}$$

where θ_i is the i-th element of $\boldsymbol{\theta}$ because as described above $h_{D,\mathrm{mix}}(v',T,x,\boldsymbol{\theta})$ depends on x only through the elements of $\boldsymbol{\theta}$. Let $\boldsymbol{P}^{\mathrm{obs}}$ be our vector of observed homogeneous pressures. Let $\boldsymbol{v}^{\mathrm{obs}}$ and $\boldsymbol{x}^{\mathrm{obs}}$ be likewise for v and x respectively (as we are considering only a single temperature T is fixed).

Let $\boldsymbol{P}_s^{\mathrm{obs}}$ be our 8-dimensional vector of observed coexistence pressures. Let $\boldsymbol{x}_l^{\mathrm{obs}}$ and $\boldsymbol{x}_g^{\mathrm{obs}}$ be the observed coexistence H_2 concentrations. Set $\boldsymbol{v}_l$ and $\boldsymbol{v}_g$ to be the 8-dimensional parameter vectors corresponding to the

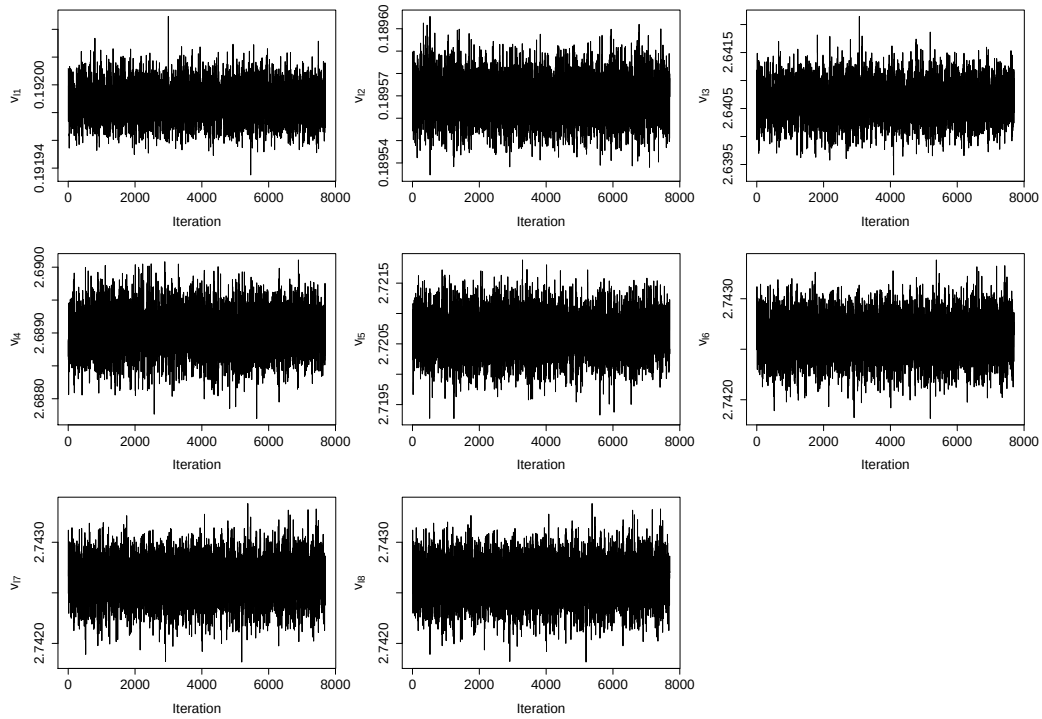


Figure 5.2: Traceplot of the sampled bubblepoint coexistence volume parameters ($\boldsymbol{v}_l$) for a single run of the PT-HMC sampler applied to the Demetriades EoS with real CO_2-H_2 mixture data at 288.15K after discarding burn-in. Traceplots for other runs are similar. We can see here that our sampler is mixing well. As repeated runs of the sampler from different starting locations find the same mode it appears we have consistent convergence.

coexistence volumes. Then our likelihood is,

$$L(\boldsymbol{\theta}, \boldsymbol{v}_l, \boldsymbol{v}_g) = \prod_{i=1}^{8} \left(P_{si}^{\text{obs}}; h_{D,\text{mix},\text{CO}_2}(v_{li}, v_{gi}, T, x_{li}, x_{gi}, \boldsymbol{\theta}), \sigma^2_{\text{mix}}\right)$$
$$\phi(0; h_{D,\text{mix}}(v_{gi}, T, x_{gi}, \boldsymbol{\theta}) - h_{\text{mix},\text{CO}_2}(v_{li}, v_{gi}, T, x_{li}, x_{gi}, \boldsymbol{\theta}), \sigma^2_{\text{mix},e})$$
$$\phi(0; h_{\text{mix}}(v_{li}, T, x_{li}, \boldsymbol{\theta}) - h_{\text{mix},\text{CO}_2}(v_{li}, v_{gi}, T, x_{li}, x_{gi}, \boldsymbol{\theta}), \sigma^2_{\text{mix},e})$$
$$\phi(0; h_{\text{mix,impurity}}(v_{li}, v_{gi}, T, x_{li}, x_{gi}, \boldsymbol{\theta}) - h_{\text{mix},\text{CO}_2}(v_{li}, v_{gi}, T, x_{li}, x_{gi}, \boldsymbol{\theta}), \sigma^2_{\text{mix},e})$$
$$\prod_{j=1}^{31} \phi(P_j^{\text{obs}}; h_{\text{mix}}(v_j^{\text{obs}}, T, x_j^{\text{obs}}, \boldsymbol{\theta}), \sigma^2_{\text{mix}}). \quad (5.51)$$

Based on the quoted uncertainty in Fandiño et al. [2015] we fix $\sigma_{\text{mix}} = 0.00045$. As in the pure CO_2 case we set $\sigma_{\text{mix},e}$ to be a tenth of the size of σ_{mix}.

We now must choose priors for the model parameters. For the model

parameters we use the same uniform priors as previously, see equations (4.15)-(4.21). Our parameters here, $a_{H_2}, \ldots, g_{H_2}$ are interpreted as the parameters were we to model the pure H_2 fluid at 288.15K. The parameter values we used previously for pure CO_2 were chosen such that they should be reasonably able to model behaviour throughout the CCS operating window, ensuring that all terms in the DEoS contribute to the model by removing areas of the parameter space where they become constant over the values of v and T in the operating window. The priors also ensure that the terms in the DEoS can take values within the whole range of P defined by the operating window. One of the boundaries of this operating window is defined by the physical properties of an ideal gas. H_2 is more volatile than CO_2 (and this is true of all impurities we would seek to model for the CCS transport problem) which means it behaves more like an ideal gas than CO_2. We would thus expect that CO_2-H_2 mixtures would also become more gas-like as the concentration of H_2 increases. Thus the CO_2-H_2 mixtures we seek to model should fall within the same operating window for which we defined our priors, having behaviour somewhere between pure CO_2 and an ideal gas. As the operating window is the same, and the prior distributions for the parameters were chosen based on this window, it is a reasonable assumption that the same priors can be used for the parameters in the case of the CO_2-H_2 mixture. Furthermore the priors used previously had hard bounds which ensured no singularities arose in the terms of the DEoS, a condition that we also require to hold here. We thus use the same priors as previously.

As we fit only a single temperature we need only consider a single critical volume, v_c. We thus have 24 total parameters; 7 model parameters, $\boldsymbol{\theta}$; 16 coexistence volume parameters, $\boldsymbol{v}_l$ and $\boldsymbol{v}_g$; and the critical volume at this temperature, v_c. Thus our model for the missing coexistence volumes in

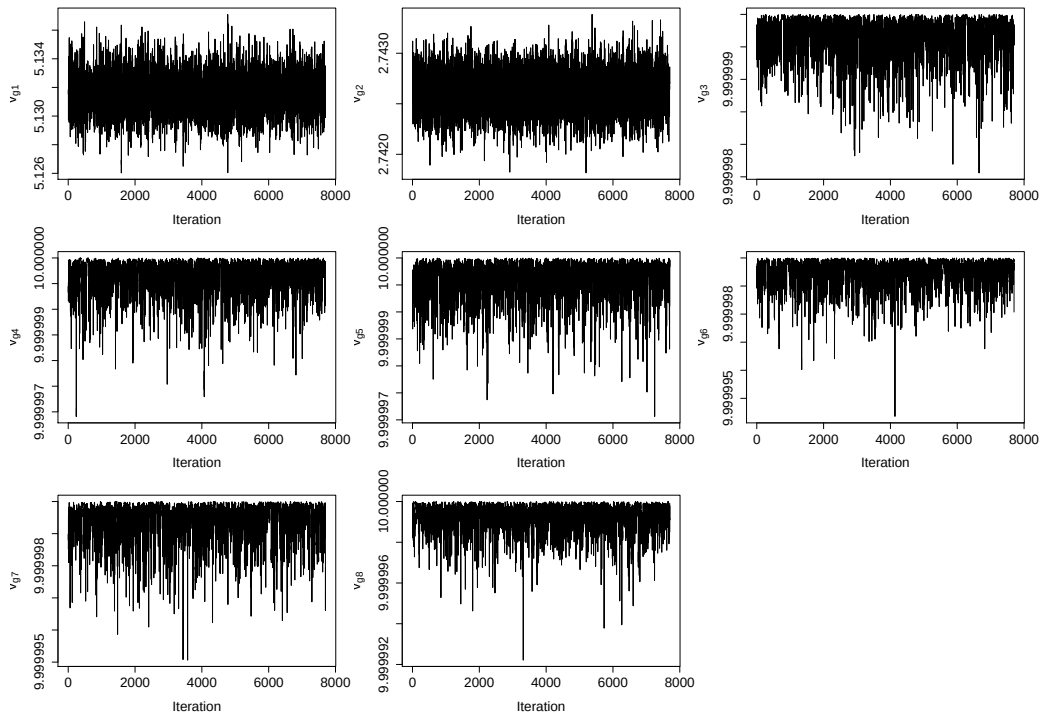


Figure 5.3: Traceplot of the sampled dewpoint coexistence volume parameters ($\boldsymbol{v}_g$) for a single run of the PT-HMC sampler applied to the Demetriades EoS with real CO_2-H_2 mixture data at 288.15K after discarding burn-in. Traceplots for other runs are similar. We can see here that our sampler is mixing well. As repeated runs of the sampler from different starting locations find the same mode it appears we have consistent convergence.

this case is,

$$p(\boldsymbol{v}_l, \boldsymbol{v}_g | v_c) = \prod_{i=1}^{8} p(v_{li}|v_c)p(v_{gi}|v_c), \tag{5.52}$$

where $p(v_{li}|v_c)$ and $p(v_{gi}|v_c)$ are as defined by equations (5.43) and (5.44) respectively.

We can now define the posterior of our parameters up to proportionality as, $\boldsymbol{\theta}, \boldsymbol{v}_l, \boldsymbol{v}_g, v_c$, as follows,

$$p(\boldsymbol{\theta}, \boldsymbol{v}_l, \boldsymbol{v}_g, v_c | \boldsymbol{P}_s^{\text{obs}}, \boldsymbol{x}_l, \boldsymbol{x}_g, \boldsymbol{P}^{\text{obs}}, \boldsymbol{v}^{\text{obs}}, \boldsymbol{x}^{\text{obs}}) \propto$$
$$L(\boldsymbol{\theta}, \boldsymbol{v}_l, \boldsymbol{v}_g)p(\boldsymbol{\theta})p(\boldsymbol{v}_l, \boldsymbol{v}_g|v_c)p(v_c). \tag{5.53}$$

Given this posterior we can now use our PT-HMC sampler to infer the parameters. We will describe the fitting and present the results from this

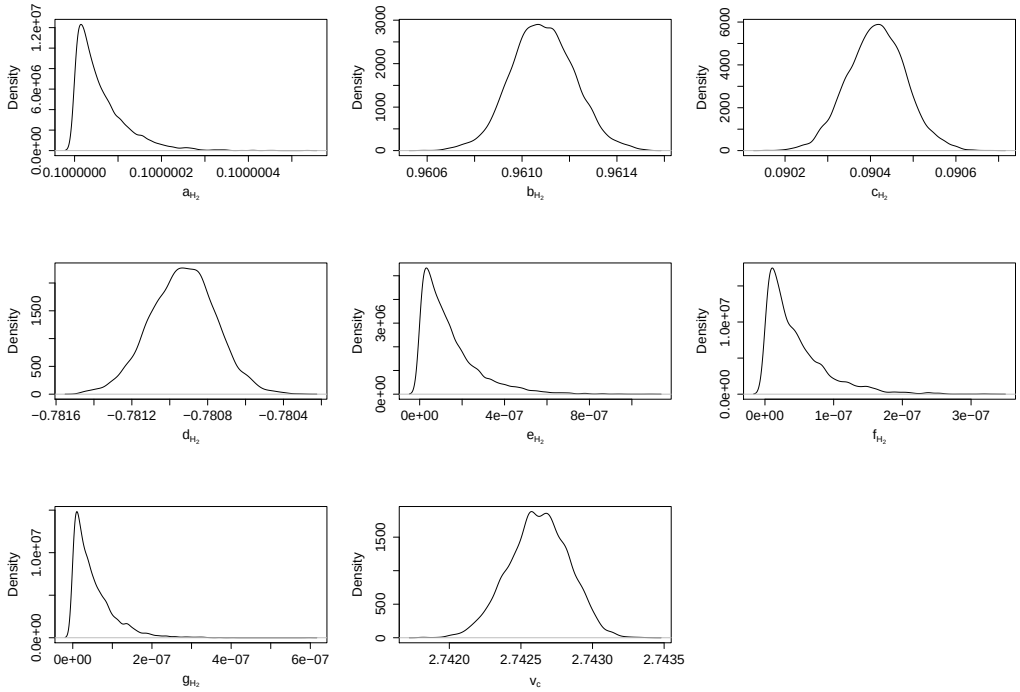


Figure 5.4: Marginal density plots of the sampled model parameters and v_c for a single run of the PT-HMC sampler applied to the Demetriades EoS with real CO_2-H_2 data at 288.15K. Only the high density region is shown.

fitting in the next section.

5.6 Fitting the Demetriades EoS for Mixtures

We fit the DEoS to the CO_2-H_2 mixture data at 288.15K using the PT-HMC sampler we developed in chapter 4. We once again temper the likelihood only, thus replica k of our sampler will have density up to proportionality of,

$$p(\boldsymbol{\theta}, \boldsymbol{v}_l, \boldsymbol{v}_g, v_c | \boldsymbol{P}_s^{\text{obs}}, \boldsymbol{x}_l, \boldsymbol{x}_g, \boldsymbol{P}^{\text{obs}}, \boldsymbol{v}^{\text{obs}}, \boldsymbol{x}^{\text{obs}}) \propto (L(\boldsymbol{\theta}, \boldsymbol{v}_l, \boldsymbol{v}_g))^{\frac{1}{\beta_k}} p(\boldsymbol{\theta}) p(\boldsymbol{v}_l, \boldsymbol{v}_g | v_c) p(v_c), \quad (5.54)$$

where β_k is the tempering level for replica k. The replicas with highest tempering level should then explore the prior. We find that as in fitting the

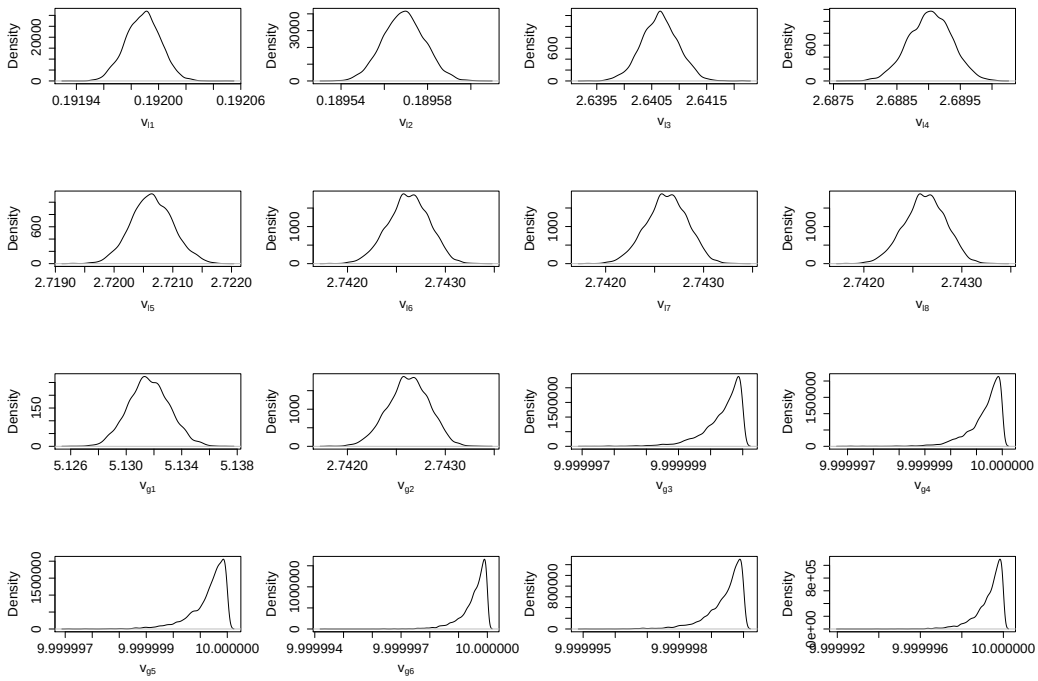


Figure 5.5: Marginal density plots of the sampled coexistence volume parameters and v_c for a single run of the PT-HMC sampler applied to the Demetriades EoS with real CO_2-H_2 data at 288.15K. Only the high density region is shown.

pure CO_2 data if we set the maximum tempering level to be $\beta = \left(\frac{3}{2}\right)^{52}$ the sampler at that level is able to explore the entire prior parameter space. We also find that a ratio of 1.5 between neighbouring tempering levels allows a good number of swaps to occur between replicas. As such we again use 53 replicas with β_k for each replica as follows,

$$\beta_k = \left(\frac{3}{2}\right)^{k-1}, \quad k = 1, \ldots, 53. \tag{5.55}$$

We run the PT-HMC sampler for 9900 iterations starting from 5 different sets of initial parameter values. The sampler was run for 9900 iterations as when performing these fits time constraints allowed us to run the sampler for only 2 days on the University of Nottingham high performance computing facility (again in parallel on 16 CPUs) and this was the number of iterations able to be performed in this time. This number of iterations is

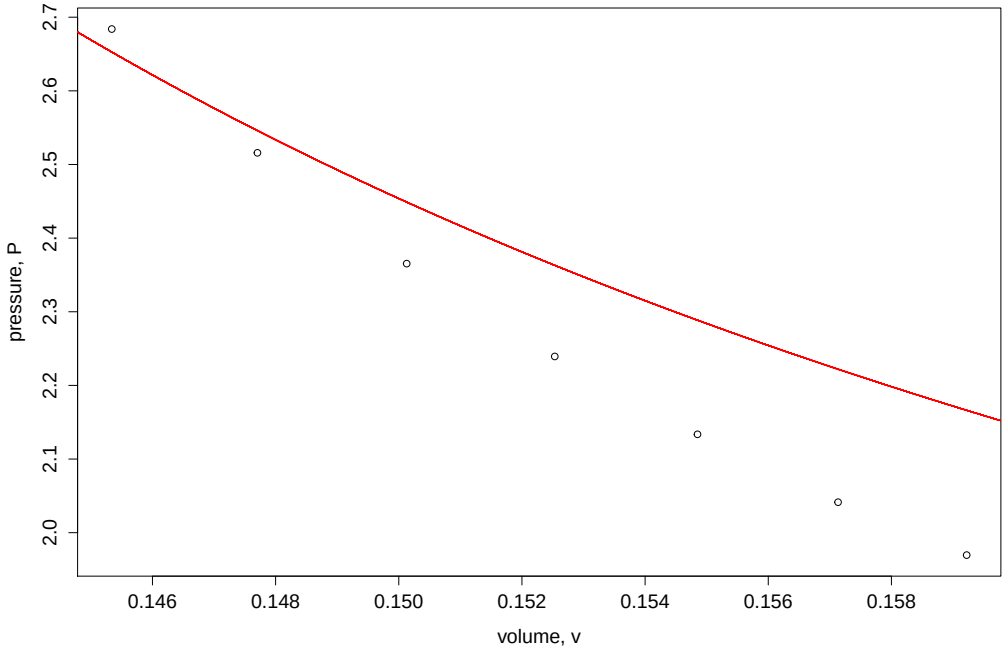


Figure 5.6: Plot of 50 posterior realisations at 288.15K and H_2 concentration of 9.97% of sample from applying PT-HMC to DEoS to real CO_2-H_2 mixture data at 288.15K. Hollow black circles are homogeneous data at 288.15K with H_2 concentration of 9.97%. Red lines are 50 posterior realisations of the pressure-volume curve.

enough for convergence to be achieved, we find that each run of the sampler from a different initial values converges to the same region of the parameter space.

In figures 5.1, 5.2 and 5.3 are the traceplots of our parameters for one of these runs of the PT-HMC sampler after discarding the first 2200 iterations as burn-in. We can see that the sampler appears to be mixing well. As each run of the sampler finds the same mode it appears we have consistent convergence of the sampler. In figures 5.4 and 5.5 we plot the high posterior probability regions of the marginal densities for the parameters. We can see that the high posterior region is a small subset of the prior parameter space. We also note that the high posterior probability regions of model parameters, a_{H_2}, e_{H_2}, f_{H_2} and g_{H_2} are at the lower bounds of their priors. Furthermore we note that the high posterior probability region v_{g3} through

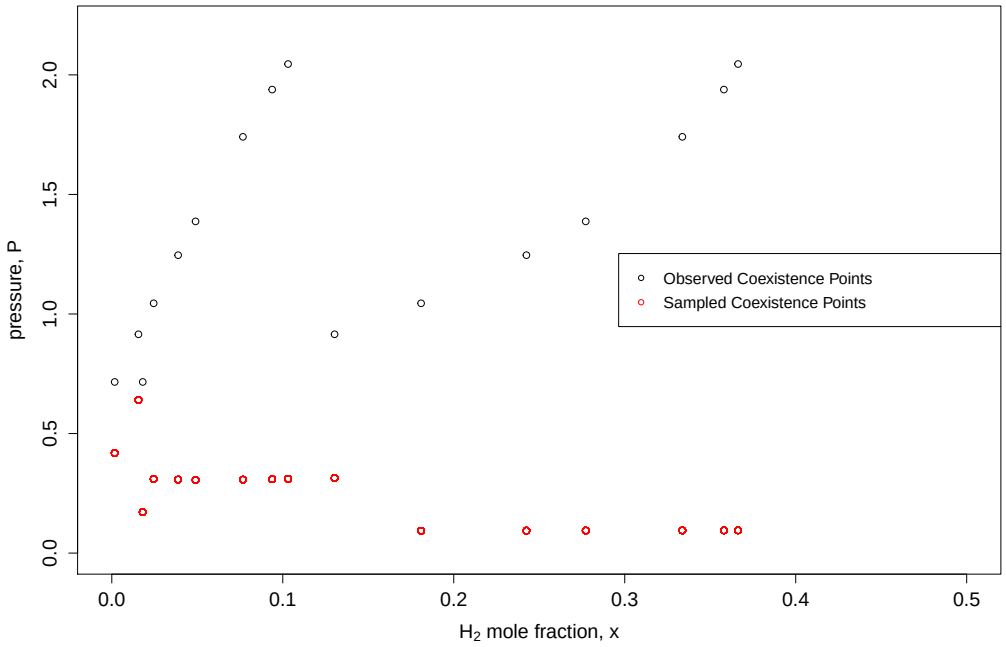


Figure 5.7: Comparison of the observed coexistence points with 50 sets of sampled coexistence points from the fit of the DEoS to real CO_2-H_2 mixture data at 288.15K. Observed coexistence points are in black. Sampled coexistence points in red.

v_{g8} are at the upper bound of their priors, nearly 10, a volume far higher than we would expect. This would suggest that the priors are either overly confident about possible values the parameters could take or the model is misspecified in some way, most likely that mixtures need to be modelled non-linearly when incorporated into the temperature rule. Our results here would suggest that a linear mixture model is too simple and lacks the flexibility to fit the data.

In figure 5.6 we plot 50 posterior realisation pressure-volume curves (red lines) at 288.15K with H_2 concentration of 9.97% against the homogeneous data at that temperature with that concentration (black circles). We can see that the posterior realisations do not describe the data well at all. Thus it seems that the DEoS model which we fitted here is not able to describe the data well. Plotting the other two concentrations for which we have data at 288.15K, 7.48% and 1.97% we see similarly poor fit. In figure 5.7

we plot 50 samples of the set of coexistence points from our fit against the observed coexistence points. In black is the observed coexistence data while the sampled coexistence points are in red. Clearly the sampled coexistence points differ significantly from the observed coexistence points.

We have here proposed a statistical model for binary mixtures of CO_2 with an impurity which accounts for both coexistence and homogeneous data, as well as being able to account for missing coexistence volumes. Fitting this model we find that, although our PT-HMC sampler consistently converges and seems to mix well having done so, the fitted results perform poorly when compared to real CO_2-H_2 data and that many of the parameters are on the boundaries of their prior distributions. This is most likely a result of a misspecified model as there are no obvious issues with the sampler. Further investigation is required to determine if the poor performance is due to the DEoS mixture model used here, if the assumption of linearity is too simple to accurately model the data, or if the issue is with our statistical model, if there is perhaps something in the way we have modelled coexistence that is leading to the poor fit. If the issue is with the DEoS mixture model used here and we can propose a mixing rule for the DEoS which combined with our statistical model fits well we would then be able to apply our statistical model to general parametric EoS for both pure CO_2 and general binary mixtures of pure CO_2 with an impurity.

We have from our fitted results here and in chapter 4 seen that it is not a simple task to propose a form for a parametric EoS which can model all of the properties of CO_2 and mixtures of CO_2 accurately. In the next chapter we propose an alternative, more flexible, means of modelling an EoS using Gaussian processes.

Chapter 6

Non-Parametric EoS

6.1 Introduction

In chapters 3, 4 and 5 we proposed a statistical model for general parametric pressure-explicit EoS, developed an MCMC algorithm to sample the parameters of this model, and fitted it to real data using the specific DEoS. In this chapter we explore an alternative approach to modelling EoS, *Gaussian processes* [Rasmussen and Williams, 2006]. Gaussian processes (GPs), which we introduced in §2.5, are a non-parametric statistical model. Compared with parametric models they are much more flexible to adapt to the data because the dimension of GPs increases with the size of the data set. As such they are sometimes considered to have infinite-parameters. If we consider the data to be the parameters of the model, the more data we use to train the GP the greater the effective dimensionality of our parameter space.

Because they learn their form from the data GPs offer several advantages when it comes to modelling EoS, compared to parametric models they require fewer assumptions, offer more flexibility and provide a natural way to characterise uncertainty in the predictions. Another advantage of GP models of special importance to modelling EoS is that the linear

transformation of a GP is itself a GP [Murray-Smith and Pearlmutter, 2005, Rasmussen and Williams, 2006] allowing us to directly incorporate derivative observations into the model.

In this chapter we introduce a free-energy-formulation GP EoS. We describe how to model observations of pressure and coexistence using this model. We then describe how to model the critical point using our GP EoS. We then perform fits of the model to NIST reference data at individual subcritical temperatures. A particular challenge is that the data are non-stationary, their distribution, the mean of the data, and the covariance between data points, is dependent on the explanatory variable, volume, v, and is not invariant to translations of the input space [Rasmussen and Williams, 2006], so a standard GP fits poorly. We however show that a transformation that accounts for non-stationarity leads to good fits for subcritical temperatures.

6.2 Modelling Gaussian Processes for EoS

As with parametric EoS we have a choice as to whether we define our GP model to be pressure-explicit or in the free-energy-formulation, see §1.4.1. We are here concerned only with pure CO_2 and so we have two explanatory variables: volume, v, and temperature, T. Then a pressure-explicit EoS will model pressure as a function of volume and temperature, $P(v, T)$ while a free-energy-formulation EoS, will model the free energy as a function of volume and temperature, $A(v, T)$. Here we will call here P and A our "response variable". A GP model assumes that values our response variable takes for different inputs of the explanatory variables are jointly normally distributed. From §2.5 to fully define a GP for pressure we need to define a mean function, $m(v, T)$, and covariance function, $k(v, T, v', T')$.

We then have

$$P(v,T) \sim \mathcal{GP}(m(v,T), k(v,T,v',T')), \tag{6.1}$$

in the pressure-explicit case or

$$A(v,T) \sim \mathcal{GP}(m(v,T), k(v,T,v',T')), \tag{6.2}$$

in the free-energy-formulation case in the notation from §2.5.

We note that we do not ever actually observe the free energy directly. Instead we observe pressure values which are the negative derivative of the free energy with respect to volume; see equation (1.1). Thus our only observations are of derivatives of the free energy and not of the free energy directly. A useful property of GP is that the linear transformation of a GP is itself a GP [Murray-Smith and Pearlmutter, 2005, Rasmussen and Williams, 2006]. This property allows us to incorporate observations of linear transformations of our response variable, including derivatives, into our model. Despite the fact that we do not observe the free energy directly we choose to model our EoS using the free-energy-formulation. This is because, as we saw in chapter 3, the fugacity constraint at coexistence depends upon the integral of P. Thus were we to use the pressure-explicit formulation this would restrict us to using covariance functions, $k(v,T,v',T')$ for which we could analytically derive their integral with respect to v and v'. Using the free-energy-formulation offers more flexibility as we can more easily differentiate covariance functions. We firstly describe how to include derivative observations in a GP, and so how to include observations of the pressure in our free-energy-formulation GP. We then describe how we model coexistence in the free-energy-formulation.

6.2.1 Modelling Pressure Observations

From (1.1) we have,

$$P = -\frac{\partial A}{\partial v}. \tag{6.3}$$

Then from the property that the linear transformation of a GP is itself a GP [Murray-Smith and Pearlmutter, 2005, Rasmussen and Williams, 2006] and as differentiation is a linear transformation we have that the covariance between two observations of pressure when,

$$A(v,T) \sim \mathcal{GP}(m_A(v,T), k_A(v,T,v',T')), \tag{6.4}$$

is,

$$k_P(v,T,v',T') = \frac{\partial^2 k_A(v,T,v',T')}{\partial v \partial v'}, \tag{6.5}$$

and the observations of pressure have mean function,

$$m_P(v,T) = -\frac{\partial m_A(v,T)}{\partial v}. \tag{6.6}$$

Then suppose we have n pressure observations which we collect in the vector,

$$\boldsymbol{P} = (P_1^{\text{obs}}, \ldots, P_n^{\text{obs}})^T. \tag{6.7}$$

For each observed pressure there is a corresponding volume and temperature measurement which we collect in the vectors,

$$\boldsymbol{v} = (v_1, \ldots, v_n)^T, \tag{6.8}$$

$$\boldsymbol{T} = (T_1, \ldots, T_n)^T. \tag{6.9}$$

We then have n x 2 design matrix,

$$\boldsymbol{X} = (\boldsymbol{v}, \boldsymbol{T}). \tag{6.10}$$

Let $\boldsymbol{x}_i$ be the i-th row of $\boldsymbol{X}$, so $\boldsymbol{x}_i = (v_i, T_i)$. We use the notation,

$$k_P(\boldsymbol{x}_i, \boldsymbol{x}_j) = k_P(v_i, T_i, v_j, T_j)^T, \tag{6.11}$$

and

$$m_P(\boldsymbol{x}_i) = m_P(v_i, T_i). \tag{6.12}$$

We assume that the pressures are observed with some error, so,

$$P_i^{\text{obs}} = -\frac{\partial A(v_i, T_i)}{\partial v} + \epsilon_i, \tag{6.13}$$

where the ϵ_i are i.i.d. normally distributed with mean 0 and variance σ^2. We set $\boldsymbol{P}^{\text{obs}} = (P_1^{\text{obs}}, \ldots, P_n^{\text{obs}})^T$. Then from §2.5 we have that

$$\boldsymbol{P}^{\text{obs}} \sim \mathcal{N}(m_P(\boldsymbol{X}), K_P(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I}_n), \tag{6.14}$$

where

$$m_P(\boldsymbol{X}) = (m_P(\boldsymbol{x}_1), \ldots, m_P(\boldsymbol{x}_n))^T, \tag{6.15}$$

and $K_P(\boldsymbol{X}, \boldsymbol{X})$ is the n x n matrix with elements,

$$K_{Pij}(\boldsymbol{X}, \boldsymbol{X}) = k_P(\boldsymbol{x}_i, \boldsymbol{x}_j), \tag{6.16}$$

and $\boldsymbol{I}_n$ is the n x n identity matrix. This is our free-energy-formulation GP model for pressure observations. We now describe how to include co-existence data in the model.

6.2.2 Modelling Coexistence Data

Suppose we have coexistence data for a given temperature T, with coexistence pressure P_s and coexistence volumes v_l and v_g. From §1.3.2 we have that for the free-energy-formulation, rather than the fugacity constraint,

we require that the Gibbs free energy is equal at each coexistence point (1.16). This constraint gives us that,

$$A(v_l, T) + P(v_l, T)v_l = A(v_g, T) + P(v_g, T)v_g. \tag{6.17}$$

Substituting in that $P(v_l, T) = P(v_g, T) = P_s$ and rearranging we have,

$$P_s = \frac{A(v_l, T) - A(v_g, T)}{v_g - v_l}. \tag{6.18}$$

So the Gibbs free energy constraint tells us that the coexistence pressure P_s is equal to the difference between the Helmholtz free energy at the coexistence points divided by the difference in the coexistence volumes. Then, as in chapter 3, we can define P_s^{obs} in three ways, as the negative derivative of the free energy at v_l, as the negative derivative of the free energy at v_g (this corresponds to the pressure being P_s at each coexistence point), and from the form derived from the Gibbs free energy constraint (6.18) (which corresponds to the pressure derived from the fugacity constraint). Then we have

$$P_s = P_l = -\frac{\partial A(v_l, T)}{\partial v}, \tag{6.19}$$

$$P_s = P_g = -\frac{\partial A(v_g, T)}{\partial v}, \tag{6.20}$$

$$P_s = P_G = \frac{A(v_l, T) - A(v_g, T)}{v_g - v_l}, \tag{6.21}$$

where P_l is the pressure at v_l, P_g is the pressure at P_g, and P_G is the pressure as derived from the Gibbs free energy constraint (6.18). These are all linear transformations of A and so can be modelled using our GP for A. We then have that the covariance between $\frac{A(v_l,T)-A(v_g,T)}{v_g-v_l}$ and $\frac{A(v_l',T')-A(v_g',T')}{v_g'-v_l'}$

is

$$k_G(v_l, v_g, T, v_l', v_g', T') = \frac{1}{(v_g - v_l)(v_g' - v_l')}(k_A(v_l, T, v_l', T') - k_A(v_l, T, v_g', T') \\ -k_A(v_g, T, v_l', T') + k_A(v_g, T, v_g', T')), \tag{6.22}$$

the mean function of $\frac{A(v_l,T)-A(v_g,T)}{v_g-v_l}$ is,

$$m_G(v_l, v_g, T) = \frac{1}{v_g - v_l}(m_A(v_l, T) - m_A(v_g, T)), \tag{6.23}$$

and the covariance between $\frac{A(v_l,T)-A(v_g,T)}{v_g-v_l}$ and $-\frac{\partial A(v',T')}{\partial v'}$ is

$$k_{GP}(v_l, v_g, T, v', T') = \frac{1}{v_g - v_l}\left(\frac{\partial k_A(v_g, T, v', T')}{\partial v'} - \frac{\partial k_A(v_l, T, v', T')}{\partial v'}\right). \tag{6.24}$$

Suppose we observe coexistence for n_c temperatures, which we collect in the vector, $\boldsymbol{T}_s = (T_{s1}, \ldots, T_{sn_c})^T$. For each temperature we have coexistence volumes,

$$\boldsymbol{v}_l = (v_{l1}, \ldots, v_{ln_c})^T, \tag{6.25}$$

$$\boldsymbol{v}_g = (v_{g1}, \ldots, v_{gn_c})^T, \tag{6.26}$$

and observed coexistence pressures, $\boldsymbol{P}_s^{\text{obs}} = (P_{s1}^{\text{obs}}, \ldots, P_{sn_c}^{\text{obs}})^T$, which we observe with some observation error,

$$P_{si}^{\text{obs}} = P_{si} + \epsilon_{si}, \tag{6.27}$$

where P_{si} is the true coexistence pressure for T_{si} and the ϵ_{si} are i.i.d. normally distributed with mean 0 and variance σ^2. We can consider each P_{si}^{obs}

is then defined in 3 ways,

$$P_{si}^{\text{obs}} = P_{li}^{\text{obs}} = -\frac{\partial A(v_{li}, T_{si})}{\partial v} + \epsilon_{si}, \tag{6.28}$$

$$P_{si}^{\text{obs}} = P_{gi}^{\text{obs}} = -\frac{\partial A(v_{gi}, T_{si}}{\partial v} + \epsilon_{si}, \tag{6.29}$$

$$P_{si}^{\text{obs}} = P_{Gi}^{\text{obs}} = \frac{A(v_{li}, T_{si}) - A(v_{gi}, T)}{v_{gi} - v_{li}} + \epsilon_{si}. \tag{6.30}$$

We can then consider that each observation of coexistence gives us three observations of pressure which we collect in the vectors,

$$\boldsymbol{P}_l^{\text{obs}} = (P_{l1}^{\text{obs}}, \ldots, P_{ln_c}^{\text{obs}})^T, \tag{6.31}$$

$$\boldsymbol{P}_g^{\text{obs}} = (P_{g1}^{\text{obs}}, \ldots, P_{gn_c}^{\text{obs}})^T, \tag{6.32}$$

$$\boldsymbol{P}_G^{\text{obs}} = (P_{G1}^{\text{obs}}, \ldots, P_{Gn_c}^{\text{obs}})^T. \tag{6.33}$$

Each of these vector of pressure observations has design matrix as follows,

$$\boldsymbol{X}_l = (\boldsymbol{v}_l, \boldsymbol{T}_s), \tag{6.34}$$

$$\boldsymbol{X}_g = (\boldsymbol{v}_g, \boldsymbol{T}_s), \tag{6.35}$$

$$\boldsymbol{X}_G = (\boldsymbol{v}_l, \boldsymbol{v}_g, \boldsymbol{T}_s). \tag{6.36}$$

Our mean and covariance functions as functions of the rows of these design matrices are then,

$$k_G(\boldsymbol{x}_{Gi}, \boldsymbol{x}_{Gj}) = k_G(v_{li}, v_{gi}, T_{si}, v_{lj}, v_{gj}, T_{sj}), \tag{6.37}$$

$$k_{GP}(\boldsymbol{x}_{Gi}, \boldsymbol{x}_j) = k_G(v_{li}, v_{gi}, T_{si}, v_j, T_j), \tag{6.38}$$

$$m_G(\boldsymbol{x}_{Gi}) = m_G(v_{li}, v_{gi}, T_{si}). \tag{6.39}$$

Then from our GP on A we have that the pressure observations due to

coexistence are jointly normal with distribution,

$$\begin{pmatrix} \boldsymbol{P}_l^{\text{obs}} \\ \boldsymbol{P}_g^{\text{obs}} \\ \boldsymbol{P}_G^{\text{obs}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m_P(\boldsymbol{X}_l) \\ m_P(\boldsymbol{X}_g) \\ m_G(\boldsymbol{X}_G) \end{pmatrix}, \boldsymbol{K}_s + \sigma^2 \boldsymbol{\Sigma}_s \right), \tag{6.40}$$

where

$$\boldsymbol{K}_s = \begin{pmatrix} K_P(\boldsymbol{X}_l, \boldsymbol{X}_l) & K_P(\boldsymbol{X}_l, \boldsymbol{X}_g) & K_{GP}(\boldsymbol{X}_G, \boldsymbol{X}_l)^T \\ K_P(\boldsymbol{X}_g, \boldsymbol{X}_l) & K_P(\boldsymbol{X}_g, \boldsymbol{X}_g) & K_{GP}(\boldsymbol{X}_G, \boldsymbol{X}_g)^T \\ K_{GP}(\boldsymbol{X}_G, \boldsymbol{X}_l) & K_{GP}(\boldsymbol{X}_G, \boldsymbol{X}_g) & K_G(\boldsymbol{X}_G, \boldsymbol{X}_G) \end{pmatrix}, \tag{6.41}$$

and

$$\boldsymbol{\Sigma}_s = \begin{pmatrix} \boldsymbol{I}_{n_c} & \boldsymbol{I}_{n_c} & \boldsymbol{I}_{n_c} \\ \boldsymbol{I}_{n_c} & \boldsymbol{I}_{n_c} & \boldsymbol{I}_{n_c} \\ \boldsymbol{I}_{n_c} & \boldsymbol{I}_{n_c} & \boldsymbol{I}_{n_c} \end{pmatrix}, \tag{6.42}$$

and where $K_G(\boldsymbol{X}_G, \boldsymbol{X}_G)$ is the n_c x n_c matrix with elements,

$$K_{Gij}(\boldsymbol{X}_G, \boldsymbol{X}_G) = k_G(\boldsymbol{x}_{Gi}, \boldsymbol{x}_{Gj}), \tag{6.43}$$

and $K_{GP}(\boldsymbol{X}_G, \boldsymbol{X}_l)$ is the n_c x n_c matrix with elements,

$$K_{GPij}(\boldsymbol{X}_G, \boldsymbol{X}_l) = k_{GP}(\boldsymbol{x}_{Gi}, \boldsymbol{x}_{lj}). \tag{6.44}$$

Note that the noise on P_{li}^{obs}, P_{gi}^{obs} and P_{Gi}^{obs} is perfectly correlated. In §2.5 we described how if it is assumed that there is no noise on the observations used to train a GP then predictions from the trained GP will interpolate those observations [Rasmussen and Williams, 2006]. Due to the noise here predictions from our GP will not interpolate the observed coexistence pressures, however because the noise is perfectly correlated they will preserve that $P_l^{\text{obs}} = P_g^{\text{obs}} = P_G^{\text{obs}}$. As such all predictions from our GP model will exactly obey the coexistence constraints. This is an improvement over the

parametric EoS model where we had to approximately fit the coexistence constraints (see §3.3.2). We set, $\boldsymbol{P}_x^{\text{obs}} = (\boldsymbol{P}_l^{\text{obs}}, \boldsymbol{P}_g^{\text{obs}}, \boldsymbol{P}_G^{\text{obs}})^T$, and

$$\boldsymbol{m}_s = \begin{pmatrix} m_P(\boldsymbol{X}_l) \\ m_P(\boldsymbol{X}_g) \\ m_G(\boldsymbol{X}_G) \end{pmatrix}. \tag{6.45}$$

Then we can define a joint model for coexistence data and homogeneous pressure observations as described in §6.2.1 as follows,

$$\begin{pmatrix} \boldsymbol{P}_x^{\text{obs}} \\ \boldsymbol{P}^{\text{obs}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{m}_s \\ m_P(\boldsymbol{X}) \end{pmatrix}, \begin{pmatrix} \boldsymbol{K}_s + \sigma^2 \boldsymbol{\Sigma}_s & \boldsymbol{K}_{sP} \\ \boldsymbol{K}_{sP}^T & K_P(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I}_n \end{pmatrix} \right), \tag{6.46}$$

where

$$\boldsymbol{K}_{sP} = \begin{pmatrix} K_P(\boldsymbol{X}_l, \boldsymbol{X}) \\ K_P(\boldsymbol{X}_g, \boldsymbol{X}) \\ K_{GP}(\boldsymbol{X}_G, \boldsymbol{X}). \end{pmatrix}. \tag{6.47}$$

We have thus fully defined a GP model for EoS which can model both homogeneous and coexistence data. We now demonstrate fits of this model to NIST reference data [National Institute of Standards and Technology, 2015] for pure CO_2 at individual subcritical temperatures.

6.3 Fitting the GP EoS Model

We fit the GP EoS model we defined in §6.2 to NIST reference data [National Institute of Standards and Technology, 2015]. This reference data comes from the Span-Wagner EoS [Span and Wagner, 1996]. Here we use this data in a similar way to the synthetic data used in chapter 4. We use it to test our fitting algorithm and to develop our specific GP model for CO_2 (i.e. choice of covariance and mean function). However as the data does not come from the model we are fitting there is no guarantee that a

good fit exists. We obtain NIST reference data for 6 temperatures, 273K, 282K, 290K, 294K, 299K, 300K. These temperatures are subcritical and as such coexistence occurs at each. The data for each of these temperatures consists of 100 homogeneous observations, consisting of a volume and pressure each, and observations of the coexistence volumes and coexistence pressure. We apply the same non-dimensionalisation (1.37)-(1.39) as in the parametric case to this data.

As the NIST reference data is derived from the SWEoS there is no observation noise on the data. However we retain the σ^2 terms in the model described in §6.2 as a nugget term [Gramacy and Lee, 2012], to prevent the covariance matrix of our model becoming computationally singular and to reflect the model as it would be applied to real data. We begin by fitting subcritical temperatures individually.

6.3.1 Fitting an Individual Subcritical Temperature

We first fit only the temperature with $T = 0.983138$ which in dimensional units corresponds to $299K$. As this temperature is subcritical the effective data we wish to model are $\boldsymbol{P}_x^{\text{obs}} = (P_l^{\text{obs}}, P_g^{\text{obs}}, P_G^{\text{obs}})^T$ and $\boldsymbol{P}^{\text{obs}} = (P_1^{\text{obs}}, \ldots, P_{100}^{\text{obs}})^T$ with design matrices,

$$\boldsymbol{X} = (\boldsymbol{v}, \boldsymbol{T}), \tag{6.48}$$

$$\boldsymbol{X}_l = (v_l, T), \tag{6.49}$$

$$\boldsymbol{X}_g = (v_g, T), \tag{6.50}$$

$$\boldsymbol{X}_G = (v_l, v_g, T), \tag{6.51}$$

where $\boldsymbol{v} = (v_1, \ldots, v_n)^T$, $\boldsymbol{T}$ is the 100-dimensional vector where every element is $T_i = T$ and v_l and v_g are the coexistence volumes. Then from

(6.46) we have that,

$$\begin{pmatrix} \boldsymbol{P}_x^{\text{obs}} \\ \boldsymbol{P}^{\text{obs}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{m}_s \\ m_P(\boldsymbol{X}) \end{pmatrix}, \begin{pmatrix} \boldsymbol{K}_s + \sigma^2 \boldsymbol{\Sigma}_s & \boldsymbol{K}_{sP} \\ \boldsymbol{K}_{sP}^T & K_P(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I}_n \end{pmatrix} \right). \tag{6.52}$$

To fully define our GP model we then need to choose mean, m_A, and covariance, k_A, functions. We choose to set the mean function of our Gaussian process to such that its negative derivative with respect to v is the ideal gas law (1.17),

$$m_A(v, T) = -T \ln v. \tag{6.53}$$

We do so because the ideal gas law describes the pressure of pure CO_2 well in the limit of large v and so incorporating it as the mean function should lead to our GP EoS following similar behaviour in this limit. The form defined in (6.53) ensures that the mean of the pressure observations in our model is the ideal gas law. A commonly used covariance function is the *squared exponential* covariance function,

$$k_A(v, T, v', T') = k_{se}(v, v') = \eta^2 \exp\left(-\frac{(v - v')^2}{l} \right), \tag{6.54}$$

where $\boldsymbol{\phi} = (\eta^2, l)^T$ are the hyperparameters of the squared exponential covariance function. We do not have any temperature dependence in our covariance function here as we are fitting a single temperature and so the temperature is fixed here. Both elements of $\boldsymbol{\phi}$ are restricted to be positive.

Having now fully defined our GP model we seek to fit the model using Bayesian parameter estimation to estimate the hyperparameters. Then as discussed in §2.5.1 we need to place a prior on our hyperparameters. In total we have three parameters we seek to estimate, $\tilde{\boldsymbol{\phi}} = (\eta^2, l, \sigma^2)$, including σ^2 along with the squared-exponential hyperparameters. We lack any strong prior information on the distributions of the hyperparameters. We thus

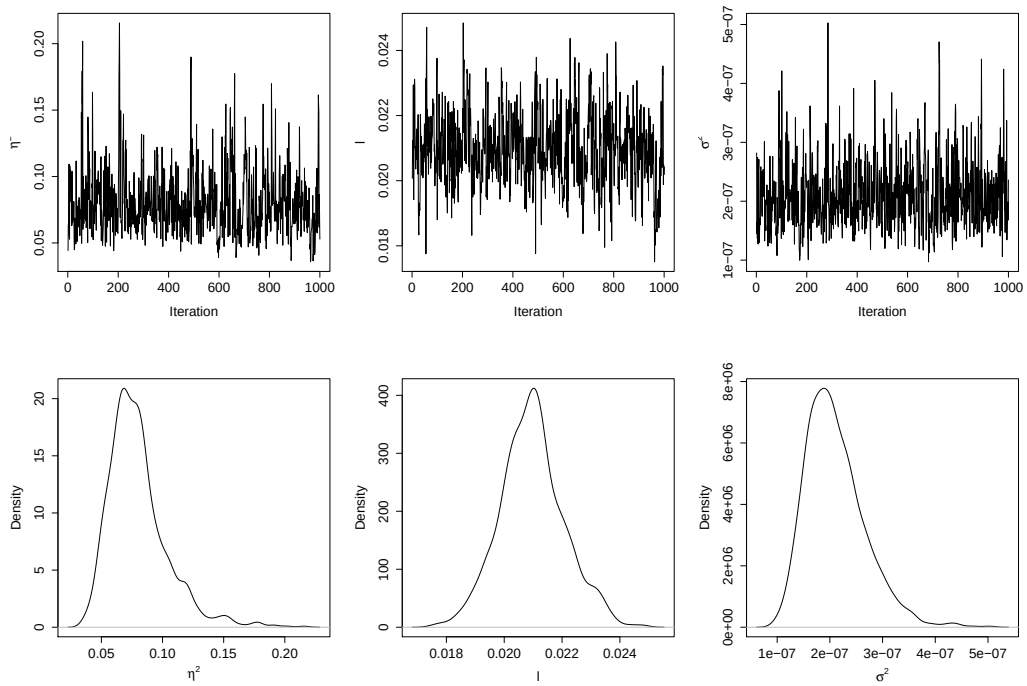


Figure 6.1: Traceplots and marginal density plots from a single run using HMC to fit the GP EoS to 294K, after discarding burn-in. We can see from the traceplots that the sampler is mixing well and appears to have converged. Plots of runs from other initial parameter values are similar.

place the following weakly informative priors on the hyperparameters,

$$\eta^2 \sim \text{half-Cauchy}(0, 5), \tag{6.55}$$

$$l \sim \text{half-Cauchy}(0, 5), \tag{6.56}$$

$$\sigma^2 \sim \text{half-Cauchy}(0, 5), \tag{6.57}$$

where the half-Cauchy distribution is the Cauchy distribution restricted to positive values. The half-Cauchy(0,5) distribution then has pdf,

$$p(x) = \frac{2}{5\pi\left(1 + \left(\frac{x}{5}\right)^2\right)}, \quad x > 0. \tag{6.58}$$

Although placing most prior mass near to 0 the half-Cauchy distribution has a heavy tail and so if the data strongly suggests the values of the parameters should be large this prior will allow for that.

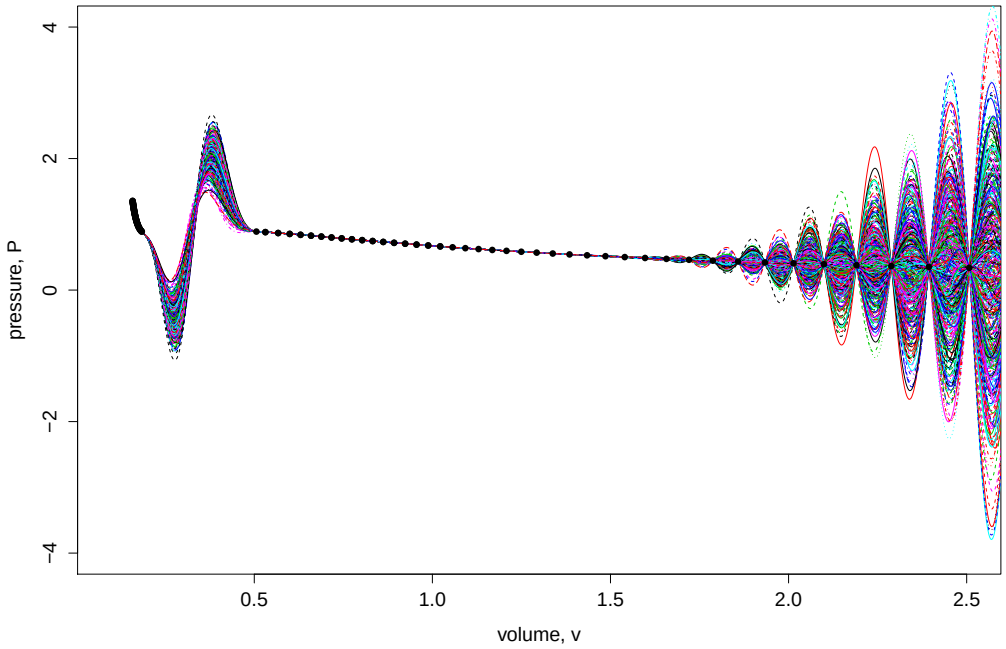


Figure 6.2: 500 posterior realisations of the pressure-volume curve from using HMC to fit the GP EoS to 294K. Coloured lines are the realisations. Solid black circles are the NIST data used to fit the model. We can see that for smaller v the realisations fit the data well but are inaccurate for large v.

Then our likelihood $L(\tilde{\boldsymbol{\phi}})$ is defined by (6.52) and with the prior $p(\tilde{\boldsymbol{\phi}})$ defined by (6.55)-(6.57) we can sample from the posterior distribution up to proportionality,

$$p(\tilde{\boldsymbol{\phi}}|\boldsymbol{X}, \boldsymbol{X}_l, \boldsymbol{X}_g, \boldsymbol{X}_G) \propto L(\tilde{\boldsymbol{\phi}})p(\tilde{\boldsymbol{\phi}}), \tag{6.59}$$

using MCMC methods. We here sample using the no-U-turn sampler extension of HMC as was described in chapter 2. Figure 6.1 presents traces and marginal densities from a single run of the no-U-turn sampler for 2000 iterations and discarding the first 1000 iterations as burn-in. We can see that the sampler is mixing well. Runs starting from other initial parameter values are similar having converged to the same region of the parameter space.

We now take a sample of 50 sets of parameters from our posterior sam-

ple. For each of these 50 sets of parameters we sample 10 realisations of the pressure-volume curve. We do so by setting $\boldsymbol{P}^*$ to be the pressure evaluated at $\boldsymbol{X}^* = (\boldsymbol{v}^*, \boldsymbol{T}^*)^T$, where $\boldsymbol{v}^*$ is a vector of 1000 volumes evenly spaced from 0.158 to 2.6, and $\boldsymbol{T}^*$ is the 1000-dimensional vector where every entry has value T. Then the distribution of $\boldsymbol{P}^*|\boldsymbol{P}_x^{\text{obs}}, \boldsymbol{P}^{\text{obs}}$ is (2.37),

$$\begin{aligned}&\boldsymbol{P}^*|\boldsymbol{P}_x^{\text{obs}}, \boldsymbol{P}^{\text{obs}} \sim \\ &\mathcal{N}\left(m_P(\boldsymbol{X}^*) + \boldsymbol{K}_{sP}^{*T}\boldsymbol{K}^{-1}(\boldsymbol{P} - \boldsymbol{m}), K_P(\boldsymbol{X}^*, \boldsymbol{X}^*) - \boldsymbol{K}_{sP}^{*T}\boldsymbol{K}^{-1}\boldsymbol{K}_{sP}^*\right), \quad (6.60)\end{aligned}$$

where

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_x^{\text{obs}} \\ \boldsymbol{P}^{\text{obs}} \end{pmatrix}, \quad (6.61)$$

$$\boldsymbol{m} = \begin{pmatrix} \boldsymbol{m}_s \\ m_P(\boldsymbol{X}) \end{pmatrix}, \quad (6.62)$$

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{K}_s + \sigma^2\boldsymbol{\Sigma}_s & \boldsymbol{K}_{sP} \\ \boldsymbol{K}_{sP}^T & K_P(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2\boldsymbol{I}_n \end{pmatrix}, \quad (6.63)$$

$$\boldsymbol{K}_{sP}^* = \begin{pmatrix} K_P(\boldsymbol{X}_l, \boldsymbol{X}^*) \\ K_P(\boldsymbol{X}_g, \boldsymbol{X}^*) \\ K_{GP}(\boldsymbol{X}_G, \boldsymbol{X}^*), \end{pmatrix} \quad (6.64)$$

and samples from this distribution give us our posterior pressure-volume curves. Figure 6.2 plots these posterior pressure-volume curve realisations, the coloured lines, against the NIST data at 294K, solid black circles. We see that for small volume the realisations fit the data well but at large volume between data points the posterior realisations deviate wildly from what we would expect. Fitting our model to the other subcritical temperatures in our NIST dataset we see similar behaviour. This is most likely because the squared-exponential covariance function is *stationary*, it depends only

on $v - v'$. Stationarity assumes that the function behaves similarly in different regions of the input space. We do not have such behaviour however, at small volume small changes in volume lead to large changes in pressure, while for large volume large changes in volume result in only small changes in pressure. We thus have *non-stationary* data. In the next section we review proposed methods for modelling non-stationary data using a GP and show how a transformation of the input allows us to define a covariance function which can model our non-stationarity here.

6.3.2 Modelling Non-stationarity

Several methods have been proposed in the literature for modelling non-stationarity in a GP context. These methods fall roughly into three groups. The first of these is to partition the input space, with different GPs trained in each region of the partition. Gramacy and Lee [2008] propose partitioning using classification and regression trees (treed GP), Kim et al. [2005] use Voronoi tessellation to divide the input space and construct piecewise GPs over these divisions, while Rasmussen and Ghahramani [2002] propose a mixture of experts method, they divide the input space through use of a gating network and in each division individual experts, here GPs, make predictions.

The second method is to transform either the output or input spaces. Snelson et al. [2004] propose a criterion for the former, whilst Schmidt and O'Hagan [2003] propose the latter, placing another GP upon the transformation of the output space itself. The final method is to use a non-stationary covariance function and several have been proposed [Paciorek and Schervish, 2004, Williams, 1998]. It is not however intuitively obvious how to define a non-stationary covariance function such that it possesses the non-stationary behaviour our model possesses.

One method to define non-stationary covariance functions could be con-

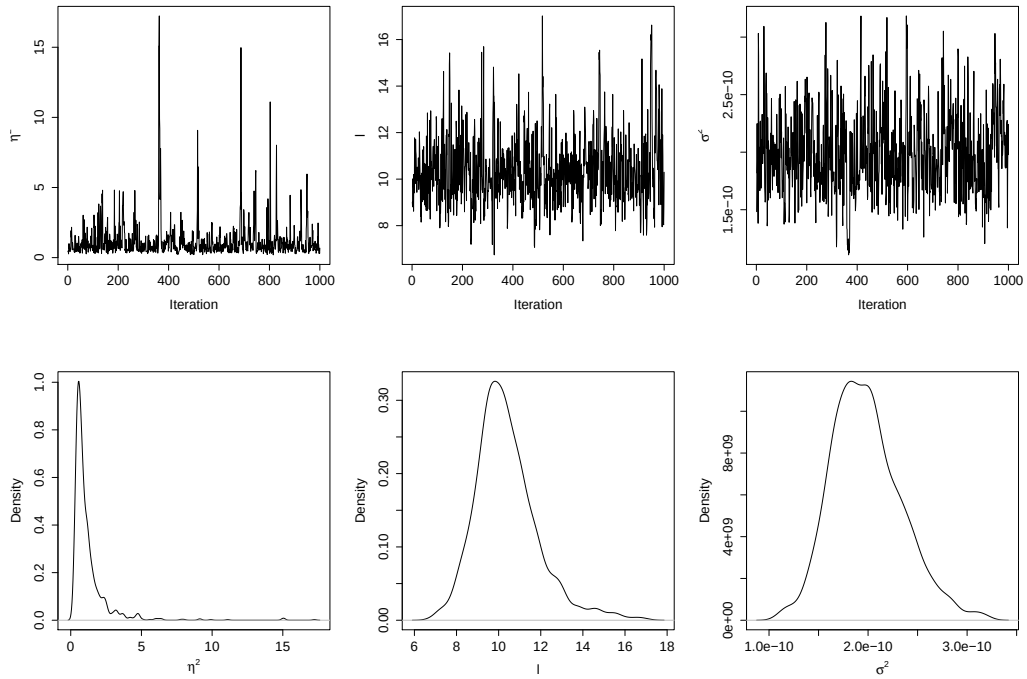


Figure 6.3: Traceplots and marginal density plots from a single run using HMC to fit the GP EoS to 294K, after discarding burn-in with non-stationary covariance function. We can see from the traceplots that the sampler is mixing well and appears to have converged. Plots of runs from other initial parameter values are similar.

sidered a combination of the second and third methods of modelling non-stationarity. Given stationary covariance function k then if we apply a non-linear transformation, $u(\boldsymbol{x})$ to $\boldsymbol{x}$ we have that $k(u(\boldsymbol{x}), u(\boldsymbol{x}'))$ is a non-stationary covariance function [MacKay, 1998]. We apply the transformation, $u(v) = \frac{1}{v}$. Applying this transformation to small v and v' leads to $u(v)$ and $u(v')$ becoming much further apart, while applying to large v and v' leads to $u(v)$ and $u(v')$ becoming closer. Then we have that $u(v)$ and $u(v')$ are far apart when the pressure changes significantly between them and close together when the pressure changes more slowly. We then set our non-stationary covariance function to be the squared-exponential

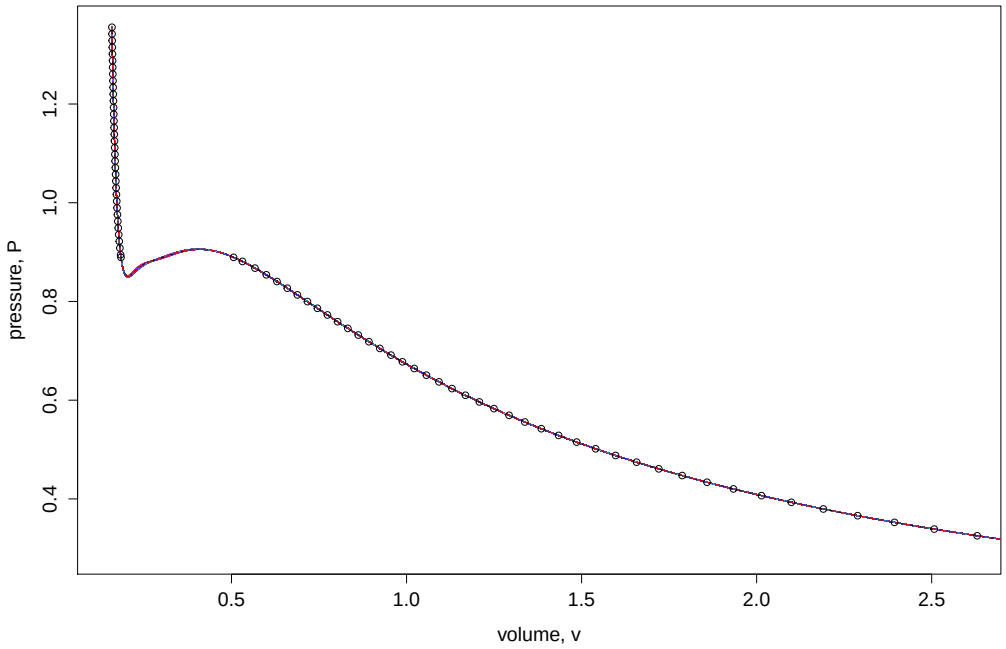


Figure 6.4: 500 posterior realisations of the pressure-volume curve from using HMC to fit the GP EoS to 294K with non-stationary covariance function. Coloured lines are the realisations. Solid black circles are the NIST data used to fit the model. We can see that the realisations fit the data well.

covariance function with this transformation,

$$\begin{aligned} k_A(v, T, v', T') &= k_{se}(u(v), u(v')) \\ &= \eta^2 \exp\left(-\frac{(\frac{1}{v} - \frac{1}{v'})^2}{l}\right). \end{aligned} \tag{6.65}$$

We once more fit the 294K NIST reference data. Then our model is as given in (6.52) and we have the same set of 3 hyperparameters to estimate. As such we use the same priors on the hyperparameters and again sample them using the no-U-turn sampler.

Figure 6.3 presents traces and marginal densities from a single run of the no-U-turn sampler for 2000 iterations and discarding the first 1000 iterations as burn-in. We once again see that the sampler is mixing well. We perform more runs starting from other initial parameter values and find they have converged to the same region of the parameter space. We once

more take a sample of 50 sets of parameters from our posterior sample, and sample 10 posterior realisations of the pressure-volume curve using each of these sets of parameters at input $\boldsymbol{X}^*$ using (6.60). Figure 6.4 plots these realisations, the coloured lines, against the NIST reference data, black circles. We can see that all of the realisations are very similar in form and fit the data well. Applying this model to the other subcritical temperatures in our dataset we get similarly good fit. Our non-stationary covariance function allows us to model EoS using GP at individual subcritical temperatures accurately.

Non-parametric EoS offer two main advantages over parametric EoS. The first, which we have seen here, is that they allow us to impose the coexistence constraints exactly rather than approximately as we did with the parametric statistical model for EoS proposed in §3.3.2. This allows for more accurate modelling of coexistence. We could further extend this by considering that we know the critical point is a saddlepoint, see §1.3.1, and so could extend our model to condition on further derivative observations at the critical point (see §7.2 for further discussion). The other advantage is that because GP learn their form from the data we might be able to avoid the issues encountered in chapter 4 where our parametric model was not flexible enough to fit well at temperatures around the critical temperature. Extensions of the fits here to multiple temperatures for pure CO_2 will enable us to see if a GP model can allow us to fit data throughout the temperature without the need to define complex temperature rules to account for the changes in the physical properties of CO_2. However while for pure CO_2 there is a large amount of accurate data, and so our GP can learn a very exact form, for mixtures data is often much sparser. GPs do not use small amounts of data efficiently, as such further investigation needs to be carried out to see how a GP model might extrapolate to the case of mixtures with only sparse data available.

Chapter 7

Conclusions

7.1 Thesis Summary and Results

Carbon capture and storage offers a potential means to reduce the emissions that contribute to climate change. Efficient transport of the CO_2 from the power stations where it is collected, to the seabed reservoirs where it will be stored, requires not just accurate predictions of the physical properties of CO_2 but accurate quantification of the uncertainty in these predictions. This calls for a need to develop rigorous statistical models for EoS, which can account for the physical properties of CO_2, and of mixtures of CO_2, and methods by which to reliably fit them. Developing such models and methods was the aim of this thesis.

We firstly developed a statistical model which can be applied to any currently existing, or any future proposed, parametric pressure-explicit EoS. To do so we developed a novel method to account for uncertainties in coexistence. This is complicated by the presence of perfectly correlated measurements on two fluid phases. We showed how it is difficult or impossible to fit coexistence data exactly and so proposed an approximate model for coexistence which can return similar fits. We also demonstrated how working in the Bayesian paradigm allows us to account for measurement

error on coexistence volume data by treating the coexistence volumes as parameters and thus infer coexistence volume information by fitting our statistical model.

We developed an MCMC sampler which is a combination of parallel tempering and Hamiltonian Monte Carlo and demonstrated how this sampler could be used to reliably and consistently fit the parametric statistical EoS model we proposed. Using this sampler we sample fro the posterior distribution of the model parameters, thus quantifying uncertainty in the parameters which then allows for quantification of the uncertainty in predictions from the model. We used our sampler to fit our statistical model for the specific case of the Demetriades EoS to real pure CO_2 data, demonstrating that good mixing is achieved and we can reliably fit the model. We found that fit had good agreement with data for most temperatures with the exception of those within the immediate vicinity of the critical point.

The CO_2 in the CCS system is usually not pure but instead a mixture of CO_2 with other fluids. As the presence of other fluids can have significant impact on the physical properties of the CO_2, and thus are of relevance to efficient pipeline operation, EoS need to be extended to model mixtures. We extended our statistical model for parametric pressure-explicit EoS to binary mixtures of CO_2 with an impurity. We demonstrated how we can take a similar approach to that in the pure CO_2 case and account for mixture coexistence constraints as what are effectively penalty terms in the model likelihood. Mixture coexistence volumes are often unrecorded, we showed how Bayesian techniques allow us to treat these missing values as parameters and so estimate them from the data, and proposed a model for the missing volumes. We fit our statistical model for parametric pressure-explicit EoS to data from the binary mixture of CO_2-H_2 using our PT-HMC sampler and the Demetriades EoS. We found that we had good mixing and could consistently fit the model extended to mixtures but that agreement

with data was poor.

Finally we explored an alternative to parametric EoS, the non-parametric Gaussian process approach. We showed how we define a Gaussian process model for both homogeneous and coexistence data and how the GP model allows us to enforce coexistence constraints exactly rather than having to make an approximation as in the parametric case. We demonstrated how we could define a non-stationary covariance function such that the GP model fit well to subcritical reference data for pure CO_2.

Previously EoS have been fit using ad hoc optimisation methods. The models and techniques developed in this thesis allow for more rigorous and consistent fitting procedures. The methods developed in this thesis allow us to rigorously quantify uncertainty in model parameters and predictions for EoS for use in CCS applications.

We now offer potential avenues of future research which could follow on from the work presented in this thesis.

7.2 Future Work

We have a fully developed statistical model for parametric pressure-explicit EoS for pure CO_2 and can consistently fit this model using the PT-HMC sampler. We fitted this model using the specific case of the Demetriades EoS. While we found good agreement with the data for most temperatures we saw that the Demetriades EoS was not flexible enough to capture the behaviour around the critical temperature. As we now have a model which can be applied to any parametric pressure-explicit EoS, and a sampler which can reliably sample from this model, we could apply our model and sampler to other proposed EoS for pure CO_2 and perform rigorous model comparison, whether through measures of out-of-sample predictive fit like cross-validation [Geisser and Eddy, 1979], or through calculation of Bayes

factors [Kass and Raftery, 1995].

In this work we focused solely on fitting EoS in the window of standard operation for a pipeline. However in the case of pipeline failure temperatures down to the triple point are of interest. Our model could be easily extended to such temperatures. There are also other quantities which are typically estimated from EoS, including heat capacity, speed of sound and viscosity, and we could investigate application of our statistical model to making predictions of these quantities.

We proposed an extension of our statistical model for parametric pressure-explicit EoS to binary mixtures of CO_2 with an impurity. It should be fairly simple to generalise this proposed model to mixtures of CO_2 with any general number of impurities. We only fit to CO_2-H_2 data in this thesis. We fit a simple model which incorporated the concentration of H_2 into the model parameters linearly, for a single temperature, and found that there was poor agreement to the data. With our model and sampler we are in a position to fit further temperatures and more complex models which introduce the H_2 mole fraction non-linearly. There are also many other impurities of relevance to CCS, for instance O_2 and N_2, for which data exists and which our model and sampler could be applied to.

We fit our GP model in chapter 6 only to individual subcritical temperatures. We could now extend this model and fit to multiple subcritical temperatures. The critical point has the property that it is a saddlepoint, see §1.3.1. The condition that the critical point is a saddlepoint can be defined in terms of derivatives. As we showed in chapter 6 we can include observations of derivatives in our GP model and so the GP approach provides a possible means to condition directly on the behaviour of pure CO_2 at the critical point. The GP approach might then offer a flexible means by which to model the change in behaviour that occurs around the critical point and which was the only data that the parametric model fit in chapter

4 could not fit. For pure CO_2 there is a large amount of accurate data, and so we might expect that our GP model might be able to learn a form such that it accounts for all of the physical properties of pure CO_2 of interest to CCS applications. However for mixtures data is often sparser. GPs do not use small amounts of data efficiently as such one possible direction of future work is to investigate how our GP model performs when extended to mixtures.

Bibliography

C. Andrieu and J. Thoms. A tutorial on adaptive MCMC. *Statistics and Computing*, 18:343–373, 2008.

A. Beskos, N. Pillai, G. Roberts, J. Sanz-Serna, and A. Stuart. Optimal tuning of the hybrid Monte-Carlo algorithm. *Bernoulli*, 19:1501–1534, 2013.

M. Betancourt and L. C. Stein. The geometry of Hamiltonian Monte Carlo. arXiv:1112.4118 [stat.ME], 2011.

M. Betancourt, S. Byrne, S. Livingstone, and M. Girolami. The geometric foundations of Hamiltonian Monte Carlo. *Bernoulli*, 23:2257–2298, 2017.

M. E. Boot-Handford, J. C. Abanades, E. J. Anthony, M. J. Blunt, S. Brandani, N. MacDowell, J. R. Fernandez, M-C. Ferrari, R. Gross, J. P. Hallett, R. S. Haszeldine, P. Heptonstall, A. Lyngfelt, Z. Makuch, E. Mangano, R. T. J. Porter, M. Pourkashanian, G. T. Rochelle, N. Shah, J. G. Yao, and P. S. Fennell. Carbon capture and storage update. *Energy Envion. Sci.*, 7:130–189, 2014.

B. Bornkamp. Approximating probability densities by iterated Laplace approximations. *Computational and Graphical Statistics*, 20:656–669, 2011.

Y.A. Cengel and M.A. Boles. *Thermodynamics: An Engineering Approach*. McGraw-Hill, 4th edition, 2001.

S. R. Cole, H. Chu, and S. Greenland. Maximum likelihood, profile likelihood and penalized likelihood: a primer. *American Journal of Epidemiology*, 179:252–260, 2014.

Comité International des Poids et Mesures. The international practical temperature scale of 1968. *Metrologia*, 5:35–44, 1969.

Comité International des Poids et Mesures. The international practical temperature scale of 1968: Amended edition of 1975. *Metrologia*, 12: 7–17, 1976.

E. de Visser, C. Hendriks, M. Barrio, M. J. Mølnvik, G. de Koeijer, S. Liljemark, and Y. Le Gallo. Dynamis CO_2 quality recommendations. *International Journal of Greenhouse Gas Control*, 2:478–484, 2008.

P. Dellaportas and D. A. Stephens. Bayesian analysis of errors-in-variables regression models. *Biometrics*, 51:1085–1095, 1995.

T. A. Demetriades, T. C. Drage, and R. S. Graham. Developing a new equation of state for carbon capture and storage pipeline transport. *Journal of Process Mechanical Engineering*, 227:117–124, 2013.

T.A. Demetriades and R.S. Graham. A new equation of state for CCS pipeline transport: Calibration of mixing rules for binary mixtures of CO_2 with N_2, O_2 and H_2. *Journal of Chemical Thermodynamics*, 93: 294–304, 2016.

S. Duane, A.D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195:216–222, 1987.

W. Dusschek, R. Kleinrahm, and W. Wagner. Measurement and correlation of the pressure, density, temperature relation of carbon dioxide. II. saturated-liquid and saturated-vapour densities and the vapour pressure along the entire coexistence curve. *J. Chem. Thermodynamics*, 22: 841–864, 1990a.

W. Dusschek, R. Kleinrahm, and W. Wagner. Measurement and correlation of the pressure, density, temperature relation of carbon dioxide. I. the homogeneous gas and liquid regions in the temperature range from 217 K to 340 K at pressures up to 9 MPa. *J. Chem. Thermodynamics*, 22: 827–840, 1990b.

D.J. Earl and M. W. Deem. Parallel tempering: Theory, applications and new perspectives. *Phys. Chem. Chem. Phys.*, 7:3910–3916, 2005.

O. Fandiño, M. Trusler, and D. Vega-Maza. Phase behavior of (CO_2-H_2) and (CO_2-N_2) at temperatures between (218.15 and 303.15)K at pressures up to 15MPa. *International Journal of Greenhouse Gas Control*, 36:78–92, 2015.

H. Fukunishi, O. Watanabe, and Takada S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.*, 116:9058–9067, 2002.

S. Geisser and W. F. Eddy. A predictive approach to model selection. *Journal of the American Statistical Association*, 74:153–160, 1979.

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. CRC Press, 2014.

C. J. Geyer. Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics: Proc. 23rd Symp. Interface*, 1991.

C. J. Geyer and E. A. Thompson. Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *Journal of the Royal Statistical Society, Series B*, 54:657–699, 1992.

R. Gilgen, R. Kleinrahm, and W. Wagner. Supplementary measurements of the pressure, density, temperature relation of carbon dioxide in the

homogeneous region at temperatures from 220 K to 360 K and pressures up to 13 MPa. *J. Chem. Thermodynamics*, 24:1243–1250, 1992.

N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140:107–113, 1993.

R. B. Gramacy and H. K. H. Lee. Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, 103:1119–1130, 2008.

R. B. Gramacy and H. K. H. Lee. Cases for the nugget in modelling computer experiments. *Statistics and Computing*, 22:713–722, 2012.

P. J. Green. Penalized Likelihood. *Encyclopedia of Statistical Sciences*, 9, 2006.

H. Haario, E. Saksman, and J. Tamminen. Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics*, 14: 375–395, 1999.

H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7:223–242, 2001.

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

M. D. Hoffman and A. Gelman. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15:1351–1381, 2014.

R. E. Kass and A. E. Raftery. Bayes factors and model uncertainty. *Journal of the American Statistical Association*, 90:773–795, 1995.

H-M Kim, B. K. Mallick, and C. C. Holmes. Analyzing nonstationary spatial data using piecewise Gaussian processes. *Journal of the American Statistical Association*, 100:653–668, 2005.

S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

O. Kunz, R. Klimeck, W. Wagner, and M. Jaeschke. *The GERG-2004 wide-range reference equation of state for natural gases*. Fortschr.-Ber. VDI, VDI-Verlag, Düsseldorf, 2007.

D. J. C. MacKay. Introduction to Gaussian processes. In C. M. Bishop, editor, *Neural Networks and Machine Learning*, pages 133–166. Springer-Verlag, 1998.

I. S. Mbalawata, S. Särkkä, M. Vihola, and H. Haario. Adaptive Metropolis algorithm using variational Bayesian adaptive Kalman filter. *Computational Statistics and Data Analysis*, 83:101–115, 2015.

M. Mohitpour, P. Seevam, K.K. Botros, B. Rothwell, and C. Ennis. *Pipeline Transportation of Carbon Dioxide Containing Impurities*. ASME Press, 2012.

R. Murray-Smith and B. A. Pearlmutter. Transformations of Gaussian process priors. In J. Winkler, N. Lawrence, and M. Niranjan, editors, *Deterministic and Statistical Methods in Machine Learning*, pages 110–123. Springer-Verlag, 2005.

National Institute of Standards and Technology. Nist chemistry webBook. `http://webbook.nist.gov/chemistry/fluids`, 2015.

R. M Neal. An improved acceptance procedure for the hybrid Monte Carlo agorithm. *Journal of Computational Physics*, 111:194–203, 1994.

R. M. Neal. Slice sampling. *Annals of Statistics*, 31:705–767, 2003.

R. M. Neal. MCMC using Hamiltonian dynamics. In *Handbook of Markov chain Monte Carlo*. CRC Press, 2011.

J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

M. Nelson, R. Graham, S. Preston, and R. Wilkinson. Personal communication, 9th June 2015.

Y. Nesterov. Primal-dual subgradient methods for convex problems. *Mathematical Programming*, 120:221–259, 2009.

H. Orbey and S. I. Sandler. *Modelling vapor-liquid equlibria: Cubic equations of state and their mixing rules.* Cambridge University Press, 1998.

C. Paciorek and M. J. Schervish. Nonstationary covariance functions for Gaussian process regression. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in neural information processing systems 16*, pages 273–280. MIT Press, 2004.

U. Paquet. *Bayesian inference for latent variable models.* PhD thesis, University of Cambridge, 2007.

D-Y Peng and D. B. Robinson. A new two constant equation of state. *Industrial and Engineering Chemistry Fundamentals*, 15:59–64, 1976.

R.T. Porter, M. Fairweather, M. Pourkashaniana, and R.M. Woolley. The range and level of impurities in CO_2 streams from different carbon capture sources. *International Journal of Greenhouse Gas Control*, 36:161–174, 2015.

H. Preston-Thomas. The international temperature scale of 1990 (ITS-90). *Metrologia*, 27:3–10, 1990.

H. Preston-Thomas, P. Bloembergen, and T. J. Quinn. Supplementary information for the international temperature scale of 1990. Bureau Internationale des Poids et Mesures, 1990.

C. R. Rao. *Linear Statistical Inference and Its Applications*. Wiley, 2nd edition, 1973.

C. E. Rasmussen and Z. Ghahramani. Infinite mixtures of Gaussian process experts. In T.G. Diettrich, S. Becker, and Z. Ghahramani, editors, *Advances in neural information processing systems 14*, pages 881–888. MIT Press, 2002.

C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning*. MIT Press, 2006.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.

G. O. Roberts and J. S. Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18:349–367, 2009.

G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability*, 7:110–120, 1997.

D. B. Rubin. Inference and missing data. *Biometrika*, 63:467–474, 1976.

R. L. Rusby. The conversion of thermal reference values to the ITS-90. *J. Chem. Thermodynamics*, 23:1153–1161, 1991.

V. S. Ryaben'kii and S. V. Tsynkov. *A theoretical introduction to numerical analysis*. CRC Press, 2006.

Y. Sanchez-Vicente, T.C. Drage, M. Poliakoff, J. Ke, and M.W. George. Densities of the carbon dioxide+hydrogen, a system of relevance to car-

bon capture and storage. *International Journal of Greenhouse Gas Control*, 13:78–86, 2013.

A. M. Schmidt and A. O'Hagan. Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65:743–758, 2003.

P N. Seevam, J. M. Race, and M. J. Downie. Carbon dioxide pipelines for sequestration in the UK: an engineering gap analysis. *Journal of Pipeline Engineering*, 6:133–146, 2007.

E. Snelson, C. E. Rasmussen, and Z. Ghahramani. Warped Gaussian processes. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in neural information processing systems 16*, pages 337–344. MIT Press, 2004.

R. Span and W. Wagner. A new equation of state for carbon dioxide covering the fluid region from the triple-point temperature to 1100K at pressures up to 800MPa. *Journal of Physical and Chemical Reference Data*, 25:1509–1596, 1996.

Stan Development Team. RStan: the R interface to Stan, version 2.11.1. `http://mc-stan.org`, 2016a.

Stan Development Team. Stan modeling language users guide and reference manual, version 2.11.0. `http://mc-stan.org`, 2016b.

D.A. Stephens and P. Dellaportas. Bayesian analysis of generalised linear models with covariate measurement error. In J.M Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 4*, pages 813–820. Oxford University Press, 1992.

K. M. Stern. Stern review: the economics of climate change. HM Treasury Department, 2006.

R. H. Swendsen and J. S. Wang. Replica Monte Carlo simulation of spin glasses. *Physical Review Letters*, 57:2607–2609, 1986.

M.-J. Tenorio, A.J. Parrot, J. Calladine, Y. Sanchez-Vicente, A.J. Cresswell, R.S. Graham, T.C. Drage, M. Poliakoff, J. Ke, and M.W. George. Measurement of the vapourliquid equilibrium of binary and ternary mixtures of CO_2, N_2 and H_2, systems which are of relevance to CCS technology. *International Journal of Greenhouse Gas Control*, 41:68–81, 2015.

C. K. I. Williams. Computation with infinite neural networks. *Neural Computation*, 10:1203–1216, 1998.