Bayesian approaches to deconvolution in well test analysis

Themistoklis Botsas

A Thesis presented for the degree of
Doctor of Philosophy

Statistics Group
Department of Mathematical Sciences
Durham University
United Kingdom

April 2020

Bayesian approaches to deconvolution in well test analysis

Themistoklis Botsas

Submitted for the degree of Doctor of Philosophy

April 2020

Abstract: In petroleum well test analysis, deconvolution is used to obtain information about the reservoir system, for example the presence of heterogeneities and boundaries. This information is contained in the reservoir response function, which can be estimated by solving an inverse problem in the well pressure and flow rate measurements.

We use a non-linear Bayesian regression model based on models of reservoir behaviour in order to make inferences about the response function. This allows us to include uncertainty for the independent variables, which is essential, since the measurements are usually contaminated with observational error. We combine the likelihood with a set of flexible priors for the response parameters, and we use Markov Chain Monte Carlo algorithms in order to approximate the posterior distribution.

We validate and illustrate the use of the algorithm by applying it to synthetic and field data sets, using a variety of tools to summarise and visualise the posterior distribution, and to carry out model selection. The results are comparable in quality to the state of the art solution, but our method has several advantages: we gain access to meaningful system parameters associated with the flow behaviour in the reservoir; we can incorporate prior knowledge that excludes non-physical results;

and we can quantify parameter uncertainty in a principled way by exploiting the advantages of the Bayesian approach.

Declaration

The work in this thesis is based on research carried out in the Department of Mathematical Sciences at Durham University. No part of this thesis has been submitted elsewhere for any degree or qualification.

Acknowledgements

I would like to thank the Imperial College Joint Industry Projects (JIP) on “New developments in well test analysis" that entrusted me with this project.

I am very grateful to my supervisors Professor Ian Hyla Jermyn and Dr. Jonathan Cumming for all their encouragement and support. Week in and week out they were there to offer me guidance. Our meetings were more enjoyable than they ought to be, and I was always looking forward to them.

I wish to thank Dr. Jochen Einbeck, who helped me from the very beginning of my PhD journey and gave me the opportunity to experience a very interesting aspect of academia.

Thank you to all the people that helped me settle in Durham, both inside and outside the department, and made the PhD life seem exciting: Pla, James, Junbin, Clare, Tim, Rob, Daniel, Job, Ting, Pui, Hugo and everyone else.

Finally, a big thank you to my parents, George and Aimilia, that were always supportive in every way imaginable, I couldn’t have done this without you.

In order to succeed, we must first believe that we can.

— Nikos Kazantzakis

Dedicated to

my parents,
Aimilia and George.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter we introduce the main terminology for well test analysis, primarily focusing on the current state of the art methodology, which is based on deconvolution. We also present a literature review on other well testing methods, and we describe the motivation behind this work.

1.1 General context

The focus of well test analysis is a system constituted of a well and a reservoir. The *well* comprises drilling and production tools, along with a *wellbore*, which is a drilled hole in the earth, designed to bring either *oil*, a complex mixture of naturally occurring hydrocarbon compounds found in rock, or *gas*, a highly compressible, naturally occurring mixture of hydrocarbon gases, primarily consisting of methane, to the surface. The *reservoir* is a body of rock with the ability to store and transmit fluids.

The wellbore and the reservoir are described by a set of properties of the geometry and geology of the system. For the reservoir, the relevant quantities are: the *permeability* of the medium k, which is a measurement of a rock's ability to transmit fluids, and is measured in millidarcy mD; the *viscosity* μ, a property of fluids that indicates their resistance to flow, measured in centipoise cp; the *porosity* ϕ, which is the percentage

of volume within rock that can contain fluids; the *total system compressibility* c_t, which is the relative change in fluid volume related to a unit change in pressure, measured in inverse pounds per square inch psi^{-1}; and the *formation thickness* h, which is measured in feet ft. For the wellbore, these properties include: the *wellbore radius* r_w, which is the distance from the middle of the wellbore to the interface between the wellbore and the formation, measured in feet ft; the *wellbore storage coefficient* C, a measure of distortion in the reservoir response due to fluid stored in the wellbore before the beginning of production, measured in barrels per pounds per square inch bbl/psi; and the *skin* S, a zone of reduced or enhanced permeability around a wellbore, damage or stimulation, respectively, which is a dimensionless quantity.

A *well test* is a period of time, during which the production of a well is recorded. During a well test, we are interested in two types of measurements. The *flow rate* q, which is a controllable variable that relates to the pace of the production and is measured with noise, usually in barrels per day (bbl/d) at surface; and the *bottomhole pressure* p, which is associated with the force distributed over the surface of the downhole, and is measured at discrete time points with noise in pounds per square inch psi. When the pressure measurements come from a gas reservoir, a conversion to pseudo-pressure needs to take place, in order to correct for the variable compressibility and viscosity of gas. For the purposes of this thesis, the general pressure will be denoted as P. The time is denoted as t and the corresponding unit of measurement is hours hrs. After the test is initiated, the change in pressure as a function of time is referred to as *transient pressure*, and is denoted as $p(t)$. During that time, the pressure transient advances from the downhole further into the reservoir, until it reaches the reservoir's boundaries. A relevant quantity that represents how far into the reservoir the transient effects have travelled during a well test is called *radius of investigation*. In addition, before starting the production, there is a constant and uniform pressure in the reservoir, called *initial pressure*,

which is also measured with noise and denoted as p_0 [1]. The pressure drop is then defined as $\Delta p(t) = p_0 - p(t)$.

There are two distinct phases in a well test: during the *drawdown* phase (Dd), the production is constant and the pressure is gradually decreasing. After the well is shut, we transition to the *buildup* phase (BU), where the flow rate is equal to 0, and the pressure is increasing. These phases can be observed in Figure 1.1.

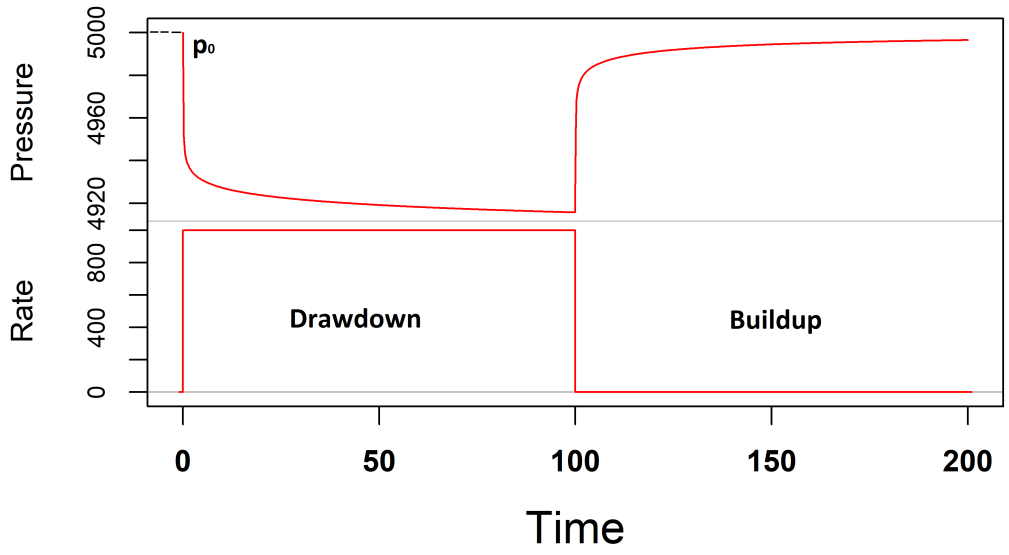


Figure 1.1: Synthetic data (bottomhole pressure and rate) from a simple well test that consists of a single drawdown Dd phase, followed by a buildup BU phase.

Well test analysis is a set of methodologies for the planning, conducting, and analysis of the results of well tests in an oil or gas reservoir. Its objective is to identify and interpret the relationship between pressure and rate, which depends on the wellbore and reservoir parameters, and then use it in order to make inferences about the reservoir properties, the well performance, the production potential, the geology, and the geometry [7].

The well test methods for analysing data are based on a variation of the diffusion

[1] In petroleum engineering literature it is usually denoted as p_i

equation that is used to describe the flow of fluid flowing in porous media [3]. Under the assumptions of negligible gravity effects, homogeneous isotropic porous medium, laminar flow, single phase and slightly compressible fluid, constant permeability, porosity, viscosity, and flow in the radial direction, the law of conservation of mass and Darcy's law (fluid flow through a porous medium) are combined and yield [52]:

$$\frac{\partial^2 P}{\partial r^2} + \frac{1}{r}\frac{\partial P}{\partial r} - \frac{1}{\eta}\frac{\partial P}{\partial t} = 0 \tag{1.1.1}$$

where, r the radius of the reservoir and η the diffusivity, a function of the permeability, the viscosity, the porosity and the total system compressibility, which will be further examined in later chapters. This equation is used to obtain closed-form solutions for simple cases, since it can be solved analytically for problems with specific initial and boundary conditions.

For example for an infinite radial reservoir and constant rate production, the solution in dimensionless terms is:

$$p_D = -\frac{1}{2} Ei\left(-\frac{r_D^2}{4t_D}\right), \tag{1.1.2}$$

where the subscript $\cdot_D$ denotes dimensionless quantity, and will be examined further in Chapter 2, $Ei(-x) = -\int_x^\infty \frac{e^{-u}}{u} du$, and the formula reflects the value of the dimensionless pressure at any given dimensionless radial distance from the wellbore r_D and any given dimensionless time t_D

Additionally, modifications of Equation 1.1.1 can model different flow behaviours. For example, by omitting the second term, we obtain the equation for linear flow, which is a behavior usually visible in channels, while if $\frac{\partial P}{\partial t} = constant$, the system is in pseudosteady state, which describes a depleted (closed) reservoir.

Finally, there are occasions where Equation 1.1.1 acts as a building block for other models. For example, using the method of images [53], a system with a fault within a radial reservoir can be replicated by a two-well system, where the wells are in

distance twice that of the distance from the active well to the boundary, whereas their pressure is described by Equation 1.1.2. This model will be further examined in Section 2.3

1.2 Well test analysis tools

A large variety of technical and mathematical tools has been used to analyse the well test data, leading up to the state of the art methods, which are based on deconvolution. An extensive history of the tools of well test analysis is described in [34].

The very first interpretation methods used in the 1950s were based on straight lines applied to semi-log data [41]. The pressure measurements of a single buildup were plotted against a logarithmic time transformation, and then, information about the well production was extracted by fitting straight lines to the curve. In the meantime, the Laplace transform was introduced as the main mathematical technique for the solution of the diffusion equation [78]. These tools were primarily focusing on results for homogeneous reservoir behaviour.

In the 1960s, the introduction of the pressure type curve analysis with [60] for both buildup and drawdown data, shifted the focus from straight lines fitting to type-curve matching, where a pressure-time log-log plot was compared against a set of predefined curves. These methods were mainly focusing on the flow behaviour near the wellbore [2]. During that time, the addition of Green's functions in well test analysis [37], managed to expand the diffusion equation solutions associated with reservoir flow problems.

In the 1970s, new interpretation methods [36] improved on the concept of type curves by introducing independent variables for the analysis, which provided more reliable results than its predecessors, and also made the expansion of the range of identifiable models possible, by allowing the detection of dual porosity reservoirs [9]. The need

of computational methods that could handle the new type curves was satisfied with the addition of the numerical Laplace inversion Stehfest algorithm [72].

In the 1980s, a pivotal point was the introduction of the pressure derivatives with [10], where the more sensitive and powerful differential of the pressure was plotted against the time on a log-log plot. This new set of type curves offered a much more detailed depiction of the reservoir flow, and eventually enhanced the identification of general heterogeneities and boundaries.

1.3 Deconvolution

The standard methods mentioned in the previous section were typically restricted to the analysis of periods of constant-rate data, like a single Dd phase in Figure 1.1, due to the corresponding simplification of the pressure-rate relationship under such conditions. Hence, variable rate data had to be treated as a collection of smaller independent data sets, resulting in inconsistencies in the analysis, arising from the dependency induced by the cumulative effect of production over time.

Deconvolution [68] emerged as a major milestone in well-testing, due to its ability to deal with variable rate tests, such as a combination of a Dd and a BU phase in Figure 1.1, in their entirety, enabling the consistent analysis of larger and richer data than was previously accessible. The ability to deconvolve entire pressure and rate histories gave access to the true radius of investigation, allowing the interpretation of the full available history, instead of small periods of data. Thus, it enables a coherent analysis of longer-term behaviour, providing greater insights into the flow behaviour of the reservoir, and potentially allows the detection of more distant geological features.

The key relationship for this inference is that, if the reservoir flow is governed by equations linear in pressure and production rate, the bottomhole pressure, $\tilde{p}$, and the production rate, $\tilde{q}$, written as functions of time t, are accurately related by a

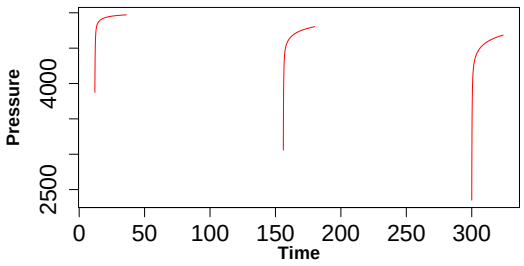
time-translation invariant linear operator g, and the following equation known as the Duhamel's principle [78] applies:

$$\begin{aligned}\Delta\tilde{p}(t) = \tilde{p}_0 - \tilde{p}(t) = (g * \tilde{q})(t) &= \int_0^\infty g(t-t')\,\tilde{q}(t')dt' \\ &= \int_0^t g(t-t')\,\tilde{q}(t')dt' \\ &= \int_0^t \tilde{q}(t-t')\,g(t')dt'\,, \end{aligned} \tag{1.3.1}$$

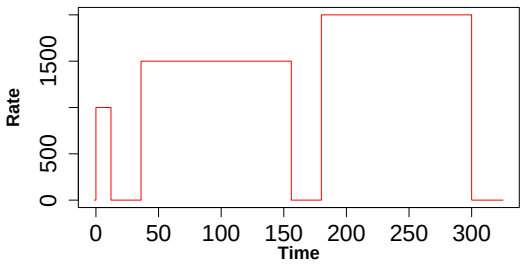
where $\Delta\tilde{p}$ is the pressure drop from an initial equilibrium pressure $\tilde{p}_0$ at $t = 0$, and use has been made of the fact that $g(t) = \tilde{q}(t) = 0$ for $t < 0$, by causality. The last line follows by symmetry. Since the rates are independent of the reservoir parameters, the *reservoir response function* g is the object of primary interest, that encapsulates all relevant well- and reservoir-specific information in a single function, and thus provides an effective summary and signature for the behaviour of a particular well in response to production.

The value $g(\Delta t)$ describes the effect on the pressure at t of the rate at $t - \Delta t$; it is also the derivative of the pressure drop at time Δt induced by a unit flow from $t = 0$ (consider $\tilde{q}(t') = 0$ in Equation 1.3.1). In principle, it can be derived as the limit of the Green function of the diffusion equation with $\tilde{q}$ as a time-varying point source.

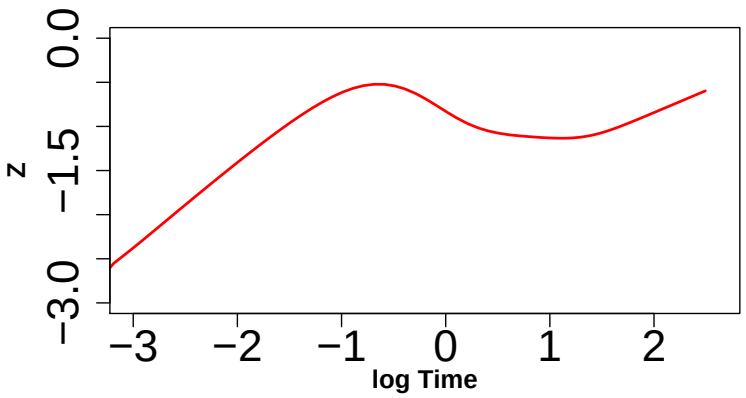
In practice, there are two key limitations which prevent the direct use of (1.3.1). First, we do not observe the entire continuous history of pressure and rate for all t, but only a collection of observations at a finite set of times. Second, $\tilde{p}$ and $\tilde{q}$ are not observed directly. Instead, measurements of these quantities, p and q, are obtained, subject to various sources of empirical error and other uncertainties. To clarify formally these distinctions, we establish our notation as follows; the bottomhole well pressure, $\tilde{p}(t)$, is observed at a series of time points t_i, giving the pressure data p_i for $i = 1, \ldots, m$. A typical truncated pressure history is shown in Figure 1.2a. The initial reservoir pressure is $\tilde{p}_0$, for which we have a single observation p_0, which typically corresponds to the $\max(p_i)$. Finally, the rate data are real value observations q_j, associated with sequential time intervals $[T_j, T_{j+1}]$, for $j = 1, \ldots, N$. We model the true rate $\tilde{q}(t)$ as



(a) Pressure $\tilde{p}$. As in a typical well test, there are long periods where the pressure is not recorded (gaps).



(b) Flow rate $\tilde{q}$.



(c) Response transformation z.

Figure 1.2: Synthetic data (pressure and rate) and corresponding z of a channel, taken from [19]. The deconvolution of the bottomhole pressure, $\tilde{p}$ (upper left), and the flow rate, $\tilde{q}$ (upper right) produces the response z (bottom).

a piecewise constant function on these same time intervals, as in Figure 1.2b, and we model the measurement process as the addition of noise to these values.

In mathematical form the above observational model is:

$$\tilde{\boldsymbol{p}}(g, \tilde{\boldsymbol{q}}, \tilde{p}_0) = \boldsymbol{p} + \epsilon'$$

$$\tilde{\boldsymbol{q}} = \boldsymbol{q} + \epsilon''$$

$$\tilde{p}_0 = p_0 + \epsilon''',$$

where ϵ', ϵ'' and ϵ''' are error terms. The model is described extensively in Section 3.2.

Typically, instead of seeking g, we consider a transformation of the response function and the time coordinate: $z(\tau) = \tau + \ln[g(e^{\tau})]$, where $\tau = \log t$. The transformation ensures the positivity of g and removes the need to enforce the causality constraint that $g(t) = 0$ for $t < 0$. It also greatly facilitates the interpretation and diagnosis of the response function; the transformed response can be plotted following the same principles used for the pressure derivative type curves in [10], with the shape and features of the response being indicative of the flow behaviour in the reservoir. An example of a response function resulting from the deconvolution process [68] is shown in Figure 1.2.

1.3.1 Common response shapes

The response, as we will show in the next chapter for particular configurations, is in practice a transformation of a solution to an extended variation of the diffusion equation in Equation 1.1.1 for some specific initial and boundary conditions. Thus, its shape and behaviour can be directly associated with particular flow regimes and reservoir features, that correspond to different conditions. These features are categorised by the times at which they appear, namely early, middle, and late times, where time of the appearance of these effects is effectively a surrogate for the distance of the physical feature from the wellbore. The distinction between the times and the key features and flow regimes are presented in Figure 1.3.

- Early-time (from a few seconds to a few hours) behaviour is dominated by the wellbore itself and its immediate surroundings. The response usually begins with a unit slope straight line, which corresponds to a constant g, representing the effect of wellbore storage – the fluid in the reservoir filling the uniform wellbore itself. This is typically followed by a characteristic 'hump', which represents the impedance to flow caused by localised damage due to the drilling of the well – a feature known as the 'skin'.

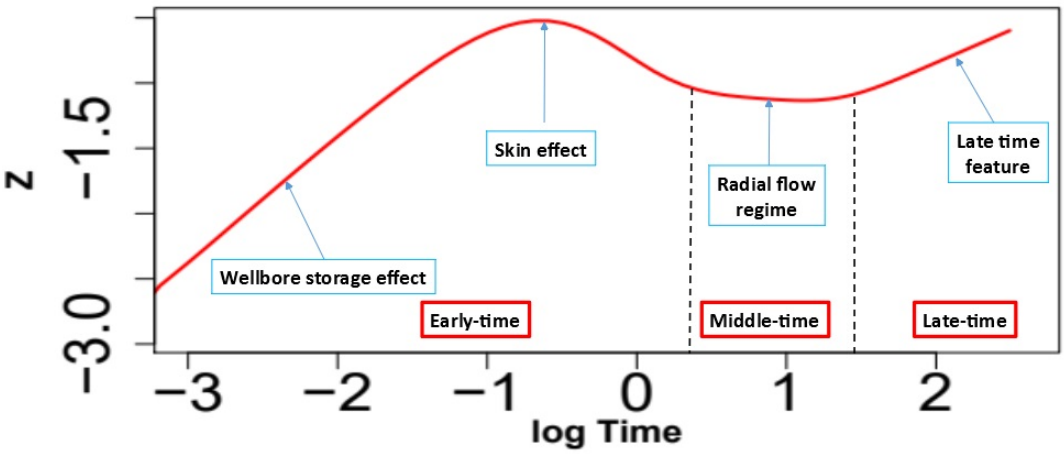
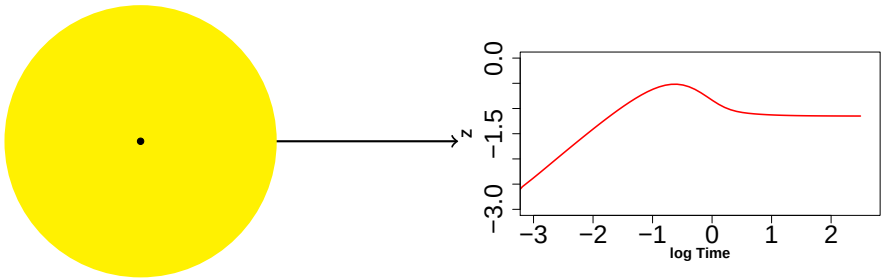


Figure 1.3: Key features and flow regimes of the response function.

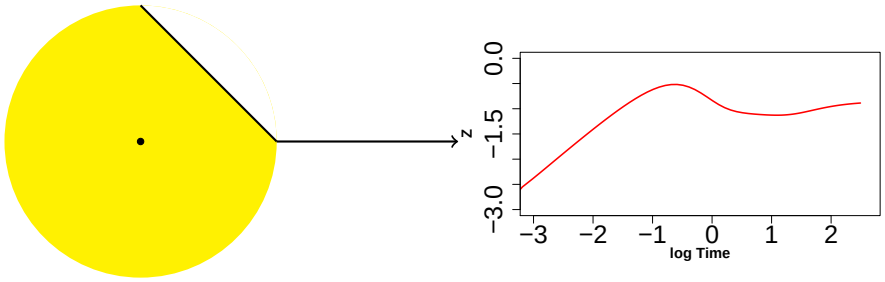
- Middle-time (a few hours or days, after production is established but after the appearance of heterogeneities and before that of boundaries) usually shows stabilisation, as the well draws fluid from the relatively homogeneous region beyond the damaged skin. The 'radial flow' regime is established, during which, fluid moves towards the well from all directions. An interesting effect appears when the wellbore is located very close to one of the reservoir's boundaries and the radial flow regime is masked. That regime is called *hemiradial flow* [73].

- Late-time (weeks or months, depending on reservoir properties and wellbore location) behaviour indicates features of the reservoir far from the wellbore, such as boundaries or other heterogeneities in the geology. An example is a characteristic flow regime, commonly known as a *dual porosity system*, which is characterised by a significant change in the diffusivity of the reservoir and manifests in the response plot as a localised deviation from the radial flow level.

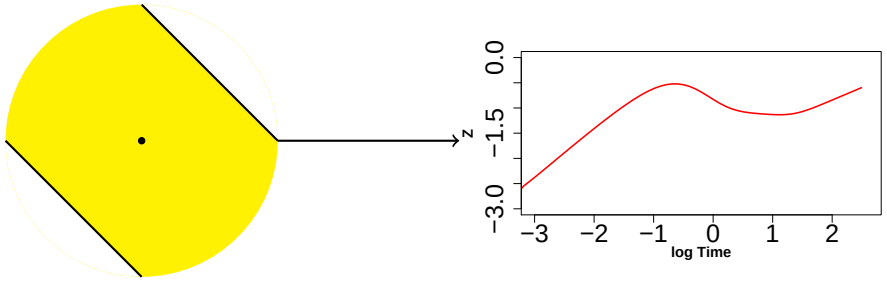
The left-hand side of Figure 1.4 shows some common reservoir boundary configurations, namely an infinite radial flow (IRF), a single fault, a channel, and a closed (depleted) reservoir, while the right-hand size shows their resulting response curves:



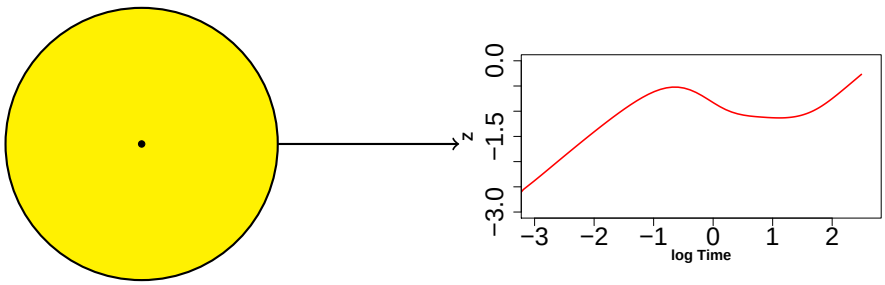
(a) IRF.



(b) Single fault.



(c) Channel.



(d) Closed (depleted) reservoir.

Figure 1.4: Indicative reservoir boundary configurations and resulting response curves. The response curves were taken from [19].

- Infinite radial flow (IRF) is a flow behaviour where the oil intake is radial. An IRF reservoir extends infinitely in all directions, without showing evidence of boundaries or heterogeneities (Figure 1.4a).
- A fault is a reservoir configuration, where, after a period of radial flow, a single no-flow boundary of infinite length appears at a fixed distance from the well, while the oil still finds its way to the wellbore from all other directions, as shown in Figure 1.4b.
- A channel is a common reservoir shape, where, after a period of radial flow, two parallel faults of infinite length appear, restricting the oil intake from those directions, as shown in Figure 1.4c. This flow regime is called linear flow.
- A closed (depleted) reservoir is another shape that occurs, after the transient pressure has reached all boundaries of the reservoir. The corresponding regime is called pseudo-steady state flow. This particular data set, comprises a bounded radial reservoir, in the middle of which lies the wellbore, as shown in Figure 1.4d. It is worth noting that every real reservoir is, in fact, closed at late times, but this regime does not necessarily appear within the radius of investigation of a well test.

These reservoirs will be further analysed in Chapter 5.

Regarding the response curve for each boundary configuration, we can observe that they are all identical for the early (wellbore storage and skin phase) and middle (radial flow phase) times, appearing in the plots as early unit slopes and 'humps' for the former and constant horizontal levels for the latter. In contrast, at late times, we observe that each set corresponds to a different late time feature. The IRF curve in Figure 1.4a carries on with the radial flow constant level until the rightmost limit of the plot. In the fault curve in Figure 1.4b, a jump at a higher level appears that leads to a second stabilisation. The channel plot in Figure 1.4c shows a late time half-unit slope. Finally, the closed model curve in Figure 1.4d introduces a unit slope.

Note that a specific response shape may not uniquely identify a reservoir feature (*e.g.* the response shape of a fault can also be achieved by reservoir heterogeneities) and that sequential features may mask one another.

1.3.2 Buildup derivative

Before we expand on the state of the art deconvolution methods, we present another commonly used well test analysis tool, which is based on the buildup derivative [10]. The aforementioned derivative is considered to provide a good indication of the middle time's *radial flow regime*, and can be used for the verification of the results of other methods, such as the ones based on deconvolution.

For the construction of the buildup derivative, we assume that, starting from the beginning of the well test ($t = 0$), we have a single, constant rate drawdown phase of duration t_p (which is an averaged version of the actual production), followed by the buildup phase. t_p is called *pseudo-production time*, and can be computed as:

$$t_p = \frac{\text{Total Production}}{\text{Last rate}} = \frac{\sum_{i=1}^{N} q_i \Delta t_i}{q_N}, \tag{1.3.2}$$

where the $q_i, \quad i = 1, \ldots, N$ denote the production rate measurements up to the buildup in question, and Δt_i the corresponding duration for each of these rate flow periods.

The buildup data and t_p are shown in Figure 1.5.

Due to the fact that there is some deviation between the drawdown and the buildup derivative, a correction needs to be applied [10], and thus, the buildup derivative is calculated with respect to the natural logarithm of the Horner time $(t_p + \Delta t)/\Delta t$.

$$\Delta p' = \frac{d\Delta p}{d \ln \frac{t_p + \Delta t}{\Delta t}} = \Delta t \frac{t_p + \Delta t}{t_p} \frac{d\Delta p}{d\Delta t} \tag{1.3.3}$$

We calculate the derivative numerically, using a weighted three point algorithm; either the *N-Point algorithm*, which is expressed as the middle point of a contiguous

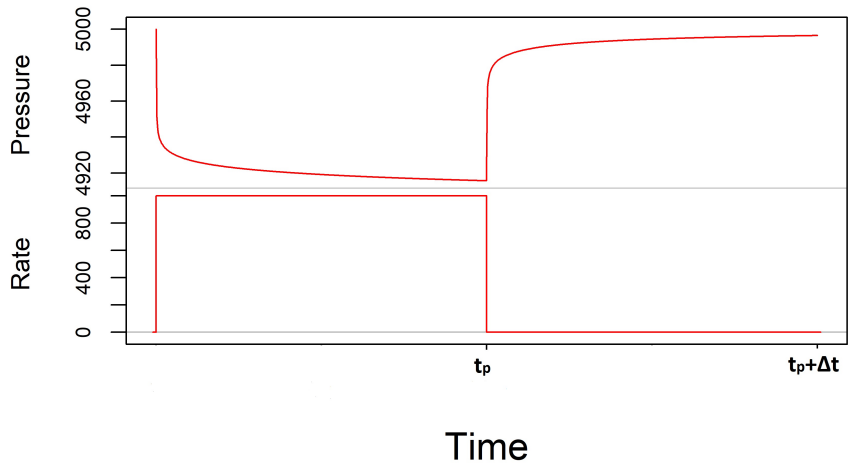


Figure 1.5: Buildup rate and pressure data, starting from the pseudo-production time t_p.

range of N points, or the *Window algorithm*, which uses the furthest points from the middle points that remain within a predefined window:

$$\left(\frac{dY}{dX}\right)_j = \frac{(X_k - X_j)\frac{Y_j - Y_i}{X_j - X_i} + (X_j - X_i)\frac{Y_k - Y_j}{X_k X_j}}{X_k - X_i} \tag{1.3.4}$$

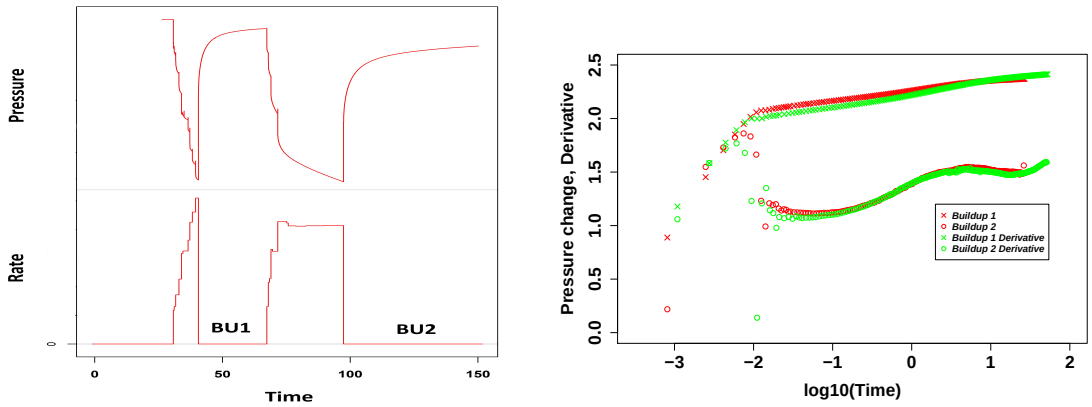
where i is the point before j, and k the point after, and the choice of these points is dictated by the respective algorithm.

Finally, the logarithm of the derivative of the pressure drop $\Delta p'$ from Equation 1.3.3 is plotted against the logarithm of the ordinary time duration Δt.

The pressure drop Δp, and respective derivatives $\Delta p'$ for the two large buildups from the oil well test data in Figure 1.6a, are shown in Figure 1.6b.

The buildup derivatives show the characteristic early time (wellbore storage and skin) and middle time (radial flow) regimes, while the late time behaviour is closer to that of a fault, followed by a slope.

As mentioned before, there is common consensus in the engineering literature that when the rate measurements are close to the true values, the buildup derivatives



(a) Field oil rate and pressure data.

(b) Buildup pressure measurements and corresponding derivatives from the oil data.

Figure 1.6: Oil data (pressure and rates), buildup pressure measurements and corresponding buildup derivatives. The data refer to an oil reservoir and are taken from [19].

provide a good approximation of the radial flow level, from which reservoir quantities such as the permeability can be determined.

In contrast, the buildup derivative has two key limitations that motivate us to use the deconvolution method instead. The first is that the pressure-derivative calculation algorithm in Equation 1.3.4 is prone to distortions. This is particularly evident in early times, where the limited number of pressure points is not sufficient for the recovery of the response. Second, a single buildup derivative is unable to work with variable-rate data, and thus, it uses only a small portion of the data available, which limits the radius of investigation and the detection of late time features. In practice, the engineers use multirate derivatives with respect to a superposition function, but the argument about the radius of investigation still stands.

1.4 Deconvolution methods

1.4.1 Methods before total least squares

There are numerous approaches to deconvolution in the literature of well test analysis. These can be classified into two categories: time domain methods and spectral do-

main methods. A detailed description of these methods and the distinction between them is given in [68].

The time domain methods include a variety of interpolation schemes for the functions g and $\tilde{q}$, as well as constraints on the shape of the response. The deconvolution problem was initially solved using a recursion relation [76]. Later, the problem was reformulated as a constrained optimisation problem, that was seeking to minimise an error measure, either based on the L_1 norm [17], or later the L_2 norm [46]. The fundamental issue of even the most contemporary among those methods is that, in practice, they cannot manage data errors above the levels of $1 - 2\%$, which is an important drawback, since the typical rate measurement error encountered in field data, is much larger.

The spectral domain methods seek to exploit the fact that the spectral transform of a convolution product is equal to the product of transforms. These methods include the use of continuous and discrete Fourier and Laplace transforms, that either use numerical inversion algorithms, such as the one proposed by Stehfest, in order to transform the response estimate back to the time domain [66], or, instead, interpret the results in the frequency domain [11]. The main upside of these methods is their simplicity and directness, while the main downside is associated with their ineffectiveness when they encounter incomplete or truncated histories. In addition, the fact that the pressure measurements are not generally equally spaced in time, which is an important requirement for some of these algorithms, makes spectral transforms unsuitable for the pressure-rate deconvolution problem.

1.4.2 Total least squares

In practical terms, the application of the state of the art deconvolution methodology for well testing is a time domain method that was first introduced in [68], as a form of separable nonlinear penalised total least squares regression and introduces a number of innovations.

Initially, the positivity of the response is achieved by the use of a logarithmic transformation, as was described in Subsection 1.3.1. The logarithmic time is denoted as $\tau = \ln t$, while the new response as $z(\tau) = \ln\{tg(t)\}$.

As a second step, interpolation schemes are introduced for the response function and the production rates. The former is represented by a piecewise linear function in log-time equispaced nodes τ_i, and is parameterised based on the slopes and intercepts: $z(\tau) = \alpha_r + \beta_r \tau$, for $\tau \in [\tau_{r-1}, \tau_r]$. The true rates $\tilde{q}$ are interpolated with a stepwise constant function.

Under these interpolation schemes, the deconvolution formula in Equation 1.3.1 takes the form

$$\begin{aligned} \Delta\tilde{p}(t) &= \int_{-\infty}^{\ln t} \tilde{q}(t - e^{\tau})\, e^{z(\tau)} d\tau \\ &= \boldsymbol{C}(z)\tilde{q} \end{aligned} \tag{1.4.1}$$

where $\boldsymbol{C}(z)$ is a matrix of convolution coefficients, whose elements depend on z and, in the case of piecewise linear response function, can be found analytically in terms of the α_r, β_r and $\tilde{q}$. Details are presented in Appendix B.

The error model is based on a penalised total least squares method. The independent parameters (production rates and initial pressure) are allowed to vary and their values are adjusted during the deconvolution process. Additionally, a curvature penalty based on the difference of the slopes is introduced, which ensures that the response curve is smooth. The error function is then the sum of the three components:

$$E = \zeta \|\tilde{p}_0 - p - \boldsymbol{C}(z)\tilde{q}\|_2^2 + \nu \|\tilde{q} - q\|_2^2 + \lambda \kappa(z)^2 \tag{1.4.2}$$

where $\|\cdot\|_2$ is the L_2 norm, $\kappa(z)$ is a measure of the curvature, and ζ, ν and λ are scalar weights that adjust the relative contributions of the respective components. In practice, they are nuisance parameters that are specified constants.

Note that the residuals of Equation 1.4.2 (the quantities within the norms) depend linearly on the rates and the initial pressure, and non-linearly on the response

parameters, which makes this a separable non-linear least squares problem, which is solved using the variable projection algorithm [6].

Finally, if the uncertainty, in the form of intervals of the response shape in each time point, is also required, it is provided using an additional Monte Carlo simulation method [19]. Simulated noise is added multiple times to the observed pressure, rate and initial pressure measurements, and each time a new deconvolution is performed. The resulting set of deconvolved responses is plotted or presented via summary statistics.

Published applications of the total least squares method are presented in [35, 73].

While it provides a general mechanism for deconvolution, this least-squares approach has a number of limitations:

1. It is highly sensitive to the specification of multiple hyperparameters.
2. Regularisation is required to ensure the response function parameters correspond to a smooth response function.
3. The flexibility of the linear spline representation admits a wider range of response functions than is physically possible.
4. It lacks a coherent approach to uncertainty analysis and quantification.

The von Schroeter algorithm spawned many variations and extensions. Most notably we refer to [49], which introduces a slightly modified version of the error model of the original paper and Ilk's B-splines representation for the response [42]. A comparison of all three can be found in [56].

Additionally, there are other new approaches, including [16], which models the quantity in question and the pressure as Gaussian processes, and combines the response model with a Bayesian statistical approach. However, as with the linear spline, the flexibility of the Gaussian process model can easily provide response functions that are not physically possible.

To address the limitations of the total least square method and its derivatives, we propose a Bayesian approach to the well testing deconvolution problem:

1. We replace the linear spline representation of the response function, whose vast flexibility can easily produce non-physical results, by appropriate parametric models, which aim to be sufficiently general, while also limiting the space of response functions.

2. We introduce a Bayesian approach, which accounts for errors on the independent variables, and inherently provides for a coherent uncertainty analysis of the response function and its parameters.

3. We can potentially use the Bayesian model to incorporate expert knowledge to the model through the prior, and further limit the space of plausible response functions, and also treat the essential and highly-sensitive hyperparameters as variance parameters in our Bayesian model, and marginalise them out of the problem.

1.5 Discussion and overview

In this chapter we introduced the general context and the main terminology of well test analysis. We presented a literature review on the different methodologies used in well testing, leading up to deconvolution, which is the technique that is going to be the main topic of this thesis. We discused the main methods used in order to tackle deconvolution, including the state of the art methodology, which is based on total least squares. Finally, we examined the limitations of those methods and we justified our motivation on using a Bayesian approach.

One of the most important challenges for any deconvolution method is the choice of a representation for the response function. In Chapter 2 we examine different models for the reservoir response function, starting from a simple model, based on

straight lines. We proceed with a model based on a rectangular reservoir, and we conclude with the main model, which is based on a radial composite reservoir.

In Chapter 3 we present a Bayesian framework for the deconvolution problem in well test analysis. We present the data formation, we propose distributions for the prior densities and we manipulate the posterior in order to derive helpful forms.

In Chapter 4 we present a variety of computational methods that can be used in association with the Bayesian context, as well as tools for assessment of these methods.

In Chapters 5 and 6 we apply the models from Chapter 2, using the statistical context and algorithms from Chapters 3 and 4, on data, and examine the results. In Chapter 5 we validate our models on four simple synthetic data sets with known solutions, while in Chapter 6, we use the radial composite model on data derived from real oil and gas reservoirs.

Finally, in Chapter 7 we conclude this thesis with a summary and discussion.

Chapter 2

Models for the response function

2.1 Introduction

The first challenge to be faced in any approach to deconvolution in well test analysis is to decide on the prior probability distribution to be placed on the response function. The model must be both flexible enough to be able to capture all the possible shapes that might be encountered in practice, but also restrictive enough that prohibits non-physical results. In particular, a model expressed in terms of the properties of the well and the reservoir would allow for a more detailed level of interpretation more closely related to the physical system than would be possible with a more generic representation, such as a spline. This would permit comparison to the results of other well test analysis methods, provide constraints on and a physical interpretation for the prior, and even provide direct information about some physical aspects of the system.

For these purposes, we introduce in this chapter a model based on a multi-region radial composite reservoir [1, 81], which has many advantages: first, through different parameter combinations the model can represent the majority of plausible response shapes. Second, since the model is a solution to the diffusion equation representing the physical fluid flow problem, it is naturally restricted to a space of only physically

sensible functional forms. Finally, the model itself is parameterised in a manner that can be associated with the flow behaviour in the reservoir [7].

In this chapter, we construct and present some simple models for the reservoir response function, designed in a manner that aims to tackle the aforementioned issues, and then, we review the literature regarding the main response model that we are going to use in the later Chapters in two steps. First we illustrate the simpler form of the two region radial composite model, and then we extend to the more general multi-region case.

Results on synthetic data for all models are presented in Chapter 5, while for the radial composite model, results on field data are shown in Chapter 6.

2.2 The straight lines model

The straight lines model is based on identifying some key features of the response function, which correspond to properties that have been widely studied in the literature and in practice. Most of them are loosely based on characteristic points from [33, 77], in which the authors introduce new techniques where straight lines from the pressure and pressure derivative curves of a simple radial flow reservoir with wellbore storage and skin are used, in order to solve directly for well and reservoir parameters.

The points in question are illustrated in Figure 2.1, and together they assemble a set of response features. Using them as model parameters, we can construct a response as a set of straight line segments, which are then treated with post-hoc smoothing techniques in order to produce a final smooth curve. They include:

i **The wellbore storage ending time** K_1, which is the theoretical point where the early time unit slope ends, and reflects the transition from drawing the oil stored in the wellbore to drawing oil from the reservoir. It depends on the well-

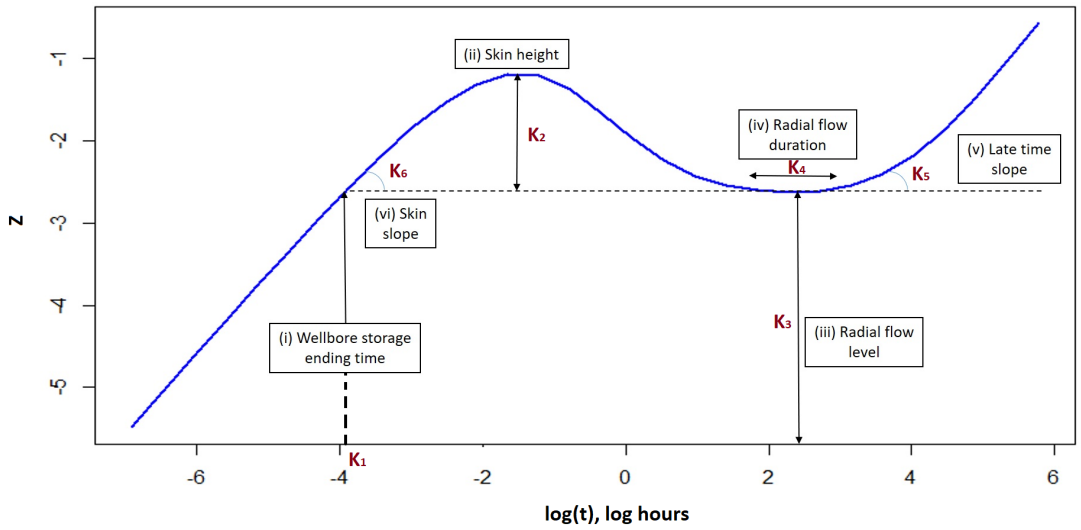


Figure 2.1: Straight line model response and parameters.

bore storage coefficient, as well as reservoir quantities, such as the permeability, viscosity, and other formation properties.

ii **The skin height** K_2, which corresponds to the size of the early 'hump' and is an indicator of the wellbore's damage, and the formation near the wellbore.

iii **The radial flow level** K_3, which is connected to attributes of the reservoir, like the permeability, viscosity, the formation thickness and the formation volume factor, and, in general, can be interpreted as the behaviour of the response curve after the reservoir production has been established and before the detection of boundaries and/or heterogeneities.

iv **The radial flow duration** K_4, which indicates the duration, during which, fluid moves towards the wellbore from all directions before it detects the presence of boundaries and/or heterogeneities.

v **The late time slope** K_5, which is connected to the nature of the flow in the reservoir at late times and is usually associated with the drop in pressure due to boundary effects.

vi **The skin slope** K_6, which is an additional regularising parameter that aims to add flexibility, in order to resemble an early time shape more similar to that of a physical response.

The piecewise linear model is:

$$z(\tau) = \begin{cases} \tau + K_3 - K_1 & \tau \leq K_1 \\ K_6\tau + K_3 - K_1K_6 & K_1 < \tau \leq K_1 + \frac{K_2}{K_6} \\ -K_6\tau + K_3 + K_1K_6 + 2K_2 & K_1 + \frac{K_2}{K_6} < \tau \leq K_1 + \frac{2K_2}{K_6} \\ K_3 & K_1 + \frac{2K_2}{K_6} < \tau \leq K_1 + K_4 + \frac{2K_2}{K_6} \\ K_5\tau + K_3 - K_1K_5 - K_4K_5 - \frac{2K_2K_5}{K_6} & K_1 + K_4 + \frac{2K_2}{K_6} < \tau. \end{cases}$$

After its construction, we fit a cubic smoothing spline, in order to derive the final smooth response.

The straight lines model has the ability to resemble some common simple response shapes, and it is also connected to some physical aspects of the wellbore - reservoir system, but it comes with many caveats: first, it is restrained to only one late time feature (slope), so it can realistically model only a handful of plausible response curves (such as IRF, channel and closed). Second, even for the early times, the fact that it is originally produced as connected straight lines, that attempt to model a smooth curve, introduces error, and can potentially produce non-physical results. Finally, even though some of the parameters (like the wellbore storage ending time and the radial flow level) have a direct connection to reservoir and wellbore features, that is not generally the case.

In order to tackle those limitations, we shifted our attention to the rectangular reservoir model.

2.3 The rectangular reservoir model

Unlike the straight lines model, the rectangular reservoir model seeks to link the model parameters directly to a simplified version of the reservoir geometry. The model is constructed as a combination of two distinct parts.

The first part is associated with the early time portion of the curve, and is derived from the exact solution to the diffusion equation for a well with wellbore storage and skin, acting on an infinite radial flow reservoir [2, 9]. In practice, that is the solution of Equation 1.1.1, but with additional inner boundary conditions that reflect the wellbore storage and skin effects. For the bottomhole pressure, in Laplace space, it is:

$$\overline{p}_D(W) \approx \frac{1}{W(W + \frac{1}{\ln \frac{2}{\gamma\sqrt{W/C_D e^{2S}}}})}$$

where $\bar{\cdot}$ denotes the Laplace transform, p_D is the dimensionless pressure drop; that is the product of the pressure drop and a quantity of reservoir parameters $\widehat{P_M}$, so that $p_D = \widehat{P_M}\Delta\tilde{p}$, W is the Laplace variable that corresponds to the dimensionless time $t_D = \widehat{T_M}t$. T_M and P_M are described in Table 2.1 and $\hat{\cdot}$ denotes the parameter without the logarithmic transformation. Those quantities and symbols will be examined further in the next section. γ is the Euler constant, S is the skin and C_D is the dimensionless wellbore storage coefficient.

This solution contributes three parameters to the model, associated with characteristics and attributes of the well and the reservoir, namely permeability, formation thickness and skin. Two of these are the scaling parameters for the time and the pressure, while the third accounts for the wellbore storage and the skin effect at early times. The same parameters are also included in the radial composite model, which is presented in the next section, where they are described extensively.

The second part of this model reflects the later time behaviour and can be described by the image method solution for a theoretical rectangular reservoir [53]. The image

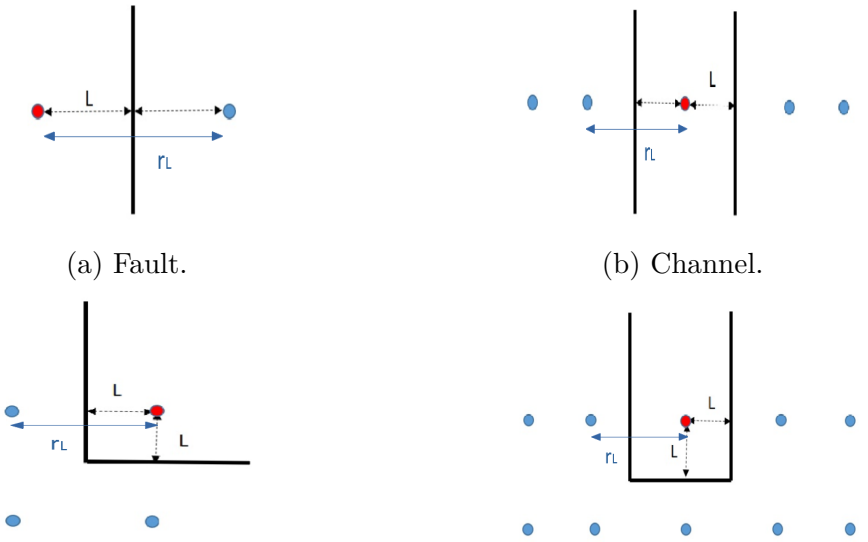
method is used in differential equation solutions in order to implement boundary conditions. In this context, it introduces image wells at specific distances from the active well, in order to produce an equivalent system that comprises no-flow boundaries, which can mimic different boundary configurations such as faults and channels. For instance, as shown in Figure 2.2a, in order to resemble a single fault at distance L from an active well in an otherwise infinite reservoir, we place an image well, with the same production rate, equidistantly (at a distance $2L$ from the active well) on the other side of the boundary. The system of the two wells generates a virtual no-flow boundary down the centre of symmetry with identical behaviour to that of a single well in the vicinity of a physical boundary. Effectively, we use Equation 1.1.1 for both wells, and the pressure in the bottomhole of the active well can be found by adding the solution in Equation 1.1.2 for the appropriate distances. Eventually the dimensionless pressure is:

$$2p_D = Ei\left(-\frac{1}{4t_D}\right) + Ei\left(-\frac{(2L_D)^2}{4t_D}\right). \tag{2.3.1}$$

Other examples of elementary fault systems are shown in Figure 2.2. In Figure 2.2b, a channel is modelled by two image wells each at a distance $2L$ from the active well, two at a distance $4L$, and so on, to infinity. In Figure 2.2c two orthogonal faults are modelled by three image wells, two at a distance $2L$ and one at a distance $2\sqrt{2}L$ from the active well, etc.

Given that the reservoir is rectangular, we can add a sufficient number of image wells at appropriate distances from the active well, and the only necessary parameters for the construction of this portion of the model will be those directly associated with the distance between the active well and the reservoir boundaries. Eventually, for a pre-specified number of image wells, all the possible boundary configurations can be described by introducing four additional parameters, that determine the exact position of the active well in the reservoir as in Figure 2.3.

Using Equation 1.1.2 (with the same manner as we did in Equation 2.3.1), the



(a) Fault. (b) Channel.

(c) Orthogonal faults. (d) U-shaped fault system.

Figure 2.2: Location of image wells for various reservoir boundary configurations.

dimensionless pressure derivative for k image wells is [73]:

$$\frac{\partial p_D}{\partial(\ln t_D)} = \frac{1}{2}\left[\sum_{j=1}^{k} \exp\left(-\frac{r_{Lj}^2}{4t_D}\right)\right],$$

where r_{Lj} are the distances from the active well to each of the image wells.

Note that in order to keep the order of the boundaries and avoid parameter degeneracies, we model the distance from the well to the closest boundary as L_1, and then we use the increments L_2 and L_3 for its closest perpendicular and parallel boundaries respectively, while the furthest perpendicular boundary is modelled as $L_1 + L_2 + L_4$.

The final response model is produced by combining the early-time Laplace solution and the late-time image method solution at their intersection, which is the point where we observe the transition from the constant level of the radial flow regime to the shape of the late time behaviour, as seen in Figure 2.4.

The rectangular reservoir model is more representative of an actual reservoir when compared to the straight line model. First, all its parameters are associated with the system: the first three with characteristic quantities of the wellbore and the reservoir, and the last four with the shape of a simplified version of the reservoir. It

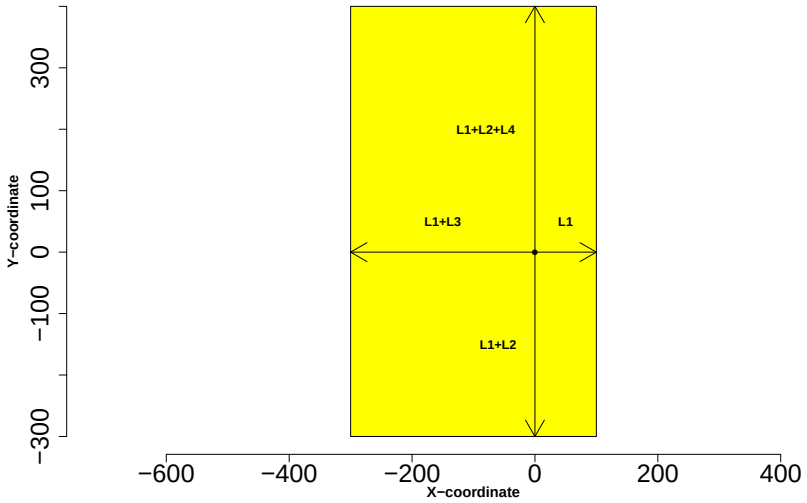
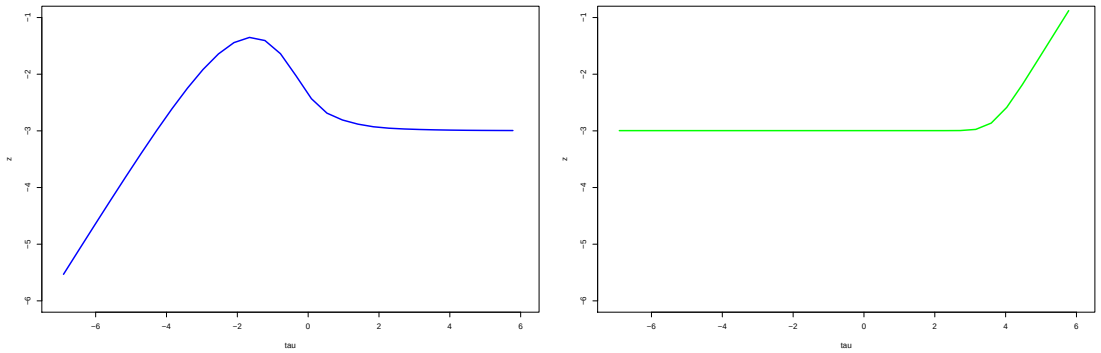


Figure 2.3: Rectangular reservoir and associated reservoir parameters.

is also connected with actual solutions to the diffusion equation, so it is restricted to physical results. Lastly, it can model a larger variety of response shapes (including the ones that correspond to a fault, and a U-shape reservoir).

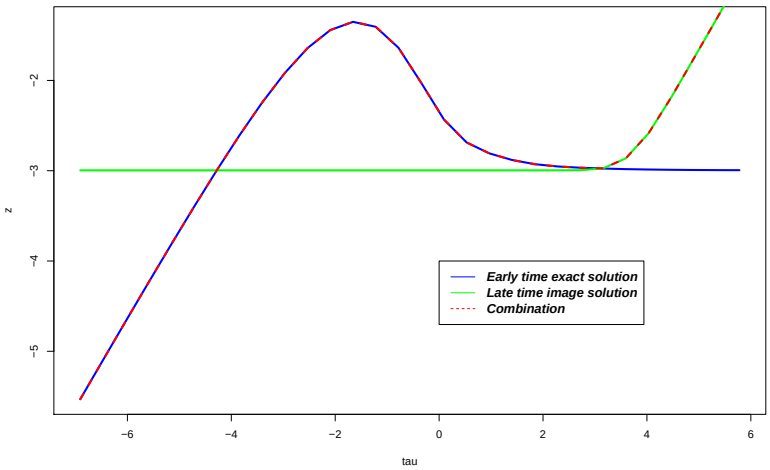
Its main limitations regard its late times flexibility, since it is restricted to only the few features (and their combinations) that a rectangular reservoir can produce. Adding to the complexity of the reservoir could potentially extend the possible features that could be modelled, but not without introducing a number of parameters that would make the inference more difficult. In addition, it would not guarantee the modelling of every plausible system, including reservoir heterogeneities. Finally, the manner in which the early time and late time portions of the model are connected introduces additional error and should give way to a more principled and smooth transition between flow regimes.

For these reasons, we introduce our third and final model, which improves on all the above mentioned issues.



(a) Early time.

(b) Late time.



(c) Rectangular reservoir model's response.

Figure 2.4: Construction of the rectangular reservoir model's response curve.

2.4 The simple radial composite model

The main response model described in this chapter is the multi-region radial composite model. Its basis is the simple radial composite model with an infinite outer region [67]. We adopt a variation based on the effective wellbore radius [22], which is equal to $r_w e^{-S}$, and corresponds to a theoretical radius of a wellbore that takes into account the skin, since it helps decreasing the number of necessary parameters. This model represents the reservoir as two porous concentric circular regions, in the centre of which lies the wellbore. The first region comprises the area between the wellbore and the boundary to the second region, which begins at this transition point and then extends infinitely as shown in Figure 2.5. The two regions are characterised by

potentially different mobility and diffusivity parameters, thus permitting a transition between different media to occur at the interface.

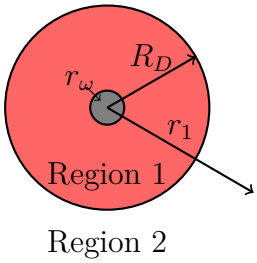


Figure 2.5: Simple radial composite reservoir model.

The simple radial composite model has six parameters – three correspond to a well with wellbore storage and skin in a homogeneous reservoir [10], and impact the early times response behaviour; and the remaining three parameterise the nature and extent of the transition at the boundary of the inner region, and primarily affect the mid- to late-time behaviour. The six parameters are summarised in Table 2.1. For reasons explained in the next chapter, a $\log_{10}$ transformation of the parameters was taken. In order to avoid confusion, for the rest of this chapter, as mentioned before, the corresponding parameters with $\widehat{\cdot}$ will denote the parameter without the logarithmic transformation.

Each of the parameters is expressible in terms of combinations of the following fundamental wellbore and reservoir properties: the permeability of the medium k, the viscosity μ, the porosity ϕ and the total system compressibility c_t, the formation thickness h, the wellbore storage coefficient C, the wellbore radius r_w, the transition radius r_1 and the wellbore skin S. The parameters k, μ, ϕ, and c_t each vary according to the region of the radial model. Primarily, it is the change in and contrast between the values of these quantities at the transition boundary which represent reservoir features such as faults and changes in rock properties and which manifest as identifiable features in the response function. The relationships between the model parameters and the fundamental parameters are given in the final column of Table 2.1. However, we note that due to the non-identifiability of these fundamental parameters we will work with the six derived parameters instead. Thus in total,

the simple radial composite model is expressed in terms of the unique set of six parameters: $\boldsymbol{\phi} = \{P_M, T_M, C_D e^{2S}, R_D, M, \eta\}$.

The governing equation for the radial composite model is expressed by dimensionless variables in two cases. For the first (inner) region where $1 \leq r_D \leq \widehat{R_D}$, it is

$$\frac{\partial^2 P_{D1}}{\partial r_D^2} + \frac{1}{r_D}\frac{\partial P_{D1}}{\partial r_D} - \frac{1}{\widehat{C_D e^{2S}}}\frac{\partial P_{D1}}{\partial t_D} = 0$$

and for the second (outer) region, where $\widehat{R_D} \leq r_D$

$$\frac{\partial^2 P_{D2}}{\partial r_D^2} + \frac{1}{r_D}\frac{\partial P_{D2}}{\partial r_D} - \frac{\widehat{\eta}}{\widehat{C_D e^{2S}}}\frac{\partial P_{D2}}{\partial t_D} = 0 \tag{2.4.1}$$

where $t_D = \widehat{T_M} t$ is the dimensionless time, $p_{Di} = \widehat{P_M}\Delta P_i$ is the dimensionless pressure drop in region i, and $r_D = \frac{r_1}{r_w e^{-S}}$ is the dimensionless radius; that is the radius scaled by the effective wellbore radius. These equations express the flow behaviour in a radial composite model that includes wellbore storage and skin effects.

The parameters $\boldsymbol{\phi}$, along with being essential for the specification of the governing equation, can also fully describe the response model, since a transformation of the diffusion equation's solution produces a response g that can then be used in the deconvolution formula Equation 1.3.1. The equation is described extensively for the multi-region case in Subsection 2.5.1.

We illustrate the effects of changing the parameters $\boldsymbol{\phi}$ on the features of the 2-region model response function in Figure 2.6. The first two parameters, T_M and P_M, correspond to a simple translation of the response function, by scaling the time and pressure to dimensionless quantities respectively, and represent the responsiveness of the well and reservoir to production, by including the permeability thickness product kh, which is a key factor in the flow potential of a well. The parameter $C_D e^{2S}$ governs the impediment to flow due to the skin S effect surrounding the well, and shows the magnitude of the wellbore storage effect. The parameter R_D corresponds to the distance from the wellbore to the inter-region transition, and so

Table 2.1: Parameters of the simple radial composite reservoir model.

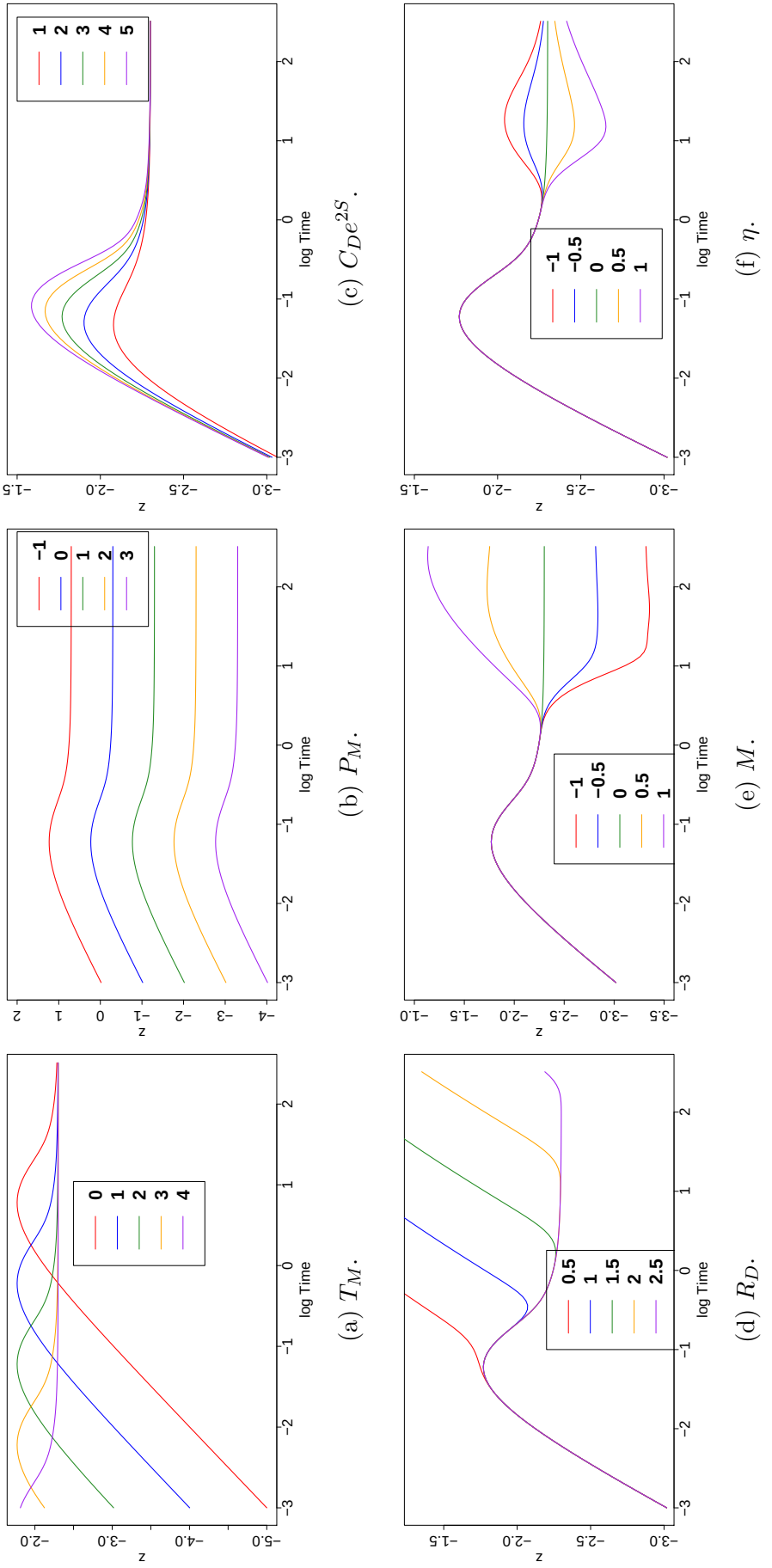
Parameter	Name	Scope	Description	Definition
P_M	Pressure match	Global	Vertical shift	$P_M = \log_{10}\left(\frac{2\pi kh}{\mu}\right)$
T_M	Time match	Global	Horizontal shift	$T_M = \log_{10}\left(\frac{2\pi kh}{\mu C}\right)$
$C_D e^{2S}$	Wellbore storage coefficient	Early time	Early time shape	$C_D e^{2S} = \log_{10}\left(\frac{Ce^{2S}}{2\pi\phi c_t h r_w^2}\right)$
R_D	Radius parameter	Transition	Dimensionless distance to the transition interface	$R_D = \log_{10}\left(\frac{r_1}{r_w e^{-S}}\right)$
M	Mobility ratio	Transition	Ratio of mobility between the 1-st and the 2-nd region	$M = \log_{10}\left(\frac{(k/\mu)_1}{(k/\mu)_2}\right)$
η	Diffusivity ratio	Transition	Ratio of diffusivity between the 1-st and the 2-nd region	$\eta = \log_{10}\left(\frac{(k/\mu\phi c_t)_1}{(k/\mu\phi c_t)_2}\right)$

affects how soon or late the impact of that transition is perceived. Finally, M and η are two parameters corresponding to the relative change in reservoir properties at the transition resulting in a shift in the response function stabilisation level, or a localised deviation from that level, and they reflect the productivity of the wellbore and the deliverability of the reservoir. Those last parameters are particularly useful, because, as will be examined further in Chapter 5, specific configurations of those can describe non-radial structures (e.g. a very specific change of mobility can produce the same response as a fault).

2.5 The multi-region radial composite model

The simple radial model with infinite outer boundary can be extended to a multi-region case [1, 81] by introducing additional concentric radial regions around the model, resulting in a model of n regions and $n-1$ transitions as shown in the Figure 2.7. The $n = 2$ case coincides directly with the simple radial composite model described above.

The multi-region model involves the same parameters as the simple radial model, but introduces an additional radius parameter, mobility ratio, and diffusivity ratio for each transition beyond the first. This gives a total of $3n$ parameters in the parameter vector $\boldsymbol{\phi} = \{P_M, T_M, C_D e^{2S}, R_{D1}, M_1, \eta_1, \ldots, R_{D(n-1)}, M_{n-1}, \eta_{n-1}\}$, where: $R_{Di} = \log_{10}\left(\frac{r_i - r_{i-1}}{r_w e^{-S}}\right)$ are the dimensionless radii increments, $M_i = \log_{10}\left(\frac{(k/\mu)_i}{(k/\mu)_{i+1}}\right)$ are the ratios of mobility between the i and $i+1$ region, and $\eta_i = \log_{10}\left(\frac{(k/\mu\phi c_t)_1}{(k/\mu\phi c_t)_{i+1}}\right)$ are the the ratios of diffusivity between the 1-st and i region. This parameterisation aims to avoid degeneracies and connect the parameters directly to the diffusion equation that will be described later. The introduction of these additional parameters and regions to the model introduces substantially more flexibility to the multi-region radial composite model beyond that of the simple radial model. With the addition of the extra regions, the model is permitted to represent sequences of multiple transitions, and so capture more complex response features than would be possible



(a) T_M. (b) P_M. (c) $C_D e^{2S}$.

(d) R_D. (e) M. (f) η.

Figure 2.6: Response for different values of each of the simple radial composite model's parameters. For the non-varying parameters the default values are 2, 2, 3, 1.5, 0 and 0 respectively (except from the case (d), where the ratio parameters are equal to 2 to aid the vizualisation of the transition).

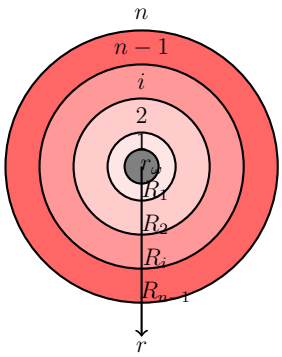


Figure 2.7: Multi-region radial composite reservoir model.

under the simple two region case. For instance, a shift of the stabilisation level, followed by a localised deviation from that level, require at least two transitions, where, in the former there is a change in mobility, and in the latter, in diffusivity. It is for these reasons that we shall adopt this multi-region model as the basis for our analysis in the subsequent sections.

2.5.1 Mathematical model for a multi-region radial composite model

The solution for the multi-region radial composite model wellbore pressure drop is expressed in terms of the parameters $\boldsymbol{\phi}$, all of which appear in the equation, the boundary conditions or the transformations to dimensionless quantities.

The mathematical model that follows consists of a governing equation, based on a variation of the diffusion equation for fluids in a radial direction in a porous medium and boundary conditions associated with the multi-region radial composite model. It is based on the following set of assumptions:

1. Negligible gravity effects, homogeneous isotropic porous medium, laminar flow, single phase and slightly compressible fluid.

2. The formation can be approximated as if it could be divided into n concentric regions. Within each region the corresponding parameters (mobility and diffusivity) remain unchanged, but are different for each region.

3. The wellbore storage and skin effects are taken into consideration.

4. The interface (transition) width is considered infinitesimal and the change of the parameters at the interface is abrupt.

The governing equation in dimensionless form for a multi-region model is the natural extension of Equation 2.4, where the pressure, radius and diffusivity ratios are dependent on the region. For the first (inner) region where $1 \leq r_D \leq \widehat{R_{D1}}$ is:

$$\frac{\partial^2 P_{D1}}{\partial r_D^2} + \frac{1}{r_D}\frac{\partial P_{D1}}{\partial r_D} - \frac{1}{\widehat{C_D e^{2S}}}\frac{\partial P_{D1}}{\partial t_D} = 0$$

For any other $i = 2, \cdots, n$ region where $\widehat{R_{D(i-1)}} \leq r_{Di} \leq \widehat{R_{Di}}$

$$\frac{\partial^2 P_{Di}}{\partial r_D^2} + \frac{1}{r_D}\frac{\partial P_{Di}}{\partial r_D} - \frac{\widehat{\eta_i}}{\widehat{C_D e^{2S}}}\frac{\partial P_{Di}}{\partial t_D} = 0. \tag{2.5.1}$$

Note that the diffusivity changes between the regions are taken into account within the governing equations as the ratios of the diffusivity of the first region over the $i+1$ region, and that the parameter $\widehat{C_D e^{2S}}$ is introduced because we include the wellbore storage and skin effects. Without these additional assumptions, the equation would be similar to the simpler version of the diffusion equation in Equation 1.1.1.

The initial condition is

$$P_{Di}(r_D, 0) = 0, \tag{2.5.2}$$

which translates to the pressure being equal to the initial pressure p_0 (equivalent to the dimensionless pressure being equal to 0) at the beginning of the test. The inner boundary conditions are

$$\frac{\partial p_D}{\partial t_D} - \frac{\partial P_{D1}}{\partial r_D} = 1, \quad r_D = 1$$
$$p_D = P_{D1} \tag{2.5.3}$$

and they reflect the pressure behaviour on the interface between the wellbore and the first region. The former takes into account the wellbore storage and skin effects, while the latter ensures that the pressure on that interface does not change. The pressure interface conditions are

$$\begin{aligned} P_{D_i} &= P_{D_{i+1}}, \quad i = 1, 2, \dots \\ \frac{\partial P_{D_{i-1}}}{\partial r_D} &= \frac{1}{\widehat{M_i}} \frac{\partial P_{Di}}{\partial r_D} \end{aligned} \tag{2.5.4}$$

where $\widehat{M_i}$ is the mobility ratio between i and $i+1$ regions. Similar to the previous boundary condition, the former equation reflects that the pressure does not change on the interface between the regions. The latter accounts for the abrupt change of mobility on the interface. Finally, the outer boundary condition, which accounts for an infinite reservoir, through the unconstrained radius of the final region, is

$$P_{Dn}(r_D \to \infty, t_{Dn}) = 0 \tag{2.5.5}$$

The solution for the governing equation is obtained using the Laplace transformation technique with the appropriate initial, boundary and interface conditions. For the dimensionless wellbore pressure drop in Laplace space, it is:

$$\overline{p}_D = G_1 I_0 \left(\sqrt{\frac{s}{\widehat{C_D e^{2S}}}} \right) + G_2 K_0 \left(\sqrt{\frac{s}{\widehat{C_D e^{2S}}}} \right) \tag{2.5.6}$$

where the constants G_i are obtained by solving the system of equations resulting from the use of boundary conditions in Laplace space and $p_D = \widehat{P_M} \Delta p$ is the dimensionless wellbore pressure drop. Specifically, from (2.5.3)-(2.5.5) we find

$$\begin{aligned} a_{1,1} G_1 + a_{1,2} G_2 &= \frac{1}{s} \\ \left(\sum_{k=-1}^{2} a_{2i,2i+k} G_{2i+k} \right) &= 0 \\ \left(\sum_{k=-1}^{2} a_{2i+1,2i+k} G_{2i+k} \right) &= 0 \end{aligned}$$

$$a_{2i,2i-1}G_{2i-1} + a_{i,i}G_{2i} = 0$$

The terms $a_{i,j}$ are the constant coefficients of the G_j in the i-th equation, and their analytic expressions are given in Table 2.2.

$a_{1,1}$	$= sI_0\left(\sqrt{\frac{s}{\widehat{C_De^{2S}}}}\right) - \sqrt{\frac{s}{\widehat{C_De^{2S}}}}I_1\left(\sqrt{\frac{s}{\widehat{C_De^{2S}}}}\right)$
$a_{1,2}$	$= sK_0\left(\sqrt{\frac{s}{\widehat{C_De^{2S}}}}\right) + \sqrt{\frac{s}{\widehat{C_De^{2S}}}}K_1\left(\sqrt{\frac{s}{\widehat{C_De^{2S}}}}\right)$
$a_{2,1}$	$= I_0(\widehat{R_{D1}}\sqrt{s})$
$a_{2,2}$	$= K_0(R_{D1}\sqrt{s})$
$a_{3,1}$	$= \widehat{M_1}\sqrt{\frac{s}{\widehat{C_De^{2S}}}}I_1(\widehat{R_{D1}}\sqrt{s})$
$a_{3,2}$	$= -\widehat{M_1}\sqrt{\frac{s}{\widehat{C_De^{2S}}}}K_1(\widehat{R_{D1}}\sqrt{s})$
$a_{2i,2i-1}$	$= I_0(\widehat{R_{Di}}\sqrt{s\widehat{\eta_{i-1}}}), \quad i = 2,3,\cdots,n-1$
$a_{2i,2i}$	$= K_0(\widehat{R_{Di}}\sqrt{s\widehat{\eta_{i-1}}}), \quad i = 2,3,\cdots,n-1$
$a_{2i,2i+1}$	$= -I_0(\widehat{R_{Di}}\sqrt{s\widehat{\eta_i}}), \quad i = 1,2,\cdots,n-1$
$a_{2i,2i+2}$	$= -K_0(\widehat{R_{Di}}\sqrt{s\widehat{\eta_i}}), \quad i = 1,2,\cdots,n-1$
$a_{2i+1,2i-1}$	$= \widehat{M_i}\sqrt{\frac{s\widehat{\eta_{i-1}}}{\widehat{C_De^{2S}}}}I_1(\widehat{R_{D1}}\sqrt{s\widehat{\eta_{i-1}}}), \quad i = 2,3,\cdots,n-1$
$a_{2i+1,2i}$	$= -\widehat{M_i}\sqrt{\frac{s\widehat{\eta_{i-1}}}{\widehat{C_De^{2S}}}}K_1(\widehat{R_{D1}}\sqrt{s\widehat{\eta_{i-1}}}), \quad i = 2,3,\cdots,n-1$
$a_{2i+1,2i+1}$	$= -\sqrt{\frac{s\widehat{\eta_i}}{\widehat{C_De^{2S}}}}I_1(\widehat{R_{Di}}\sqrt{s\widehat{\eta_i}}), \quad i = 1,2,\cdots,n-1$
$a_{2i+1,2i+2}$	$= \sqrt{\frac{s\widehat{\eta_i}}{\widehat{C_De^{2S}}}}K_1(\widehat{R_{Di}}\sqrt{s\widehat{\eta_i}}), \quad i = 1,2,\cdots,n-1$

Table 2.2: The constant coefficients $a_{i,j}$ for the multi-region radial composite reservoir model.

Since the solution is in Laplace space, in order to transition from the pressure to the pressure derivative, the derivative Laplace transform property needs to be applied:

$$\overline{\frac{\partial p_D}{\partial t_D}} = \overline{sp}_D - p_D(t_D = 0) = \overline{sp}_D,$$

Using the Stehfest algorithm for numerical Laplace inversion [72] we derive the dimensionless wellbore pressure drop in the time domain $\frac{\partial p_D}{\partial t_D}$. Finally the response is:

$$z(\tau) = \log\left(tg(t)\right) = \log\left(t\frac{\partial p}{\partial t}\right) = \log\left(t_D\frac{\partial p_D}{\partial t_D}/\widehat{P_M}\right) = \log\left(t_D\right) + \log\left(\frac{\partial p_D}{\partial t_D}\right) - log(\widehat{P_M}), \tag{2.5.7}$$

where $\tau = \ln t$.

$$
\begin{aligned}
a_{1,1} &= sI_0\left(\sqrt{\tfrac{s}{\widehat{C_De^{2S}}}}\right) - \sqrt{\tfrac{s}{\widehat{C_De^{2S}}}}I_1\left(\sqrt{\tfrac{s}{\widehat{C_De^{2S}}}}\right)\\
a_{1,2} &= sK_0\left(\sqrt{\tfrac{s}{\widehat{C_De^{2S}}}}\right) + \sqrt{\tfrac{s}{\widehat{C_De^{2S}}}}K_1\left(\sqrt{\tfrac{s}{\widehat{C_De^{2S}}}}\right)\\
a_{1,3} &= 0\\
a_{2,1} &= I_0(\widehat{R_{D1}}\sqrt{s})\\
a_{2,2} &= K_0(\widehat{R_{D1}}\sqrt{s})\\
a_{2,3} &= -K_0(\widehat{R_{D1}}\sqrt{s\widehat{\eta_1}})\\
a_{3,1} &= \widehat{M_1}\sqrt{\tfrac{s}{\widehat{C_De^{2S}}}}I_1(\widehat{R_{D1}}\sqrt{s})\\
a_{3,2} &= -\widehat{M_1}\sqrt{\tfrac{s}{\widehat{C_De^{2S}}}}K_1(\widehat{R_{D1}}\sqrt{s})\\
a_{3,3} &= \sqrt{\tfrac{s\widehat{\eta_1}}{\widehat{C_De^{2S}}}}K_1(\widehat{R_{D1}}\sqrt{s\widehat{\eta_1}}), \quad i = 1, 2, \cdots, n-1
\end{aligned}
$$

Table 2.3: The constant coefficients $a_{i,j}$ for the simple radial composite reservoir model.

For the special case of the two-region radial composite model all the initial, boundary and interface conditions of the multi-region radial composite model specification apply for $i = 2$, reducing to a simpler problem, and the terms $a_{i,j}$ are given in Table 2.3.

The simple radial composite model and its multi-region extension comprise all the desirable qualities that we require to proceed to the formulation of the statistical model: first, it produces a response, based on a solution to the diffusion equation, given a specific configuration of initial and boundary conditions. As a result, the response curves are always physical, which is an advantage against the piecewise linear response of the total least squares method and our straight lines model. Second, the combination of the late time parameters, as well as the ability of the multi-region version to resemble different features, offer flexibility, that permits the model to mirror the majority of all plausible curves, which is a big improvement compared to our two previous models. Finally, the parameterisation of the model is made in terms of parameters that are all connected with the reservoir flow behaviour, an attractive property that will aid the interpretation of the results.

2.6 Summary

In this chapter we introduced three different models to serve the purpose of modelling the reservoir response function. First, we started with the straight lines model. Its construction from only six key points associated with the system offered simplicity, but at the same time it generated a model too restrictive to resemble even some common reservoir features. We proceeded with the rectangular reservoir model, which was the first attempt to associate the parameters with the flow behaviour in the reservoir. It suffered from similar issues, since, even though it was more flexible than its predecessor, it could model only a handful of features that corresponded to a few reservoir configurations. Finally, we settled on the radial composite model and its multi-region extension, which combined flexibility and, as a solution to the diffusion equation, the ability to associate the model parameters to the flow behaviour in the reservoir, with the additional advantage of producing guaranteed physical solutions.

In the next chapter we will present a Bayesian statistical model for the deconvolution problem in well test analysis.

Chapter 3

Bayesian statistical model

3.1 Introduction

In this chapter we develop a Bayesian model for the well test data, centred around the assumptions of Gaussian noise for all the quantities and that the convolution model of Equation 1.3.1 holds. The well pressure, $\tilde{p}$, is observed with noise at times $\boldsymbol{t} = (t_i)$, idealised as point measurements, giving pressure data $\boldsymbol{p} = (p_i)$, for $i = 1, \ldots, m$. We have a single observation with noise, p_0, for the initial reservoir pressure $\tilde{p}_0$. The rate is modelled as periods of constant flow, $\tilde{\boldsymbol{q}} = (\tilde{q}_j)$, defined over known time intervals $\boldsymbol{T} = ([T_j, T_{j+1}])$, which are observed with noise as rate values $\boldsymbol{q} = (q_j)$, for $j = 1, \ldots, N$. We denote the piecewise constant function given by $\boldsymbol{T}$ and $\tilde{\boldsymbol{q}}$, with zeroes elsewhere, as $\tilde{\boldsymbol{q}}(t)$. That is:

$$\tilde{\boldsymbol{q}}(t) = \begin{cases} \tilde{q}_i, & t \in [T_j, T_{j+1}] \\ 0, & \text{otherwise} \end{cases}.$$

We are interested in the response g, the true rates $\tilde{q}$ and true initial pressure $\tilde{p}_0$. Our data comprise the observed pressures $\boldsymbol{p}$, observed rates $\boldsymbol{q}$ and observed initial pressure p_0, as well as other prior information K (including *e.g.* $\boldsymbol{T}$).

Applying the Bayes Theorem gives us the following expression for the conditional probability:

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$
$$\propto P(x|\theta)P(\theta), \tag{3.1.1}$$

where x is the collection of data and θ the set of parameters, in which we are interested.

In this context, and bearing in mind Equation 1.3.1, and other prior knowledge K, Equation 3.1.1 becomes:

$$P(g(\boldsymbol{\phi}), \tilde{\boldsymbol{q}}, \tilde{p}_0 \mid \boldsymbol{p}, \boldsymbol{q}, p_0, K) \propto P(\boldsymbol{p} \mid \tilde{p}_0, g(\boldsymbol{\phi}) * \tilde{q}, K)\, P(\boldsymbol{q} \mid \tilde{\boldsymbol{q}}, K)$$
$$\times P(p_0 \mid \tilde{p}_0, K)\, P(g(\boldsymbol{\phi}) \mid K)\, P(\tilde{\boldsymbol{q}} \mid K)\, P(\tilde{p}_0 \mid K)\,, \tag{3.1.2}$$

where we have used various independences to be described later. As stated, g will be given by a parametric model based on an explicit solution to the diffusion equation, more specifically the radial composite model, as described in Section 2.5, with parameters $\boldsymbol{\phi}$, so that the distributions on g will in fact be on $\boldsymbol{\phi}$. In the next section, we describe this model in detail.

3.2 The data model

We begin by assuming a multivariate Gaussian distribution for the observed pressures:

$$\boldsymbol{p} \mid \boldsymbol{\phi}, \tilde{\boldsymbol{y}}, \sigma_p^2 \sim \mathrm{N}\left(\tilde{\boldsymbol{p}}(\boldsymbol{\phi}, \tilde{\boldsymbol{y}}), \boldsymbol{\Sigma}_p\right). \tag{3.2.1}$$

where $\tilde{\boldsymbol{y}} = (\tilde{\boldsymbol{q}}, \tilde{p}_0)$ is the vector of true values of independent variables.

Here $\tilde{\boldsymbol{p}}(\boldsymbol{\phi}, \tilde{\boldsymbol{y}}) = \tilde{p}_0 - \boldsymbol{C}(\boldsymbol{\phi})\tilde{\boldsymbol{q}}$ is the m-vector of true convolved pressures obtained by evaluating Equation 1.3.1 at times $\boldsymbol{t}$ using the reservoir parameters $\boldsymbol{\phi}$, the true rate values $\tilde{\boldsymbol{q}}$ and the true initial pressure $\tilde{p}_0$. Note that the time points $\boldsymbol{t}$ are not equally spaced, so that the obvious Fourier transform method for computing the

convolution is not available, and instead, we follow Subsection 1.4.2 assuming a linear spline z. We shall formally define the form of $\tilde{\boldsymbol{p}}$ below. The variance of the normal distribution is $\boldsymbol{\Sigma}_p$ which represents the observational error in the pressure data. Typically, we make an assumption of conditional independence given the other parameters resulting in the diagonal form $\boldsymbol{\Sigma}_p = \sigma_p^2 \boldsymbol{I}$.

The magnitude of the pressure error variances on the diagonal is informed by the performance of the pressure gauges, which are well-documented in the literature [19]. For a typical well test, it is considered that the accuracy is within ± 5 psi [19], though our simulations suggested larger errors which lead us to represent the pressure standard deviation parameter as $\sigma_p \sim \mathrm{U}(0, 5)$. As mentioned in Chapter 1, when dealing with pressure measurements from a gas reservoir, a conversion to pseudo-pressure needs to take place. This has an effect on the pressure standard deviation, which, according to our simulations with real gas data and expert judgement, can lead to an upper boundary of up to 10 times the magnitude. Thus, $\sigma_p \sim \mathrm{U}(0, 50)$ is more appropriate in this case.

Conditional on the true rate values, $\tilde{\boldsymbol{q}}$, we assume a further multivariate Gaussian model for the observed rate values:

$$\boldsymbol{q} \mid \tilde{\boldsymbol{q}}, \sigma_q^2 \sim N(\tilde{\boldsymbol{q}}, \boldsymbol{\Sigma}_q) \tag{3.2.2}$$

where $\boldsymbol{\Sigma}_q$ is the rate error variance matrix. Similarly to the pressure model, we assume conditional independence, which results to the diagonal form $\boldsymbol{\Sigma}_q = \sigma_q^2 \boldsymbol{I}$. Experience indicates that rate measurements are generally of a lower level of quality, with associated uncertainty of up to 10% of their magnitude [19]. There are many reasons for an error of this size, but principally the accuracy of measurement instruments is substantially lower, and more importantly, the rate values are often not directly measured. For instance, in cases where there are only a few directly measured rates then the rest of the production rate history is estimated indirectly from information gained elsewhere during the operation of the reservoir, leading to a large amount of additional uncertainty. In literature, it is common to model this

type of information (knowledge about the error percentage) with a log-normal error model. Instead, here we use a multivariate Gaussian for simplicity and in order to facilitate calculations as will be shown in Subsection 3.4.1. It is worth noting, though, that the simplicity and flexibility of the full error model means that such a change would be straight forward.

For rate values that are directly observed, we use a Gaussian prior for σ_q, with small variance proportional to the magnitude of the rate data: $\sigma_q \sim N\left(0.05q_m, (0.005q_m)^2\right)$, where we set $q_m = \max(\boldsymbol{q})$, or we keep the value of the variance fixed to a constant $\sigma_q = 0.05q_m$.

For rate values that are not directly observed, we split $\boldsymbol{\Sigma}_q$ into two components: one for the measured values with a variance parameter prior as above; and a second for the unobserved rates with a substantially increased prior variance, based on the level on confidence (usually 10 to 20 times larger, as in the example of Section 6.2). In that way, we reflect the increasing uncertainty and lack of confidence for those values. For reasons discussed above, fundamental differences between the behaviour of oil and gas mean that the units of production for oil and gas can differ in scale up to many orders of magnitude, which has to be reflected in changes in many of the parameter priors.

Finally, we adopt a Gaussian model for the initial pressure value p_0, namely:

$$p_0 \mid \tilde{p}_0, \sigma_{p_0}^2 \sim \mathrm{N}(\tilde{p}_0, \sigma_{p_0}^2), \tag{3.2.3}$$

which is centred on the true initial pressure $\tilde{p}_0$, with a standard deviation of σ_{p_0}, which is usually fixed to 10, again informed by expert judgement, our simulations, and the fact that initial pressure is usually recorded during an observation period before production, which usually leads to different confidence levels than the rest of the pressure measurements [19].

To complete our specification, we construct a model for the true pressure values

based on the convolution relationship Equation 1.3.1 for the pressure drop:

$$\tilde{\boldsymbol{p}}(\boldsymbol{\phi}, \tilde{\boldsymbol{y}}) = \tilde{p}_0 \mathbf{1} - \boldsymbol{c}(g(\boldsymbol{\phi}), \tilde{\boldsymbol{q}}), \tag{3.2.4}$$

where $\mathbf{1}$ is a vector of ones and the function $\boldsymbol{c}(\cdot, \cdot)$ denotes the vectorised application of the convolution integral to obtain true pressures at each of the pressure observation times. In general, this is a complex and non-trivial calculation requiring numerical integration. However, in the case where we treat the rate function as a stepwise constant, and the response function as a linear spline, a number of simplifications can be made [68]. First, the convolution operation is linearised into a matrix-vector product, simplifying the model to the form:

$$\tilde{\boldsymbol{p}}(\boldsymbol{\phi}, \tilde{\boldsymbol{y}}) = \tilde{p}_0 - \boldsymbol{C}(\boldsymbol{\phi})\tilde{\boldsymbol{q}}, \tag{3.2.5}$$

where $\boldsymbol{C}(\boldsymbol{\phi})$ is a $(m \times N)$ matrix of convolution coefficients depending on $\boldsymbol{\phi}$. Second, the elements of $\boldsymbol{C}$ can be found analytically given the response function parameters $\boldsymbol{\phi}$. Thus, given values of $\tilde{p}_0$, $\boldsymbol{\phi}$, and $\tilde{\boldsymbol{q}}$, our approach to the evaluation of $\tilde{p}$ comprises the following steps:

1. Given $\boldsymbol{\phi}$, evaluate the response function $g(\boldsymbol{\phi})$.
2. Linearly interpolate $g(\boldsymbol{\phi})$ to obtain a linear spline representation.
3. Evaluate $\boldsymbol{C}(\boldsymbol{\phi})$ from the interpolated response and stepwise rate structure (see Appendix B).
4. Use this information to calculate $\tilde{p}$ from Equation 3.2.5.

Overall, this general setup is comparable to that of a separable nonlinear errors-in-variables regression model [21]. The pressure can be viewed as the dependent variable with a nonlinear relationship with the reservoir parameters, $\boldsymbol{\phi}$, but a linear relationship with the true rate and initial pressure. The true rates and initial pressure are then the independent variables which are subject to re-evaluation due to the uncertainty in their values.

3.3 The prior

Given the likelihood from the previous section, we now require a prior distribution over the collection of parameters $\boldsymbol{\theta} = (\boldsymbol{\phi}, \tilde{\boldsymbol{y}}, \boldsymbol{\sigma})$, where $\boldsymbol{\phi}$ are the reservoir parameters, $\tilde{\boldsymbol{y}}$ are the true independent data variables, and $\boldsymbol{\sigma} = (\boldsymbol{\Sigma}_p, \boldsymbol{\Sigma}_q, \sigma_{p_0}^2)$ are the variance parameters. We separate these into two subsets: first, the reservoir model parameters, $\boldsymbol{\phi}$, comprise the characteristics and properties of the system of the wellbore and the reservoir. Second, the remaining parameters $\boldsymbol{\lambda} = (\tilde{\boldsymbol{y}}, \boldsymbol{\sigma})$ represent characteristics and properties of the data generation processes. For simplicity, we adopt an assumption of independence for the individual elements of the prior:

$$P(\boldsymbol{\theta}) = P(\boldsymbol{\phi}, \tilde{\boldsymbol{y}}, \boldsymbol{\sigma}) = P(\boldsymbol{\phi})P(\tilde{\boldsymbol{y}}, \boldsymbol{\sigma}),$$

and so we consider the prior distributions for the two parameter subsets separately.

3.3.1 Prior for model parameters

Absent of other information, we assume independence for each of the reservoir and system parameters in $\boldsymbol{\phi}$, producing a prior distribution of the following form:

$$P(\boldsymbol{\phi}) = P(P_M)P(T_M)P(C_D e^{2S}) \prod_{i=1}^{n-1} P(R_{Di})P(M_i)P(\eta_i),$$

for a multi-region radial composite model with n regions. For each component of $\boldsymbol{\phi}$, we approach prior specification from the perspective of eliminating non-physical and implausible values as robustly as possible. When possible, choices for priors are derived from information in geological and geophysical studies of the corresponding parameters, and supplemented by expert knowledge from petroleum engineers. Alternatively, when the parameters of our model do not equate to quantities of particular geological interest, we rely instead on a synthesis of expert judgement, inspection of results from the analysis of other well tests in the literature, and knowledge of the sensitivity of the response function to the parameters as seen in Figure 2.6.

Table 3.1: Parameters of the simple radial composite reservoir model and their corresponding prior densities.

Parameter	Name	Scope	Description	Oil reservoir prior	Gas reservoir prior
P_M	Pressure match	Global	Vertical shift	$\mathrm{N}(1.5, 0.2^2)$	$\mathrm{N}(-2.5, 0.15^2)$
T_M	Time match	Global	Horizontal shift	$\mathrm{N}(2, 0.2^2)$	$\mathrm{N}(0.5, 0.15^2)$
$C_D e^{2S}$	Wellbore storage coefficient	Early time	Early time shape	$\mathrm{Ga}(1, 0.2)$	$\mathrm{Ga}(1, 0.2)$
R_{Di}	Radius parameter	Transition	Dimensionless distance to transition	$N(2, 1^2)$	$N(2, 1^2)$
M_i	Mobility ratio	Transition	Ratio of mobility between regions	$\mathrm{N}(0, 1^2)$	$\mathrm{N}(0, 1^2)$
η_i	Diffusivity ratio	Transition	Ratio of diffusivity between regions	$\mathrm{N}(0, 1^2)$	$\mathrm{N}(0, 1^2)$

First, the parameters T_M and P_M are the time and pressure scale parameters respectively. In logarithmic scale, they correspond to horizontal and vertical shifts of the response function, as shown in Figure 2.6. Thus, a $\log_{10}$ transformation was taken. Of the reservoir parameters, T_M and P_M were those most heavily dependent to choice of data units, and so different priors are required for problems involving oil and those involving gas. After inspecting plausible ranges for the values, by fitting the model and analysing various field oil datasets, we chose Gaussian priors $T_M \sim \mathrm{N}(2, 0.2^2)$ and $P_M \sim \mathrm{N}(1.5, 0.2^2)$, which, from analysing and fitting the model to gas reservoir datasets in a similar way, become $T_M \sim \mathrm{N}(0.5, 0.15^2)$ and $P_M \sim \mathrm{N}(-2.5, 0.15^2)$.

From literature, it is known that the magnitude of the effect of the parameter $C_D e^{2S}$ is logarithmic in scale [7]. Values below 0.7 correspond to a stimulated well, between 0.7 and 3 refer to approximately a zero effect skin condition, whereas values above 3 are associated with a damaged wellbore. The prior we use is $C_D e^{2S} \sim \mathrm{Ga}(1, 0.2)$, since, in practice, the absence of a significant skin effect and the case of stimulated wells would be known to the analyst and the prior could be adjusted accordingly. In addition, the fact that values close to 5 (the mean of the distribution) seem more common, whereas increasingly larger values, which suggest more heavily damaged wellbore are less likely, led us to use this distribution.

The radii parameters are key in determining the position at which any boundary effects manifest in the response function. For consistency, we also applied a $\log_{10}$ transformation to the region radii parameters, which also corresponds to an intuitive scale of variation in the sensitivity analysis of the models, since the scaled values can more easily reflect the distances, where the features appear on the response curve (see Figure 2.6). A key consideration with the radii parameters is that the radius values are monotonically increasing in distance from the wellbore. To respect this ordering, we consider R_{D1} to be equal to the first radius r_1, scaled by the effective wellbore radius $r_w e^{-S}$, with the R_{Di} for $i > 1$ representing positive increments in distance from the previous boundary. In reality, we have very little knowledge of the R_{Di}, rather than the assumption that there is reasonable distance between the

different transitions (for example, in dimensionless terms that could translate to a value of 100, which leads to the value $\log_{10}(100) = 2$ for the parameters R_{Di}). Thus, we use, in distributional terms, $R_{Di} \sim \mathrm{N}(2, 1^2)$, for $i \geq 1$. It is worth noting that because of the log scaling, large negative values of R_{Di} translate to the two regions getting very close to each other.

Finally, for the mobility and storativity ratio parameters, the motivation for using the logarithmic scale came from the expectation that a ratio value and its reciprocal should have equal probability. Thus, we adopt a prior with a mean value of 0, which corresponds to maintaining the same values for the parameters between regions (e.g. $M = 0 \implies \log_{10}\left(\frac{(k/\mu)_1}{(k/\mu)_2}\right) = 0 \implies (k/\mu)_1 = (k/\mu)_2$), and to avoid degenerate solutions, when the ratio values grow exceptionally large, we restrict ourselves to a relatively modest variance. A symmetric distribution around 0 also reflects an absence of prior knowledge of whether the parameters will increase or decrease beyond the transition. Using a Gaussian prior density gives $M_i \sim \mathrm{N}(0, 1^2)$ and $\eta_i \sim \mathrm{N}(0, 1^2)$. The standard deviation of 1 translates to one mobility being 10 times larger than the other (because of the logarithmic transformation), which seems reasonable according to our simulations.

In Figure 3.1, we present corresponding plots for the median, the 95% and the 99% credible intervals of the oil priors presented in this section. Since we know that the system comprises a wellbore acting on an oil reservoir, we can significantly reduce the plausible shape ranges for T_M and P_M compared to Figure 2.6, where we make no such assumption. The $C_D e^{2S}$ prior range could potentially be narrowed down, but that would require expert knowledge for the specific geological structure of the wellbore-reservoir system. The last three parameters are typically the ones informed from the 'unseen' structure of the reservoir and, therefore, the corresponding priors reflect that lack of knowledge, with the median values being the ones dictated by the assumptions we made above. For the radius parameter, the lower end of the prior corresponds to response shapes where the first transition masks the early time behaviour, while, the upper end, reflects a model whose transitional effect does

not appear within the radius of investigation, since the pressure transient has not reached any boundary. For the M_i and η_i priors, the 0 median implies that the two regions have the same diffusivity value, and therefore we do not observe any change during the transition, while, as mentioned above, positive or negative values reflect whether the response is going to move upwards or downwards. Specifically for the M_i parameter, the very large values are actually big jumps that do not stabilise within the radius of investigation, and therefore appear as slopes. The peculiar behaviour for the smaller values corresponds to a steady state flow regime, which appears when the pressure transient reaches a constant pressure boundary.

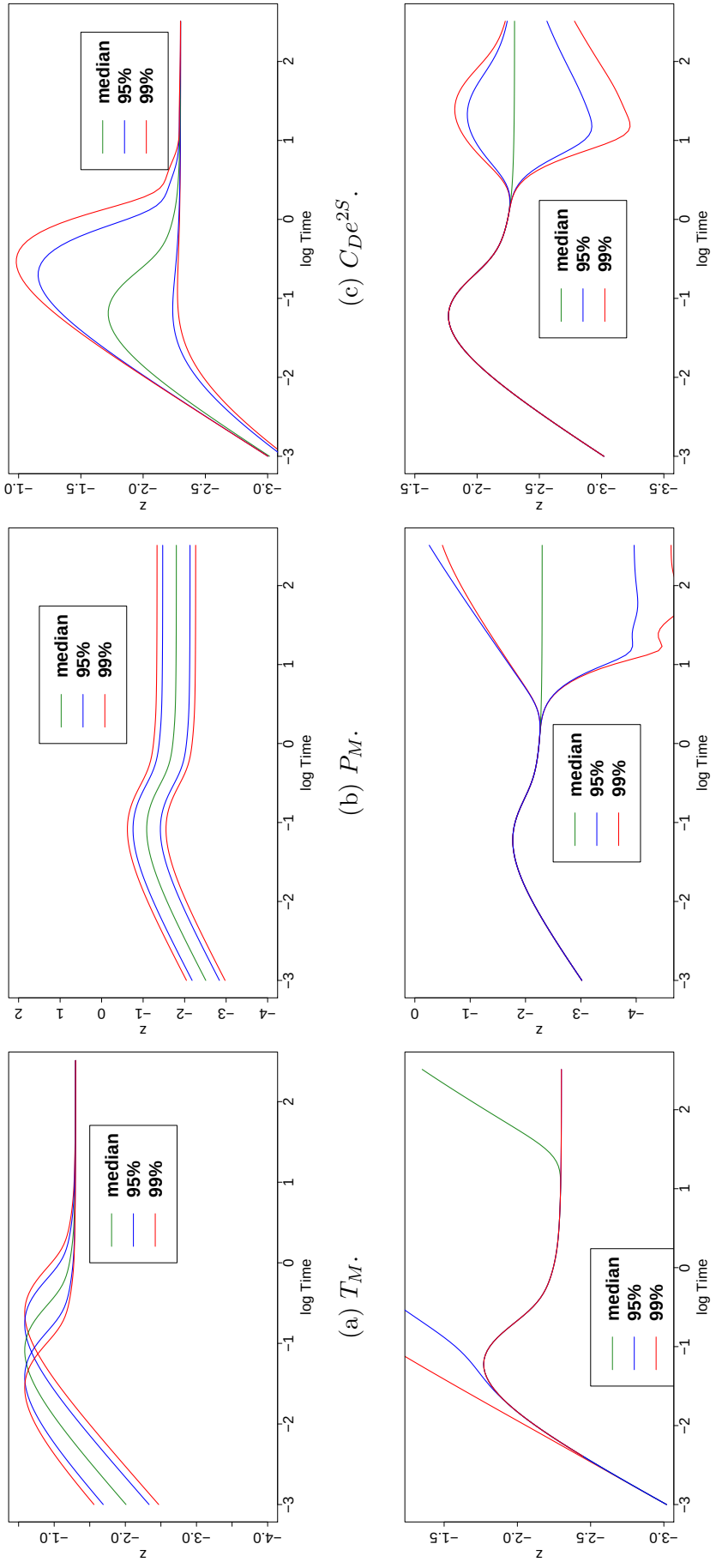
3.3.2 Prior for data parameters

Again, for simplicity we assume independence for the individual true rate values giving the prior. This assumption is based on the fact that, for long enough periods, the correlation between the rates is not very significant. Mathematically, that is:

$$P(\tilde{\boldsymbol{q}}) = \prod_{j=1}^{N} P(\tilde{q}_j)$$

In general, we choose a uniform distribution for the individual values, where, for a production well, the lower support is always non-negative to ensure the positivity of the rates, while for an injection well the upper support is always non-positive. Without this constraint, it may be possible for rates to switch sign, which would correspond to a well moving from production to injection, which would be an undesirable property, except from occasions where such a change would be known to the analyst and not a random event. Again, the rate behaviour is sensitive to whether we are dealing with an oil or a gas reservoir, which leads us to use different, almost non-informative priors for each case; we use $\tilde{\boldsymbol{q}} \sim \prod_{j=1}^{N} \mathrm{U}(0, 10000)$ for the former and $\tilde{\boldsymbol{q}} \sim \prod_{j=1}^{N} \mathrm{U}(0, 10)$ for the latter, unless there is explicit information about the test. A version of the above mentioned prior is used in Section 6.2.

Another form of uniform prior that can be used, is an improper uniform prior. Since



(a) T_M. (b) P_M. (c) $C_D e^{2S}$.

(d) R_D. (e) M. (f) η.

Figure 3.1: Response for the median, the 95% and the 99% credible intervals of the oil reservoir priors.

the $\tilde{\boldsymbol{q}}$ are constrained by the $\boldsymbol{q}$ measurements in the data model, the posterior will be proper. This prior is used in Chapter 6.

Another possibility is to adopt a Gaussian prior of the form: $\tilde{\boldsymbol{q}} \sim \prod_{j=1}^{N} \mathrm{N}(q_e, \boldsymbol{\Sigma}_{qe})$, where q_e is any fixed number and $\boldsymbol{\Sigma}_{qe}$ a covariance matrix with appropriate form. In this case (as well as in the improper prior case), this can lead to a conjugate analysis and an efficient factorisation of the posterior (see Subsection 3.4.1).

Finally, we follow a similar path for the true initial pressure $\tilde{p}_0$, using either a uniform distribution of the form $\tilde{p}_0 \sim \mathrm{U}(0, 10000)$ for oil reservoirs and $\tilde{p}_0 \sim \mathrm{U}(0, 5000)$ for gas reservoirs, an improper uniform prior, or a Gaussian distribution centred on the initial pressure measurement, in the same manner as the rates.

3.4 Posterior

In general, the derivation of a closed-form expression for the posterior distribution is not possible, regardless of the choice of prior, due to the complex form of the response and the nonlinear manner in which the $\boldsymbol{\phi}$ parameters are combined. Consequently, all the model parameters $\boldsymbol{\theta}$ need to be estimated indirectly using computational methods, such as Metropolis type MCMC algorithms.

There are, however, some ways in which we can exploit the properties of the Bayesian context.

The posterior density can be factorised into two components, where the first comprises the marginal distribution of $\boldsymbol{\phi}$ and the variance parameters, while the second is the distribution of the independent variables, namely $\tilde{\boldsymbol{q}}$ and $\tilde{p}_0$, conditioned on the rest of the parameters. Denoting the collection of observed well test data as $\boldsymbol{x} = (\boldsymbol{p}, \boldsymbol{q}, p_0)$ and the collection of all the variance parameters as $\boldsymbol{\sigma} = (\boldsymbol{\Sigma}_p, \boldsymbol{\Sigma}_q, \sigma^2_{p_0})$, we can write:

$$P(\boldsymbol{\phi}, \tilde{\boldsymbol{y}}, \boldsymbol{\sigma} \mid \boldsymbol{x}) = P(\boldsymbol{\phi}, \boldsymbol{\sigma} \mid \boldsymbol{x}) P(\tilde{\boldsymbol{y}} \mid \boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{x}).$$

There are two key advantages to this factorisation. First, it permits quick and easy

sampling from the Gaussian distribution of the true rates and initial pressure; second, the dimensionality of the space that requires MCMC sampling, no longer depends on the length of the time series of measurements. In practice, sampling only the $\boldsymbol{\phi}$ from the MCMC has a significant positive effect on the convergence, and therefore, it results to a significant reduction in the time needed to generate a given number of samples of $(\boldsymbol{\phi}, \boldsymbol{\sigma}, \tilde{\boldsymbol{y}})$; in the case that one is interested only in $\boldsymbol{\phi}$, the reduction is even more significant.

In addition, the essential and highly-sensitive variance hyperparameters $\boldsymbol{\sigma}$ in the Bayesian model could be considered nuisance parameters with non-informative priors and marginalised out of the problem (which would also change the prior distributions of those parameters from the ones mentioned in Section 3.2). Those parameters are indispensable for the construction and the functionality of the model, but not necessary for the analysis. Using marginalisation, the model still includes the effect and the uncertainty of the variance parameters, but without the need to make explicit inference. This marginalisation is described in Subsection 3.4.2, but, for reasons mentioned in the same Section, they are not used for the applications.

3.4.1 Conditional posterior for the true rates and initial pressure

Theorem 1. *The posterior distribution in Equation 3.1.2 can be factorised as:*

$$P(\boldsymbol{\phi}, \tilde{\boldsymbol{y}}, \boldsymbol{\sigma} \mid \boldsymbol{x}) = P(\boldsymbol{\phi}, \boldsymbol{\sigma} \mid \boldsymbol{x}) P(\tilde{\boldsymbol{y}} \mid \boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{x}).$$

where

$$P(\tilde{\boldsymbol{y}} \mid \boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{x}) \sim N(\boldsymbol{A}^{-1}\boldsymbol{b}, \boldsymbol{A}^{-1}) \,, \tag{3.4.1}$$

with $\boldsymbol{A}$ and $\boldsymbol{b}$ are given by

$$\boldsymbol{A} = \begin{bmatrix} m\sigma_p^{-2} + \sigma_{p_0}^{-2} & -1_m^T I_{\sigma_p^{-2}} \boldsymbol{C}(\boldsymbol{\phi}) \\ -\boldsymbol{C}(\boldsymbol{\phi})^T I_{\sigma_q^{-2}} 1_m & I_{\sigma_q^{-2}} + \boldsymbol{C}(\boldsymbol{\phi})^T I_{\sigma_p^{-2}} \boldsymbol{C}(\boldsymbol{\phi}) \end{bmatrix},$$

and

$$\boldsymbol{b} = \left(\sum_{i=1}^{m} \boldsymbol{p}\sigma_p^{-2} + p_0\sigma_{p_0}^{-2}, \boldsymbol{q}^T I_{\sigma_q^{-2}} - \boldsymbol{p}^T I_{\sigma_p^{-2}} \boldsymbol{C}(\boldsymbol{\phi}) \right)^T ,$$

where 1_a *is the* a*-dimensional vector with all elements equal to* 1*, and*

$$P(\boldsymbol{\phi}, \boldsymbol{\sigma} \mid \boldsymbol{x}) \propto |2\pi \boldsymbol{A}|^{\frac{1}{2}} \exp\Big\{ -m \ln(\sigma_p^2) - N \ln(\sigma_q^2) - \ln(\sigma_{p_0}^2) \\ - \frac{1}{2} \left(\boldsymbol{p}^T I_{\sigma_p^{-2}} \boldsymbol{p} + \boldsymbol{q}^T I_{\sigma_q^{-2}} \boldsymbol{q} + {p_0}^2 \sigma_{p_0}^{-2} \right) + \operatorname{tr}(\boldsymbol{b}^T \boldsymbol{A}^{-1} \boldsymbol{b}) \Big\} .$$

Proof. In what follows, I_a denotes the diagonal matrix with a in the main diagonal. The size of the matrix is dictated by a. Specifically, $a = \sigma_p$ corresponds to a $m \times m$ matrix, $a = \sigma_q$ to a $N \times N$ matrix and $a = \sigma_{p_0}$ is a 1×1 matrix (scalar). 1_a denotes the a-dimensional vector with all elements equal to 1. All vectors $\boldsymbol{v}$ are column vectors and $\boldsymbol{v}^T$ is the corresponding row vector. For generality, we assume that all the variance quantities are considered model parameters. The standard form for the likelihood $L = P(\boldsymbol{x} \mid \boldsymbol{\phi}, \tilde{\boldsymbol{y}}, \boldsymbol{\sigma})$ that takes into account errors for the rates and initial pressure including Equation 3.2.1-Equation 3.2.3 is

$$\begin{aligned} L = & \left(\frac{1}{2\pi\sigma_p^2} \right)^{\frac{m}{2}} \exp\left(-\frac{1}{2\sigma_p^2} \sum_{i=1}^{m} \{ \tilde{p}_0 - p_i - [\boldsymbol{C}(\boldsymbol{\phi}) \tilde{q}_i] \}^2 \right) \\ & \times \left(\frac{1}{2\pi\sigma_q^2} \right)^{\frac{N}{2}} \exp\left(-\frac{1}{2\sigma_q^2} \sum_{i=1}^{N} (q_j - \tilde{q}_j)^2 \right) \\ & \times \left(\frac{1}{2\pi\sigma_{p_0}^2} \right)^{\frac{1}{2}} \exp\left(-\frac{(p_0 - \tilde{p}_0)^2}{2\sigma_{p_0}^2} \right) \end{aligned}$$

The logarithm of the likelihood is

$$\log L = -\frac{m}{2} \log(2\pi) - m \log(\sigma_p) - \frac{1}{2} \{ \tilde{p}_0 - \boldsymbol{p} - [\boldsymbol{C}(\boldsymbol{\phi}) \tilde{\boldsymbol{q}}] \}^T I_{\sigma_p^{-2}} \{ \tilde{p}_0 - \boldsymbol{p} - [\boldsymbol{C}(\boldsymbol{\phi}) \tilde{\boldsymbol{q}}] \}$$

$$
\begin{aligned}
&- \frac{N}{2}\log(2\pi) - N\log(\sigma_q) - \frac{1}{2}(\boldsymbol{q}-\tilde{\boldsymbol{q}})^T I_{\sigma_q^{-2}}(\boldsymbol{q}-\tilde{\boldsymbol{q}}) \\
&- \frac{1}{2}\log(2\pi) - \log(\sigma_{p_0}) - \frac{1}{2}(p_0-\tilde{p}_0)^T I_{\sigma_{p_0}^{-2}}(p_0-\tilde{p}_0) \\
=&c - \frac{1}{2}\{\tilde{p}_0 - \boldsymbol{p} - [\boldsymbol{C}(\boldsymbol{\phi})\tilde{\boldsymbol{q}}]\}^T I_{\sigma_p^{-2}}\{\tilde{p}_0 - \boldsymbol{p} - [\boldsymbol{C}(\boldsymbol{\phi})\tilde{\boldsymbol{q}}]\} \\
&- \frac{1}{2}(\boldsymbol{q}-\tilde{\boldsymbol{q}})^T I_{\sigma_q^{-2}}(\boldsymbol{q}-\tilde{\boldsymbol{q}}) - \frac{1}{2}(p_0-\tilde{p}_0)^T I_{\sigma_{p_0}^{-2}}(p_0-\tilde{p}_0),
\end{aligned}
$$

where c is a combination of constants and variance parameters:

$$
c = -\frac{m}{2}\log(2\pi) - m\log(\sigma_p) - \frac{N}{2}\log(2\pi) - N\log(\sigma_q) - \frac{1}{2}\log(2\pi) - \log(\sigma_{p_0}).
$$

Expanding the rest of the terms leads to

$$
\begin{aligned}
\log L = c - \frac{1}{2}\Bigg(&\tilde{p}_0 1_m^T I_{\sigma_p^{-2}} 1_m \tilde{p}_0 + \boldsymbol{p}^T I_{\sigma_p^{-2}}\boldsymbol{p} + \tilde{\boldsymbol{q}}^T \boldsymbol{C}(\boldsymbol{\phi})^T I_{\sigma_p^{-2}} \boldsymbol{C}(\boldsymbol{\phi})\tilde{\boldsymbol{q}} \\
&- 2\tilde{p}_0 1_m^T I_{\sigma_p^{-2}}\boldsymbol{p} - 2\tilde{p}_0 1_m^T I_{\sigma_p^{-2}}\boldsymbol{C}(\boldsymbol{\phi})\tilde{\boldsymbol{q}} + 2\boldsymbol{p}^T I_{\sigma_p^{-2}}\boldsymbol{C}(\boldsymbol{\phi})\tilde{\boldsymbol{q}} \\
&+ \boldsymbol{q}^T I_{\sigma_q^{-2}}\boldsymbol{q} + \tilde{\boldsymbol{q}}^T I_{\sigma_q^{-2}}\tilde{\boldsymbol{q}} - 2\boldsymbol{q}^T I_{\sigma_q^{-2}}\tilde{\boldsymbol{q}} \\
&+ p_0 I_{\sigma_{p_0}^{-2}} p_0 + \tilde{p}_0 I_{\sigma_{p_0}^{-2}}\tilde{p}_0 - 2\tilde{p}_0 I_{\sigma_{p_0}^{-2}} p_0\Bigg).
\end{aligned}
$$

Eventually,

$$
\log L = c - \frac{1}{2}\left(\tilde{\boldsymbol{y}}^T A \tilde{\boldsymbol{y}} - 2b^T\tilde{\boldsymbol{y}} + \boldsymbol{p}^T I_{\sigma_p^{-2}}\boldsymbol{p} + \boldsymbol{q}^T I_{\sigma_q^{-2}}\boldsymbol{q} + \frac{p_0^2}{\sigma_{p_0}^2}\right), \tag{3.4.2}
$$

where

$$
\begin{aligned}
A &= \begin{bmatrix} 1_m^T I_{\sigma_p^{-2}} 1_m + I_{\sigma_{p_0}^{-2}} & -1_m^T I_{\sigma_p^{-2}}\boldsymbol{C}(\boldsymbol{\phi}) \\ -\boldsymbol{C}(\boldsymbol{\phi})^T I_{\sigma_p^{-2}} 1_m & I_{\sigma_q^{-2}} + \boldsymbol{C}(\boldsymbol{\phi})^T I_{\sigma_p^{-2}}\boldsymbol{C}(\boldsymbol{\phi}) \end{bmatrix} \\
&= \begin{bmatrix} \frac{m}{\sigma_p^2} + \sigma_{p_0}^{-2} & -1_m^T I_{\sigma_p^{-2}}\boldsymbol{C}(\boldsymbol{\phi}) \\ -\boldsymbol{C}(\boldsymbol{\phi})^T I_{\sigma_p^{-2}} 1_m & I_{\sigma_q^{-2}} + \boldsymbol{C}(\boldsymbol{\phi})^T I_{\sigma_p^{-2}}\boldsymbol{C}(\boldsymbol{\phi}) \end{bmatrix}
\end{aligned}
$$

is a $(N+1)\times(N+1)$ matrix, which depends on $\boldsymbol{\phi}$, $\boldsymbol{\sigma}$ and m. The $(N+1)$-dimensional column vector b is then given by

$$b^T = \left(\boldsymbol{p}^T I_{\sigma_p^{-2}} 1_m + p_0 I_{\sigma_{p0}^{-2}}, \quad \boldsymbol{q}^T I_{\sigma_q^{-2}} - \boldsymbol{p}^T I_{\sigma_p^{-2}} \boldsymbol{C}(\boldsymbol{\phi})\right)^T$$
$$= \left(\frac{\sum_{i=1}^m p_i}{\sigma_p^2} + \frac{p_0}{\sigma_{p0}^2}, \quad \boldsymbol{q}^T I_{\sigma_q^{-2}} - \boldsymbol{p}^T I_{\sigma_p^{-2}} \boldsymbol{C}(\boldsymbol{\phi})\right)^T.$$

By 'completing the square' in Equation 3.4.2

$$\log L = c - \frac{1}{2}\Bigg(\tilde{\boldsymbol{y}}^T \boldsymbol{A} \tilde{\boldsymbol{y}} - 2\boldsymbol{b}^T \tilde{\boldsymbol{y}} + \boldsymbol{b}^T \boldsymbol{A}^{-1} \boldsymbol{b} \\ - \boldsymbol{b}^T \boldsymbol{A}^{-1} \boldsymbol{b} + \boldsymbol{p}^T I_{\sigma_p^{-2}} \boldsymbol{p} + \boldsymbol{q}^T I_{\sigma_q^{-2}} \boldsymbol{q} + \frac{{p_0}^2}{\sigma_{p0}^2}\Bigg). \quad (3.4.3)$$

Considering $\boldsymbol{\Sigma} = \boldsymbol{A}^{-1}$ and $\boldsymbol{\mu_c} = \boldsymbol{A}^{-1}\boldsymbol{b}$,

$$\tilde{\boldsymbol{y}}^T \boldsymbol{A} \tilde{\boldsymbol{y}} - 2\boldsymbol{b}^T \tilde{\boldsymbol{y}} + \boldsymbol{b}^T \boldsymbol{A}^{-1} b = (\tilde{\boldsymbol{y}} - \boldsymbol{\mu_c})^T \boldsymbol{\Sigma}^{-1} (\tilde{\boldsymbol{y}} - \boldsymbol{\mu_c}).$$

Equation 3.4.3 becomes:

$$\begin{aligned}\log L =& c - \frac{1}{2}\Bigg[(\tilde{\boldsymbol{y}} - \boldsymbol{\mu_c})^T \boldsymbol{\Sigma}^{-1} (\tilde{\boldsymbol{y}} - \boldsymbol{\mu_c}) \\ & - {\boldsymbol{\mu_c}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu_c} + \boldsymbol{p}^T I_{\sigma_p^{-2}} \boldsymbol{p} + \boldsymbol{q}^T I_{\sigma_q^{-2}} \boldsymbol{q} + \frac{{p_0}^2}{\sigma_{p0}^2}\Bigg] \\ =& c - \frac{1}{2}(\tilde{\boldsymbol{y}} - \boldsymbol{\mu_c})^T \boldsymbol{\Sigma}^{-1} (\tilde{\boldsymbol{y}} - \boldsymbol{\mu_c}) - \frac{1}{2}\log|2\pi\boldsymbol{\Sigma}| \\ & - \frac{1}{2}\left(-{\boldsymbol{\mu_c}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu_c} + \boldsymbol{p}^T I_{\sigma_p^{-2}} \boldsymbol{p} + \boldsymbol{q}^T I_{\sigma_q^{-2}} \boldsymbol{q} + \frac{{p_0}^2}{\sigma_{p0}^2}\right) + \frac{1}{2}\log|2\pi\boldsymbol{\Sigma}|, \quad (3.4.4)\end{aligned}$$

which indicates a Gaussian conditional probability for $\tilde{\boldsymbol{y}}$ (given a conjugate or improper prior):

$$P(\tilde{\boldsymbol{y}}|\boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\sigma}) \sim N(\boldsymbol{\mu_c}, \boldsymbol{\Sigma}).$$

Finally, to derive the marginal posterior distribution of $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}$, we include the effect of the prior, and we integrate Equation 3.4.4 with respect to $\tilde{\boldsymbol{y}}$:

$$\begin{aligned}
\log P(\boldsymbol{\phi}, \boldsymbol{\sigma}|\boldsymbol{x}) &= \log \int P(\boldsymbol{\phi}, \tilde{\boldsymbol{y}}, \boldsymbol{\sigma}|\boldsymbol{x}) d\tilde{\boldsymbol{y}} \\
&= \log \int LP(\tilde{\boldsymbol{y}}) P(\boldsymbol{\phi}) P(\boldsymbol{\sigma}) d\tilde{\boldsymbol{y}} + c' \\
&= c'' - \frac{1}{2}\left(-{\boldsymbol{\mu_c}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu_c} + \boldsymbol{p}^T I_{\sigma_p^{-2}} \boldsymbol{p} + \boldsymbol{q}^T I_{\sigma_q^{-2}} \boldsymbol{q} + \frac{{p_0}^2}{\sigma_{p_0}^2}\right) \\
&+ \frac{1}{2}\log|2\pi\boldsymbol{\Sigma}| + \log P(\boldsymbol{\phi}) + \log P(\boldsymbol{\sigma}) \\
&= c'' - \frac{1}{2}\left(\boldsymbol{p}^T I_{\sigma_p^{-2}} \boldsymbol{p} + \boldsymbol{q}^T I_{\sigma_q^{-2}} \boldsymbol{q} + \frac{{p_0}^2}{\sigma_{p_0}^2}\right) \\
&+ \mathrm{tr}({\boldsymbol{\mu_c}}^T \boldsymbol{\mu_c} \boldsymbol{\Sigma}^{-1}) + \frac{1}{2}\log|2\pi\boldsymbol{\Sigma}| + \log P(\boldsymbol{\phi}) + \log P(\boldsymbol{\sigma}),
\end{aligned}$$

where c' comes from the normalising constant of the Bayes theorem, and $c'' = c + c'$. By exponentiating, we derive:

$$P(\boldsymbol{\phi}, \boldsymbol{\sigma}|\boldsymbol{x}) \propto e^{c''} e^{-\frac{1}{2}\left(\boldsymbol{p}^T I_{\sigma_p^{-2}} \boldsymbol{p} + \boldsymbol{q}^T I_{\sigma_q^{-2}} \boldsymbol{q} + \frac{p_0^2}{\sigma_{p_0}^2}\right)} e^{\mathrm{tr}({\boldsymbol{\mu_c}}^T \boldsymbol{\mu_c} \boldsymbol{\Sigma}^{-1})} |2\pi\boldsymbol{\Sigma}|^{\frac{1}{2}} P(\boldsymbol{\phi}) P(\boldsymbol{\sigma}).$$

A different way to derive the same result is to use the fact that $P(\boldsymbol{x}, \tilde{\boldsymbol{y}}|\boldsymbol{\phi}, \boldsymbol{\sigma})$ follows a multivariate normal distribution and eventually derive the main results as the conditional $P(\tilde{\boldsymbol{y}}|\boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\sigma})$ and the marginal $P(\boldsymbol{x})$, by using standard distributional results: $P(\boldsymbol{x}, \tilde{\boldsymbol{y}}|\boldsymbol{\phi}, \boldsymbol{\sigma}) P(\boldsymbol{\phi}, \boldsymbol{\sigma}) = P(\boldsymbol{x}|\tilde{\boldsymbol{y}}, \boldsymbol{\phi}, \boldsymbol{\sigma}) P(\tilde{\boldsymbol{y}}|\boldsymbol{\phi}, \boldsymbol{\sigma}) P(\boldsymbol{\phi}, \boldsymbol{\sigma}) = P(\boldsymbol{x}) P(\tilde{\boldsymbol{y}}|\boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\sigma})$

□

The significance of this factorisation is that sampling from the marginal posterior density $P(\boldsymbol{\phi}, \boldsymbol{\sigma}|\boldsymbol{x})$ requires MCMC sampling for only the response and variance parameters, which for an n-region radial composite model are $3n+3$ in total, while if the remaining $N+1$ rate and initial pressure parameters are also required, they can be sampled more efficiently from the conditional Gaussian distribution $P(\tilde{\boldsymbol{y}}|\boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{x})$ (as long as $P(\tilde{\boldsymbol{y}})$ is a conjugate Gaussian prior, or its effect has been marginalised).

3.4.2 Marginalising the variances

As mentioned in Section 3.2, there is, generally, expert judgement for the values of σ_q and σ_{p_0}. In addition, using this procedure for σ_p prohibits us from using the marginalisation of the rates and initial pressure from Subsection 3.4.1, which is usually more beneficial. For these reasons, we will not show results using the following marginalisation, but we still present this method here as potentially useful in similar problems, where marginalising the variance is a priority.

Theorem 2. *Given inverse Gamma priors for* $\boldsymbol{\sigma}$*:*

$$P(\tilde{\boldsymbol{q}}|\boldsymbol{q}) \propto P(\tilde{\boldsymbol{q}}) t_{2a_q}\left(\tilde{\boldsymbol{q}}, \frac{b_q}{a_q} I_N\right)$$
$$P(\boldsymbol{\phi}|\boldsymbol{p}, \tilde{\boldsymbol{q}}) \propto P(\boldsymbol{\phi}) P(\tilde{\boldsymbol{q}}) t_{2a_p}\left(\boldsymbol{C}(\boldsymbol{\phi})\tilde{\boldsymbol{q}}, \frac{b_p}{a_p} I_m\right)$$

and

$$P(\tilde{p}_0|p_0) \propto P(p_0) t_{2a_{p_0}}\left(p_0, \frac{b_{p_0}}{a_{p_0}}\right),$$

where $t_v(\cdot,\cdot)$ *is the multivariate t distribution with* v *degrees of freedom, and* a_i *and* b_i *are the hyperparameters of the inverse Gamma prior for* σ_i*.*

Proof. In order to marginalise the variance parameters of the model, we assign inverse Gamma priors, which is the conjugate prior for the variance of a Gaussian distribution. In what follows, we are going to show the marginalisation for the rate variance σ_q^2. The same extends to the cases of the pressure σ_p^2 and initial pressure $\sigma_{p_0}^2$ variance.

We begin by defining the two quantities in question; the data model for the true rates $\tilde{\boldsymbol{q}}$ is

$$P(\boldsymbol{q}|\sigma_q^2, \tilde{\boldsymbol{q}}) = \left(\frac{1}{2\pi\sigma_q^2}\right)^{\frac{N}{2}} e^{-\frac{1}{2\sigma_q^2}(\boldsymbol{q}-\tilde{\boldsymbol{q}})^T(\boldsymbol{q}-\tilde{\boldsymbol{q}})},$$

and the rate variance σ_q^2 prior is

$$P(\sigma_q^2) = IG(a_q, b_q) = \frac{b_q^{a_q}}{\Gamma(a_q)} \left(\frac{1}{\sigma_q^2}\right)^{a_q+1} e^{-\frac{b_q}{\sigma_q^2}}.$$

The goal is to derive the marginal probability of $\tilde{\boldsymbol{q}}$. For this, we integrate over σ_q^2

$$\begin{aligned} P(\tilde{\boldsymbol{q}}|\boldsymbol{q}) &= \int_0^\infty P(\tilde{\boldsymbol{q}}, \sigma_q^2|\boldsymbol{q}) d\sigma_q^2 \\ &\propto \int_0^\infty P(\boldsymbol{q}|\tilde{\boldsymbol{q}}, \sigma_q^2) P(\tilde{\boldsymbol{q}}) P(\sigma_q^2) d\sigma_q^2 \\ &= P(\tilde{\boldsymbol{q}}) \int_0^\infty P(\boldsymbol{q}|\tilde{\boldsymbol{q}}, \sigma_q^2) P(\sigma_q^2) d\sigma_q^2. \end{aligned} \tag{3.4.5}$$

First, we multiply the two quantities

$$\begin{aligned} P(\boldsymbol{q}|\tilde{\boldsymbol{q}}, \sigma_q^2) P(\sigma_q^2) &= \left(\frac{1}{2\pi\sigma_q^2}\right)^{\frac{N}{2}} e^{-\frac{1}{2\sigma_q^2}(\boldsymbol{q}-\tilde{\boldsymbol{q}})^T(\boldsymbol{q}-\tilde{\boldsymbol{q}})} \frac{b_q^{a_q}}{\Gamma(a_q)} \left(\frac{1}{\sigma_q^2}\right)^{a_q+1} e^{-\frac{b_q}{\sigma_q^2}} \\ &= \frac{b_q^{a_q}}{\Gamma(a_q)} \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} \left(\frac{1}{\sigma_q^2}\right)^{\frac{N}{2}+a_q+1} e^{-\frac{1}{\sigma_q^2}\left(b_q + \frac{(\boldsymbol{q}-\tilde{\boldsymbol{q}})^T(\boldsymbol{q}-\tilde{\boldsymbol{q}})}{2}\right)}. \end{aligned}$$

Then, by integrating over σ_q^2 we compute the integral

$$\begin{aligned} &\int_0^\infty P(\boldsymbol{q}|\tilde{\boldsymbol{q}}, \sigma_q^2) P(\sigma_q^2) d\sigma_q^2 \\ =&\int_0^\infty \frac{b_q^{a_q}}{\Gamma(a_q)} \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} \left(\frac{1}{\sigma_q^2}\right)^{\frac{N}{2}+a_q+1} e^{-\frac{1}{\sigma_q^2}(b_q + \frac{(\boldsymbol{q}-\tilde{\boldsymbol{q}})^T(\boldsymbol{q}-\tilde{\boldsymbol{q}})}{2})} d\sigma_q^2 \\ =&\frac{b_q^{a_q}}{\Gamma(a_q)} \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} \int_0^\infty \left(\frac{1}{\sigma_q^2}\right)^{\frac{N}{2}+a_q+1} e^{-\frac{1}{\sigma_q^2}(b_q + \frac{(\boldsymbol{q}-\tilde{\boldsymbol{q}})^T(\boldsymbol{q}-\tilde{\boldsymbol{q}})}{2})} d\sigma_q^2 \\ =&\frac{b_q^{a_q}}{\Gamma(a_q)} \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} \Gamma\left(a_q + \frac{N}{2}\right) \left[b_q + \frac{(\boldsymbol{q}-\tilde{\boldsymbol{q}})^T(\boldsymbol{q}-\tilde{\boldsymbol{q}})}{2}\right]^{-(\frac{N}{2}+a_q)} \\ =&\frac{b_q^{a_q}\Gamma(a_q + \frac{N}{2}) b_q^{-\frac{N}{2}-a_q}}{\Gamma(a_q) 2\pi^{\frac{N}{2}}} \left[1 + \frac{(\boldsymbol{q}-\tilde{\boldsymbol{q}})^T \frac{a_q}{b_q} I_N (\boldsymbol{q}-\tilde{\boldsymbol{q}})}{2a_q}\right]^{-\left(\frac{N+2a_q}{2}\right)}. \end{aligned}$$

The last line denotes the density of the multivariate t distribution $t_{2a_q}\left(\tilde{\boldsymbol{q}}, \frac{b_q}{a_q} I_N\right)$. From Equation 3.4.5:

$$P(\tilde{\boldsymbol{q}}|\boldsymbol{q}) \propto P(\tilde{\boldsymbol{q}}) t_{2a_q}\left(\tilde{\boldsymbol{q}}, \frac{b_q}{a_q} I_N\right). \tag{3.4.6}$$

Following the same steps, we derive:

$$P(\boldsymbol{\phi}|\boldsymbol{p},\tilde{\boldsymbol{q}}) \propto P(\boldsymbol{\phi})P(\tilde{\boldsymbol{q}})t_{2a_p}\left(\boldsymbol{C}(\boldsymbol{\phi})\tilde{\boldsymbol{q}}, \frac{b_p}{a_p}I_m\right)$$

and

$$P(\tilde{p}_0|p_0) \propto P(p_0)t_{2a_{p_0}}\left(p_0, \frac{b_{p_0}}{a_{p_0}}\right).$$

Eventually:

$$\begin{aligned}
P(\boldsymbol{\phi},\tilde{\boldsymbol{y}}|\boldsymbol{x}) =& \int_0^\infty \int_0^\infty \int_0^\infty P(\boldsymbol{\phi},\tilde{\boldsymbol{y}},\boldsymbol{\sigma}|\boldsymbol{x})d\sigma_p^2 d\sigma_q^2 d\sigma_{p_0}^2 \\
\propto& \int_0^\infty \int_0^\infty \int_0^\infty P(\boldsymbol{x}|\boldsymbol{\phi},\tilde{\boldsymbol{y}},\boldsymbol{\sigma})P(\boldsymbol{\phi},\tilde{\boldsymbol{y}},\boldsymbol{\sigma})d\sigma_p^2 d\sigma_q^2 d\sigma_{p_0}^2 \\
=& P(\boldsymbol{\phi})P(\tilde{\boldsymbol{q}})P(\tilde{p}_0)\int_0^\infty P(\boldsymbol{p}|\tilde{\boldsymbol{q}},\boldsymbol{\phi},\sigma_p^2)P(\sigma_p^2)d\sigma_p^2 \\
& \int_0^\infty P(\boldsymbol{q}|\sigma_q^2,\tilde{\boldsymbol{q}})P(\sigma_q^2)d\sigma_q^2 \int_0^\infty P(p_0|\sigma_{p_0}^2,\tilde{p}_0)P(\sigma_{p_0}^2)d\sigma_{p_0}^2 \\
=& P(\boldsymbol{\phi})P(\tilde{\boldsymbol{q}})P(\tilde{p}_0)P(\boldsymbol{p}|\boldsymbol{\phi},\tilde{\boldsymbol{q}})P(\boldsymbol{q}|\tilde{\boldsymbol{q}})P(p_0|\tilde{p}_0) \\
=& P(\boldsymbol{\phi},\tilde{\boldsymbol{y}})P(\boldsymbol{x}|\boldsymbol{\phi},\tilde{\boldsymbol{y}}).
\end{aligned}$$

□

The final distribution fully incorporates the uncertainty of the variance parameters $\boldsymbol{\sigma}$ in the model, even though they are not explicitly included in the formula. It should be mentioned that in practice, the above theorem reflects standard results for a normal-inverse-gamma setup.

3.5 Summary

In the chapter we presented a Bayesian model for the well test deconvolution problem. We included errors for the rates and initial pressure measurements and we introduced prior distributions, both for the parameters of the radial composite model, and for

the data parameters. Finally, we took advantage of the Bayesian context in order to simplify the posterior distribution in various ways and derived improved forms.

In the next chapter we are going to demonstrate a variety of algorithms that can be used to summarise the posterior distribution, along with their advantages and limitations.

Chapter 4

Computational methods

4.1 Introduction

When attempting to apply the statistical model of Chapter 3, we encounter two problems that require our attention: first, there is the issue of approximating the posterior distribution, which, depending on the complexity of its form, can be a challenging task. Second, we sometimes require a point estimate that indicates the 'best fit' of the model given a specific data set. That can be particularly difficult since the model possesses a complex covariance structure. In order to solve those issues, we need to resort to computational methods.

Markov Chain Monte Carlo (MCMC) methods have emerged in the last few decades as the most popular algorithmic class for extracting information from models based on a Bayesian approach. MCMCs aim to recover the posterior density by approximating it, using a set of rules. There is a rich and vast literature regarding those methods [18], as well as associated convergence and efficiency diagnostics. Alternatively, there are other methods, such as the Laplace approximation, that seek to provide a simple framework in order to summarise the posterior.

Regarding parameter estimation, the issue of model fitting is usually tackled by popular and diverse optimisation algorithms. In the context of Bayesian statistics,

these algorithms can be used as an analogue to maximum likelihood estimation methods, which in the Bayesian context seek to find the maximum a posteriori probability estimate (MAP). This estimate can either be used for inference directly, or as a starting point, that initialises the algorithms that aim to summarise the posterior as a whole. Alternatively, an MCMC can also be used in place of optimisation.

In this chapter, we give a brief overview of these algorithmic classes, focusing on MCMC methods, which are the algorithms mainly used in the later chapters. It is worth noting that because of some characteristics of the posterior distribution such as multimodality (e.g. in Figure 5.14), high correlation (e.g. in Figure 6.7) and nonlinear features (e.g. in Figure 6.4), which are going to be evident in the results of Chapter 5 and Chapter 6 the choice of an appropriate MCMC algorithm that can deal with those issues is very important. The algorithms, diagnostics and criteria described in this section were implemented in open source R packages [71, 40].

4.2 Markov Chain Monte Carlo theory

In many cases, the normalising constant of the Bayes theorem (the denominator in Equation 3.1.1) cannot be explicitly computed, and thus, the derivation of a closed-form expression for the posterior distribution is not possible. Under these circumstances, the only way to fully exploit the benefits of the Bayesian context and summarise the model uncertainty, is to apply numerical methods, typically MCMC algorithms [13, 25]. The main advantage of these algorithms is that they only require a function proportional, and not necessarily equal, to the density of interest, and therefore the normalising constant can be omitted from the computations.

An MCMC algorithm comprises two parts: The *Monte Carlo* sampling method and the *Markov Chain* probabilistic model.

Monte Carlo techniques are used for numerical integration. Given a probability distribution, $\pi(x)$, whose density is known only up to a constant, and a function $f(x)$, the expectation $E_\pi(f(x))$ is given by:

$$E_\pi(f(x)) = \int_{-\infty}^{\infty} f(x)\pi(X)dx,$$

and given n independent random samples $X = \{X_1, X_2, \ldots, X_n\}$ from $\pi(x)$, it can be approximated by the sum

$$E_\pi(f(x)) = \frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

By law of large numbers, as n grows larger, the approximation becomes more accurate.

The second part of an MCMC algorithm is the Markov Chain, a probabilistic model based on the Markov property, which states that the next sample of a sequence of random variables X depends only on the current state. Mathematically, that is:

$$P(X_{t+1}|X_1, X_2, \ldots, X_t) = P(X_{t+1}|X_t),$$

where $P(X_{t+1}|X_t)$ is called the *transition probability*. Given a transition from state i to state j, and based on the stationarity property of the Markov chains, according to which this quantity is independent of t, the corresponding transition probability is denoted as:

$$P_{ij} = P(X_{t+1} = j|X_t = i).$$

Finally, the Markov Chain model possesses two important properties: the first is called *ergodicity* [30], and states that there is a unique stationary distribution $\pi(\cdot)$, such that:

$$P_{ij}(t) \to \pi(j), \quad \text{as} \quad t \to \infty.$$

The second is called *reversibility* and states that:

$$\pi(i)P(X_{t+1} = j|X_t = i) = \pi(j)P(X_{t+1} = i|X_t = j).$$

Combining the two properties, we can construct a Markov Chain with stationary dis-

tribution equal to the posterior, following acceptance rules based on the reversibility property. Then, we can approximate this distribution using Monte Carlo sampling. Note that by construction, these samples are correlated.

4.2.1 Metropolis-Hastings algorithm

One of the first MCMC algorithms introduced in literature was the Metropolis-Hastings [54], which was later reviewed and popularised in [14]. It is a very general and simple algorithm that is often used as a building block for more complex and powerful MCMCs.

The algorithm seeks to create a Markov Chain with stationary distribution equal to the function of interest $\pi(\cdot)$. It starts from an *initial vector* of parameters, otherwise known as a *starting point*, X_0, and for the current state X_t, a *candidate* Y is sampled from a *proposal distribution* $q(\cdot|X_t)$. Note that the candidate depends only on the current state of the chain.

The next step is to choose whether the candidate is going to be accepted or rejected. That is based on the *acceptance probability* $a(X_t, Y)$, for which:

$$a(X_t, Y) = \min\left\{1, \frac{\pi(Y)q(X_t|Y)}{\pi(X_t)q(Y|X_t)}\right\}. \tag{4.2.1}$$

The derivation of a is based on the reversibility property, and the fact that the transition probability can be written as the product of the proposal distribution and the acceptance probability: $P(X_{t+1} = Y|X_t) = q(Y|X_t)a(X_t, Y)$.

After this calculation, a random number u is generated from the uniform distribution U$[0, 1]$. If $u \leq a$, the candidate is accepted and we set $X_{t+1} = Y$. If $u > a$, the candidate is rejected and the chain remains in the same state $X_{t+1} = X_t$.

Note that, for the acceptance ratio in Equation 4.2.1, if $\pi(\cdot)$ is a posterior density in a Bayesian context, the normalising constants cancel each other. Additionally, if

the proposal $q(\cdot|X_t)$ is symmetric, Equation 4.2.1 simplifies to:

$$a(X_t, Y) = \min\left\{1, \frac{\pi(Y)}{\pi(X_t)}\right\}. \tag{4.2.2}$$

Finally, the most popular and widely used variation of a symmetric, Metropolis-Hastings in a vector space, comprises a proposal distribution of the form

$$q(Y|X_t) = \mathrm{N}(X_t, C_0) \tag{4.2.3}$$

and is called *Random Walk Metropolis (RWM)*. In practice, that is a multivariate Gaussian distribution with mean equal to the current state vector and fixed covariance C_0. The popularity of the RWM comes from the ease of its implementation and sampling. A two-dimensional example can be seen in Figure 4.1.

Computational challenges

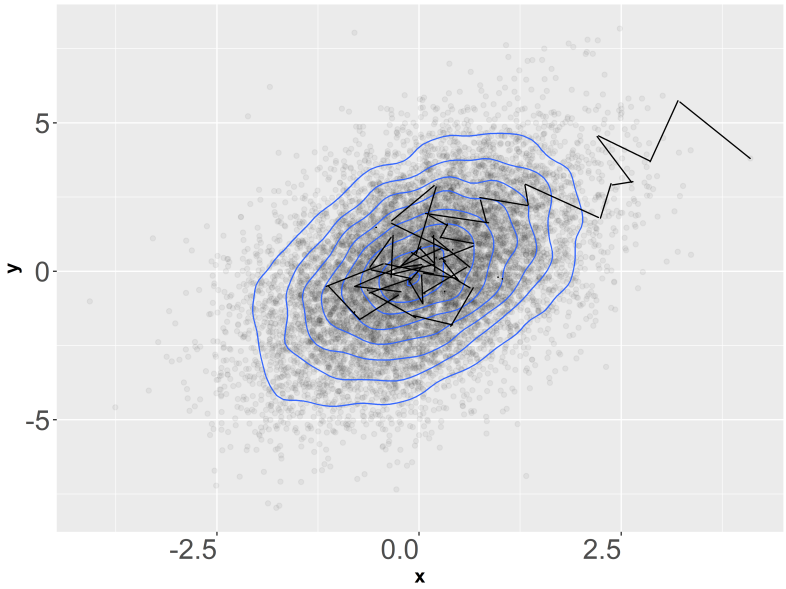


Figure 4.1: Two-dimensional random walk Metropolis, starting from a random point.

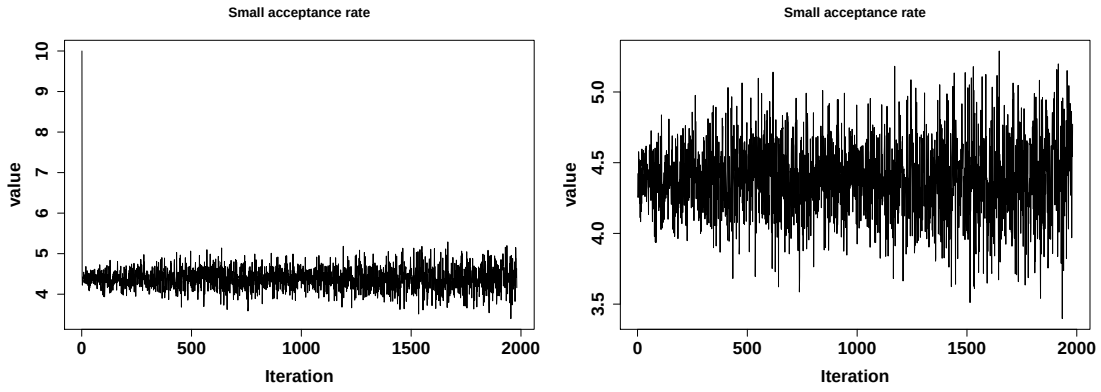
In practice, there are two distinct phases in a Metropolis-Hastings algorithm. This distinction generalises to other MCMCs:

- In the first phase, the algorithm starts making transitions from the initial vector. Since its choice is usually arbitrary, that point can be far from the stationary

distribution, and therefore the first iterations aim to find the distribution in question, by scanning the whole parameter space.

- After the samples move towards a region of high probability, the algorithm seeks to recover the stationary distribution's shape. It achieves that by accepting more transitions to candidates of high density, and less of low density.

A common practice when dealing with MCMCs is to dispose those early iterations of the first phase, in order to diminish the influence of the starting value. Discarding the '**burn-in**' phase seeks to remove the effect of the starting point and, instead, focus only on the converged iterations. In that way, the remaining chain is independent of the initial vector and the posterior distribution can be properly recovered. In Figure 4.2a, we show the trace of a RWM marginal distribution starting from a random initial point. In Figure 4.2b it is the same trace, after we discard the burn-in.



(a) Trace of a marginal density before discarding the burn-in.

(b) Trace of a marginal density after discarding the burn-in.

Figure 4.2: The effect of burn-in on a trace of a marginal density.

An alternative way of solving the aforementioned issue is to run a **two-phase simulation**, using two separate MCMCs. The first algorithm takes the place of optimisation, where the main goal is to reach the MAP. The second uses the MAP approximation of the first algorithm as a starting point, thus eliminating the values of low probability.

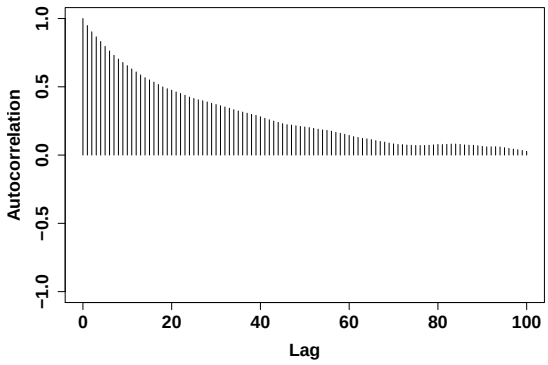
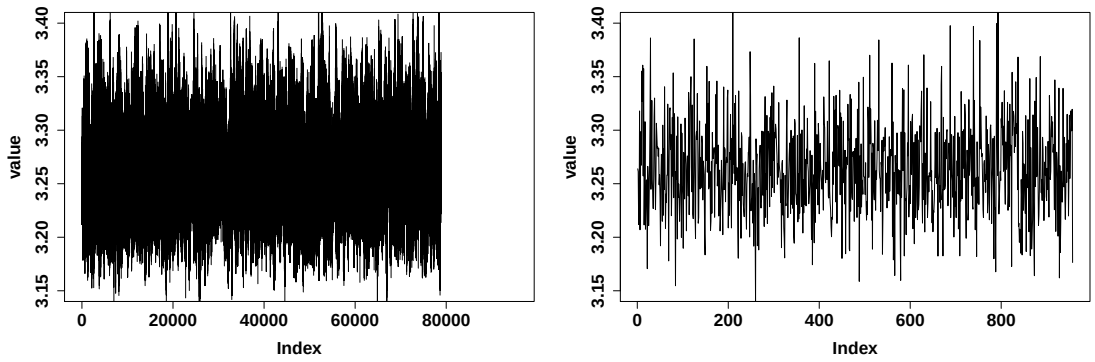


Figure 4.3: Autocorrelation plot.

A different issue comes from the fact that, as mentioned before, by construction the iterations of the MCMC are not independent. It is very common in applications to see the effect of the dependence of previous iterations in each sequence, even after convergence has been achieved, which can potentially cause computational storage and efficiency issues. A diagnostic tool for the dependence of the iterations is the autocorrelation, which refers to the level of similarity between the chain and a lagged version of the chain over a specific number of steps. In the corresponding plot in Figure 4.3, each line shows the dependence of one iteration to the previous 1 to 100 iterations. A common way of dealing with this issue, is **thinning** the sequence, by keeping only every other i-th iteration and discarding the rest. In that way, we preserve fewer iterations that still include the same amount of information for the density in question as the whole chain. In Figure 4.4 we show the effect of thinning on the trace of a RWM marginal distribution.

Two other issues associated with the RWM that need attention, include the number of candidates that get accepted, and the structure of the covariance of the random walk C_0. If the proposal variance is larger than the variance of the distribution in question, the algorithm can easily propose states that lie outside the stationary distribution. As a result, most candidates are going to be rejected, which causes the flat sections of the trace shown in Figure 4.5a, and the chain will take a long time to collect enough samples that fully construct the density. In contrast, small covariance

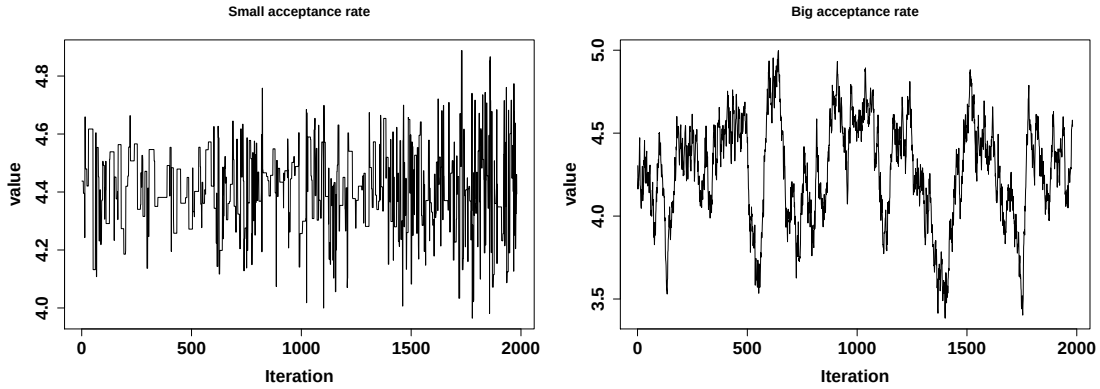


(a) Trace of a marginal density before thinning.

(b) Trace of a marginal density after thinning.

Figure 4.4: The effect of thinning on a trace of a marginal density.

corresponds to candidates close to the current state, which causes the algorithm to accept many similar candidates. Because of that, the chain is going to take a lot of time to converge as in Figure 4.5b.



(a) Trace of a marginal density with a small acceptance rate.

(b) Trace of a marginal density with a large acceptance rate.

Figure 4.5: The effect of the acceptance rate on a trace of a marginal density.

It has been determined that the maximum efficiency acceptance rate for a univariate distribution is approximately 44%, while for a general multivariate case with d parameters it is 23.4% [63]. For a RWM, this requires tuning the C_0 appropriately before running the algorithm, which is not always easy.

In order to overcome this issue, we can use **adaptive algorithms**. An adaptive MCMC [39] is an algorithm that is modified while it is running ('on the fly'). Those

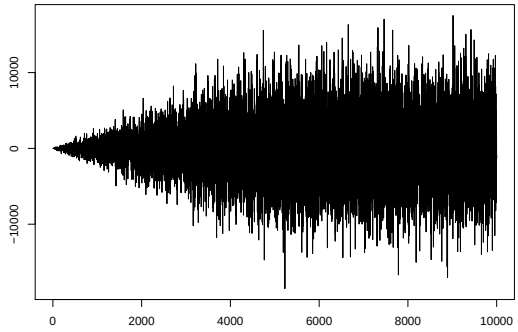


Figure 4.6: Trace of an adaptive MCMC. The Figure is inspired by a similar Figure in [64].

algorithms acquire information about the covariance structure by using updating rules that depend on previous steps. The trace of the marginal distribution of an adaptive MCMC is shown in Figure 4.6, where the variance of the proposals starts with a small value and it increases until it convergences after approximately 5,000 iterations.

Lastly, another challenge that we may face when dealing with MCMCs, is the problem of *multimodality*, which appears when the target distribution has more than one mode (for example in the case of a mixture model as in Figure 4.7). Algorithms such as the RWM, based on a strict covariance structure, are unable to provide the necessary flexibility that would allow us to simultaneously detect multiple modes and capture the shape of the target distribution. Solving this problem usually requires more complex algorithms, which we are going to introduce later in this chapter.

4.2.2 The Gibbs sampler

The Gibbs sampler [27] is a special case of the Metropolis-Hastings, where we accept all the candidates. Suppose we want to sample the parameter i of a d-dimensional vector $X_{t+1} = (X_{t+1}^{(1)}, X_{t+1}^{(2)}, \ldots, X_{t+1}^{(i)}, \ldots, X_{t+1}^{(d)})$. In order to produce one sample from the Gibbs algorithm, we need to sample each parameter of the vector individually at time $t+1$ from its corresponding full conditional distribution. For the i-th parameter,

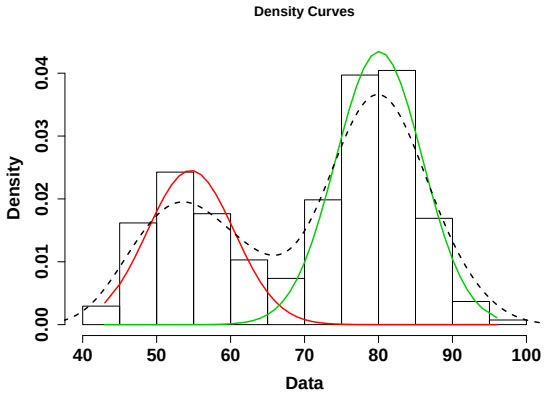


Figure 4.7: Multimodal density (dashed curve) constructed by two Gaussian densities (red and green curves) and approximated by a histogram.

we can achieve this by conditioning the parameters from 1 to $i-1$ on their X_{t+1} state, and the parameters from $i+1$ to d on their X_t state. In mathematical form, that is:

$$\pi(X_{t+1}^{(i)}|X_{t+1}^{(1)}, X_{t+1}^{(2)}, \ldots, X_{t+1}^{(i-1)}, X_t^{(i+1)}, \ldots, X_t^{(d)}).$$

We carry on with this approach until all d parameters are sampled, and we construct the X_{t+1}. A two dimensional example of four samples from Gibbs is shown in Figure 4.8. The $x^{(1)}$ coordinate is sampled first, and the $x^{(2)}$ second. Together they construct the next point in the chain. This sampling scheme is called *componentwise sampling*.

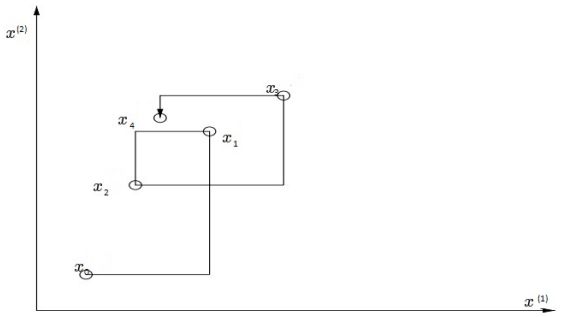


Figure 4.8: Two-dimensional Gibbs sampler; each vertex with a circle is a sample. When the chain is in equilibrium, each vertex (circled or not) is a sample from the target distribution.

The advantage of the Gibbs sampler over the RWM is that it generally converges more

quickly, since, by construction, it accepts all the candidates. The main disadvantages are the computational time required to produce the same number of samples, which is proportional to the number of parameters, the fact that since every parameter is updated individually the algorithm cannot capture the covariance structure of the target distribution, and finally, the fact that the conditional distributions are not always accessible.

In case the full conditionals are not available in closed form for one, some, or even all of the parameters, we can adopt a Metropolis step, where we use a Metropolis-type proposal, as described in the previous section, and we accept or rejected it using the Metropolis acceptance probability in Equation 4.2.1. This variation of the algorithm is called Metropolis-within-Gibbs [29].

Both the RWM and the Gibbs samplers are very easy to implement. Their main disadvantage is that, in order to use them, we need to specify the size of the proposals for all the parameters a priori. In the case of the RWM that requires knowledge about the covariance matrix C_0, in the case of the standard Gibbs, information about the hyperparameters that specify the full marginals, and in the case of the Metropolis-within-Gibbs, the variance for each of the individual univariate proposals.

In most cases, this information is not available, and thus, we need to use adaptive variations of those algorithms.

4.3 Alternative MCMC Algorithms

We group a selection of MCMC algorithms into three classes: the first comprises algorithms based on the standard Metropolis-Hastings MCMC and its adaptive variations. The second group contains the 'hit-and-run' algorithms, which sample new candidates by choosing a random directions in $\mathbb{R}^k$. Finally, the last group includes a set of algorithms derived from an evolutionary algorithm called differential evolution.

The performance of these algorithms on the multi-region radial composite model of Chapter 2 is presented in Subsection 6.3.2.

4.3.1 The Metropolis variations class

Adaptive Metropolis (AM) [39] tackles the fixed covariance limitation, by using the sample covariance of the whole chain to update the covariance of the proposal distribution. The covariance adaptation uses the assumption that the optimum acceptance rate for multivariate distributions is 23.4% [63]. The transition probability is given by

$$q(Y|X_t) = \mathrm{N}(X_t, C_t),$$

where C_t is fixed to C_0 for the burn-in period and

$$C_t = \frac{2.38^2}{d}\left(\mathrm{cov}(X_0, \ldots, X_t) + \epsilon I_d\right)$$

afterwards. The constant $\frac{2.38^2}{d}$ is associated with the acceptance rate [63], ϵ is a small constant that ensures ergodicity [39], and I_d is the d-dimensional identity matrix.

Using the covariance information, the AM algorithm converges much faster than the RWM for problems where there is not a-priori knowledge about the parameters' correlation structure.

Adaptive Mixture Metropolis (AMM) [64] is a variation of AM, where the proposal is a weighted mixture of an adaptive multivariate and a symmetric univariate proposal:

$$q(Y|X_t) = (1-\beta)\mathrm{N}(X_t, (2.38^2\,\mathrm{cov}(X_0, \ldots, X_t)/d) + \beta\mathrm{N}(X_t, 0.1^2 I_d/d),$$

where β is a small positive constant (usually $\beta = 0.05$) that ensures ergodicity [64].

The AMM and AM algorithms are very similar both in terms of their premise and results. Their main difference is that the AMM proposal density is a Gaussian mixture, and offers the additional flexibility of adjusting the parameter β in order

to favour either the adaptive component (lower values of β) or a more symmetric covariance (higher values of β).

Delayed Rejection Adaptive Metropolis (DRAM) [38] is an extension to AM, where whenever a proposal is rejected, the algorithm attempts one or more alternate proposals (usually with a scaled down covariance), which requires a correction for the conditional acceptance probability. The first step of this algorithm is identical to AM, while if the candidate was not accepted, a second candidate is proposed using a proposal distribution of the form $q_2(Y_2|Y_1, X_t) = \mathrm{N}(X_t, C_t')$, where C_t' is the product of C_t and a constant, usually 1/10. This smaller covariance produces candidates closer to the current state, which are more likely to be accepted. The second step of DRAM has an acceptance rate equal to

$$a_2(X_t, Y_1, Y_2) = \min\left\{1, \frac{\pi(Y_2)q_1(Y_2|Y_1)q_2(X_t|Y_1, Y_2)[1 - a_1(Y_2, Y_1)]}{\pi(X_t)q_1(Y_1|X_t)q_2(Y_2|Y_1, X_t)[1 - a_1(X_t, Y_1)]}\right\},$$

where a_1 is the acceptance rate of the first step, Y_1 and Y_2 are the first and second step candidates respectively, and q_1 and q_2 are the corresponding proposal distributions.

By accepting second step candidates, DRAM can update the covariance more quickly. Thus, it generally converges faster than AM and AMM. On the other hand, since it needs to propose new candidates and calculate a new acceptance ratio when the first candidate gets rejected, it takes longer to produce the same number of iterations.

Robust Adaptive Metropolis (RAM) [79] is another extension to AM that aims to achieve the acceptance rate from [63] by shrinking or expanding the proposal covariance. The candidate is produced as:

$$q(Y|X_t) = \mathrm{N}(X_t, S_t U_{t+1})$$

where U_{t+1} is an independent random vector from a standard normal distribution, and S_t is a lower-diagonal matrix with positive diagonal elements, that satisfies the equation:

$$S_{t+1}S_{t+1}^T = S_t\left(I_d + \eta_t(a - a^*)\frac{U_{t+1}U_{t+1}^T}{\|U_{t+1}\|^2}\right)S_t^T.$$

In the above equation, η_t is a step size sequence decaying to zero and a^* is the target acceptance ratio.

The RAM algorithm simultaneously uses a coerced acceptance rate and adapts to the shape of the target distribution. Therefore, it is more likely to detect multimodality than AM, AMM or DRAM. Its main downside is that, in practice, the adaptation can be very slow.

Adaptive Metropolis-within-Gibbs (AMWG) [64] is an adaptive algorithm that uses componentwise proposals and, therefore, each parameter is updated individually in each iteration. The proposal of each parameter changes in a way that optimises the associated acceptance rate, using the univariate version of the optimum acceptance ratio [65]. The proposal density is:

$$q(Y|X_t) = \mathrm{N}(X_t, \sigma^2),$$

where σ^2 is increased or decreased in each step so that the acceptance ratio becomes $a^* = 0.44$.

The AMWG algorithm possesses all the advantages of the Metropolis-within-Gibbs sampler (it convergences more quickly by accepting all the candidates, and it does not require the full conditionals), with the additional benefit of being adaptive. Thus, it is not restricted to the initial variance assigned to each parameter. It also carries some of the disadvantages, since it takes much time per iteration, and it cannot capture the target covariance structure.

4.3.2 The Hit-and-Run class

Hit-and-Run Metropolis (HRM) [70, 30] is an MCMC algorithm that constructs its candidates by initially sampling a direction randomly on the unit d-dimensional sphere, and then, choosing a uniformly-distributed distance r_t.

Initially, we generate d independent normal random variates z_i, and we set the

vector:

$$e_t = \frac{z_i}{\sqrt{\sum_{j=1}^{d} z_j^2}}.$$

Afterwards, we consider the proposal density

$$q(Y|X_t) = \mathrm{N}(X_t, diag(r_t e_t)),$$

where $r_n \sim \mathrm{U}(0,1)$ and $diag(\cdot)$ is a matrix with the vector $\cdot$ in its diagonal and zeros everywhere else.

HRM is a fast algorithm thanks to the simplicity of its candidate generation. In addition, the fact that it does not rely on the covariance, but samples the transition distance instead, it is more suitable for multimodal target densities.

Adaptive Hit-and-Run Metropolis (HARM) is a variation of the previous algorithm, that uses the Robbins–Monro process [23] for the transition probability, which modifies the algorithm and makes it adaptive. After it samples the direction on the surface of the unit sphere, it generates a distance along the direction of the constrained space, based on the optimum acceptance ratio. The proposal density becomes:

$$q(Y|X_t) = \mathrm{N}(X_t, diag(se_t)),$$

where s is a uniformly-distributed distance, for which $s \sim \mathrm{U}(0,\tau)$ applies. The value of τ depends on whether the previous candidate was successful. Specifically:

- For successful candidates:

$$\tau = \tau + (\tau/(a^*(1-a^*)))(1-a^*)/n$$

- For unsuccessful candidates:

$$\tau = |(\tau - (\tau/(a^*(1-a^*)))a^*/n)|,$$

where $a^* = 0.234$ according to [63].

HARM has the advantages of HRM, with the additional benefit of being adaptive.

In addition, since it does not require calculations based on the empirical covariance, which can potentially be expensive, depending on the number of parameters, it is faster than AM and AMM.

Componentwise Adaptive Hit-and-Run Metropolis (CHARM) is similar to HARM in every aspect, except from the fact that it uses componentwise proposals and therefore each parameter is updated individually in each iteration. It uses $a^* = 0.44$ for the univariate context, as suggested by [65].

CHARM as a sampling algorithm has the same advantages as HRM, regarding the ability to handle multimodality and its speed, along with the fast convergence that componentwise algorithms offer. Its main disadvantage is, as with all Gibbs-type samplers, that it cannot capture the correlation structure of the parameters. Because of those advantages and limitations, it is ideal for the optimisation step, when we use a two-phase simulation, but not necessarily for summarising the target distribution.

4.3.3 The Differential Evolution Markov Chain class

Differential evolution (DE) [74, 58], is an evolutionary computational method for multidimensional optimisation that uses a population of candidate solutions, generates new solutions by combining those candidates, keeps only the ones that show improvement and, finally, repeats, constructing new generations until it gets to a satisfactory solution.

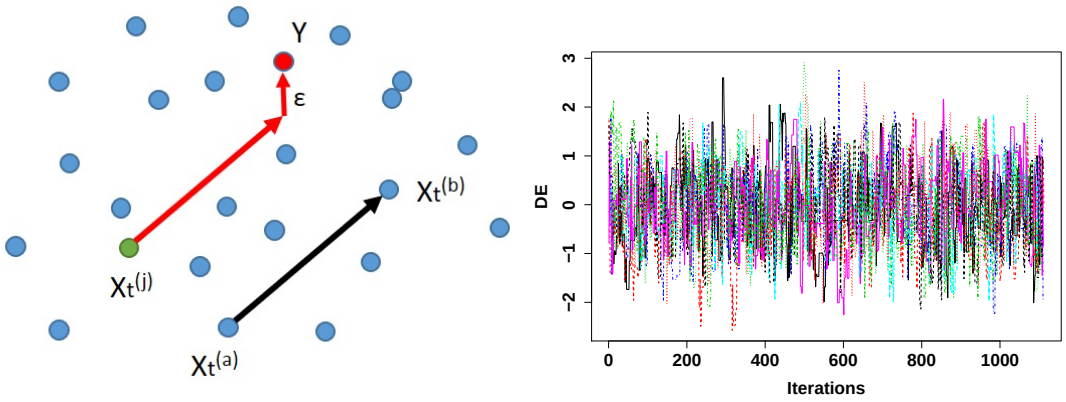
The MCMC analogue to DE is the simple **Differential Evolution Markov Chain (DEMC)** [75], where the candidate solutions correspond to N parallel chains, while the proposal for each chain is derived from the remaining $N - 1$ chains, which designs the proposal distribution in each step by updating scale and orientation and, therefore, the covariance structure information. Specifically, for N d-dimensional chains, the candidate new state $Y^{(j)}$ for chain j at step t is:

$$q(Y^{(j)}|{X_t}^{(j)}) = \mathrm{N}({X_t}^{(j)}, diag(f(a, b, \gamma_r, \varepsilon))) \tag{4.3.1}$$

$$f(a, b, \gamma_r, \varepsilon) = \gamma_r(X_t{}^{(a)} - X_t{}^{(b)}) + \varepsilon,$$

where $X_i{}^{(a)}$ and $X_i{}^{(b)}$ are the current states of two other, randomly selected chains a and b respectively, γ_r is a tuning parameter used for the optimal acceptance ratio, equal to $2.38/\sqrt{2d}$, in accordance with [65], and ε is a draw from a d-dimensional Gaussian distribution with small variance, whose main purpose is to ensure the ergodicity of the chain.

The way the candidate is constructed is shown in Figure 4.9a, while a parameter trace of $10,000$ iterations for 9 chains are shown in Figure 4.9b.



(a) DEMC candidate construction; The candidate Y is constructed according to Equation 4.3.1.

(b) Posterior marginal trace, using the DEMC algorithm for 9 chains.

Figure 4.9: DEMC algorithm.

The way the chains interact with each other in order to construct the candidate makes this algorithm adaptive. At the same time, it does not require empirical covariance calculations, so it is faster than AM and AMM.

Its main downside is that, because the candidate solutions lie on an N-dimensional space, the necessary number of chains for the algorithm must be larger than d, otherwise, the proposed candidates might not be flexible enough to scan the whole parameter space. According to [75], $N = 2d$, which, when dealing with problems that comprise many parameters, can be computationally inefficient.

Differential Evolution Markov Chain with snooker updater and fewer chains algorithm (DEzs) [12] is an extension to the aforementioned algorithm,

based on two alterations: first, in a randomly selected 90% portion of the chain, sampling the difference vectors from the history of the chains is allowed. The history of all chains is then recorded to a single matrix Z, regardless of which chain they were originated from. This procedure reduces the necessary number of chains. The proposal formula is:

$$q(Y^{(j)}|{X_t}^{(j)}) = \mathrm{N}({X_t}^{(j)}, diag(f(ta, tb, \gamma_r, \varepsilon)))$$
$$f(ta, tb, \gamma_r, \varepsilon) = \gamma_r(X_{ta} - X_{tb}) + \varepsilon,$$

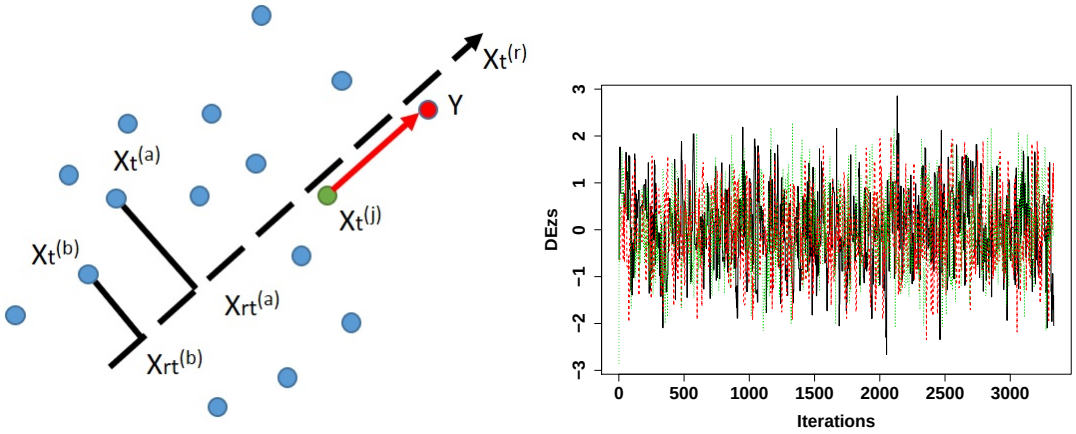
where X_{ta} and X_{tb} are randomly (uniformly and without replacement) selected rows from Z. Second, for the remaining 10% portion of the chain, the proposal is constructed using a snooker update; in order to update the t step of the j-th chain, a chain r in state ${X_t}^{(r)}$ is selected, along with two other chains a and b at states ${X_t}^{(a)}$ and ${X_t}^{(b)}$ respectively, which are then projected orthogonally on to the line ${X_t}^{(j)} - {X_t}^{(r)}$. The projections are ${X_{rt}}^{(a)}$ and ${X_{rt}}^{(b)}$. Finally, the proposal $Y^{(j)}$ is:

$$q(Y^{(j)}|{X_t}^{(j)}) = \mathrm{N}({X_t}^{(j)}, diag(f(a, b, rt, \gamma_s))) \tag{4.3.2}$$
$$f(a, b, rt, \gamma_s) = \gamma_s({X_{rt}}^{(a)} - {X_{rt}}^{(b)}),$$

where $\gamma_s = 2.38/\sqrt{2}$ is the univariate version of the optimal scaling mentioned. In the case of the DEzs algorithm, where the snooker update is combined with sampling from the past of the chains, the chains chosen for the update are replaced by sampled rows from the matrix Z. The advantage of the snooker update is that it permits larger jumps. Therefore the algorithm can jump easily between modes, thus permitting the sampling from a multimodal density.

The construction of the snooker update is shown in Figure 4.10a, while a parameter trace of $10,000$ iterations for 3 chains of the DEzs algorithm are shown in Figure 4.10b.

In the context of the deconvolution problem, as formulated in Chapter 2 by the multi-



(a) Snooker update construction; The candidate Y is constructed according to Equation 4.3.2.

(b) Posterior marginal trace, using the DEzs algorithm for 3 chains.

Figure 4.10: DEzs algorithm.

region radial composite model and in Chapter 3 by the Bayesian context, we expect non-trivial and potentially strong correlations between the posterior parameters. Therefore, Gibbs-type samplers will not be suitable, at least for summarising the target distribution, since all the parameters need to be updated simultaneously. Second, this covariance structure is not known a priori, therefore the algorithm also needs to be adaptive in order to gradually update the covariance structure as samples from the chain are accumulated. Finally, we have reason to expect multimodality of the posterior, as different combinations of reservoir parameter values can yield similar response functions; therefore, the algorithm also needs to be able to detect the presence of more than one modes.

An algorithm that can tackle all those challenges is the DEzs. It takes into account the correlation structure, and also adapts it 'on the fly' by constructing the candidates through combinations of interconnected chains. In addition, it can handle multiple modes through the snooker update. Because of this, it is the main algorithm that we are going to use for the results.

4.4 Convergence diagnostics

One of the main challenges, when using any MCMC of the previous section, is to determine whether the algorithm has converged. In this section, we present some popular diagnostics, that are used to establish that convergence.

A first step to investigate whether an adaptive algorithm is efficient, and whether the chain has converged, is to inspect the convergence of the adaptation. For this, we use the **empirical covariance matrix**, and we analyse the progress of its elements graphically. For t iterations of the d-dimensional state X, the empirical covariance matrix Cov is the $d \times d$ matrix:

$$Cov = \frac{1}{t-1} \sum_{i=1}^{t} (X_i - \overline{X})(X_i - \overline{X})^T,$$

where

$$\overline{X} = \frac{1}{t} \sum_{i=1}^{t} X_i$$

is the mean vector. During the initial stages of an adaptive algorithm, we usually observe big changes in the elements of the covariance matrix, which suggest that the adaptation is taking place. After a while, when the algorithm has acquired information about the covariance structure, the elements of the matrix stabilise. This stabilisation suggests that the adaptation has converged. An example is shown in Figure 4.11a, where the empirical covariance plot suggests stabilisation of the adaptation after 2,000 iterations.

The **Geweke plot** [28] for Markov chains is based on a test for equality of the means of the first and last part of the chain, called Geweke diagnostic. The premise of the test is to replicate a simple two-sample equality of means test, based on the standard Z-score, which is the difference between the two sample means divided by the estimated standard error, after taking into account autocorrelation:

$$z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}},$$

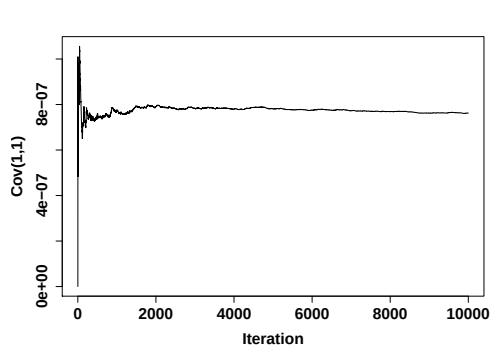
where n_i is the number of iterations and σ_i^2 the variance in each sample, for $i = 1, 2$.

We also use the assumption that the two parts of the chain are asymptotically independent.

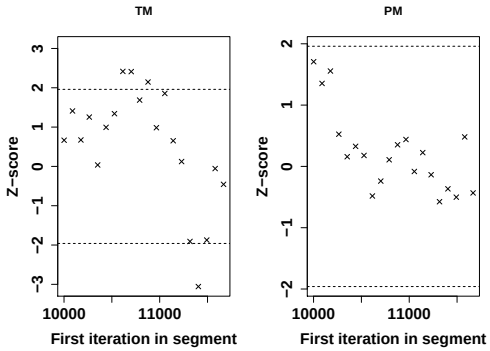
In practice, after we discard the 'burn-in' iterations, we split the first half of the remaining chain into two parts: the first, is divided into $n - 1$ segments, and then the Geweke diagnostic is repeatedly calculated; first using all chain iterations, then after discarding the first segment, the first two segments, and so on. Examples are shown in Figure 4.11b; in the first, there are four Z-scores outside the dashed lines that denote the 95% confidence interval, which suggests that convergence may not have been achieved. In the second, all scores are within the interval, which indicates that the chain has converged.

The **Cumulative quantile plot** shows the evolution of the sample median (50-th percentile), 2.5-th and 97.5-th percentiles as a function of the number of iterations. Again, we expect volatility in the first stages of the adaptation, when the adaptation is still taking place, that leads to stabilisation for all three percentiles, when the chain has converged. In contrast to the Geweke plots, the cumulative quantile plot analyses the convergence, by showing both the evolution of the location (in the sense of the median), and the scale, through the other percentiles. An example is shown in Figure 4.11c, where all percentiles converge after 10,500 iterations (in this plot, we have already discarded the burn-in phase).

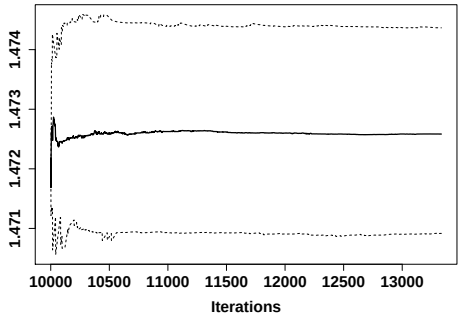
Finally, the **Gelfand Diagnostic** [24] is a graphical convergence diagnostic for Markov chains, where we overlay marginal posterior distributions in the form of density curves. These curves are constructed using samples from different sections of the same chain. For example, a chain, split into k equal parts, corresponds to k density curves, constructed by the respective section of the chain. If all densities are similar, a case can be made for convergence. An example of three densities from different parts of a single converged chain is shown in Figure 4.11d.



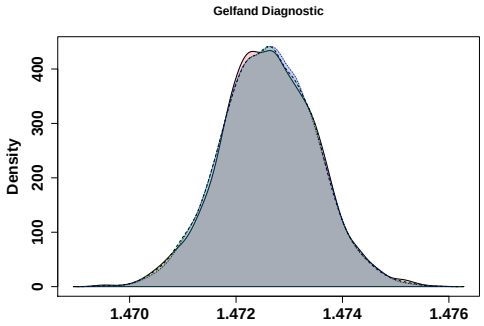
(a) Empirical covariance plot.



(b) Geweke plots.



(c) Cumulative quantile plot.



(d) Gelfand plot.

Figure 4.11: Convergence diagnostic plots.

4.5 Model Selection

In order to determine the number of transitions for the multi-region radial composite model in the later chapters, we use model selection methods: first, we examine a variety of different information criteria that compare statistical models for a given set of data, and report a relative goodness-of-fit measure. Then, we present the Bayes factor, which is a variation of the classic hypothesis testing, and is widely popular with Bayesian problems.

4.5.1 Information criteria

The **Akaike information criterion (AIC)** [4] is a likelihood-based criterion that favours goodness of fit, while it penalises complexity, based on the number of model parameters. For d parameters θ over the parameter space Θ and data y, the maximum

likelihood estimate (MLE) is

$$\widehat{\theta}_{MLE} = \arg\max_{\theta\in\Theta}[P(y|\theta)].$$

The maximum value of the likelihood function $\widehat{L}$ is given as the likelihood value for the MLE:

$$\widehat{L} = P(y|\widehat{\theta}_{MLE}).$$

Finally, the AIC formula is:

$$AIC = 2d - 2\ln\widehat{L} = 2d - 2\ln\left[P(y|\widehat{\theta}_{MLE})\right],$$

where d is the number of model parameters, and smaller values of the AIC, as well as the rest of the criteria in this section, correspond to better models.

In practice, the formula favours the models with higher maximum likelihood values (better fit), while penalising the number of parameters (model complexity).

In the Bayesian context, an equivalent quantity to the maximum likelihood estimate is the maximum a posteriori estimate (MAP), which reflects the mode of the posterior.

$$\widehat{\theta}_{MAP} = \arg\max_{\theta\in\Theta}[P(\theta|y)],$$

and the AIC formula becomes:

$$AIC = 2d - 2\ln\left[P(\widehat{\theta}_{MAP}|y)\right].$$

Since it uses information related only to the MLE, the AIC is a criterion that compares the mode, or the 'best fit' of the models.

The **Bayesian information criterion (BIC)** [69] is another criterion, similar to the AIC. The difference between the two is the BIC uses a different formula that also takes into account the number of observations n for the penalty term:

$$BIC = \ln(n)d - 2\ln\{\arg\max_{\theta\in\Theta}[P(y|\theta)P(\theta)]\}.$$

In general the BIC gives similar results to the AIC, and it also serves the same purpose, which is to compare penalised modes. The main difference is that if the 'true model', from which the data were generated, is part of the selection process, the BIC will select it with probability 1 for large enough data [5]. In practice, we do not have access to the true model, so the difference between the two criteria is mostly academic.

The **Deviance information criterion (DIC)** [26] is a criterion that takes into account summary statistics of the deviance:

$$D(\theta) = -2\log P(y|\theta). \tag{4.5.1}$$

It is widely used for models, where the information comes from MCMC samples, but has some limitations, since it lacks consistency and it is not invariant to reparameterisation.

The DIC formula is:

$$DIC = \overline{D(\theta)} + \mathrm{Var}(D(\theta))/2,$$

where $\overline{D(\theta)}$ and $\mathrm{Var}(D(\theta))$ denote the posterior mean and the variance of the deviance in Equation 4.5.1 with respect to θ.

The main advantage of the DIC among the other information criteria mentioned in this section, is that it includes information coming from the whole chain. Therefore, it is more suitable for problems associated with the Bayesian context. It is also very easy to compute from the output of an MCMC. Its main disadvantage regards the restrictive assumption of an approximately Gaussian posterior, which leads us to seek a more robust method for Bayesian model selection.

4.5.2 Bayes factor

In order to perform model comparison, using the **Bayes factors** [43], we initially consider two hypotheses H_1 and H_2. We also consider a set of data y that has arisen under one of those hypotheses with probabilities $P(y|H_1)$ and $P(y|H_2)$ respectively, given priors $P(H_1)$ and $P(H_2)$. We consider the Bayes theorem in order to derive the *posterior opinions*, which are the hypotheses after consideration of the data. This leads to:

$$P(H_j|y) = \frac{P(y|H_j)P(H_j)}{\sum_{k=1}^{2} P(y|H_k)P(H_k)}, \quad j = 1, 2.$$

The ratio of the hypotheses' probabilities is:

$$\frac{P(H_1|y)}{P(H_2|y)} = \frac{P(y|H_1)}{P(y|H_2)} \frac{P(H_1)}{P(H_2)},$$

where the quantity

$$BF_{21} = \frac{P(y|H_1)}{P(y|H_2)} \tag{4.5.2}$$

is referred to as the Bayes factor of H_1 against H_2.

In the case that the hypotheses are associated with distributions with no free parameters, BF_{21} corresponds to the *likelihood ratio*. In contrast, when there are unknown parameters (which is the case in the general Bayesian context), $P(y|H_i)$ can be calculated by integrating over the parameter space. That leads to:

$$P(y|H_i) = \int_{\Theta} P(y|\theta_i, H_i)P(\theta_i|H_i)d\theta_i. \tag{4.5.3}$$

The quantity in Equation 4.5.3 corresponds to the marginal likelihood, since it is obtained by integrating the joint density of data and parameters over the latter. The marginal likelihood of an MCMC output can be calculated using numerical methods [15].

After we make the appropriate calculations and we take into account the formula in Equation 4.5.2, we can make a decision on whether there is enough evidence against H_1, in a manner similar to conventional hypothesis testing, using Table 4.1 [43].

$2 \ln BF_{21}$	BF_{21}	Evidence against H_1
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
>10	>150	Very strong

Table 4.1: Interpretation of the Bayes factor values.

Note that the fact that small positive values of the Bayes factors do not provide enough evidence against H_1, is essentially equivalent to the complexity penalty term of the information criteria.

In case we need to compare more than two models, we can either compare all of them pairwise, or calculate the *weighted Bayes factor*, which takes the marginal likelihoods of all the alternative hypotheses with respect to the model with the smaller marginal likelihood. In that sense, we compare all the candidate models against the 'best' model and examine the levels of evidence against it.

Even though the Bayes factor is not very easy to compute, it has some significant advantages. First, it includes information for the whole model, and not only the mode such as the AIC and the BIC. Second, it does not require strict assumptions for the posterior as the DIC. Finally, by using marginalisation of the parameter space, instead of summary statistics, the Bayes factor is the most principled way of model selection among the ones presented in this chapter. However, it should be noted that Bayes factors can be very sensitive to the choice of prior distribution, and, thus, they should be used cautiously.

4.6 Optimisation methods

As mentioned in the beginning of this chapter, conventional optimisation methods can be associated with the Bayesian context. Specifically, they can be used in order to provide us with a global maximum of the logarithm of the unnormalised joint posterior density, which can then either be used as a starting point for an MCMC method or as the first step for Laplace approximation, a method described in the

next section. Because of the nature of the multi-region radial composite model, there are various reasons that cause most optimisation techniques to perform poorly: the complicated posterior structure, the possible degeneracies that come from overparameterisation, when more regions than required are added, as well as the potential multimodality of the posterior density, make it challenging for algorithms that use derivative information such as Newton-Raphson [55] and Levenberg-Marquardt [48, 51] to act efficiently. Moreover, stochastic methods like particle swarm optimisation [45] tend to take a very long time to find a sufficient solution. This last issue can potentially be tackled with the addition of expert knowledge that could decrease the search space by a significant margin.

In the context of this thesis, we found that optimisation algorithms do not provide substantial results, and instead we prefer using the two-phase simulation with MCMCs performing the optimisation when it is required.

4.7 Laplace approximation

An alternative to the Markov Chain Monte Carlo algorithms for summarising posterior distributions, when the closed form of the density is not available, is the Laplace approximation or Laplace method [47], which is a family of asymptotic techniques used to approximate integrals of smooth functions, by appropriate Gaussian densities.

We consider the posterior distribution $P(\theta|x)$, where θ are the parameters of interest, x the data and $\widehat{\theta}_{MAP}$ is the maximum a posteriori estimate. We denote the logarithm of the un-normalised posterior as

$$g(\theta) = \log\left[P(x|\theta)P(\theta)\right]. \tag{4.7.1}$$

We approximate g using a second order Taylor series expansion around the MAP:

$$g(\theta) \approx g\left(\widehat{\theta}_{MAP}\right) + g'\left(\widehat{\theta}_{MAP}\right)\left(\theta - \widehat{\theta}_{MAP}\right) - \frac{1}{2}\left(\theta - \widehat{\theta}_{MAP}\right)^T g''\left(\widehat{\theta}_{MAP}\right)\left(\theta - \widehat{\theta}_{MAP}\right), \tag{4.7.2}$$

where the gradient at the MAP is:

$$g'(\widehat{\theta}_{MAP}) = \frac{\partial \log\left[P(x|\theta)P(\theta)\right]}{\partial\theta}|\theta = \widehat{\theta}_{MAP},$$

and the Hessian at the MAP is:

$$g''(\widehat{\theta}_{MAP}) = \frac{\partial^2 \log\left[P(x|\theta)P(\theta)\right]}{\partial\theta^2}|\theta = \widehat{\theta}_{MAP}.$$

Since $\widehat{\theta}_{MAP}$ is the mode, $g'(\widehat{\theta}_{MAP}) = 0$, and $g''(\widehat{\theta}_{MAP})$ is positive definite.

The un-normalised posterior can be approximated by exponentiating Equation 4.7.2:

$$P(x|\theta)P(\theta) \approx P\left(x|\widehat{\theta}_{MAP}\right) P\left(\widehat{\theta}_{MAP}\right) \exp\left\{-\frac{1}{2}\left(\theta - \widehat{\theta}_{MAP}\right)^T g''\left(\widehat{\theta}_{MAP}\right)\left(\theta - \widehat{\theta}_{MAP}\right)\right\}. \tag{4.7.3}$$

Note that the d-dimensional Gaussian density of a random variable θ with mean equal to $\widehat{\theta}_{MAP}$, and covariance $\left[g''\left(\widehat{\theta}_{MAP}\right)\right]^{-1}$ is:

$$\frac{1}{(2\pi)^{\frac{d}{2}}}\left|g''\left(\widehat{\theta}_{MAP}\right)\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\theta - \widehat{\theta}_{MAP}\right)^T g''\left(\widehat{\theta}_{MAP}\right)\left(\theta - \widehat{\theta}_{MAP}\right)\right\}. \tag{4.7.4}$$

From Equation 4.7.3 and Equation 4.7.4, we conclude that the posterior of θ can be approximated by:

$$\theta \sim N\left(\widehat{\theta}_{MAP}, \left[g''\left(\widehat{\theta}_{MAP}\right)\right]^{-1}\right). \tag{4.7.5}$$

In Figure 4.12, we plot the posterior marginal Gaussian density of a Laplace approximation and the corresponding histogram from the MCMC. Note that the results agree, which is not always the case as we will see in the next chapters.

The main advantage of the Laplace approximation compared to an MCMC run is that it is in general much faster; since it requires an optimisation step and a

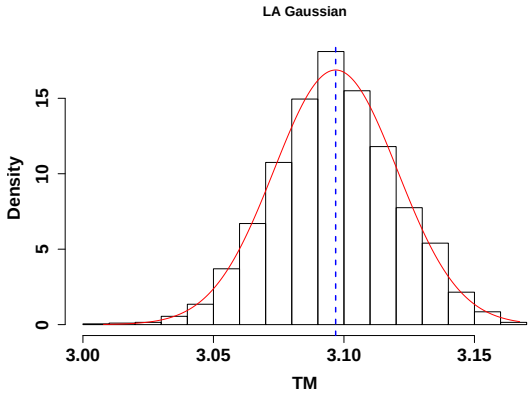


Figure 4.12: Histogram of the marginal posterior from an MCMC and corresponding Laplace approximation density curve. The dashed blue line denotes the MAP.

numerical Hessian approximation step, such as Richardson's extrapolation [61], it is easier to compute than an MCMC. However, in cases where a good optimisation technique is not available, this advantage is irrelevant.

In addition, the Laplace approximation is inappropriate for posterior densities that are not approximately multivariate normal, like heavily skewed or multimodal distributions.

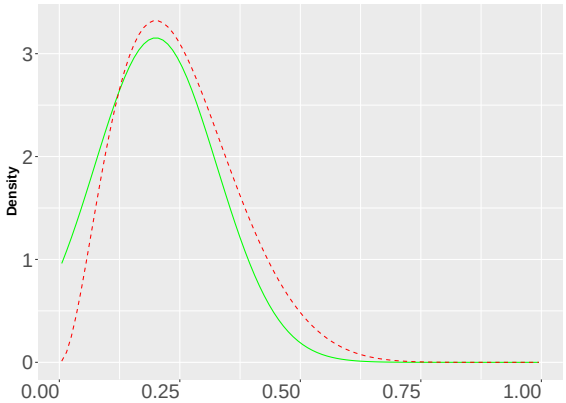


Figure 4.13: Sampling importance resampling (SIR); the green curve corresponds to an importance function that approximates the target density (dashed red curve).

A possible way of overcoming this issue is by using the Sampling Importance Resampling (SIR) technique [31], which is a method of gaining information for a distribution that we cannot sample directly, given a proposal distribution that we can

access. The steps of SIR are as follows:

- **Sampling** from the known distribution. In the case of the Laplace approximation, that corresponds to the Gaussian approximation of the posterior.
- Calculating the **importance weights**; first we calculate the posterior densities of the samples from the previous step, then, the corresponding densities from the importance function, which is the Gaussian approximation, and finally, we take the ratios:
$$w = \frac{P_{post}(X_i)}{P_{Gauss}(X_i)},$$
where X_i are the samples from the previous step.
- **Resampling** from the weighted samples with replacement. In that way, the probability of obtaining a sample is proportional to its weight, and the new samples are samples from the target density.

An example of SIR is shown in Figure 4.13, where the red dashed curve is the target density, while the green curve is a Gaussian distribution, that takes the role of the importance function.

In practice, the Laplace approximation includes the following steps:

1. We use an optimisation algorithm, in order to find the MAP.
2. We use an algorithm, in order to numerically approximate the Hessian on the MAP.
3. We take some random samples from the multivariate Gaussian distribution with mean equal to the MAP and covariance equal to the inverse of the aforementioned Hessian.
4. Additionally, in case a multivariate Gaussian distribution is not a good approximation of the posterior, we can improve it by using Sampling Importance Resampling (SIR), using the multivariate Gaussian distribution from the previous step as the importance function.

4.8 Computational methods for the results chapters

The general procedure that we follow for the next chapters, where we use computational methods in order to apply the Bayesian methodology of Chapter 3 on synthetic and real data, is as follows: first, we apply a two-phase simulation, where we initially run a few iterations (usually $20,000$) of the componentwise adaptive Hit-and-Run Metropolis (CHARM) algorithm. Our results showed that this algorithm surpasses conventional optimisation techniques when applied to the radial composite models. Then, we run our main MCMC, which is the Differential Evolution Markov Chain with snooker updater and fewer chains algorithm (DEzs), using the largest posterior value point obtained from the previous step as the initial vector. This algorithm is adaptive, it updates all the parameters simultaneously, and therefore recovers the correlation structure, and can also handle multimodality. After we retrieve the samples, we discard the burn-in, which accounts for the iterations where the covariance structure gets updated, and thinning for efficiency. We assess convergence using the methods from Section 4.4 and we summarise the results using a variety of visualisation methods. In Chapter 6, where we show results for real data, we also perform model selection using the techniques from Section 4.5.

Finally, in the same chapter, we compare the performance of the MCMC algorithms from Section 4.3, and we show results of the Laplace method from Section 4.7.

4.9 Summary

In this chapter we introduced the Markov Chain Monte Carlo theory, and we presented a selection of algorithms, including the popular Metropolis-Hastings and Gibbs samplers. We mentioned various computational challenges and ways to tackle them. We presented convergence diagnostics and model selection methods. We discussed why optimisation techniques are not suitable for the deconvolution problem and we

demonstrated how the Laplace approximation method works. Finally, we gave an overview of the steps of our method and how we plan to implement them.

In the next chapter we apply our Bayesian statistical configuration using the radial composite model, as well as the other preliminary models, on synthetic data and we present and discuss the results.

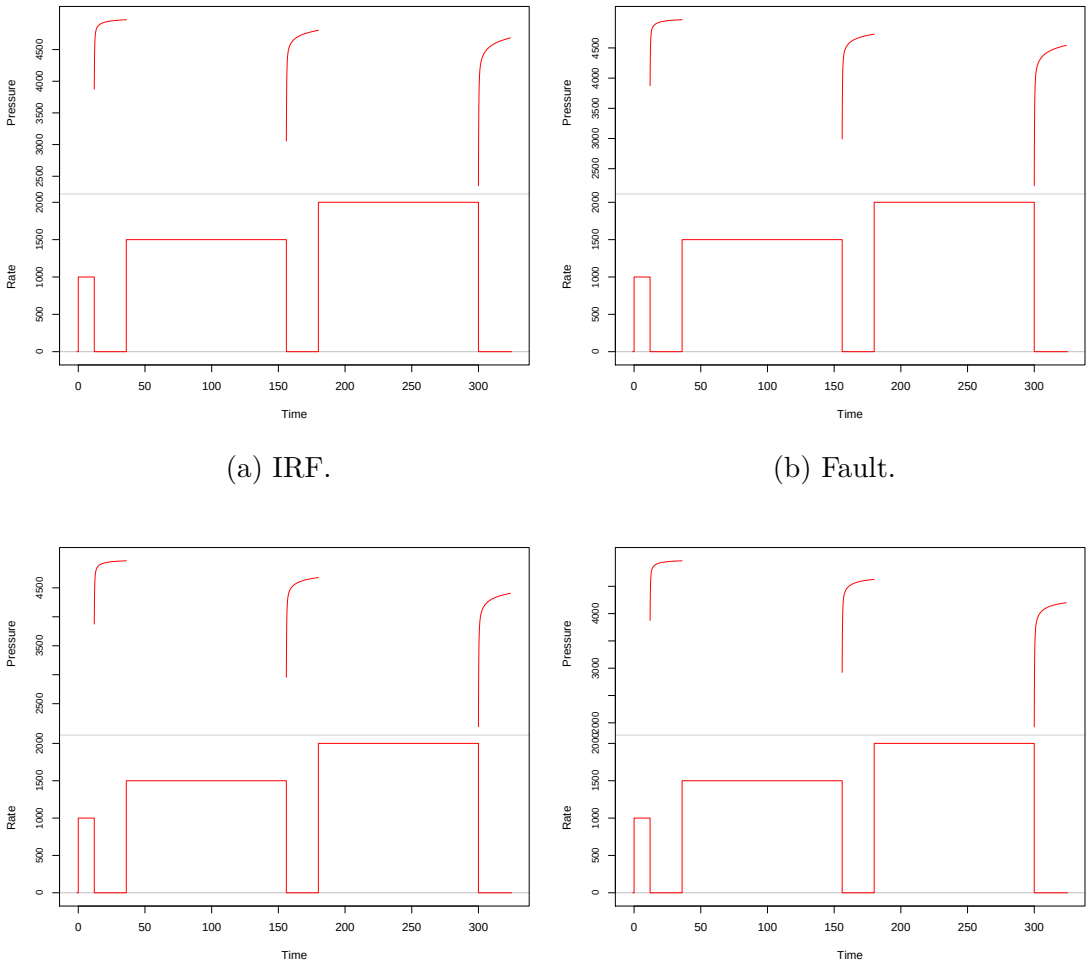
Chapter 5

Analysis of synthetic data

In this chapter we validate our method by applying it to synthetic problems with known solutions. We use the response models from Chapter 2, the statistical model from Chapter 3 and the DEzs algorithm, which we described in Chapter 4 on four synthetic data sets, first presented in Subsection 1.3.1. The data sets are from [19] and were provided by the collaborators without access to the underlying models and simulators. The purpose of this Chapter is to examine the effect of applying the radial composite model to simple synthetic examples derived from simple reservoir boundary configuration models, thus validating and illustrating our methodology. In each set we examine the MCMC output, the convergence and efficiency of the chain, and some additional visualising tools.

5.1 Synthetic data sets

The four data sets are generated from synthetic reservoirs with known flow behaviours. This allows for verification of the validity of our approach on a known solution problem. Each data set corresponds to a well test of 320 hours, where there are three periods of production (drawdowns) at constant but different rate values, and three no flow periods (buildups). A total of 272 pressure measurements are generated by convolution of the rate with the known response. To represent potential observational



(a) IRF.

(b) Fault.

(c) Channel.

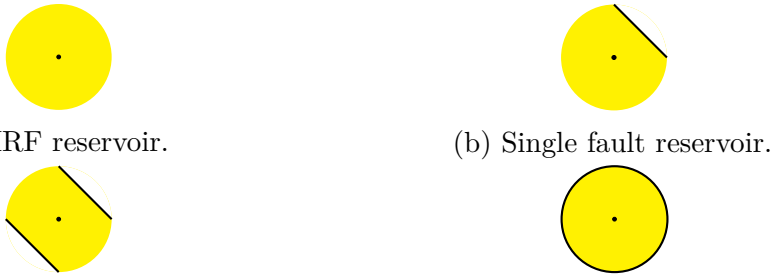
(d) Closed.

Figure 5.1: True rate and pressure for the synthetic data sets as provided. The rates are identical, but the pressure is different in each case (even though it is indistinguishable in this scale).

error, the pressure measurements include additive random noise with known standard deviation $\sigma_p = 5$. The corresponding parameter is fixed to that value in our model.

The data for the simulated sets are shown in Figure 5.1. Even though the pressure histories are different by construction, the plots are indistinguishable from each other in this scale, which is one of the main reasons that modern well test analysis interpretation tools use the pressure derivative and its transformation (in Figure 5.3), which facilitate visualisation.

Each set is associated with a simple reservoir boundary configuration, shown in



(a) IRF reservoir.

(b) Single fault reservoir.

(c) Channel reservoir.

(d) Closed reservoir.

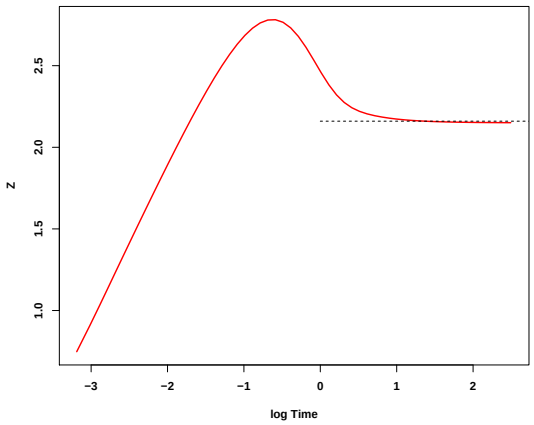
Figure 5.2: Reservoir boundary configurations of the synthetic data sets.

Figure 5.2 and a corresponding response curve feature, shown in Figure 5.3. They were all examined in Subsection 1.3.1.

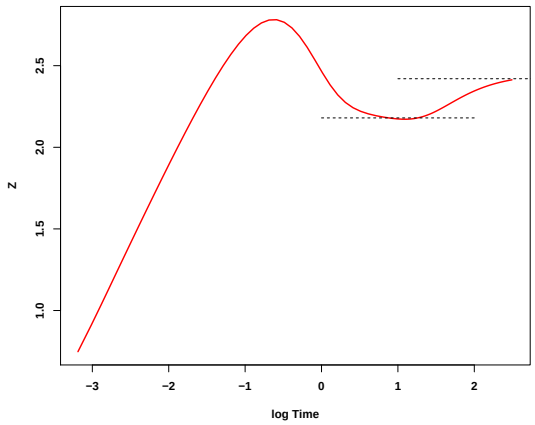
5.2 Model setup and priors

Our primary focus with this analysis is to verify that our methods are able to capture the reservoir response behaviour, even though some of the reservoir boundary configurations, like the fault and the channel, are not products of radially composite reservoirs, and to determine how well the uncertainty about the parameters can be resolved from the data given only weakly informative priors. For this analysis, we treat the rate values and initial pressure as known and equal to their measured values ($\tilde{q} = q$ and $\tilde{p}_0 = p_0$), and so our parameters of interest are just the parameters of the radial composite response function, $\boldsymbol{\phi}$. For the response parameters, we choose vaguer priors than those recommended in Chapter 3: a Gamma prior for $C_D e^{2S}$ and broad uniform priors for the remaining parameters. Specifically, we choose a radial composite model with 1 transition (2 regions), motivated by the known complexity of the system, and our prior choices were $T_M \sim \mathrm{U}(-1, 5)$, $P_M \sim \mathrm{U}(-1, 5)$, $C_D e^{2S} \sim \mathrm{Ga}(1, 0.2)$, $R_D \sim \mathrm{U}(0, 10)$, $M \sim \mathrm{U}(-10, 30)$ and $\eta \sim \mathrm{U}(-10, 30)$.

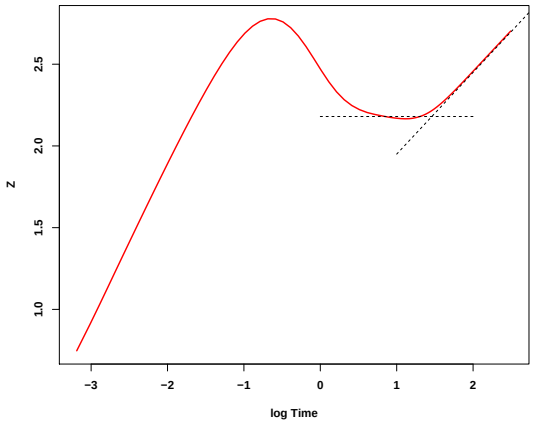
After 20,000 iterations from the CHARM algorithm (as described in Section 4.8), which are not presented here, the main results consist of 666,667 iterations for each of three parallel chains from the DEzs algorithm. Thinning of 50 was applied, and



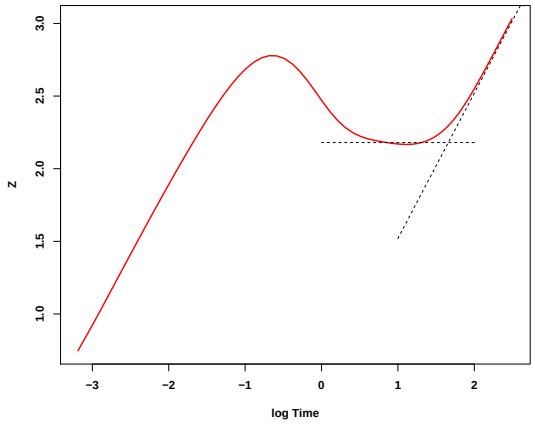
(a) IRF.



(b) Fault.



(c) Channel.



(d) Closed.

Figure 5.3: Response curves for the synthetic data sets, including the corresponding stabilisation and slope lines.

the first 1,000 iterations were discarded as burn-in, leaving 37,002 iterations to be used for inference.

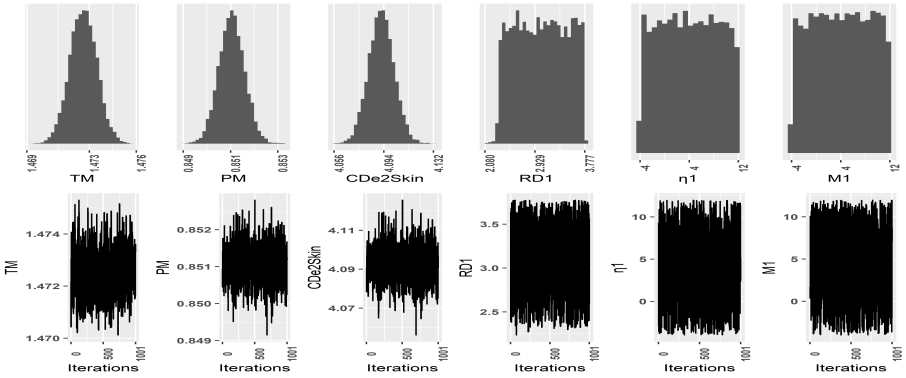
The results took several hours to be produced, mainly due to the computation of the convolution integral in Equation 1.4.1 and the numerical Laplace inversion using the Stehfest algorithm [72]. Possible ways to tackle those issues are presented in Chapter 7.

5.2.1 MCMC output

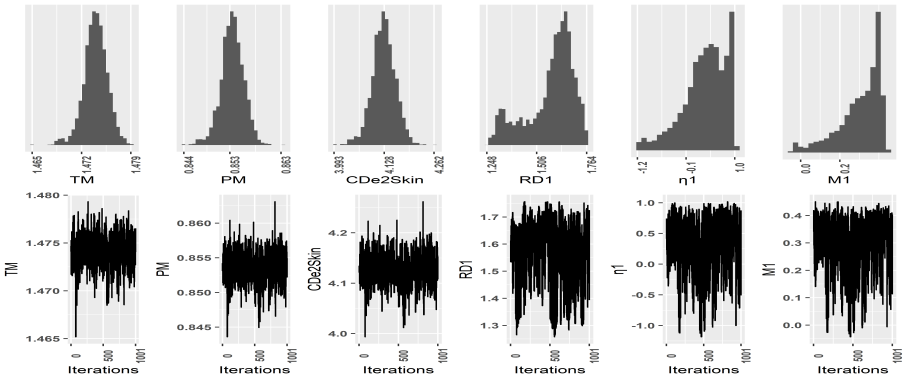
The histograms and traces of the posterior MCMC samples are shown in Figure 5.4, while the summary statistics are presented in Table 5.1. The first three parameters (T_M, P_M and $C_D e^{2S}$) have been identified with a small posterior variance and approximately same ranges in all four data sets. These parameters correspond to the position and the early time shape of the response function, and as such, we would expect they could be identified from the data.

In contrast, there are significant dissimilarities between the late time parameters. For the IRF model, the radius and ratio posterior marginal distributions have a uniform shape with very wide range. Given that the radial flow regime is the default for the inner region of the radial composite model, these results indicate that the effect of the transition has been somehow rendered irrelevant for this model, which is expected, since there is no change in the late time behaviour. Specifically, the transition takes place *beyond the rightmost boundary* of the plot, and thus, regardless of the late time parameter values, the visible part of the plot reflects the common IRF stabilisation.The fault indicates that all three late time posterior marginals are spike-shaped, negatively skewed with relatively small ranges, and they also exhibit a second mode. Those results suggest that, for the fault model, the data identify the late time parameters quite precisely, but also that there are two distinct configurations that are perceived by the model as similar, with respect to the response. These different configurations are going to be examined later in this chapter. The channel and the closed models appear to have similar behaviour, with the histogram of R_D exhibiting a narrow peak, albeit with a longer tail to the right, while the histograms of M and η have a very wide range and approximately uniform posterior shape, indicating an inability to resolve specific values for the parameters of the transition, when dealing with a reservoir configuration that corresponds to a late time slope response feature.

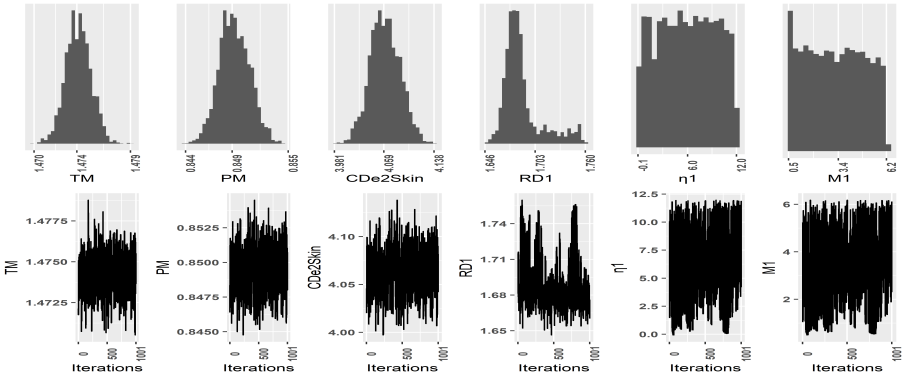
In practice, those results illustrate that the early time shape, as well as the distance that a feature appears are identified without significant difficulty from the model



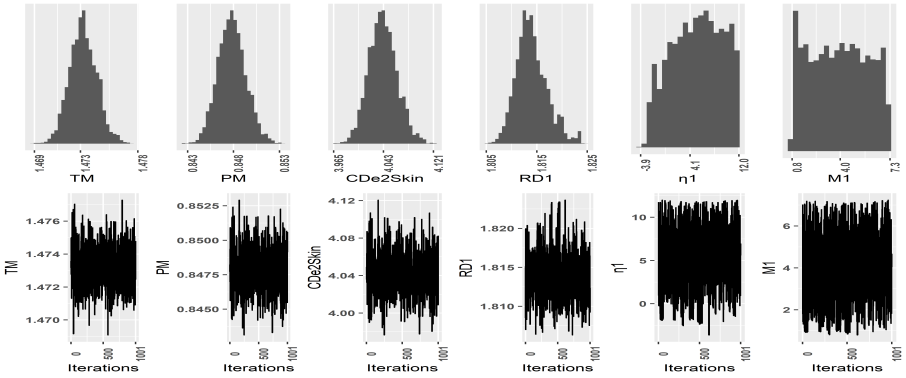
(a) IRF.



(b) Fault.



(c) Channel.



(d) Closed.

Figure 5.4: Marginal posterior densities in the form of histograms and traces from the MCMC for the synthetic data sets.

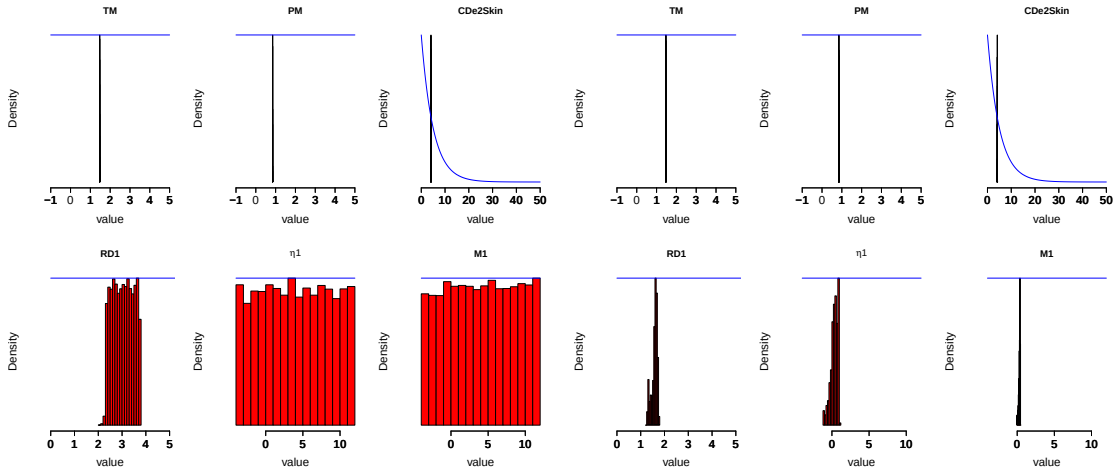
IRF	min	25%	mean	75%	max	sd
T_M	1.47	1.47	1.47	1.47	1.48	0.00
P_M	0.85	0.85	0.85	0.85	0.85	0.00
$C_D e^{2S}$	4.06	4.09	4.09	4.10	4.13	0.01
R_D	2.08	2.67	3.04	3.41	3.78	0.43
η	-4.00	0.07	4.02	7.98	12.00	4.61
M	-4.00	0.15	4.11	8.11	12.00	4.60
Fault	min	25%	mean	75%	max	sd
T_M	1.47	1.47	1.47	1.48	1.48	0.00
P_M	0.84	0.85	0.85	0.85	0.86	0.00
$C_D e^{2S}$	3.99	4.11	4.13	4.15	4.26	0.03
R_D	1.25	1.52	1.58	1.66	1.76	0.12
η	-1.20	0.06	0.32	0.71	1.02	0.48
M	-0.05	0.25	0.30	0.39	0.46	0.11
Channel	min	25%	mean	75%	max	sd
T_M	1.47	1.47	1.47	1.47	1.48	0.00
P_M	0.84	0.85	0.85	0.85	0.85	0.00
$C_D e^{2S}$	3.98	4.04	4.06	4.08	4.14	0.02
R_D	1.65	1.67	1.69	1.69	1.76	0.02
η	-0.08	3.05	5.94	8.88	12.00	3.45
M	0.49	1.75	3.21	4.64	6.21	1.67
Closed	min	25%	mean	75%	max	sd
T_M	1.47	1.47	1.47	1.47	1.48	0.00
P_M	0.84	0.85	0.85	0.85	0.85	0.00
$C_D e^{2S}$	3.96	4.03	4.04	4.06	4.12	0.02
R_D	1.80	1.81	1.81	1.82	1.82	0.00
η	-3.88	2.05	5.22	8.57	12.00	3.99
M	0.76	2.28	3.92	5.52	7.28	1.87

Table 5.1: Summary statistics for the posteriors distributions of $\boldsymbol{\phi}$ for the synthetic data sets.

(the exception of the IRF will be explored and justified later), while the 'ease' of the identification of a late time feature depends on its nature.

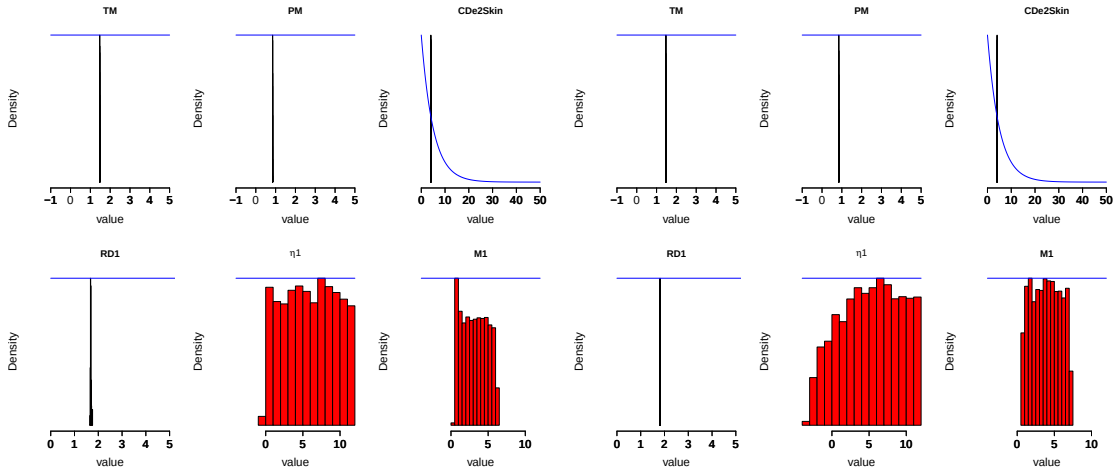
Finally, it is worth noting that the shapes of the MCMC traces are the first indicators that the chains have converged. The issue of convergence will be examined later in this chapter.

Overlaying the histograms with the matching prior densities in Figure 5.5, shows that the posterior marginal distributions for the early time parameters are substantially narrower than our vague prior densities, which indicates that the data identifies these quantities quite precisely in all cases. In contrast, the late time parameters have



(a) IRF.

(b) Fault.



(c) Channel.

(d) Closed.

Figure 5.5: Posterior marginal histograms from the MCMC and corresponding priors for the synthetic data sets.

a different behaviour, depending on the reservoir: The IRF model, comprises the widest posterior marginal ranges, when compared to the other reservoir boundary configurations. The radius parameter shows considerably larger values than the other models, while the ratio parameters cover the whole prior range. The lack of information from the ratio parameters, along with the large values of R_D, leads us to assume that the model is overparameterised, and thus the transition lies outside the radius of investigation, where the late time parameters do not have any actual effect on the response curve.

In the fault case, all the posterior marginals of the late time parameters have very narrow range, indicating that reservoir boundary configurations which correspond

to a second stabilisation on the response plot, can identify the late time parameters from the level of stabilisation.

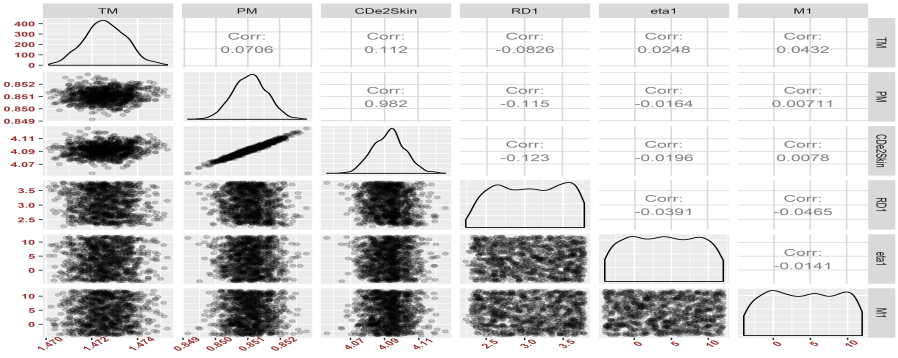
The channel histograms present a narrow radius, but wide ratio ranges. Both η and M posterior histograms comprise positive values, with the former covering the whole positive prior subspace, and the latter a significant portion. The results suggest that, for the reservoir models that experience a late time slope in the response curve (such as the half-unit slope of a channel), the posterior marginal densities of the ratio parameters are only weakly identifiable from the data, and the model is unable to find specific values for these transition parameters.

Finally, the closed reservoir model is similar to the channel, reinforcing the belief that the aforementioned results are associated with models that correspond to a response with a late time slope. One significant difference is that the η parameter covers a larger portion of the prior range, including negative values.

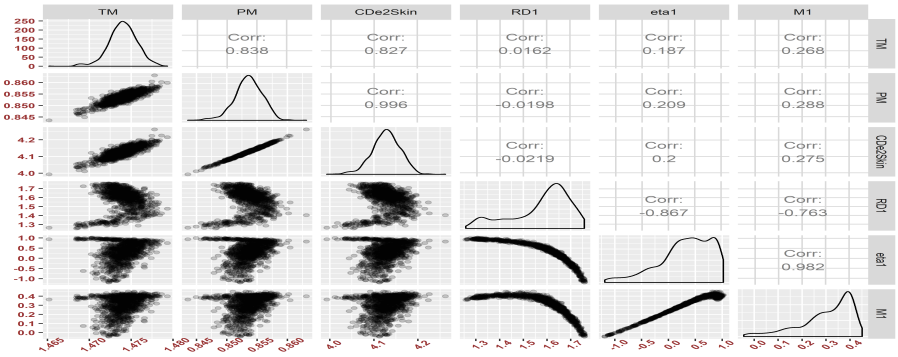
It is worth noting that in both the channel and the closed reservoir cases we can observe that the posterior behaviour of η is localised relative to the prior, but in a non-Gaussian way. As the scatterplots in Figure 5.6 indicate, this parameter is heavily correlated with M, and thus a possible explanation is that the form of the posterior η parameter is actually dictated by the posterior of M, and the values that the prior of the latter allows.

The joint posterior marginals, summarised by scatterplots, and the Pearson correlation coefficients in Figure 5.6 reveal some substantial correlations, lending credence to the hypothesis that slope type features do not have a uniquely identifiable representation in a radially-structured model.

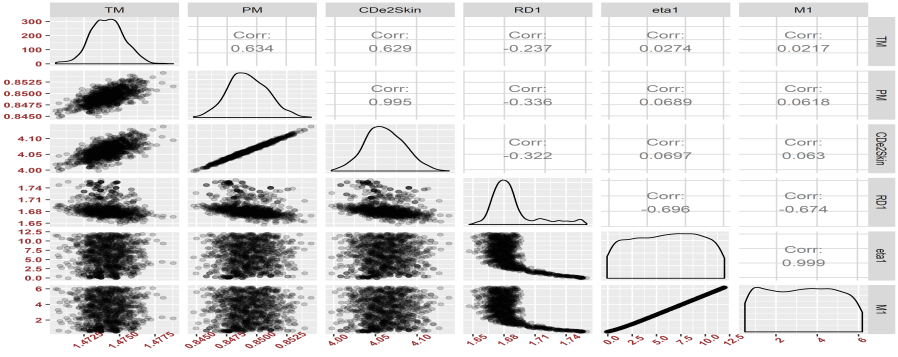
We note first, for all four data sets, the strong correlation between P_M and $C_D e^{2S}$ – larger P_M values shift the whole response curve downwards, while larger $C_D e^{2S}$ values increase the height of the skin 'hump'. Thus, P_M and $C_D e^{2S}$ appear to compensate for each other with minimal overall impact on the response function. Additionally, with the exception of the IRF model, the positive correlation for the early-time



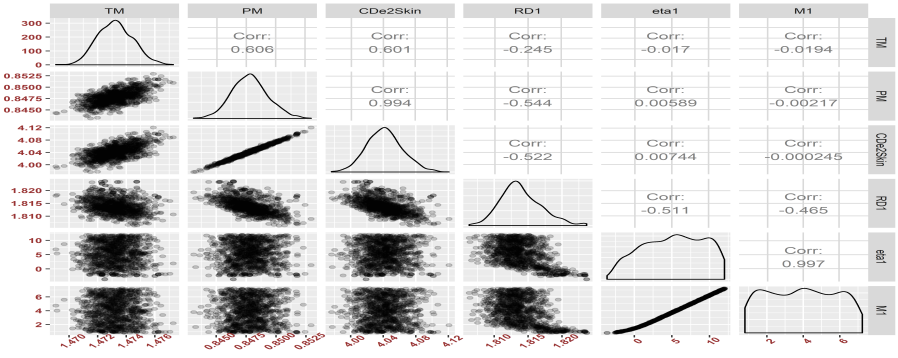
(a) IRF.



(b) Fault.



(c) Channel.



(d) Closed.

Figure 5.6: Joint posterior marginals, in the form of scatterplots and Pearson correlation values for the synthetic data sets.

parameters extends, in a weaker form, to T_M, which implies that the horizontal shift of the response can also compensate for the effects of P_M and $C_D e^{2S}$; larger values of T_M shift the curve to the left, causing the skin 'hump' to appear earlier, which, in consequence, causes $C_D e^{2S}$ to increase.

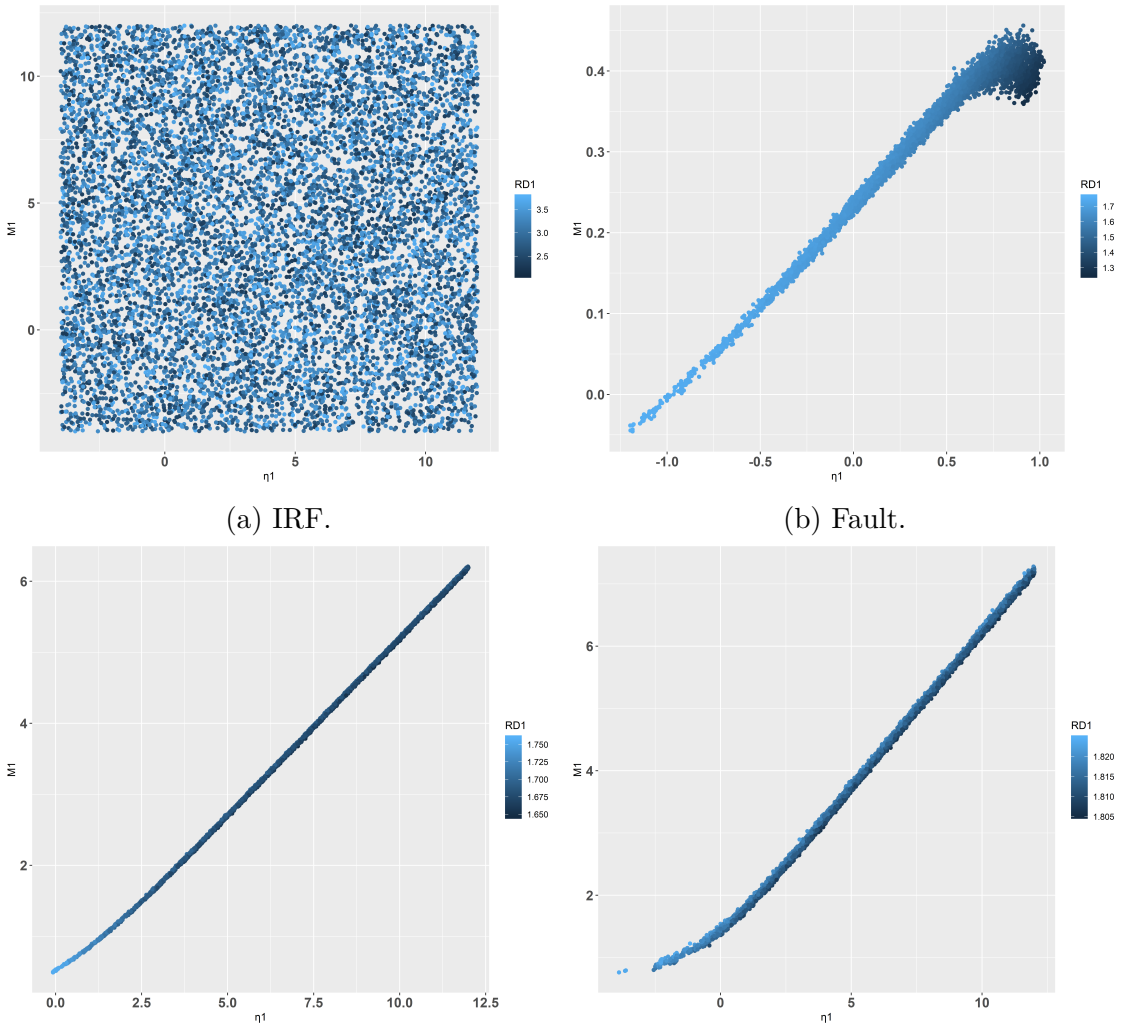
With the exception of the IRF model, where the pairwise correlation between P_M and $C_D e^{2S}$ is the only notable correlation, the late time parameter scatterplots show interesting behaviour. Specifically, the most important result is the strong correlation between η and M. These two parameters govern the late-time shape of the response curve, and suggest that some late time behaviours are being captured via parameter choices restricted along a narrow line in (η, M) space. Individually, these parameters could not be identified, however η and M are jointly quite heavily constrained.

In Figure 5.7, we present the scatterplots between the ratio parameters; η is in the x-axis, M in the y-axis, while R_D is also present in the form of different shades of blue, where dark blue corresponds to small R_D values and light blue to larger values.

For the IRF, the scatterplot suggests that there is no correlation between any of the three late time parameters. Thus, it differentiates itself from the models that comprise late time features. In the other three cases, we observe strong positive correlations between the two ratio parameters, but also negative correlation between the radius and the ratio parameters, with two distinct behaviours for each model.

In the fault case, there is strong positive correlation between the ratio parameters, which leads to a second, deviating behaviour for large η and M values, where, for a very narrow range, there is no correlation. This result agrees with our observation in Figure 5.4b, that there is a second mode, that apparently corresponds to a different behaviour for the late time parameters.

For the scatterplots of both the channel and the closed reservoir cases, we also observe interesting behaviours. In both models, most of the samples lie along a narrow line of (M, η), while the R_D parameter maintains relatively low values. We

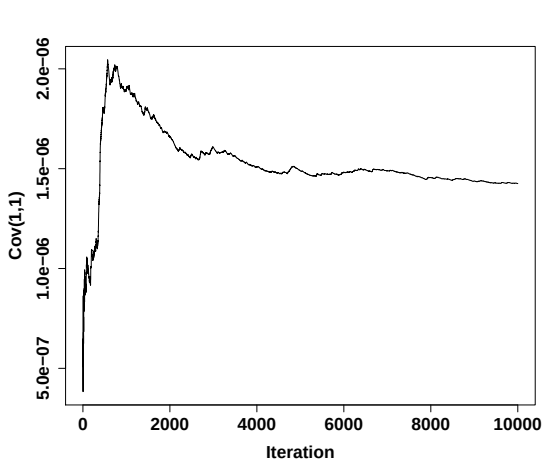


(a) IRF. (b) Fault.

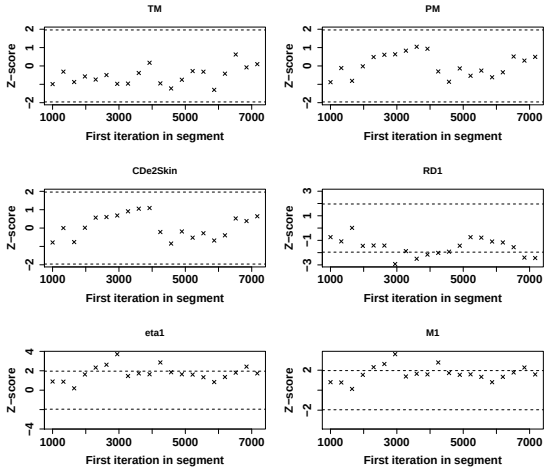
(c) Channel. (d) Closed.

Figure 5.7: Joint posterior marginals, in the form of scatterplots for the transition parameters M and η of the synthetic data sets, coloured based on R_D.

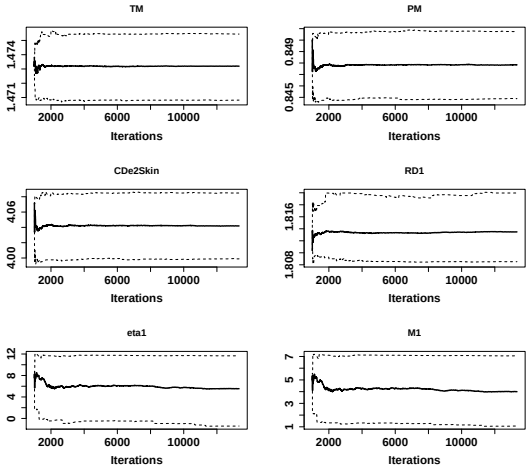
fitted a simple linear model, and the outcome was a half unit slope that passes from the point $(0, 0.2)$ for the channel, and a similar slope that passes from $(0, 1.2)$ for the closed case. This was a novel discovery that, first, differentiates the two cases from each other, and, second, shows that regardless of the value of the late time slope in the response curve, we obtain the same slope for the ratio parameters' joint density. Finally, a second action is visible for the smaller values of the ratio parameters, that appears to correspond to the largest radius parameter values, and indicates a possible second late time behaviour.



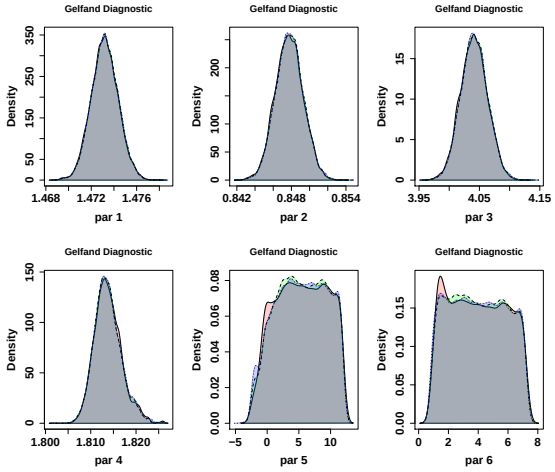
(a) Covariance element $(1, 1)$.



(b) Geweke plots.



(c) Cumulative quantile plots.



(d) Gelfand plots.

Figure 5.8: Convergence diagnostic plots for the closed data set.

5.2.2 Convergence and efficiency diagnostics

In Figure 5.8, we use a variety of tools, first described in Chapter 4, in order to assess the convergence of the chains. Since the results are similar for all four cases, we focus on the analysis of the closed reservoir model.

In Figure 5.8a we present the top left element of the empirical covariance matrix, which corresponds to the variance of the T_M parameter. The plot shows a very large initial spike, which stabilises very quickly, signifying the effectiveness of the adaptation. Small volatility is expected along the whole span, while the algorithm is making small adjustments throughout the run of the chain. The stabilisation of the adaptation, usually hints that the chain has also converged, since the covariance

matrix remains approximately constant, which indicates that the corresponding posterior marginal has settled on its final shape.

The Geweke plots for the closed case in Figure 5.8b show most of the scores within the appropriate region with only a small number of exceptions for the late time parameters, where many of them lie on the edge of the 95% confidence interval. The exceptions are more likely contributed to the complex shape of the posterior marginals, which we described in Figure 5.7d, since the test is designed for approximately normal densities.

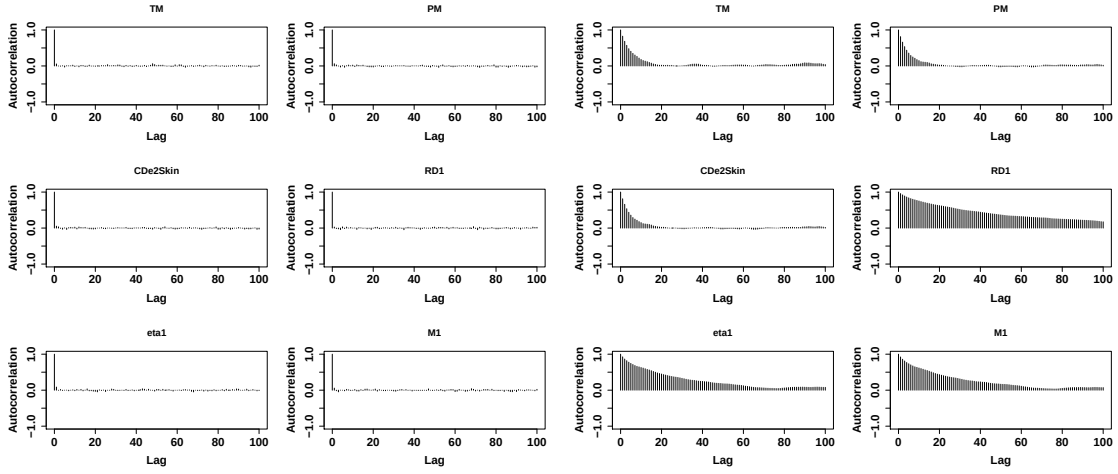
The cumulative quantile plots in Figure 5.8c show that the median, the 2.5-th and the 97.5-th percentiles of the parameters in all cases stabilise to a single value after the first $5 - 10\%$ of the samples, which, as described in Chapter 4 indicates that both the average and the variance have stabilised.

The Gelfand diagnostic plots in Figure 5.8d indicate that all the parameters have stabilised. Parameters that cannot be identified marginally, such as M and η for the channel and the closed cases, show slower signs of convergence through the small deviations between the densities of the different portions of the MCMC chain.

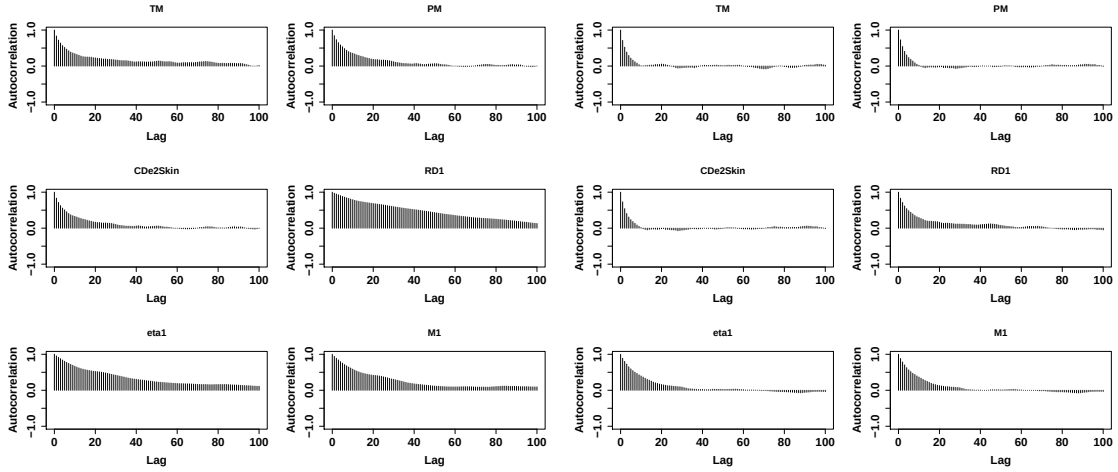
Figure 5.9 shows the autocorrelation plots for lag values up to 100. We observe that even after the thinning of 50 that we applied at the beginning of the analysis, with the exception of the IRF model, there is still strong evidence of dependence even after a large number of iterations, especially for the late time parameters. This indicates slow mixing, a possible reason for which could, again, be that the corresponding posterior marginals were identified as bimodal.

5.2.3 Uncertainty plots

The resulting response functions from the MCMC posterior samples are shown in Figure 5.10, as light blue curves; the MAP response is shown in dark blue, and the true channel response is overlaid in red.



(a) IRF. (b) Channel.

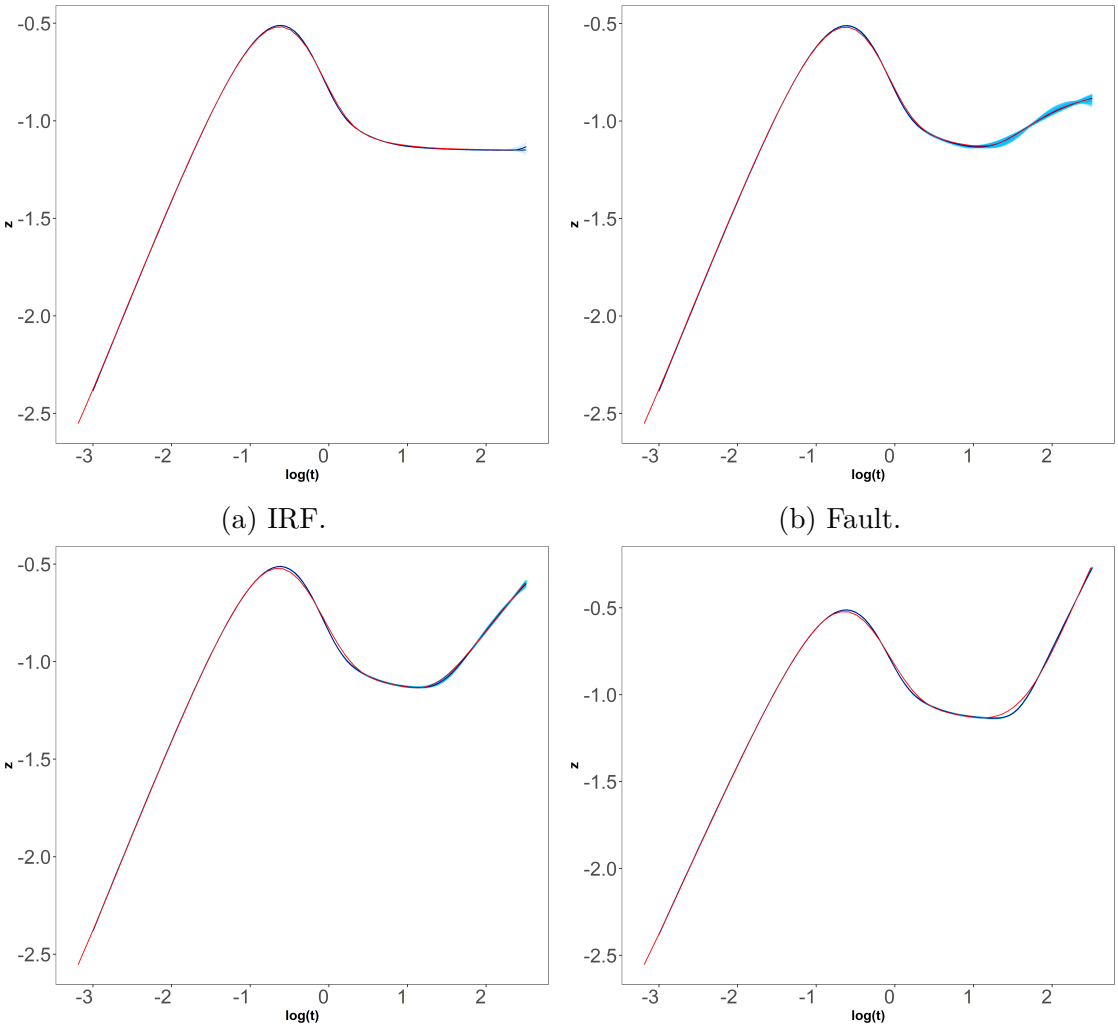


(c) Fault. (d) Closed.

Figure 5.9: Autocorrelation plots for the parameters of the synthetic data sets.

The narrow range of the posterior response function samples for all four cases indicates that the early time behaviour of the response is well identified from the data, which agrees with the results from the histograms.

Regarding the late time portion of the curves, all samples show the characteristic limiting behaviour expected of the corresponding reservoirs, thereby validating the response shapes. It is worth noting that in all cases the uncertainty of the response curves is very narrow, which, for most reservoirs, is counter-intuitive, when we take into account the large parameter uncertainty. Therefore, we can conclude that a large posterior range for the parameters does not necessarily correspond to a similar uncertainty of response shapes.



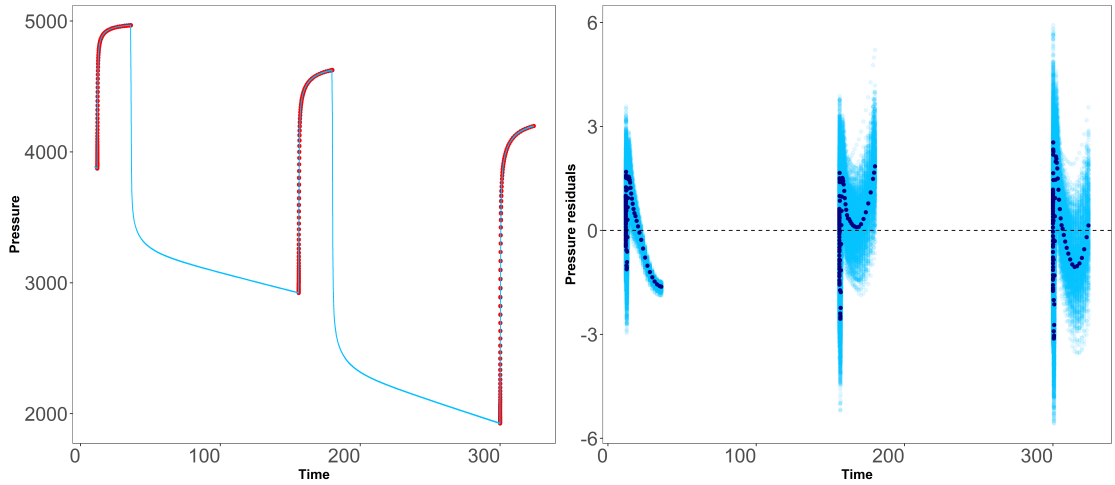
(a) IRF. (b) Fault.

(c) Channel. (d) Closed.

Figure 5.10: Uncertainty response plots for the synthetic data sets.

As with the data, the plots of the pressure histories derived from the MCMC samples are indistinguishable for the four different reservoir boundary configurations. In Figure 5.11a we present only the resulting pressure history for the closed reservoir, where the uncertainty is very narrow, and in the measurement scale it does not show any significant deviation from the true values of the underlying synthetic model. It is interesting to note that the deconvolution recovers the whole pressure history, including the unobserved portions.

In order to derive more information, we need to look at the *pressure residuals plot* in Figure 5.11b, where we can examine the residuals between the pressure points produced by the posterior samples and the corresponding true values. Most of the



(a) Pressure plot. The red dots show the true pressure, while the light blue pressure lines are sampled from the posterior of the parameters.

(b) Pressure residuals. At each time point, one light blue pressure residual point is sampled from the posterior of the parameters. The dark blue points indicate the MAP pressure residuals.

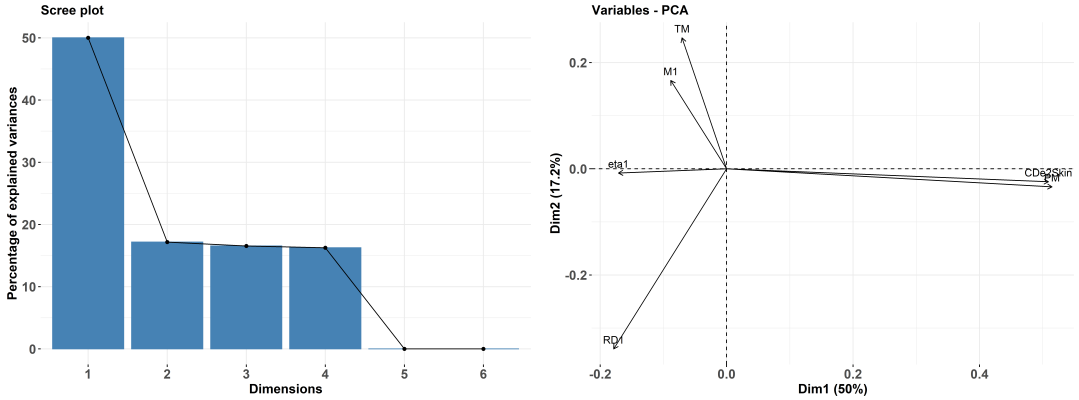
Figure 5.11: Uncertainty plots for the closed data set.

samples lie within ±3 psi of the true pressures, which is well within usual tolerances for such analyses [19].

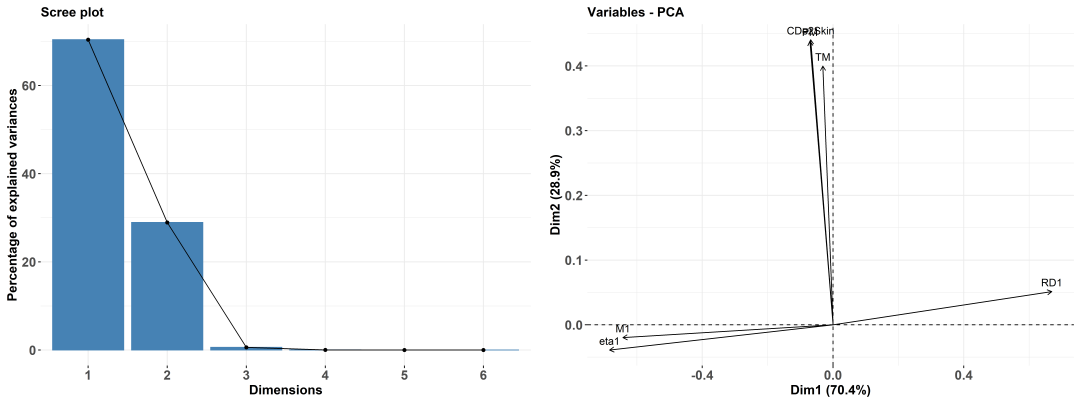
5.2.4 Principal Component Analysis

Principal Component Analysis (PCA) is a procedure that aims, among other things, to provide an insight into the the main sources of variation of a model. It does that by constructing new uncorrelated variables, called principal components, as linear combinations of the original parameters. The first principal component shows the direction of the maximal amount of variance (accounts for the largest amount of information), the second component, which is orthogonal to the first, for the second largest, and so on. It is worth noting that, by construction, the total number of principal components is equal to the number of the original parameters. In this section, we carry out PCA using the posterior correlation matrix. The goal is to detect which groups of model parameters contribute more to the system's variation.

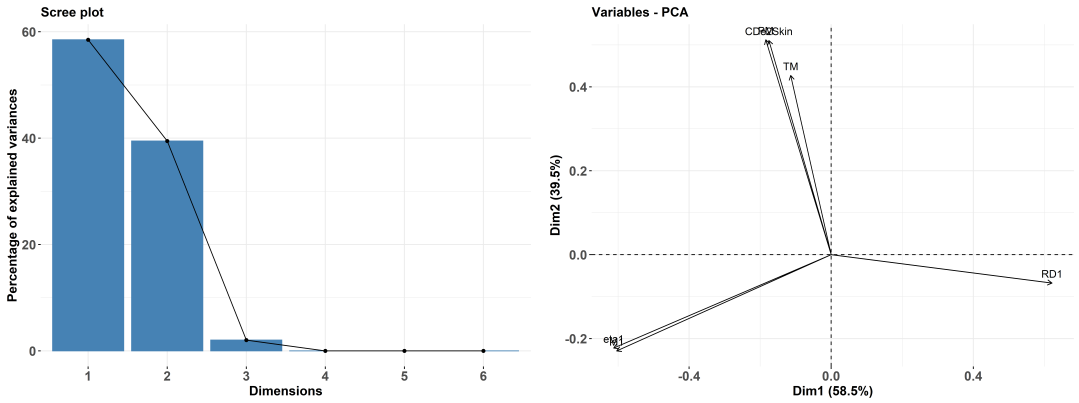
The scree plots on the left column of Figure 5.12 reveal the amount of information of each of the components. Once again, the IRF constitutes the only exception by



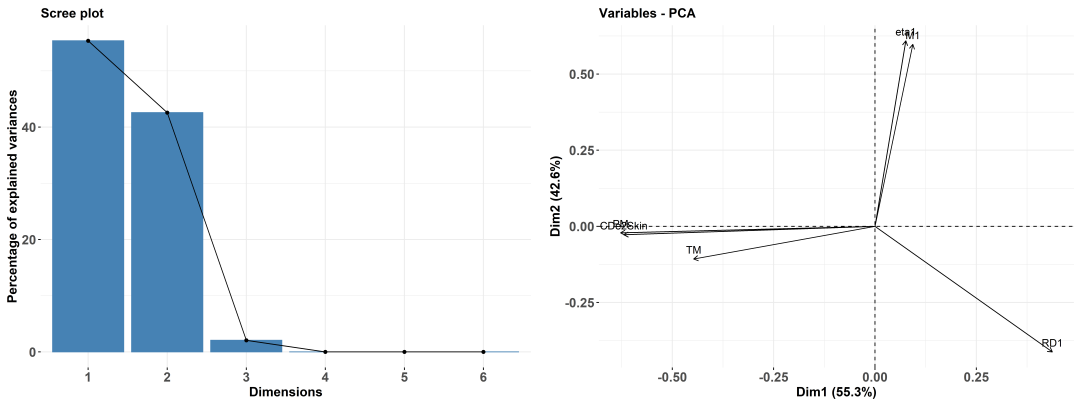
(a) IRF.



(b) Fault.



(c) Channel.



(d) Closed.

Figure 5.12: PCA results (scree plots on the left and loading plots on the right) for the synthetic data sets.

having a main principal component that contributes to $\sim 50\%$ of the total variation, and three secondary components that comprise $\sim 15\%$ of variation each. The fault, channel and closed reservoir model show similar behaviour; they all include two main principal components, a dominant and an auxiliary, that together account for more than 95% of the total system variation.

The loading plots on the right column, examine further how strongly the original parameters influence the first two principal components, and therefore, help us draw a connection between the parameters and the sources of variation in the model. The x-axis value of each vector denotes the effect of the corresponding parameter for the first component, and the y-axis value for the second principal component. Note that it is not important whether the value is positive or negative. In addition, the angles between the vectors indicate the correlation between the corresponding parameters.

For the IRF model, we observe that the P_M and $C_D e^{2S}$ parameters, which were the only ones strongly correlated as we saw in Figure 5.6a, are those dictating the main source of variation. Therefore, this model is mostly influenced by the early time shape of the response, while the effect of the late time parameters is much smaller.

For the other three models we observe similar results; there is strong indication of the existence of two dominant sources of variation (effectively two-dimensional posterior), one that comprises all three early time parameters, and one that includes the three transition parameters. In the latter, the radius' vector is in a different quadrant from the ratios' vectors, denoting the negative correlation that we examined in Figure 5.7. Thus, there is clear indication, that the total amount of model variation comes from two distinct groups of parameters that correspond to the early and the late time behaviour respectively. This relates to the scatterplots in Figure 5.6, where, with the exception of Figure 5.6a, we can clearly see the two distinct groups. We should acknowledge the fact that, although the fault and the channel loading plots show that the dominant component is influenced mainly by the late time parameters, the closed reservoir indicates the opposite behaviour. In this model, though, the scree plot suggests that the first two components contribute a very similar amount of

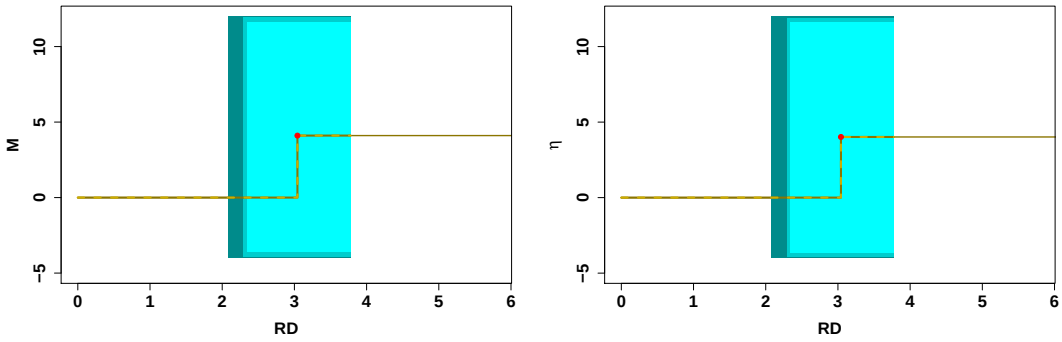
variation, and therefore the difference is not substantial.

5.2.5 Step plots

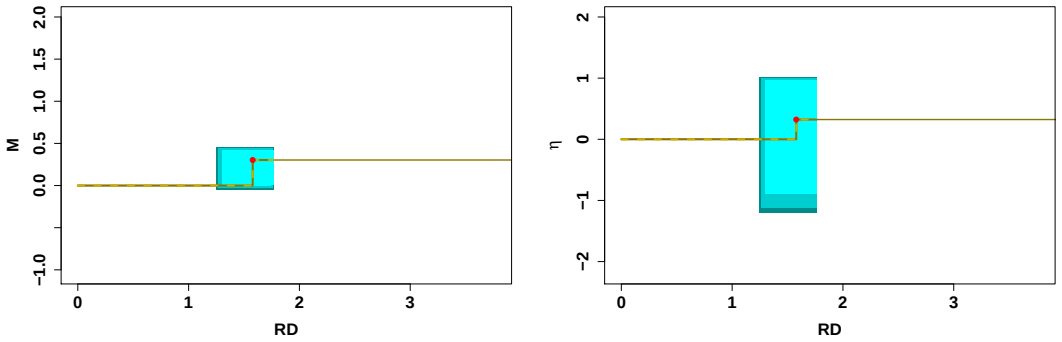
Figure 5.13 presents an alternative representation of the posterior for the reservoir parameters, showing how mobility and diffusivity change with radial distance from the wellbore. Specifically, defining $\rho_i = \log \sum_{j=1}^{i} e^{R_{Dj}} = \log\left(\frac{r_i}{r_w e^{-S}}\right)$, which corresponds to constructing the dimensionless radii from the incremental parameters, and $m_i = \sum_{j=1}^{i-1} M_j = \log\left(\frac{(k/\mu)_1}{(k/\mu)_i}\right)$, which are the mobility ratios of the first transition with respect to the i transition, we plot the piecewise constant functions taking $[\rho_{i-1}, \rho_i)$ to m_i and η_{i-1} respectively. In these plots, a solid line indicates the posterior mean, while the coloured bands depict the posterior sample range (dark blue), 99% credible posterior intervals (middle blue), and 95% credible posterior intervals (light blue). Since we use a one transition radial composite model for the synthetic results, the solid line shows only one level change, and the credible intervals correspond to rectangles. In general terms, the step changes in these parameters correspond to the features in the response function after the early-time effects have taken place. The plots on the left show the changes in mobility and the ones on the right the changes in diffusivity.

For the IRF model, the uncertainty about the distance from the wellbore that the transition takes place and the uncertainty about the magnitude of that change in terms of mobility and diffusivity are very large, especially when compared to the other models. For example, in Table 5.1 we saw that the mean of R_D is close to 3, while in all the other cases it was smaller than 2. That is another indication that the transition does not actually affect the response.

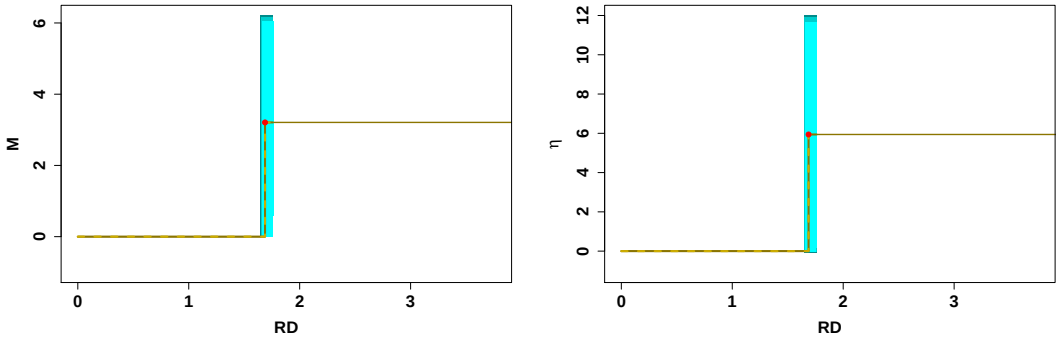
For the fault, the radius uncertainty is relatively large, while the ratio parameters' uncertainty is much narrower than the rest of the models. This suggests that for models, such as the fault, that exhibit a different stabilisation level on the response curve, the values of the mobility and diffusivity ratios are well defined, a conclusion



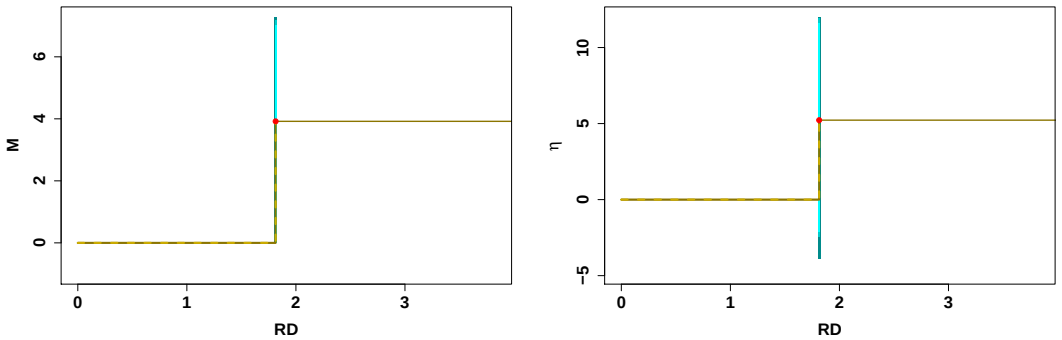
(a) IRF.



(b) Fault.



(c) Channel.



(d) Closed.

Figure 5.13: Transition parameters step plots for the synthetic data sets.

that agrees with the analysis of Figure 5.5b.

The channel and closed reservoir models show a similar behaviour: first, the radius uncertainty is very narrow, which indicates that, for models where a late time slope appears in the response, the transition distance is strictly defined. In contrast, the uncertainty is very large for both M and η, which, as we saw in Figure 5.6, happens because those parameters can be identified jointly, but are only weakly identifiable individually.

Finally, it is worth reminding that large uncertainty for some of the parameters does not correspond to large uncertainty in the late time behaviour of the corresponding response plots in Figure 5.10.

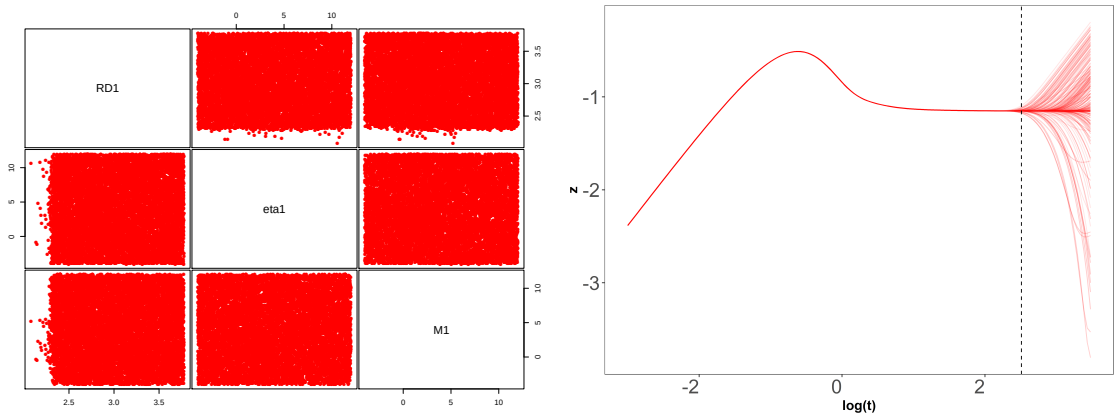
5.2.6 Extended response

Using the samples from the MCMC in order to extend the response curves can provide insight and help visualise the 'unseen' behaviour of the system, such as particular reservoir behaviours and how likely they are.

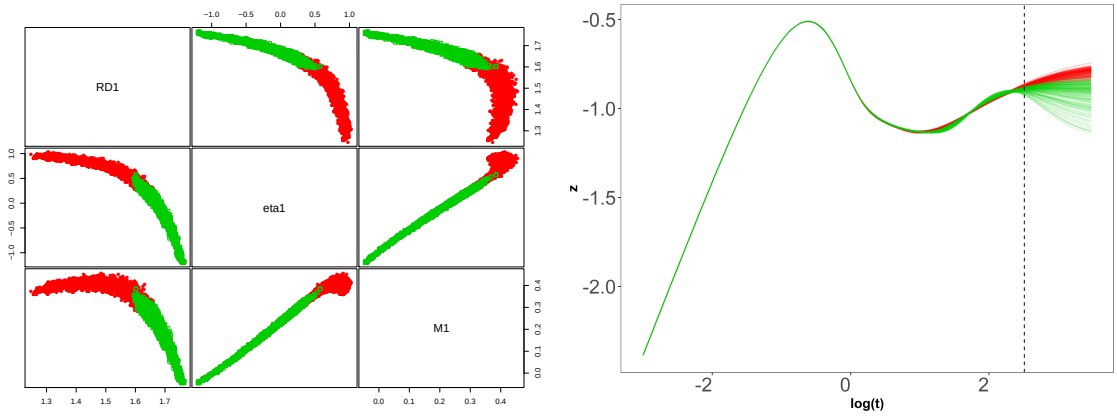
The left-hand side of Figure 5.14 shows the joint posterior marginals from Figure 5.6, that correspond to the transition parameters. The right-hand side shows the corresponding response curves, produced from the posterior MCMC samples, plotted over an extended range, beyond the duration of the well test.

For the IRF model, we observe that the transition does in fact appear on the beginning of the extended period. Since there is no information regarding the shape of the response, we observe curves covering a large range of plausible shapes. That result justifies the large uncertainty associated with the late time parameters, since, because it is unobserved, the values of the transition are irrelevant.

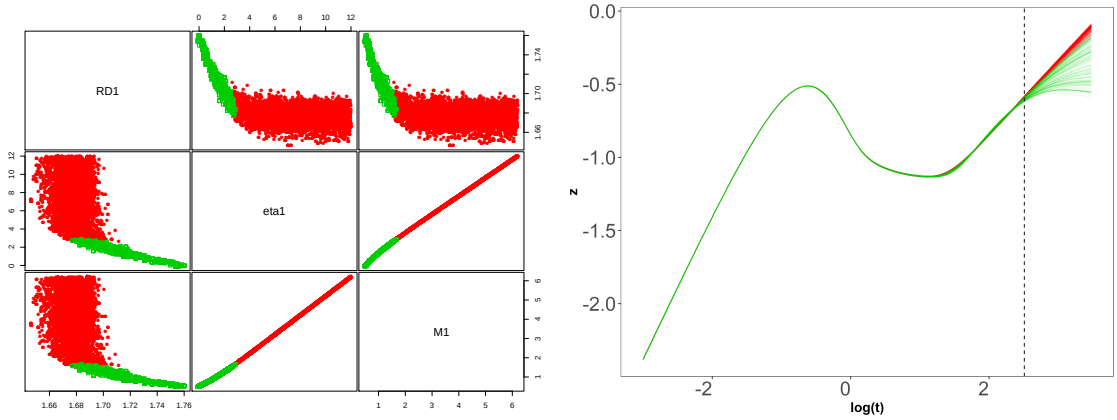
For the rest of the models there are two separate regimes, associated with the bimodal behaviour we uncovered in Figure 5.7. The red points in the scatterplot correspond to smaller values of R_D, while the green points are associated with larger values of R_D. The value of R_D used in each case to distinguish the groups was assigned



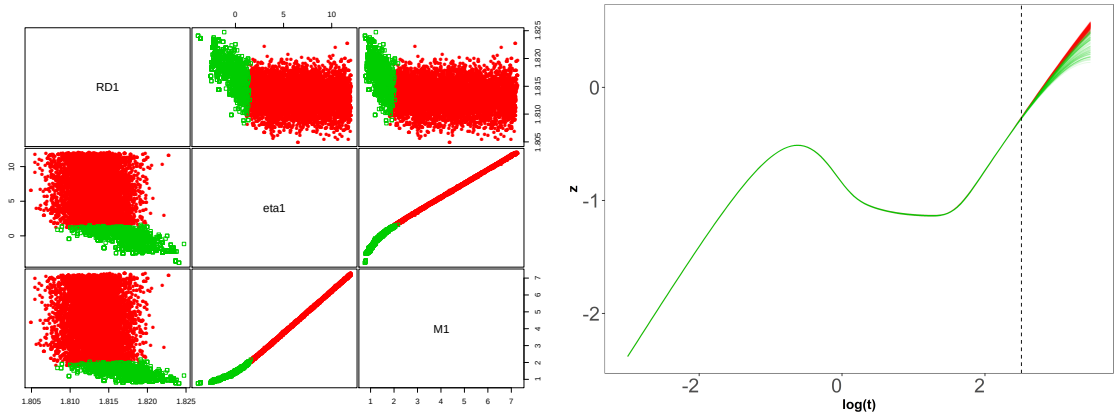
(a) IRF.



(b) Fault.



(c) Channel.



(d) Closed.

Figure 5.14: Posterior marginal scatterplots and extended response uncertainty plots, coloured based on the values of R_D.

based on where the corresponding scatterplots were visibly changing behaviour in the scatterplots. In all three remaining cases, there is a clear physical explanation for the two regimes, best explained by reference to the corresponding extended response curves.

For the fault response function, the red curves indicate the characteristic jump that leads to a second stabilisation level of a slope. In contrast, the green curves show another characteristic flow regime, commonly known as a *dual porosity system*, which was first mentioned in Subsection 1.3.1.

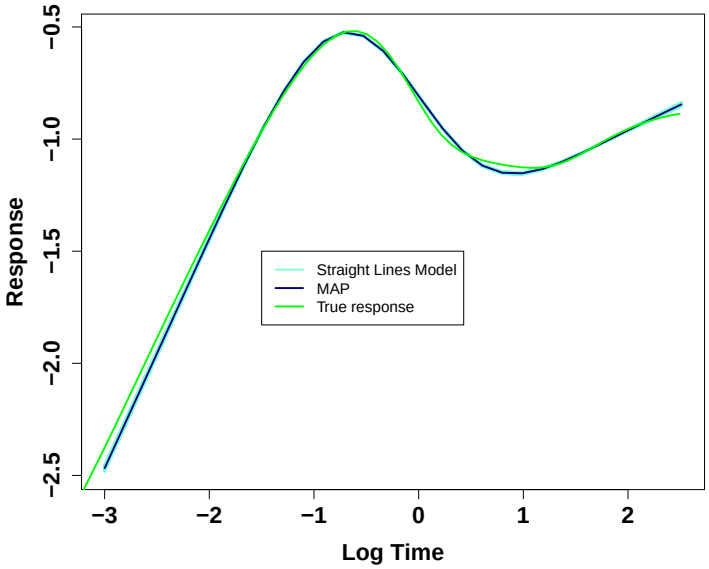
For the channel and the closed reservoir models, we observe similar results; the red curves correspond to the limiting slope for each case; half-unit slope for the former and unit slope for the latter. For the regime produced by the green curves, the limiting slope is an illusion produced by the window of observation; in reality, there is a jump in the response function to a constant higher level, and the values of the late time parameters are tightly coupled; the smaller the jump, the later the transition must take place in order to push it out of the observation window.

5.3 Results for other deconvolution methods

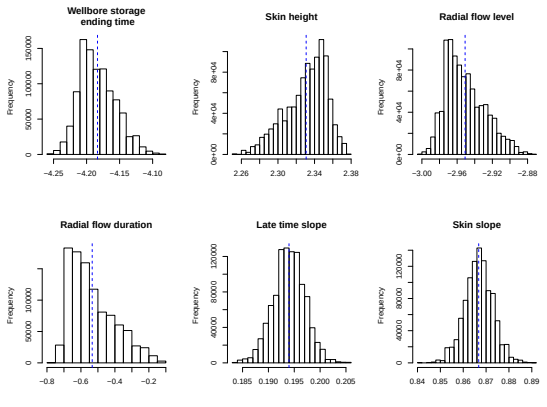
As described in the corresponding sections, the methods we examined before the radial composite in Chapter 2, are quite restrictive, and have limitations that prohibit them to cope with the complexity of field data. Nevertheless, they provide interesting results when applied to the simple synthetic data sets from this chapter.

5.3.1 Results: Straight lines

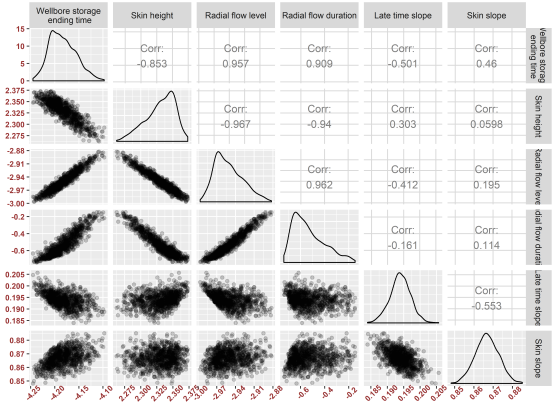
To illustrate the straight lines model from Section 2.2, we applied the Bayesian deconvolution methodology to the fault synthetic data set that we analysed in the previous sections. We ran 300,000 iterations of the AMM algorithm, and we discarded 200,000 of those as burn-in. We did not use thinning. The reason we changed the



(a) Uncertainty response plot.



(b) Posterior marginal histograms.



(c) Posterior joint marginal scatterplot.

Figure 5.15: Results of the straight lines model for the synthetic fault data set.

procedure of our analysis, was that the simplicity of the straight lines model did not require the complexity of an algorithm, such as the DEzs, in order to summarise the posterior. Thus, the much simpler, and faster AMM was adequate to perform the optimisation step, as well as the step of summarising the posterior, at once. The results were produced in less than an hour.

The results shown in Figure 5.15a indicate that the posterior samples translated to response shapes in light blue, cover only a very narrow region, which is at the most part, indistinguishable from the true response and the MAP. There are, though, some

deviations in the initial unit slope of the wellbore storage phase and the stabilisation of the radial flow regime. This is caused by the structure of the straight lines model, which, even after the smoothing treatment, is not perfectly adequate to replicate the smooth late time behaviour of the fault.

Most of the posterior marginal parameter histograms in Figure 5.15b only cover a very small area of the sample space. This is expected, since the parameters of this model are associated with points and distances of the response plot, and their uncertainty actually reflects the uncertainty of the response curves (in contrast to the radial composite model).

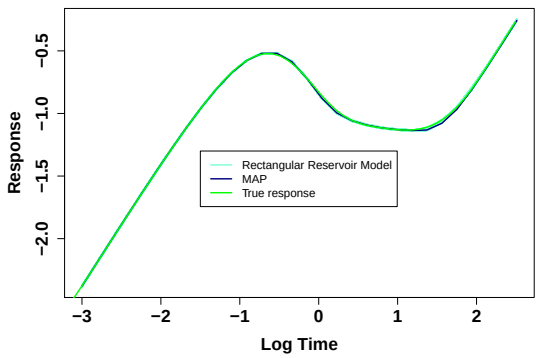
The scatterplots of the joint marginal posterior densities in Figure 5.15c suggest that there are many strong correlations among the parameters. That is, again, justified by the fact that the parameters are directly connected with the shape of the response, and therefore, changes of one parameter can usually be adjusted by changes in another. For example we can observe that similar effects can be produced, by either increasing the radial flow level, or by both decreasing the Skin height and increasing the radial flow duration.

5.3.2 Results: Rectangular reservoir

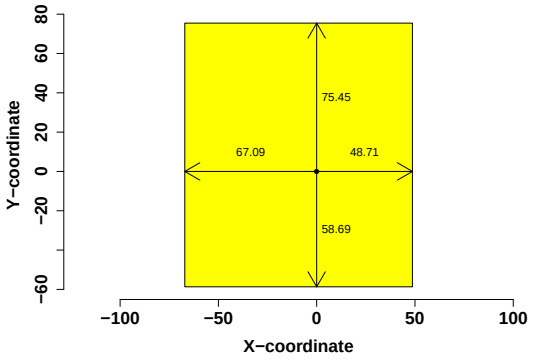
In this subsection, we present results for the synthetic closed reservoir data set, given the rectangular reservoir model, that we described in Section 2.3. We also apply the Bayesian statistical model, using the same algorithmic procedure as in the straight lines model. The results were again produced in less than an hour.

Once again, the results of the response uncertainty plot in Figure 5.16a, show that both the curves from the MCMC and the MAP, are very similar to the true response solution, with only a very small deviation when the late time unit slope begins. This is probably caused by the manner in which the early and the late time shapes of the response are connected, which was described in the analysis of Figure 2.4.

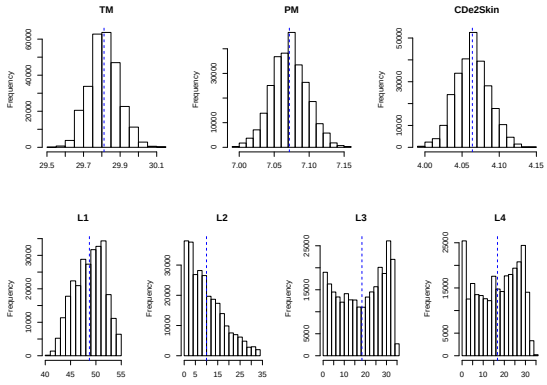
The MAP reservoir geometry in Figure 5.16b shows that all the boundaries appear



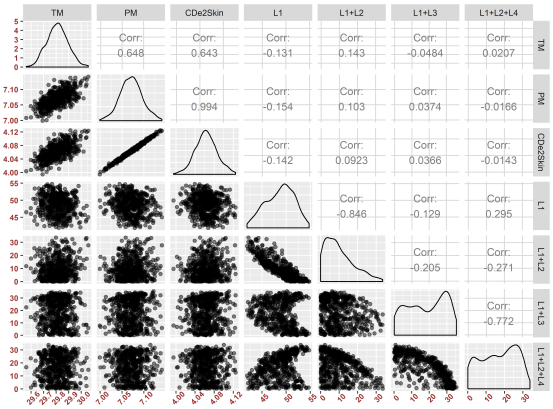
(a) Uncertainty response plot.



(b) Rectangular reservoir, constructed from the corresponding parameters.



(c) Posterior marginal histograms.



(d) Posterior joint marginal scatterplot.

Figure 5.16: Results of the rectangular reservoir model for the synthetic fault data set.

in approximately similar distances from the well, which is almost central in the reservoir. That is a relatively similar configuration to the radial closed reservoir that it is trying to replicate. It is worth noting that the precise geometry of the reservoir and the location of the wellbore are unique aspects of this model.

The early time posterior parameter histograms in Figure 5.16c are approximately Gaussian and comprise a very narrow range, suggesting that there is no much uncertainty regarding the early time response shape. The reservoir boundary distance parameters behave differently, having a wider range and more complex forms. This is likely caused by the parameterisation and the connection of the late time parameters with each other.

Finally, the joint posterior marginals in Figure 5.16d show similar correlations for

the early time parameters with the radial composite model. That includes the strong positive correlation between P_M and $C_D e^{2S}$ and the weaker positive correlations between T_M and the aforementioned parameters. This result goes back to the specification of those parameters, that comprise the same role in the diffusion equation, but for different configurations.The early time parameters are uncorrelated with the distance parameters, due to the form of the parameterisation of this model and the connection of the early time and late time portions of the curve through the 'connecting point'. Finally, the reservoir distance parameters include complicated correlations with each other, which, as with the corresponding histograms, is contributed to the parameterisation.

5.4 Comparison of the different models

The straight lines and rectangular reservoir models have very simple forms. Because of that, combining them with the Bayesian statistical model is straight forward, most of the MCMC algorithms can summarise them easily, and the results are produced very quickly, when compared to the radial composite model. In addition, they are much easier to interpret. The straight lines model comprises parameters, whose effect can directly be observed on the response plot, and the rectangular reservoir model can always be associated with the corresponding reservoir plot, providing an effortless interpretation of the system.

Their main disadvantage is their lack of flexibility. Even when applied to the simple synthetic data sets, we can observe deviations from the true response. This problem becomes more clear when they are applied to more complex systems, where they have to encounter subsequent response features. A solution would be to add more parameters, but that would make the models more complicated and less interpretable. In addition, it is evident from the analysis that there is a deep connection of all the radial composite model's parameters with the physical aspects of the problem, which makes it far superior to the other models.

5.5 Summary

In this chapter, we validated the radial composite model using four synthetic data sets with known solutions. We analysed the results from the MCMC, we assessed convergence and efficiency, we introduced new visualising tools, such as the uncertainty plots and the step plots, we performed principal component analysis, and we investigated the results of the extended response. We also showed results from our other models, we compared them with the radial composite model and discussed their limitations.

In the next chapter, we are going to illustrate and study the effect of the radial composite model on real data.

Chapter 6

Analysis of field data

In this chapter we apply our Bayesian approach to deconvolution, using the response models of Chapter 2, to two real data sets, derived from an oil and a gas reservoir respectively. Because of the nature of real data, we face some additional challenges, which we did not encounter in the previous chapter: first, the inclusion of an unknown level of noise. Second, the fact that the true solution, including the boundary conditions and the reservoir behaviour, is also unknown. Finally, we need to take into account the source of the data, and possibly make different prior choices, depending on whether they are derived from an oil or a gas reservoir. The oil field data set is from [19], while the gas field data set from [35]. The purpose of this Chapter is to examine the effect of applying the multi-region radial composite model to real examples, try to tackle the challenges that our method encounters, and attempt to make a connections with results from Chapter 5.

In each data set we examine the MCMC output, the uncertainty plots, and some additional visualising tools. We also show results for the Laplace approximation and the buildup derivative, and we compare the different algorithms from Chapter 4.

6.1 Example: field data from an oil reservoir

The first application concerns an oil field data set that comprises 2,273 pressure observations, measured in non-equidistant time points and 22 production rates over approximately 150 hours, as shown in Figure 6.1.

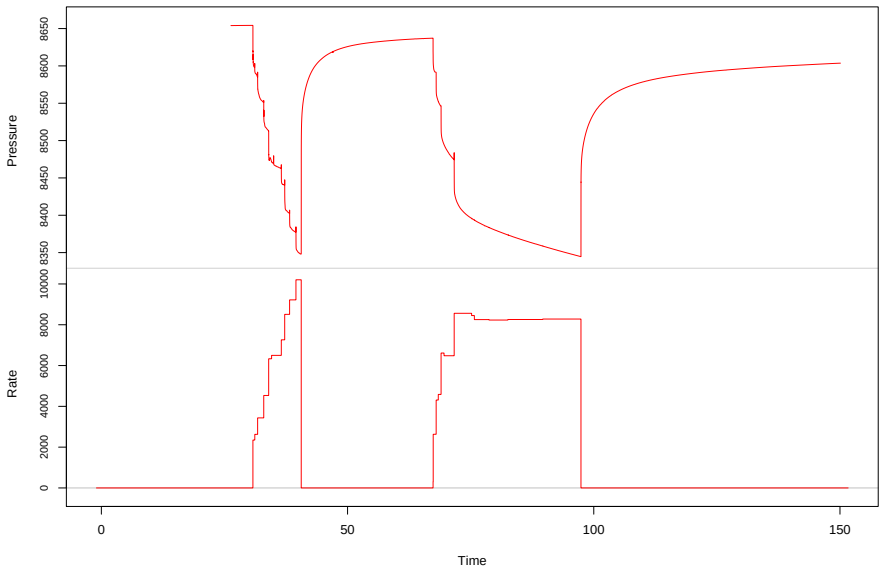


Figure 6.1: Field oil rate and pressure measurements.

Since the complexity of the system is not known, we consider four possible radial composite models from Section 2.5, with 1–4 transitions (2–5 regions respectively). The true rates and initial pressure are unknown, therefore, we use the DEzs algorithm for $\boldsymbol{\phi}$ and σ_p (marginal model), and we sample from the conditional Gaussian the $\tilde{\boldsymbol{q}}$ and $\tilde{p}_0$, as described in Chapter 3. σ_q is fixed to $0.05 \max(q)$ and σ_{p_0} is fixed to 10 in accordance with [19].

For the response parameters, we consider the prior set for oil reservoirs, described in Chapter 3, which consists of: $T_M \sim \mathrm{N}(2, 0.2^2)$, $P_M \sim \mathrm{N}(1.5, 0.2^2)$, $C_D e^{2S} \sim \mathrm{Ga}(1, 0.2)$, $R_{Di} \sim \mathrm{N}(2, 1^2)$, for $i > 1$, $M_i \sim \mathrm{N}(0, 1^2)$ and $\eta_i \sim \mathrm{N}(0, 1^2)$.

Finally, we use the DEzs algorithm with three parallel chains, for 1,666,667 iterations each, which, after thinning of 50 and discarding the first 10,000 iterations as burn-in, leaves 70,002 samples for inference.

6.1.1 MCMC output

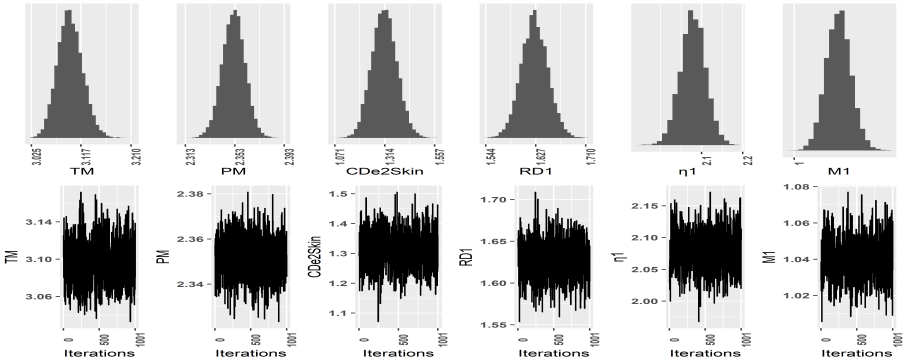
The traces and one dimensional posterior marginals in the form of histograms for the first six parameters for each of the four models are shown in Figure 6.2. The corresponding summary statistics are collected and presented in Table 6.1. We observe that all the histograms indicate approximately Gaussian posterior distributions. The traces appear to indicate convergence, but it is clear that the more transitions (and parameters) in the radial model, the slower the convergence.

The summary statistics suggest that the variance, and therefore the uncertainty, for the early time parameters is approximately doubled for the three and four transition models, even though they are still small relative to the priors, which indicates that there is a possible change in the response behaviour when the third transition is introduced. On the same note, the $C_D e^{2S}$ parameter exhibits an increase in value with all transitions, but especially with the addition of the third transition, which reinforces the previous point.

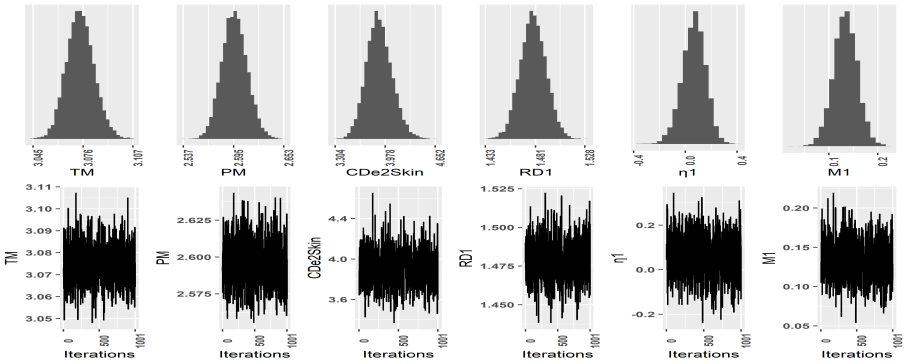
Additionally, the parameters that correspond to the radii increments are reasonably large, with the possible exception of the last increment in the four transition model, suggesting that, at least in the other three cases, there is no degeneracy arising from a trivial region. The tables make it difficult to interpret the late time parameter behaviour using only the statistics, and, thus, we will attempt to extract more information using visualisation tools later in this chapter.

Overlaying the posterior histograms with the corresponding priors in Figure 6.3 shows that the marginals have very narrow posterior range, in comparison to the much broader priors, indicating that uncertainty on those parameters has been substantially reduced.

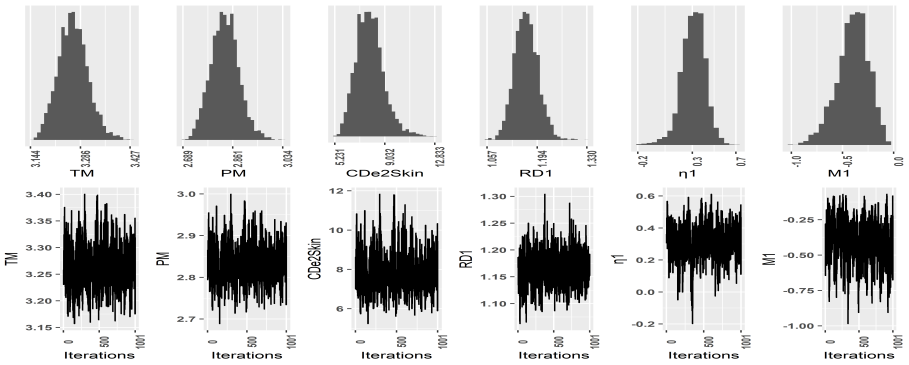
It is worth noting that even though the posterior marginals for the early time parameters tend to considerably high values relative to their priors, the specification of the latter plays an important role for those parameters, since it prohibits non-physical values and degenerate results, as will become more clear later in this chapter.



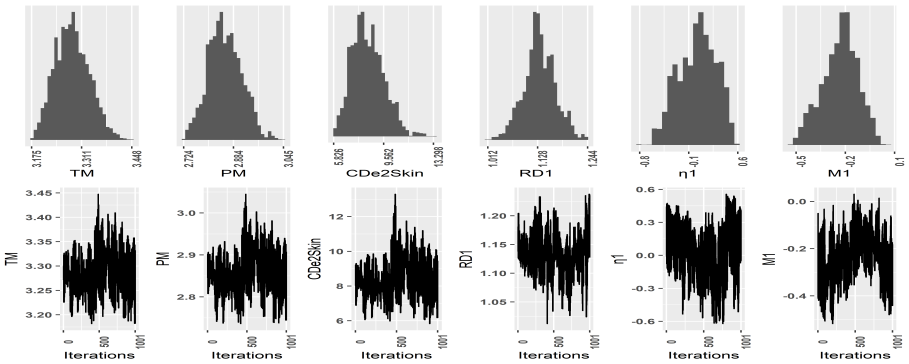
(a) 1 transition.



(b) 2 transitions.



(c) 3 transitions.

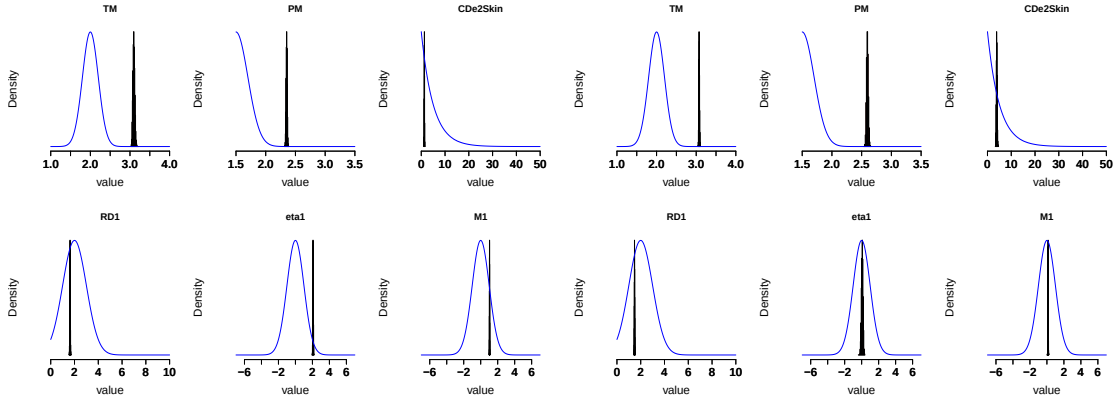


(d) 4 transitions.

Figure 6.2: Marginal posterior densities in the form of histograms and traces for the first six response parameters of the oil data set.

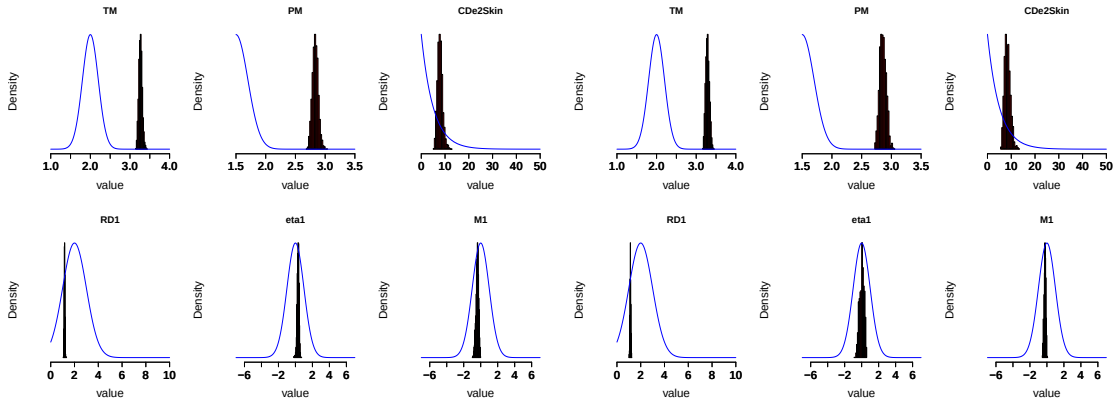
1 transition	min	25%	mean	75%	max	sd
T_M	3.02	3.08	3.10	3.11	3.21	0.02
P_M	2.31	2.35	2.35	2.36	2.39	0.01
$C_D e^{2S}$	1.07	1.27	1.31	1.35	1.56	0.06
R_{D1}	1.54	1.61	1.63	1.64	1.71	0.02
η_1	1.95	2.06	2.08	2.10	2.19	0.03
M_1	1.00	1.03	1.04	1.05	1.08	0.01
2 transitions	min	25%	mean	75%	max	sd
T_M	3.05	3.07	3.07	3.08	3.11	0.01
P_M	2.54	2.59	2.60	2.61	2.65	0.01
$C_D e^{2S}$	3.30	3.78	3.90	4.01	4.65	0.17
R_{D1}	1.43	1.47	1.48	1.49	1.53	0.01
R_{D2}	2.12	2.23	2.25	2.27	2.40	0.03
η_1	-0.37	0.01	0.07	0.13	0.39	0.09
η_2	0.41	1.89	2.26	2.62	4.47	0.54
M_1	0.02	0.12	0.13	0.15	0.22	0.02
M_2	1.28	1.80	1.97	2.12	3.04	0.23
3 transitions	min	25%	mean	75%	max	sd
T_M	3.14	3.24	3.27	3.29	3.43	0.04
P_M	2.69	2.80	2.83	2.87	3.03	0.05
$C_D e^{2S}$	5.23	7.14	7.90	8.51	12.83	1.05
R_{D1}	1.06	1.14	1.16	1.19	1.33	0.03
R_{D2}	0.21	0.66	0.78	0.90	1.32	0.19
R_{D3}	1.89	2.00	2.03	2.06	2.25	0.05
η_1	-0.20	0.24	0.31	0.39	0.70	0.11
η_2	0.44	1.02	1.11	1.22	1.55	0.15
η_3	-3.76	-0.86	-0.18	0.52	3.04	0.98
M_1	-0.99	-0.49	-0.40	-0.29	-0.03	0.16
M_2	0.56	0.80	0.90	0.98	1.41	0.14
M_3	0.76	1.24	1.53	1.75	3.13	0.38
4 transitions	min	25%	mean	75%	max	sd
T_M	3.18	3.25	3.29	3.32	3.45	0.05
P_M	2.72	2.82	2.86	2.89	3.05	0.05
$C_D e^{2S}$	5.83	7.53	8.40	9.20	13.30	1.21
R_{D1}	1.01	1.11	1.13	1.16	1.24	0.04
R_{D2}	0.43	1.06	1.23	1.45	1.82	0.30
R_{D3}	1.94	2.12	2.24	2.36	2.64	0.15
R_{D4}	-0.54	1.96	2.30	2.71	3.96	0.59
η_1	-0.83	-0.17	0.02	0.22	0.57	0.27
η_2	-0.20	0.39	0.65	0.88	1.42	0.33
η_3	-2.45	-1.01	-0.78	-0.58	0.75	0.37
η_4	-2.88	-0.08	0.43	1.01	2.95	0.84
M_1	-0.55	-0.30	-0.23	-0.15	0.05	0.11
M_2	0.47	0.64	0.71	0.75	1.03	0.10
M_3	-0.05	1.44	1.68	2.00	2.92	0.50
M_4	-1.52	-0.37	0.01	0.38	2.03	0.58

Table 6.1: Summary statistics for the posteriors distributions of $\boldsymbol{\phi}$ for the one, two, three and four transition models.



(a) 1 transition

(b) 2 transitions



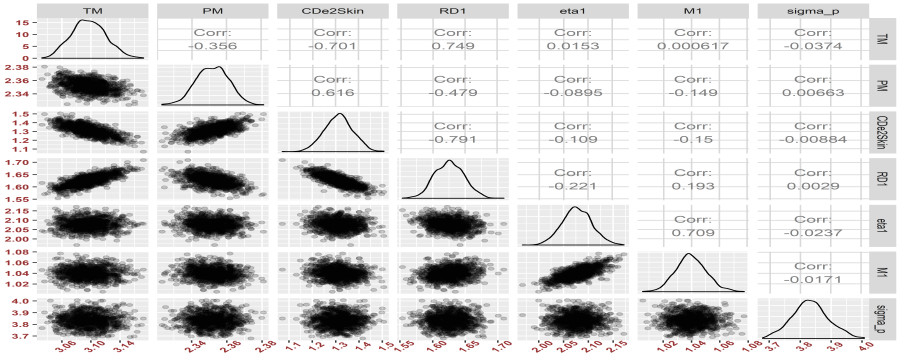
(c) 3 transitions

(d) 4 transitions

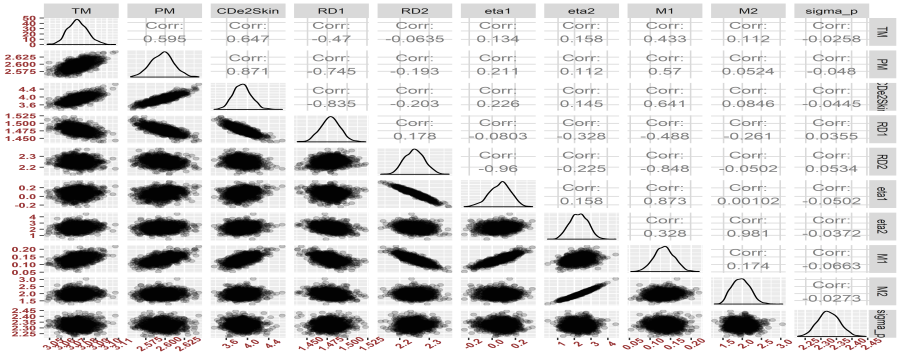
Figure 6.3: Posterior marginal histograms from the MCMC and corresponding priors for the first 6 response parameters.

Figure 6.4 shows the two-dimensional posterior marginal distributions as scatterplots. The results show a more complicated correlation structure when compared to the synthetic cases. The early time structure for the one transition model seems significantly different from the rest, with the positive correlation between P_M and $C_D e^{2S}$ being the only common feature. For the rest of the models, and especially the three and four transition models, the strong positive correlation for the early-time parameters extends to T_M, which implies that the horizontal shift of the response has a much more significant role in compensating for the effects of P_M and $C_D e^{2S}$, and thus, suggests possible degeneracies among those parameters.

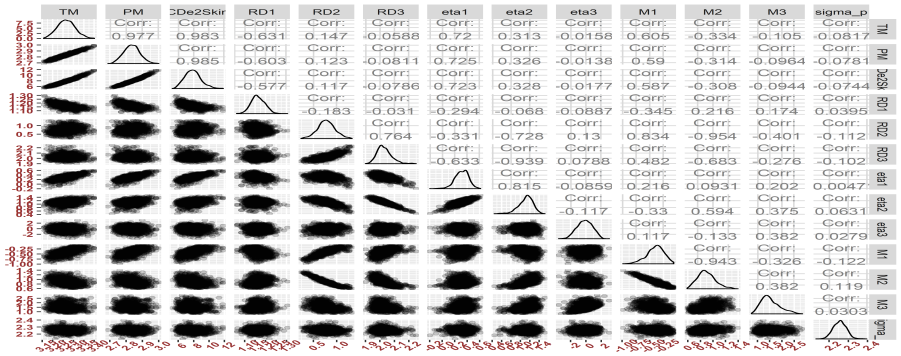
There are also numerous late-time parameter correlations, some of which are reasonably interpretable. For example, in the two transition model, we can see a strong positive correlation between η_2 and M_2, for a significant range of those parameters,



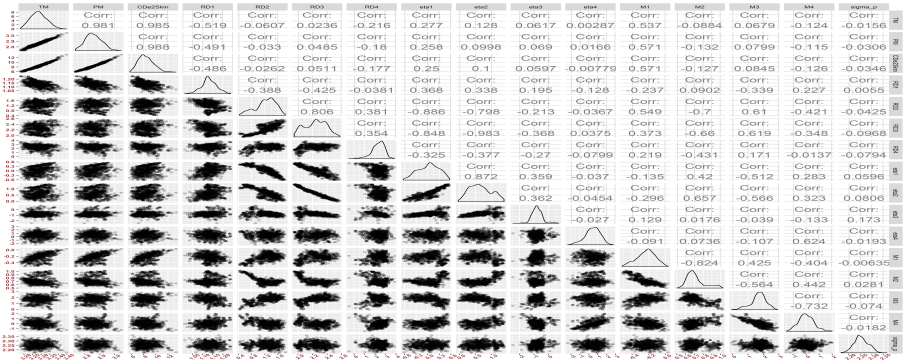
(a) 1 transition.



(b) 2 transitions.



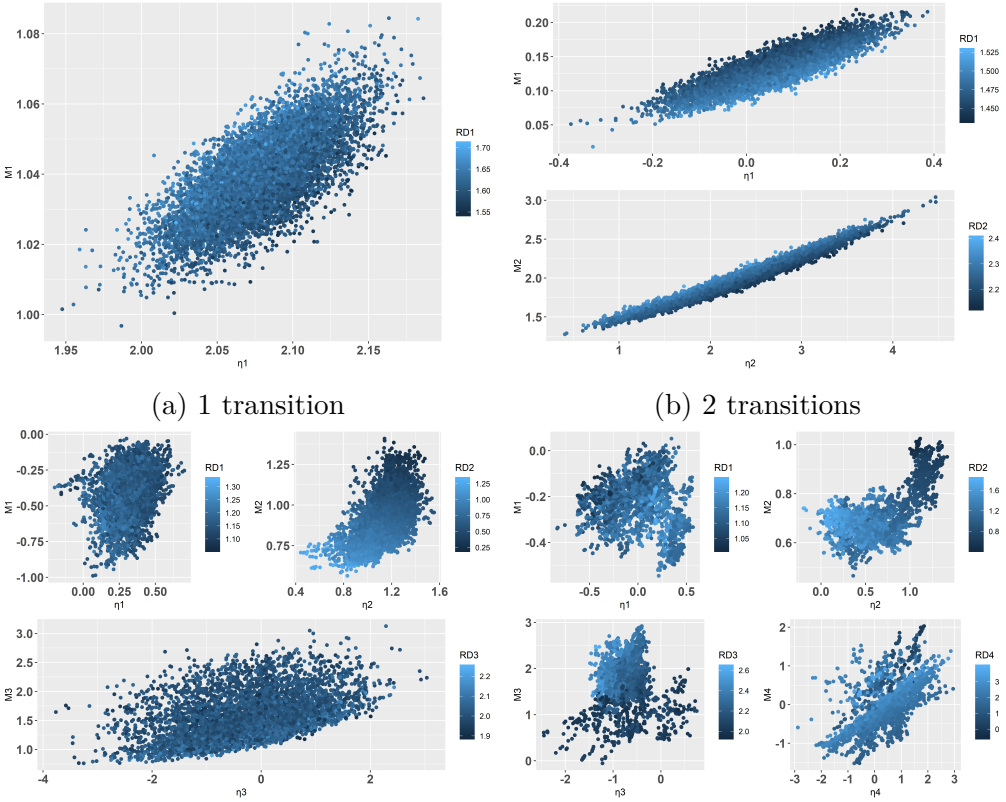
(c) 3 transitions.



(d) 4 transitions.

Figure 6.4: Joint posterior marginals, in the form of scatterplots and Pearson correlation values (with the exception of the 4 transitions model) for the response parameters and σ_p.

which suggests a late time slope, similar to the ones we examined in the corresponding synthetic cases. In contrast, for the three transition model we detect negative correlations between M_2 and the parameters M_1 and R_{D2}, which indicate that larger in magnitude second transitions appear earlier and correspond to smaller first transitions.



(a) 1 transition (b) 2 transitions

(c) 3 transitions (d) 4 transitions

Figure 6.5: Joint posterior marginals, in the form of scatterplots for the transition parameters, coloured based on R_D.

In Figure 6.5 we present the scatterplots between the η and M parameters for each transition in each model. The R_D parameter is also present in the form of shades of blue, where a lighter shade corresponds to a higher value.

Our first observation is that the addition of new regions makes the correlations between the parameters more complicated. The single scatterplot for the one transition model does not correspond to any of the common reservoir boundary behaviours that we examined in the previous chapter, but the very narrow posterior marginal range of both the η and M parameters suggests that the model is very confident

about the result. This could indicate that the late time behaviour does not correspond to a slope, but could also be a sign that this model might be too restrictive for this data set.

For the two transition model, we observe linear relationships for each of the sets of the transition parameters. As we saw in the previous chapter, this might be an indicator of a jump or a slope. It is worth noting that the last transition in this model is very close to the half-unit slope that passes from the point $(0, 1.2)$, which we identified in Chapter 5 as the behaviour of a closed reservoir.

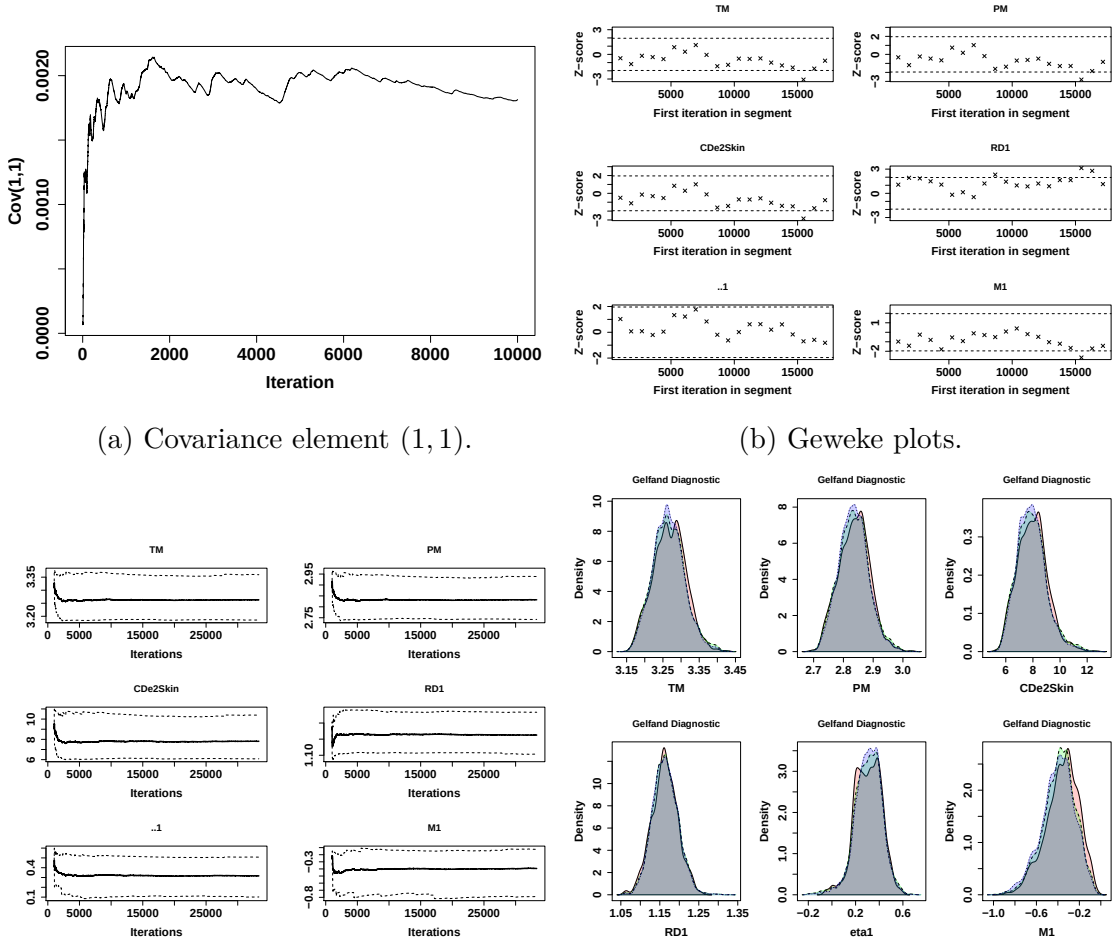
For the three transition model, a direct interpretation is harder. We can see that the radius has an important role on the second transition, which is inversely proportional to the size of the transition, showing again that similar results can be produced by different configurations of the parameters of the penultimate region. Regarding the last transition, even though the scatterplot does not strictly indicate linear correlation, the result is not far from the aforementioned closed reservoir configuration.

Finally, all the four transition model's scatterplots have very complicated forms, indicating nonlinear behaviour, thus prohibiting easy interpretation.

6.1.2 Convergence and efficiency diagnostics

Regarding the chain convergence, we indicatively present convergence and efficiency diagnostics for selected parameters of the three transition model in Figure 6.6.

The top left element of the empirical covariance matrix in Figure 6.6a shows an approximate final stabilisation level, which indicates that the adaptation has stabilised. Comparing this with the synthetic results in the previous chapter, and taking into account that this chain has more than twice the size, suggests that the convergence is significantly slower. The addition of the extra regions, and the complexity of the real data, are the main factors that contribute to this result.



(a) Covariance element $(1, 1)$.

(b) Geweke plots.

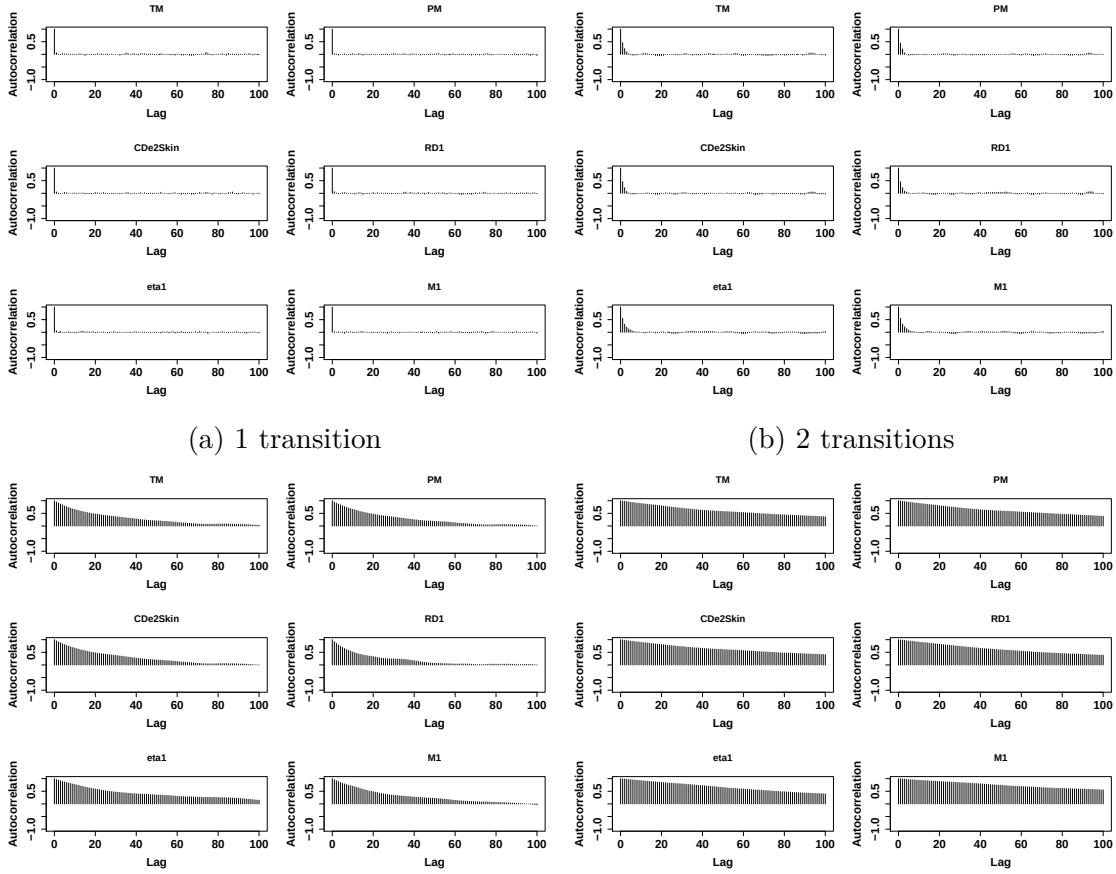
(c) Cumulative quantile plots.

(d) Gelfand plots.

Figure 6.6: Convergence diagnostic plots for the three transition model.

The Geweke plots in Figure 6.6b are, with a very few exemptions, within the acceptance region. A possible explanation for the ones outside could be the poor efficiency of the algorithm. Another interesting point is that the late time parameter Geweke plots show arguably better signs of convergence than the corresponding synthetic results in Figure 5.8b. The reason is that the marginal posteriors of the three transition model are unimodal and closer to approximate Gaussian distributions than the corresponding synthetic closed reservoir parameters.

Similarly, the cumulative plots in Figure 6.6c show a median stabilisation for all the parameters from the early iterations. Regarding the variance, which is associated with the other percentiles, we observe quick convergence for most of the parameters, with some exceptions like in M_1. This can happen sometimes in the case of heavy-



(a) 1 transition

(b) 2 transitions

(c) 3 transitions

(d) 4 transitions

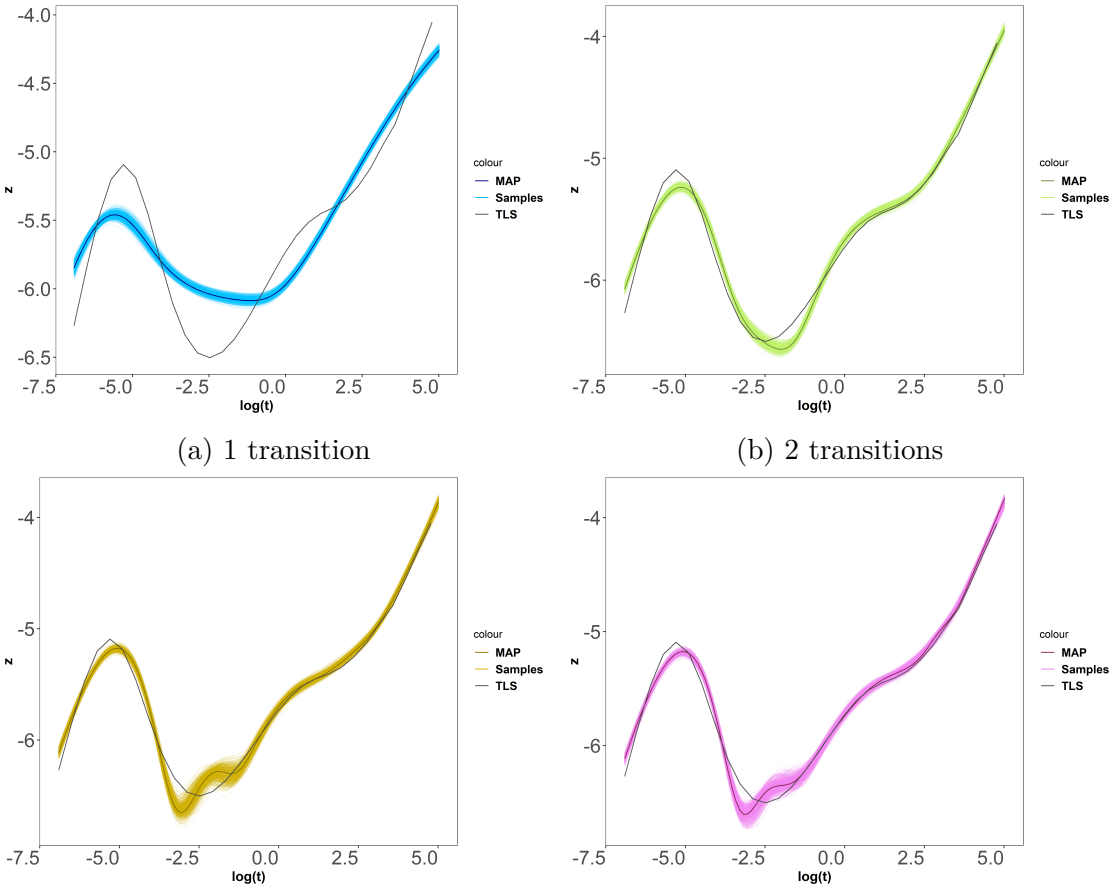
Figure 6.7: Autocorrelation plots for the first six response parameters.

tailed marginal posterior distributions.

Finally, the densities of the Gelfand plots in Figure 6.6d are fairly similar, therefore they do not provide enough evidence to dispute the fact that they are derived from converged samples.

In Figure 6.7 the autocorrelation, after thinning, for the one and two transition models shows little dependence on previous iteration, which suggests that the algorithm is very efficient. The complexity of the new parameters and their complicated correlation structure for the three and especially the four transition models, on the other hand, shows an increasing number of dependent iterations, and thus poor algorithmic efficiency. In some occasions that may be an indicator of over-parameterisation, or an artefact of the complexity of the system.

Uncertainty plots



(a) 1 transition

(b) 2 transitions

(c) 3 transitions

(d) 4 transitions

Figure 6.8: Response uncertainty plots.

In Figure 6.8 we compare the uncertainty plots for the four response models, including the curve from the MAP, in darker colour, and the curve obtained from the Total Least Squares (TLS) method with the piecewise linear response model from Subsection 1.4.2, in black.

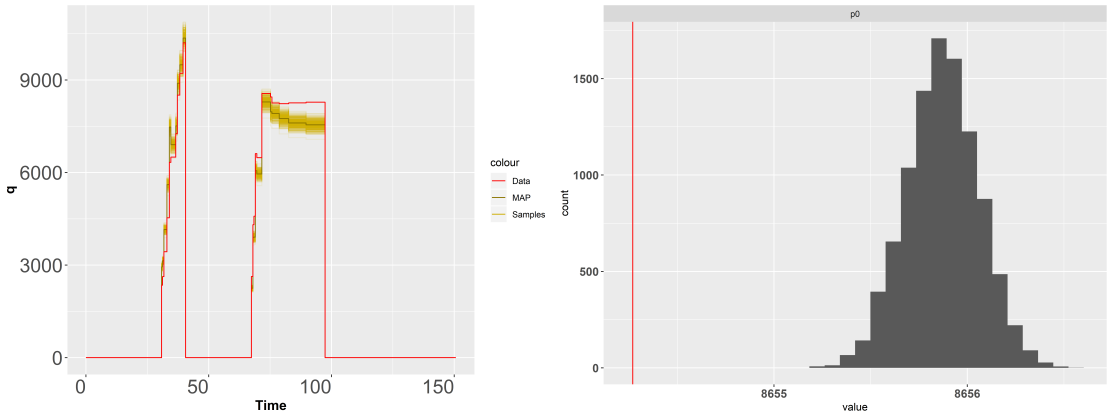
First, we observe that the one transition model looks vastly different from the rest, and more simplistic than the TLS curve. That is an indication of under-parameterisation, perhaps suggesting that one transition is not sufficient to properly capture all the response features.

The two transition model is probably the closest to the TLS result, agreeing with it at late times, and only slight deviating from it at early and middle times. Those deviations could be an artefact of the TLS method's parameterisation (piecewise

linear function with a curvature penalty), which could be quite different from the actual solution, whereas the radial composite model's nature allows only physical solutions.

The main difference of the three transition models' response, when we compare it with the two transition model's response and the TLS curve, is that we see an extra feature at approximately -2 $\log hrs$, where we can observe an additional localised deviation from the radial flow level. Whether this is a genuine feature or overfitting of the model, will be examined later in this chapter. On the contrary, the addition of the fourth transition in the last model produces a very similar response shape to the three transitions model, and the fact that there is no significant change perhaps indicates that the four transitions model is over-parameterised.

A final observation is that the two, three and four transition models show the same late time behaviour, which corresponds to a sequence of a large jump that leads to a stabilisation and a late time unit slope. That reinforces the belief that the one transition model is too simplistic to appropriately model this data set.



(a) Rate uncertainty plot. (b) $\tilde{p}_0$ uncertainty plot.

Figure 6.9: Uncertainty plots for the three transition model.

The posterior marginals for the data parameters are very similar for all four models. Due to the likelihood structure and the form of the deconvolution formula, those parameters are not affected much by the complexity of the response shape, but rather by its vertical position, which, as the response uncertainty plots suggest, is

similar in all cases. Thus, we again focus on the three transition model and we show the results in Figure 6.9.

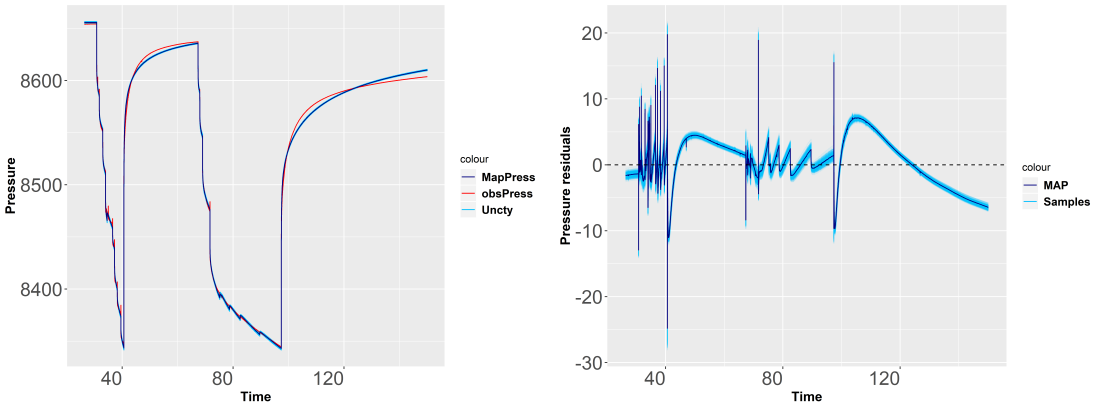
Regarding the rate samples in Figure 6.9a, we can see significant deviation from the measured values, although the general structure remains. This is consistent with the known scale of errors in these measurements, and very common when dealing with real data sets.

The $\tilde{p}_0$ marginal in Figure 6.9b suggests that the true initial pressure is larger than the corresponding measurement, but the difference, which is around 2psi, is within the expected range.

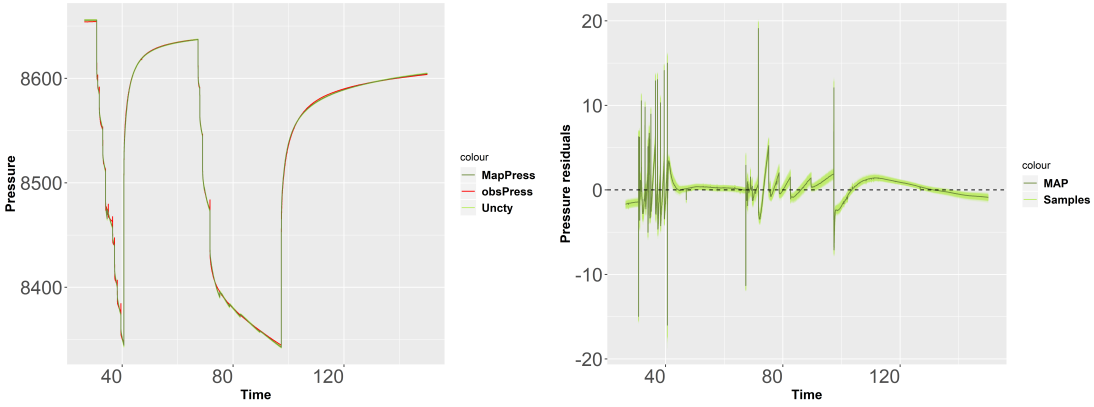
The plots on the left-hand side of Figure 6.10, include the posterior pressure history, produced by the convolution of the posterior samples of the response and rates, along with the pressure measurements in red. On the right-hand side we show the corresponding pressure residuals.

Similarly to the synthetic data results, the pressure uncertainty plots do not show any difference between the models and the data on the measurement scale, except from the one transition model, where there is significant deviation. That is another indication that this model is under-parameterised.

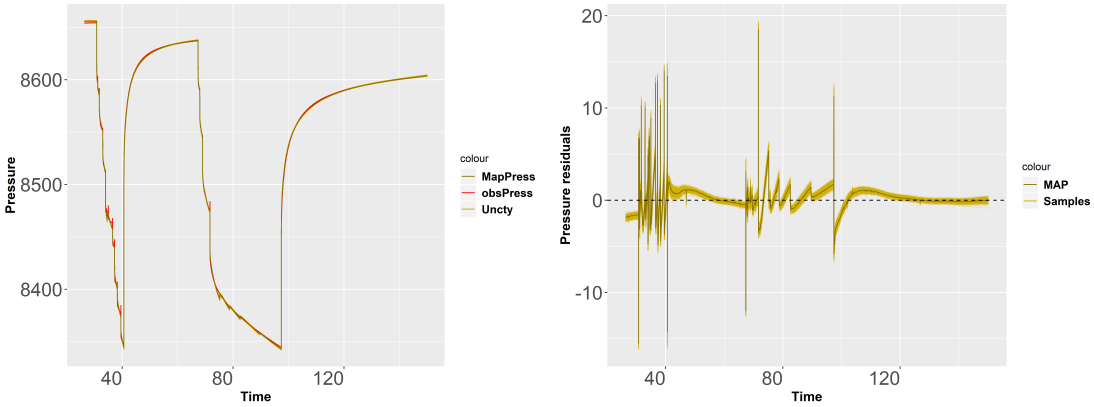
Comparison of the pressure residuals on the other hand, shows that, for the rest of the models, the variance σ_p lies within a range of $2 - 2.5$ psi, with the exception of some peaks that appear when there is a sudden change in production. That is a reasonable pressure measurement error for an oil field reservoir. We can also observe that the residuals decrease in magnitude as we add transitions (from 1 to 3), since the extra flexibility allows for a better fit. The residuals of the four transition model look similar to the equivalent plot of the three transition model, which is a consequence of the similarity of the corresponding response functions.



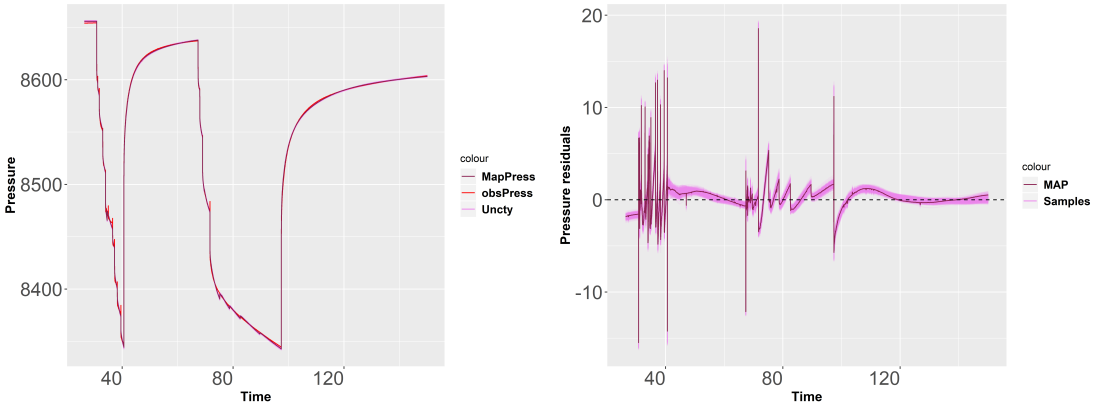
(a) 1 transition.



(b) 2 transitions.



(c) 3 transitions.



(d) 4 transitions.

Figure 6.10: Pressure and pressure residuals uncertainty plots.

6.1.3 Principal Component Analysis

In Figure 6.11 we apply Principal Component Analysis to the chains of the samples of the response parameters and σ_p, as discussed in the previous chapter.

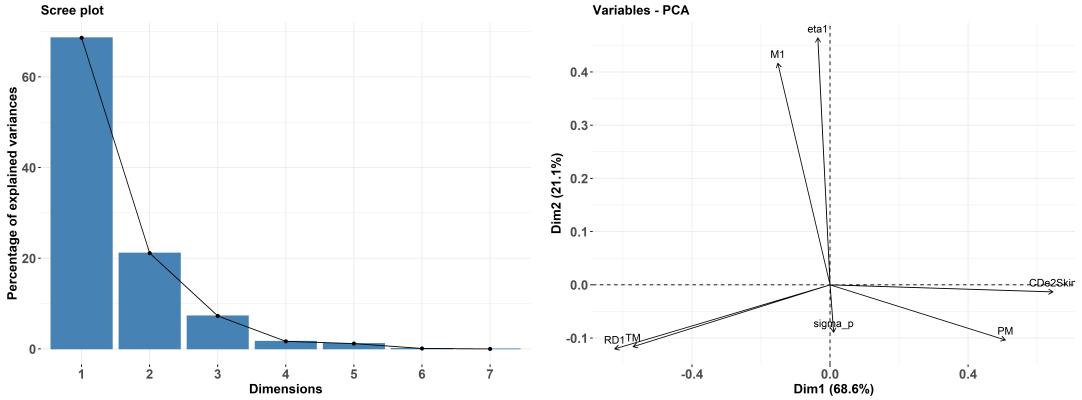
The scree plots suggest that there is one dominant component that ranges from $50\% - 70\%$ for the four models, and an auxiliary component that accounts for 20% for the total variation, with the exception of the three transition model, where it accounts for 40% of the total variation. Only for the two transition model, the percentage of the third component is above 10%. Therefore, in all cases, the corresponding two-dimensional loading plots are good indicators of which parameters are the main contributors of the total system variation.

For the one transition model, the early time and ratio parameters are dictating the main source of variation, while only the late time ratio parameters are the main contributors of the second component. Combining this result with the information from the response plot in Figure 6.8a, we suspect that the early and middle times vary simultaneously, while the cause of the second source of variation is just the size of the transition at late times. This could be suggesting that the model does not provide the necessary flexibility to model the data.

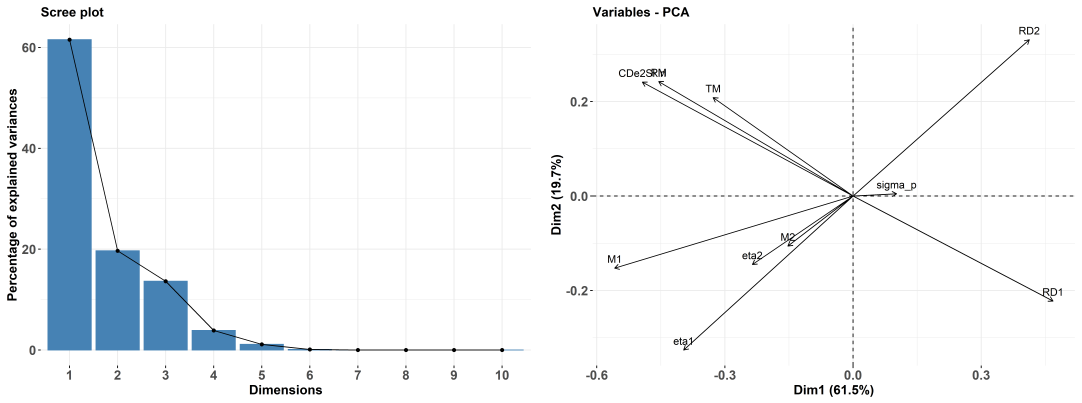
For the two transition model, all the parameters are contributing almost equally to the two components, which suggests that the early and late time portions of the response explain a similar amount of the system variation.

Finally, for the three and four transition models we observe similar behaviour, namely that the main contributors of the first component are the mobility ratio parameters, along with other late time parameters, which suggests that the late time shape is mainly dictating the variation of the model, whereas the early time behaviour is contributing primarily to the second component.

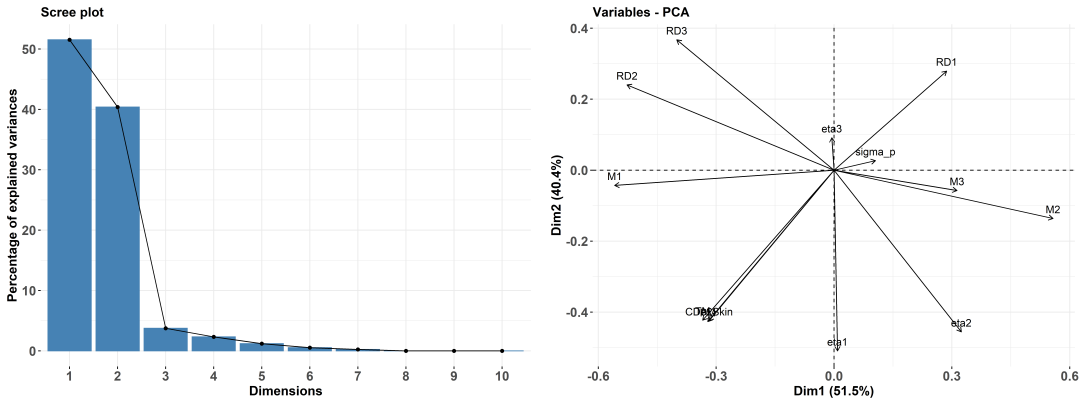
Thus, as we add regions, we observe that the main source of variation in the model moves from the early time shape (when the model is more restrictive) to the late time behaviour (when the model is very flexible).



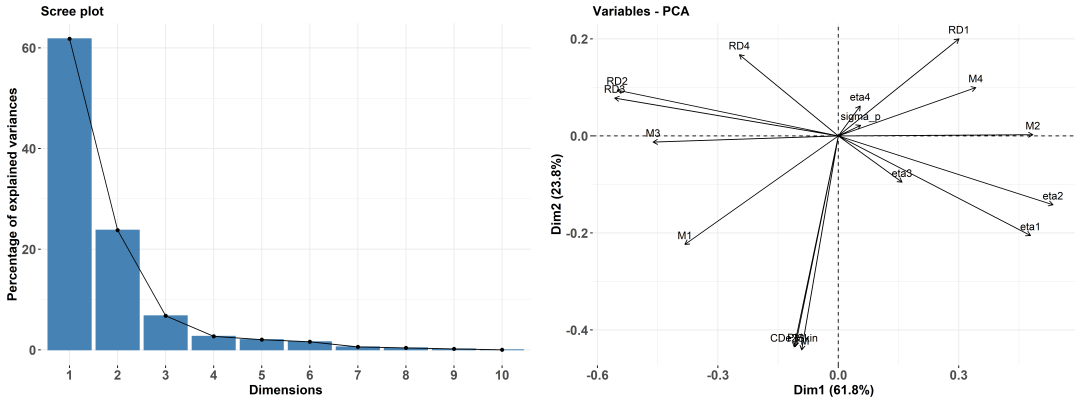
(a) 1 transition.



(b) 2 transitions.



(c) 3 transitions.



(d) 4 transitions.

Figure 6.11: PCA results (scree plots on the left and loading plots on the right).

6.1.4 Step plots

In Figure 6.12, we use step plots to better visualise the changes of the mobility m_i and the diffusivity η_i, based on their radial distance from the wellbore ρ_i. The change in mobility is shown on the y-axis of the left-hand side and in diffusivity on the y-axis of the right-hand side. We can see how those parameters change with the addition of one region at the time, and what that means for the response curve.

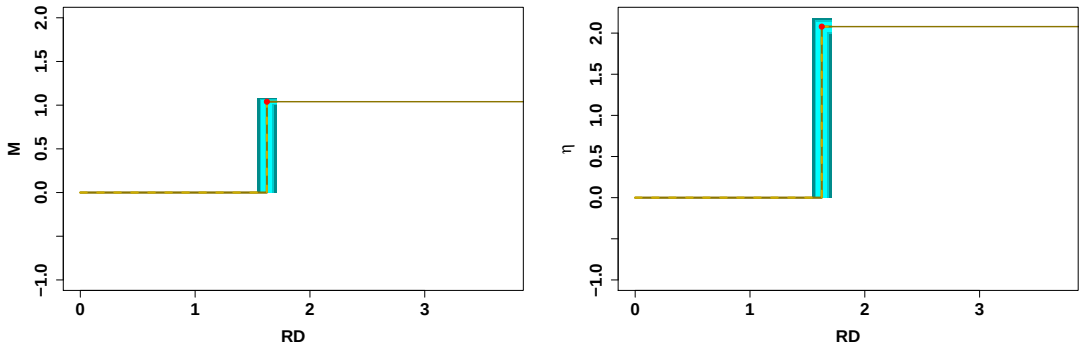
Starting from the one transition model, we notice that the radius uncertainty is very narrow, which indicates that the model is confident about the distance that the transition takes place, while the other parameters' uncertainty is substantially larger. The relatively small m values indicate that the late time response shape corresponds to a jump rather than a slope.

The addition of the second transition introduces a new feature before the one transition feature. The new one is characterised by small positive values of mobility and diffusivity, which suggests a relatively small jump before the final feature. The flexibility of the extra region also adds to the uncertainty of the mobility of the final feature. Thus, it reflects a late time slope rather than a jump.

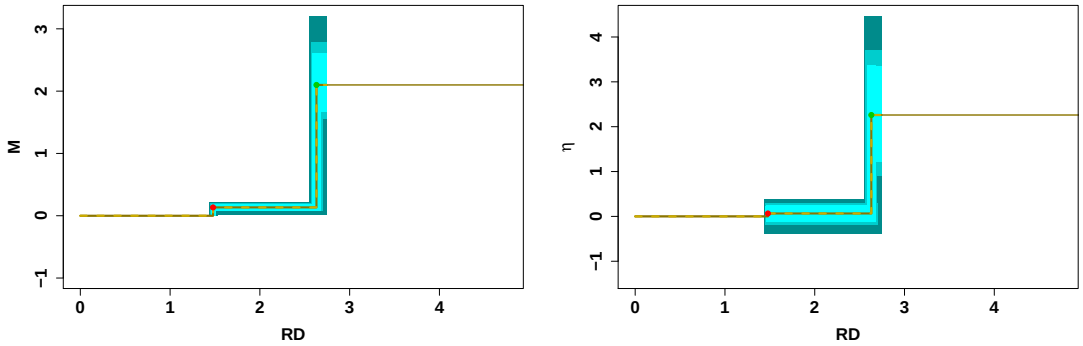
The third transition model adds its new feature before the other two. It includes small negative values of m and small positive values of η, which in Figure 6.8c takes the form of the localised deviation before the jump. This further flexibility causes the uncertainty of the final diffusivity to increase significantly.

The last model adds the fourth transition after the final transition of the previous model, and seemingly merges those two. The shape of the step plots remain almost unchanged, with the exception of the radius uncertainty of the last two transitions that has increased significantly, since there are now two different regions that effectively have the same behaviour. That is the reason why the four transition response curve looks very similar to the three transition curve.

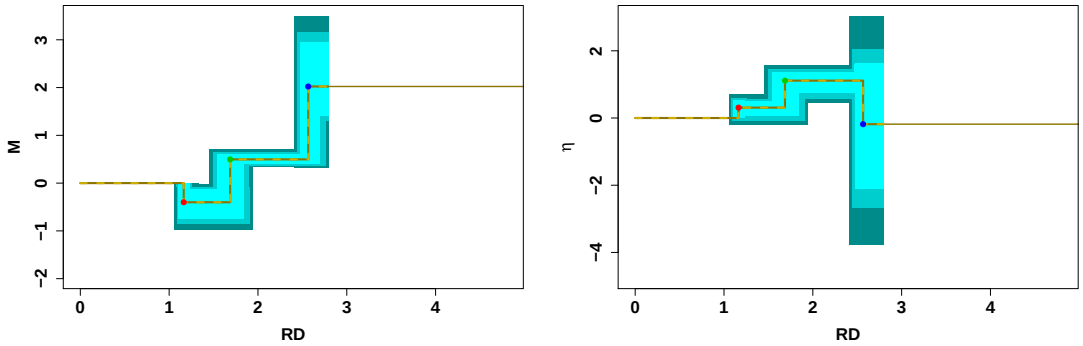
Finally, it is worth mentioning that, as in the synthetic case, it becomes evident that large uncertainty on late time parameters does not directly translate to large



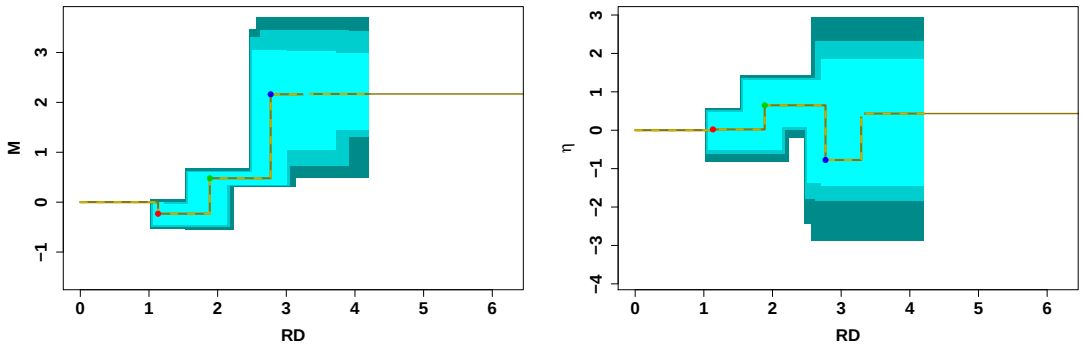
(a) 1 transition.



(b) 2 transitions.



(c) 3 transitions.



(d) 4 transitions.

Figure 6.12: m and η step plots.

Criterion	1	2	3	4
AIC	12781.58	10559.51	10454.77	10457.10
BIC	12832.06	10631.62	10548.51	10572.47
DIC	12781.36	10559.41	10453.53	10459.58

Table 6.2: Information criteria results for the models with one to four transitions. Red indicates the smallest value (best model) for each criterion.

uncertainty for the response curve.

6.1.5 Model selection

In Section 4.5 we presented a selection of information criteria that can be used for model selection. According to the analysis, AIC and BIC compare the posterior modes, while DIC, using the deviance, includes information from the whole MCMC chain.

Table 6.2 and Figure 6.13 show the information criteria results for the one to four transition models. All criteria demonstrate similar results: as expected, the one transition model performs considerably worse than the rest of them, which seems to confirm that this model is under-parameterised. The other three models comprise values that are closer to each other, indicating that they are fairly similar in terms of performance. According to all criteria, the three transition model performs better than the two transition model, which signifies that the addition of the third transition provides the model with enough flexibility to find a substantially better fit. On the contrary, the fact that the values of the three transition criteria are better than the corresponding four transition values, suggests that the addition of the fourth transition did not change the model enough to justify the extra complexity. This is why the response curves for those models look very similar.

The Bayes factors in Table 6.3 and Figure 6.13d are, as we established in Chapter 4 the most principled methods for performing model selection for results derived from MCMC. For all possible pairs, the Bayes factors show that there is strong evidence against the model with fewer transitions, with only exception the pair of three and

	1	2	3	4
1	1			
2	0	1		
3	0	5.6563e-24	1	
4	0	1.5433e-22	27.285	1

Table 6.3: Bayes factors for all possible pairs of the models, where H_1 is on the column and H_2 on the row. Blue indicates that H_1 is better.

four transition models. For this comparison, the Bayes factor suggests that there is not enough evidence against the three transition model. Those results agree with the model comparison criteria, and effectively suggest that the three transition model is the best among those four in terms of describing this data set.

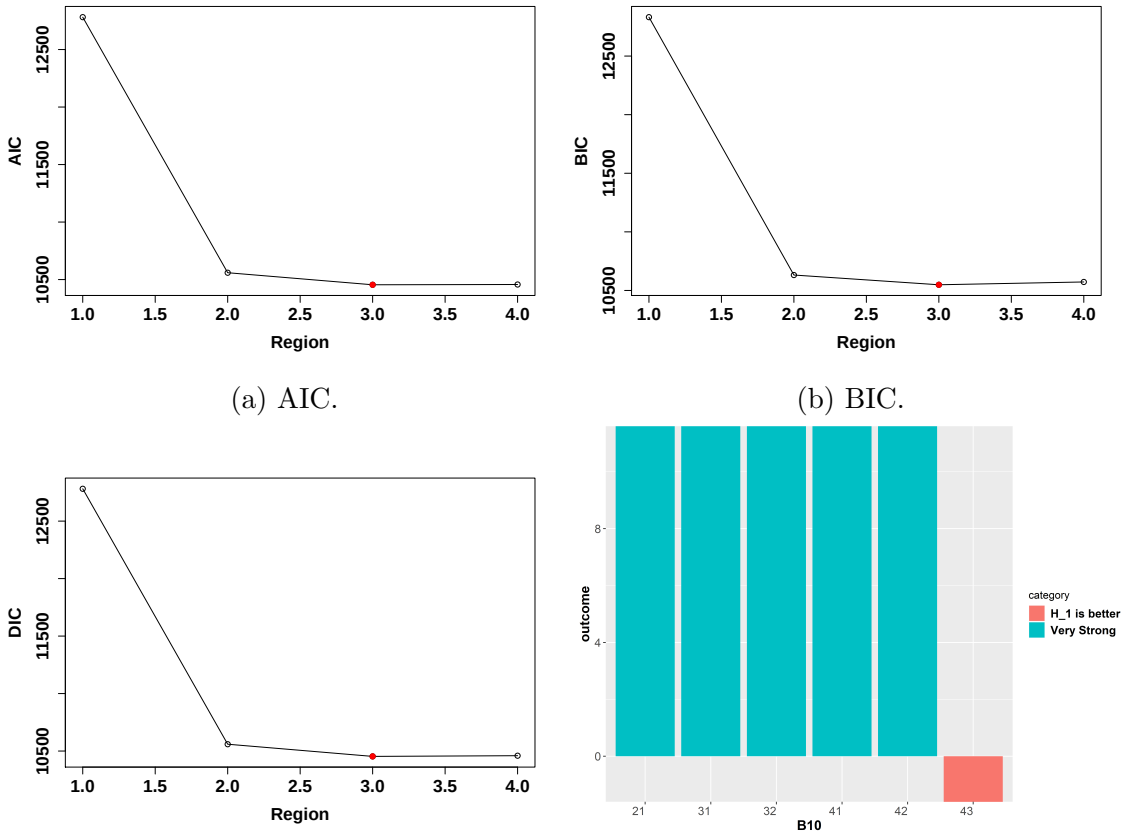
6.1.6 Laplace approximation

We apply the Laplace method in the data, as described in Section 4.7, for one to four transitions.

In each of the four sets of plots in Figure 6.14, the first rows show the histograms from the DEzs algorithm for the three early time parameters, the second rows show the respective Laplace approximations as histograms, and the third rows show the respective SIR simulations. For the first and the last rows we also overlay the Laplace approximation posterior marginals, in the form of the red smooth densities for comparison.

In the left-hand side of Figure 6.15, we plot the corresponding response uncertainty plots, composed of samples from the Gaussian distribution of the Laplace approximation samples, and on the right-hand side, the corresponding SIR samples.

According to the simulations, the approximation that uses the Gaussian density provides results fairly similar to the DEzs samples for the first three transition models, while, for the four transition model, where the MCMC posterior is heavily skewed, there is considerable difference. In contrast, the SIR samples give similar



(a) AIC.

(b) BIC.

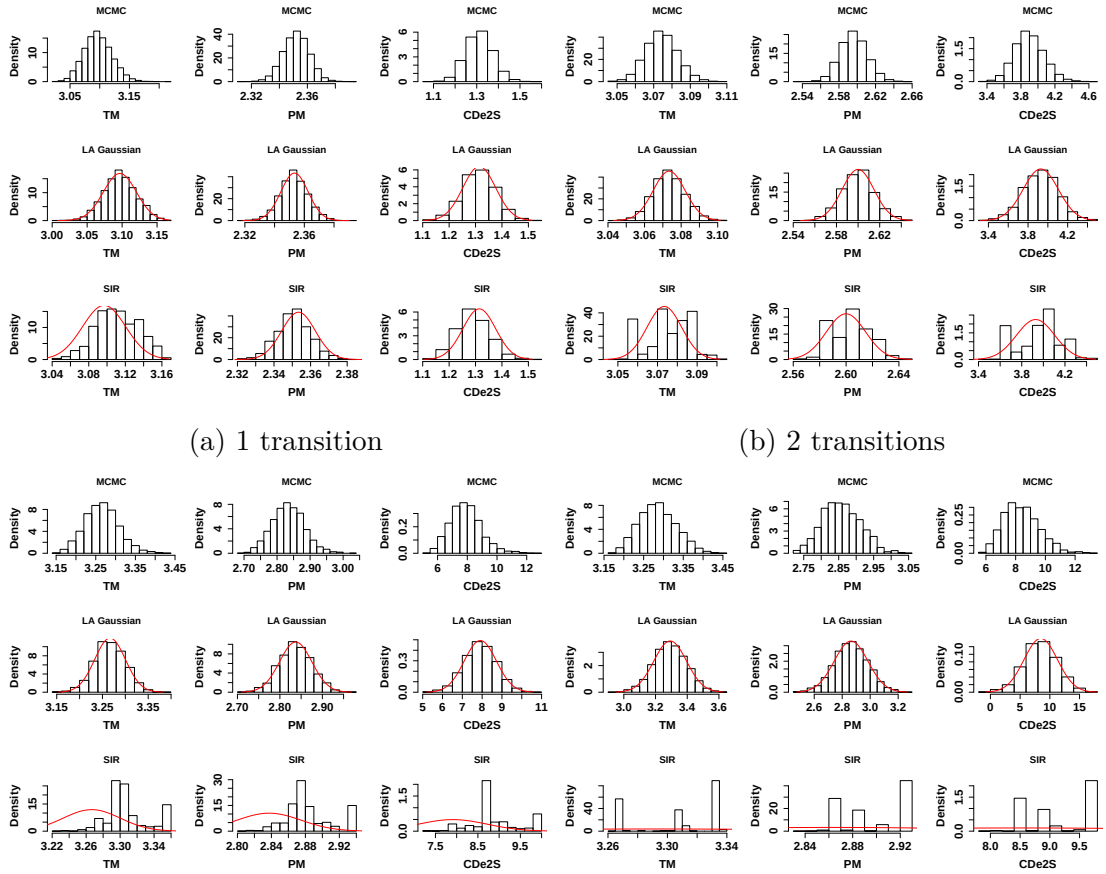
(c) DIC.

(d) Bayes factors.

Figure 6.13: Model selection plots.

results for one transition, considerably different for two transitions, and very different for three and four transitions, where the complexity of the models has increased.

When we compare the response plots with the plots in Figure 6.8, we can see that, as the number of parameters increases, the difference between the corresponding response shapes grows larger. One and two transition curves are similar for all cases. For the three transition curves the Gaussian approximation is still similar to the MCMC responses plot, but the SIR sample curves are significantly fewer in number. Large differences can be seen for the last model, where the Laplace approximation response curves cover a much larger space than their MCMC equivalent plots, while the SIR curves a much smaller space. Those results agree with Figure 6.14d, where the range of the posterior marginals of the Laplace method and the SIR are substantially larger and smaller than the corresponding MCMC ranges respectively. It is also



(a) 1 transition

(b) 2 transitions

(c) 3 transitions

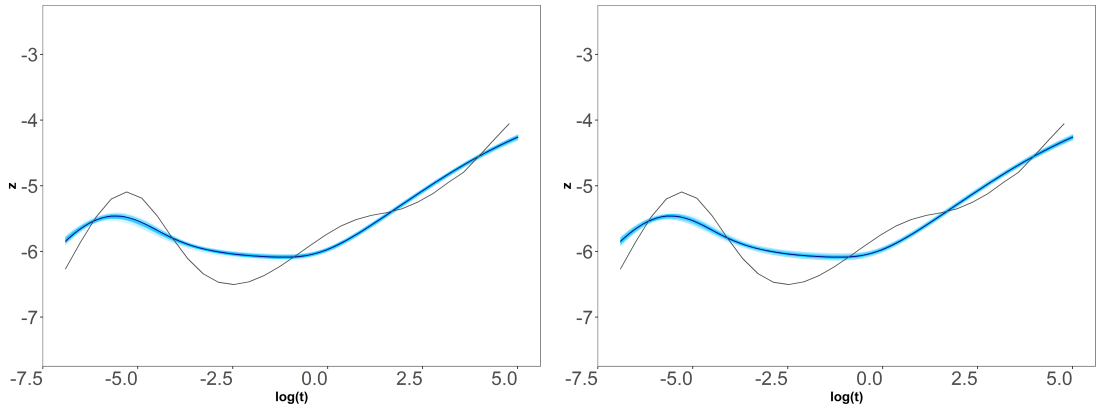
(d) 4 transitions

Figure 6.14: MCMC, Gaussian approximation and SIR histograms of the early time parameters.

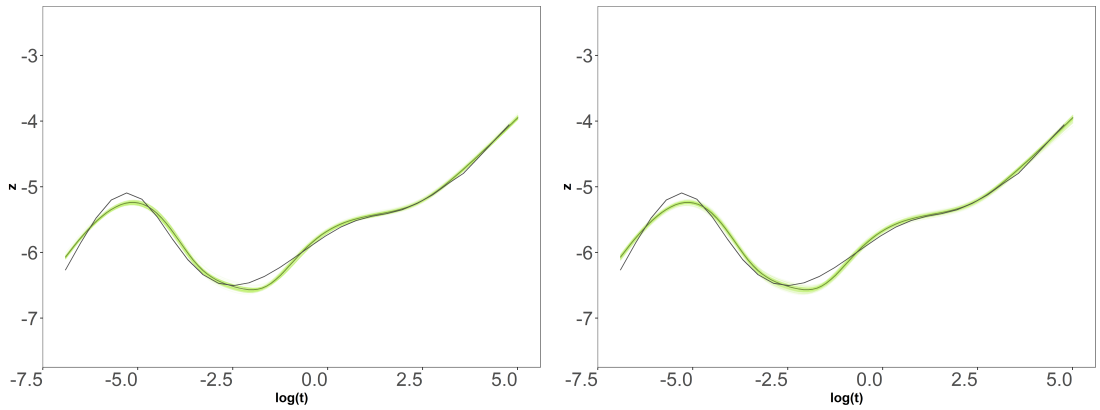
interesting that the structure of the SIR histogram includes gaps, which indicates that the SIR method in this case (as well as some of the other cases) needs many more samples to produce a decent posterior approximation.

The main issue of the Laplace approximation is that, in complex cases, like the four transition model, a Gaussian approximation of all the marginals is not necessarily sufficient, while for the SIR samples, a complicated covariance structure among the parameters is very hard to capture.

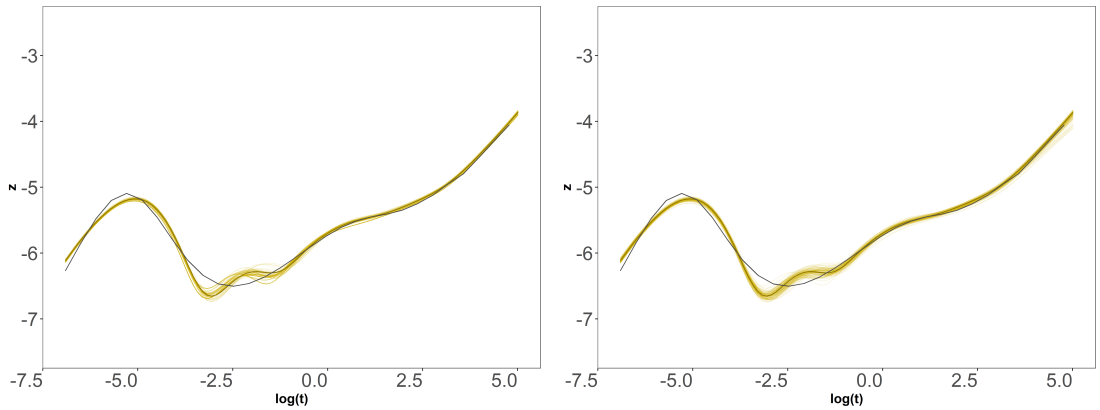
In addition to the issues described above, there is not, in general, an optimisation method that finds the MAP of this model quickly. Therefore, we prefer to use the MCMC methods, which are more robust and able to handle the optimisation, correlation structure and different posterior marginal forms more efficiently than the Laplace approximation and the SIR.



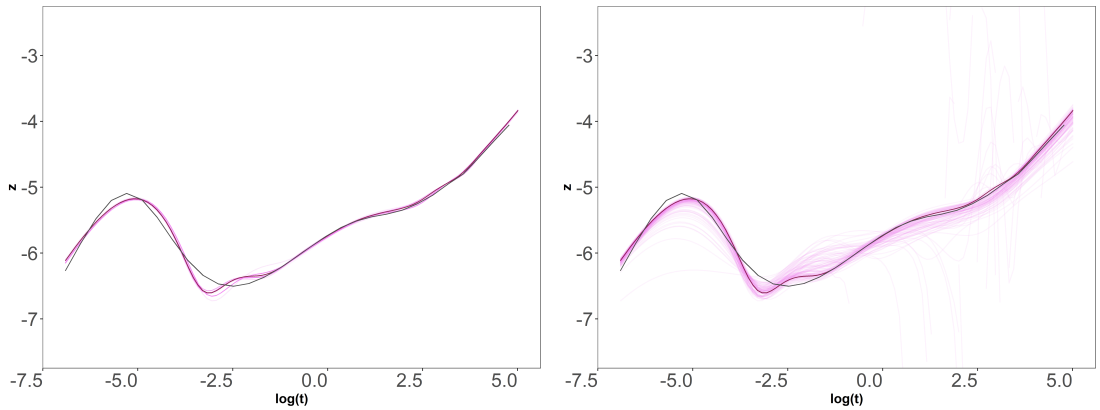
(a) 1 transition



(b) 2 transitions



(c) 3 transitions



(d) 4 transitions

Figure 6.15: Response uncertainty plots from the samples of the Laplace Gaussian approximation (left) and the SIR method (right).

6.1.7 Buildup derivative

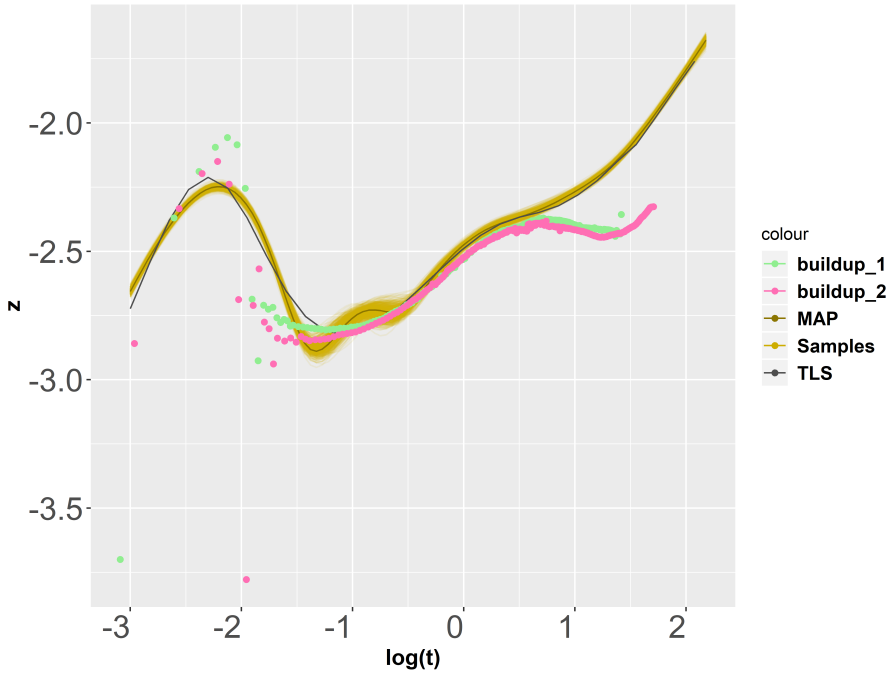


Figure 6.16: Buildup derivatives for the oil data, the three transition uncertainty response and the TLS response.

Another way to validate the model is by comparing its shape to the shape of the buildup derivative, which we described in Chapter 1. The buildup derivative is a long-established well testing tool that petroleum engineers mainly use to derive information for the early time response behaviour, and more specifically, for the radial flow regime. The observation error usually affects the numerical derivative and the wellbore storage/skin phases may be distorted. In addition, each derivative corresponds to a different buildup, and thus, some differences among the levels are common. Nevertheless, it is expected that the most prominent flow regimes should be visible. In Figure 6.16, we overlay the uncertainty plot of the three transition model, the response plot from the MAP and the TLS curve, and the derivative from the two main buildup phases.

During the very early stages, the buildup derivatives do not agree with each other. This happens because of the aforementioned issues and shows that we cannot trust the information we derive from them for this section of the response. During the

middle times, we derive the most important information we can extract from the buildup derivatives. We can see that the two curves lie mostly within the uncertainty range of the response and are also close to the TLS curve. This agreement is an indicator that our result and the results from the buildup derivatives do not contradict each other. Finally, at late times, there is significant deviation from the buildup derivatives and the curves from the TLS and our method. Since it is known that the buildup derivative results provide limited information about the late time behaviour, as was explained in Subsection 1.3.2, this is not a reason for concern.

6.1.8 Effect of the prior on the early time parameters

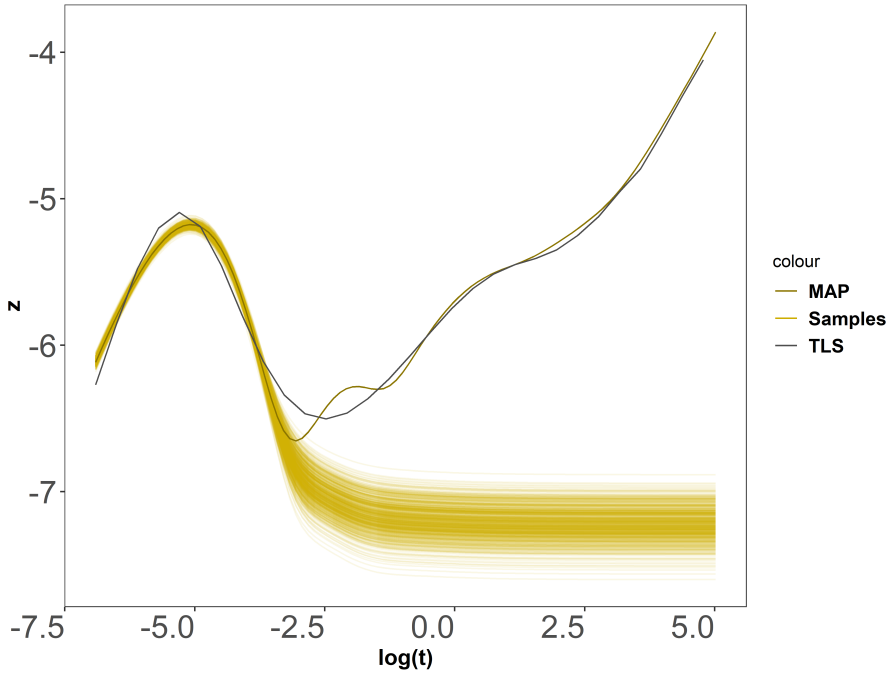


Figure 6.17: Equivalent IRF response uncertainty plot, based on the three transition model's early time parameters. Specifically, the three early time parameters are taken from the three transition model's samples, while the late time parameters correspond to an IRF model ($M_i = \eta_i = 0$, which also makes R_{Di} irrelevant).

In Subsection 6.1.1 and Subsection 6.1.2 we observe some peculiar results once the third transition was added. First, the scatterplots of the early time parameters indicate uncharacteristically strong positive correlations for those marginals. Second,

the autocorrelation results show a significant decrease of efficiency for the three and four transition models.

The justification for those results comes from the fact that the effect of the first transition appears very early and, thus, it masks the effect of the radial flow regime. In terms of the response curve, the stabilisation that we usually observe in middle times is covered by the transition feature. In Figure 6.17, we display the response uncertainty plot of an IRF model that consists of the same posterior marginals for the early time parameters. The response of the MAP and the TLS are overlaid for comparison. The results show that the IRF level is much lower than expected, but since the effect of the first transition comes early for the three and four transition model, we never actually see this behaviour.

In practice that means that all early time parameters can have very large values, and produce very similar shapes, since the actual portion of the response that changes is no longer visible.

In terms of the radial composite model, that effect corresponds to a region very close to the wellbore that causes a very steep change of the mobility and diffusivity values. A common situation where that happens in practice is when the wellbore is located close to one of the reservoir's boundaries, which is called *hemiradial flow regime*, and was presented in Subsection 1.3.1. That regime prohibits the engineers from learning the correct values for the early time parameters. In situations like this, the likelihood on its own is very uninformative for these parameters, and the only way to restrict them to configurations that correspond to physical results, is through their prior distributions, which signifies the importance of that choice.

6.1.9 Alternative priors

In some occasions the engineers might have other sources of information from different methods, geological data and expert knowledge, that claim, for instance, that the reservoir show homogeneous behaviour close to the wellbore. In this occasion, they

would expect the response to stabilize to the radial flow level early on, rather than fluctuate as in Figure 6.8c. This might imply that the three transition model presented in the previous sections was actually overfitting the data, and the additional feature that appeared in the early times was just an artefact of noisy measurements.

This information can be incorporated to the model, by using a different set of priors, where the main focus would be on a larger low boundary on the R_{D1} prior, which would ensure that the first region is placed at a long enough distance from the wellbore.

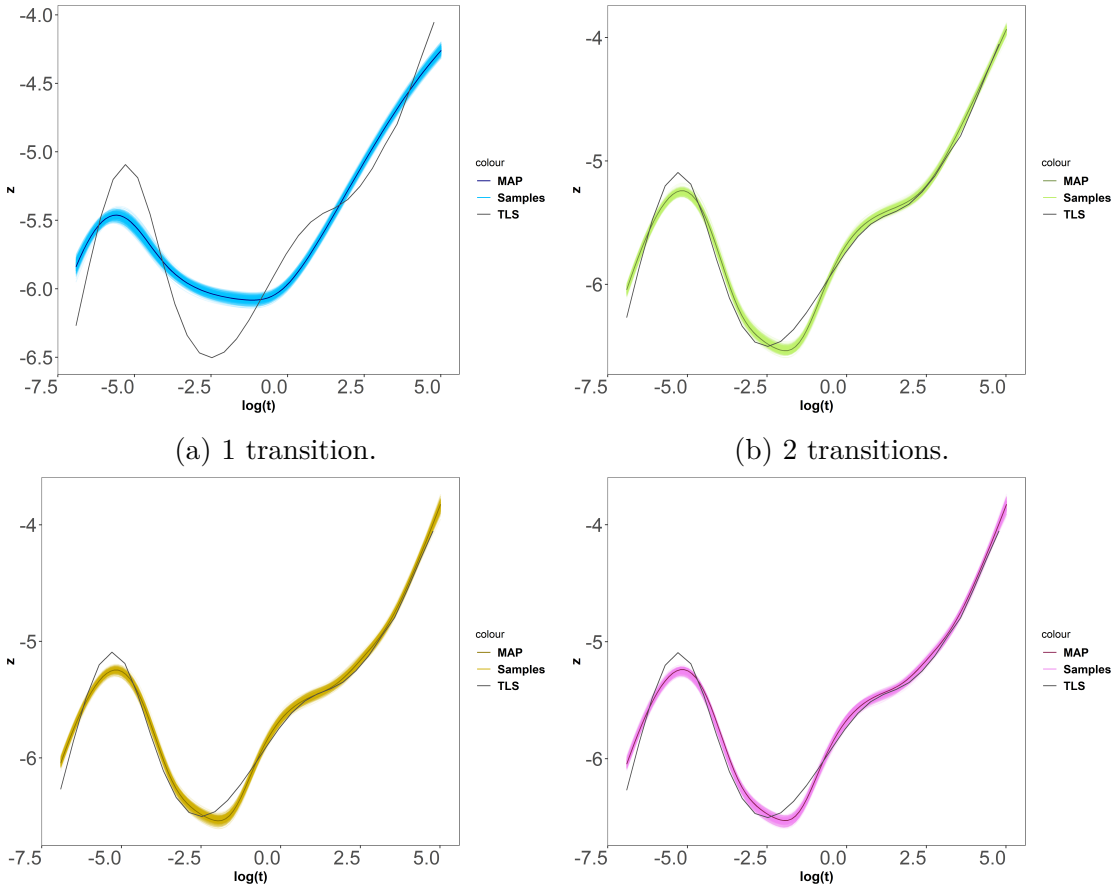
For this subsection, we use an alternative prior configuration, based on the aforementioned comments. Our new choices are: $T_M \sim \mathrm{N}(2, 0.2)$, $P_M \sim \mathrm{N}(1.5, 0.2)$, $C_D e^{2S} \sim \mathrm{Ga}(1, 0.2)$, $R_{D1} \sim \mathrm{U}(3.3, 10)$, $R_{Di} \sim \mathrm{N}(4, 2)$, for $i > 1$, $M \sim \mathrm{N}(0, 2)$ and $\eta \sim \mathrm{N}(0, 2)$.

The analysis was repeated with the same choices of algorithms and configurations. We present a selection of the most interesting results.

Uncertainty plots

In Figure 6.18 we observe that the one and two transition model response uncertainty plots look very similar to the corresponding plots from the previous analysis. That was expected, since the main objective of the new prior configuration was to ensure that the first transition does not mask the radial flow regime, which was not happening for those two models.

However, we can observe significant differences regarding the response shapes of the three and four transition models, which now are very similar to the two transition model's response. The three sets of curves show a sequence of features: the radial flow regime, the big jump, and the approximately unit late time slope, while the early appearance of heterogeneity is no longer evident. Whether the addition of the third transition is still significant or not is no longer clear from the response plots.



(a) 1 transition.

(b) 2 transitions.

(c) 3 transitions.

(d) 4 transitions.

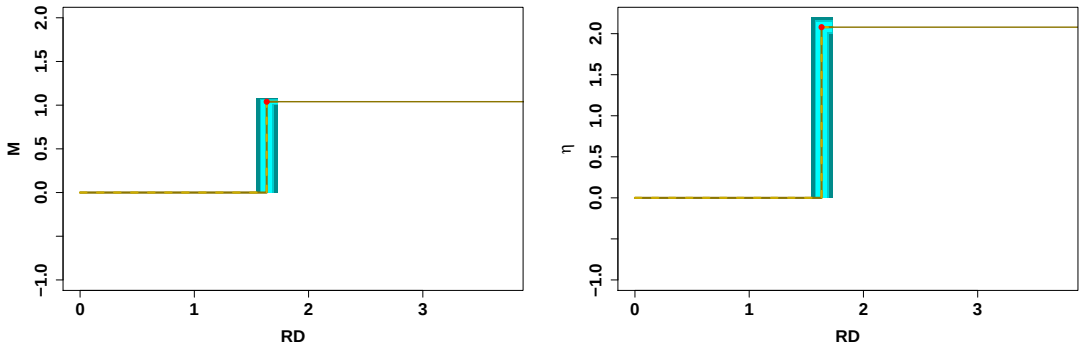
Figure 6.18: Response uncertainty plots.

Step plots

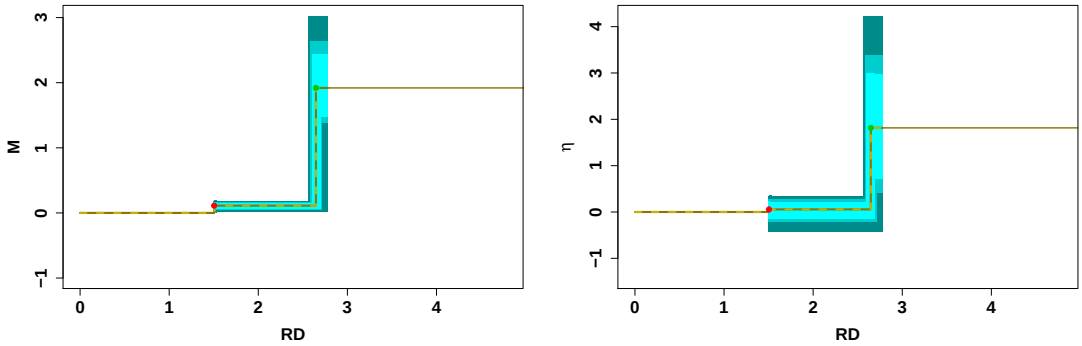
Looking at the step plots in Figure 6.19, we can see that the plots of the one and two transition models are similar to those from the previous analysis.

The extra region of the three transition model appears into sight in between the transitions of the two transition model, changing slightly the values of m and η, and mainly adding to their uncertainty, and thus indicating another feature, very similar to the initial jump. As we saw in Figure 6.18, this feature is not visible in the response plots.

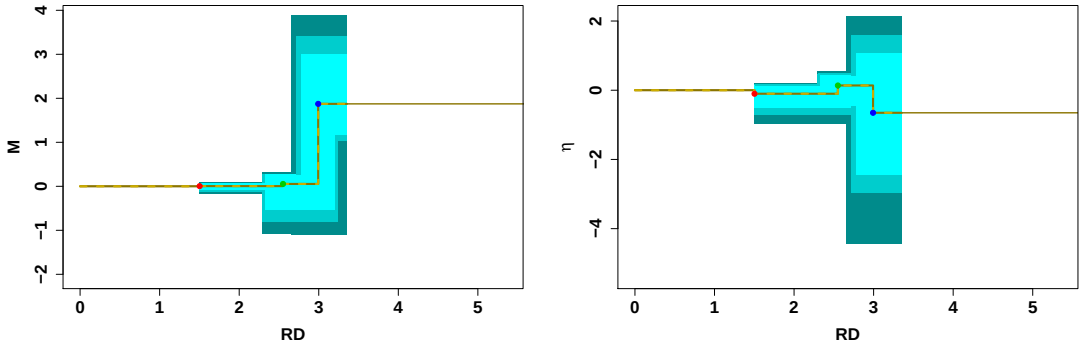
Finally, the addition of the fourth transition appears very close to the last transition of the previous model, and affects primarily the uncertainty of the radial distance from the wellbore ρ, producing similar results for the response curves of the two models. This is a situation where the effect of one transition masks the effect of



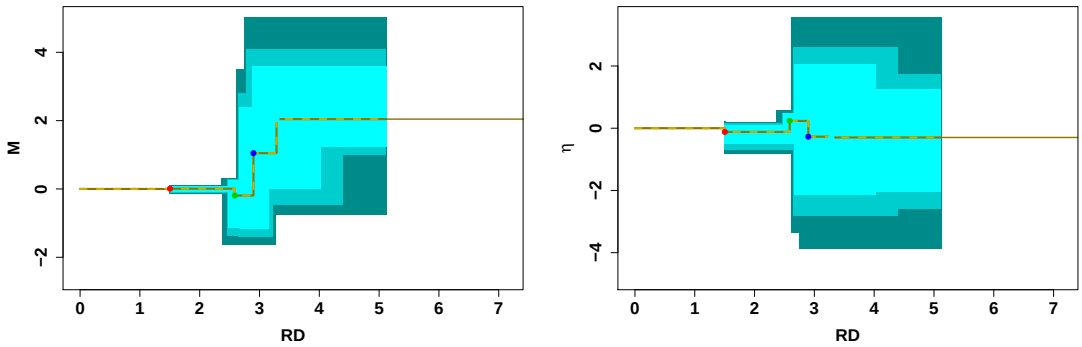
(a) 1 transition.



(b) 2 transitions.



(c) 3 transitions.



(d) 4 transitions.

Figure 6.19: M and η step plots.

Criterion	1	2	3	4
AIC	12744.96	10518.90	10478.12	10489.80
BIC	12795.43	10591.00	10571.86	10605.17
DIC	12744.77	10519.35	10475.80	10487.69

Table 6.4: Information criteria results for the models with one to four transitions. Red indicates the smallest value (best model) for each criterion.

	1	2	3	4
1	1			
2	0	1		
3	0	9.4185e-09	1	
4	0	1.1495e-08	1.2205	1

Table 6.5: Bayes factors for all possible pairs of the models, where H_1 is on the column and H_2 on the row. Blue indicates that H_1 is better.

another, indicating overparameterisation.

Model selection

In Table 6.4 and Table 6.5 we present the results for the model selection information criteria and the Bayes factors respectively. We observe that, once more, all the methods choose the three transition model, which suggests that even though the response curves look similar, the effect of the third transition, and the small extra feature it produces, are not negligible.

It is worth noting that the values of the criteria for the two and three transition models, as well as the respective Bayes factor, suggest a much smaller relative difference between these models than the corresponding difference with the previous prior configuration.

6.2 Example: field data from a gas reservoir

The gas field example consists of 922 irregularly measured pressure data, and a production rate history which is mostly estimated, with the exception of a few

observed measurements at the end. In total, we have 96 rate values over 430 hours. The initial pressure is also not measured, but estimated. The data can be seen in Figure 6.20.

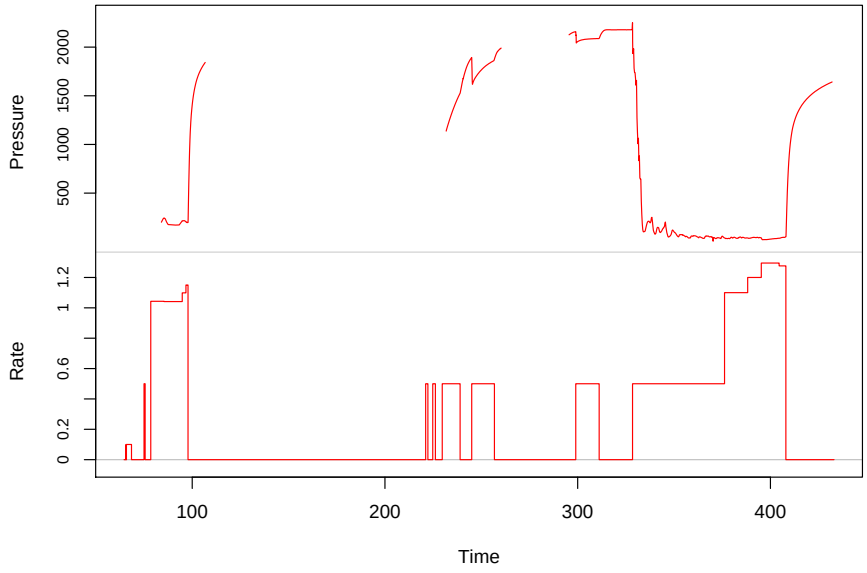


Figure 6.20: Gas data rate and pressure measurements.

We use the radial composite model with one transition, and the Bayesian context that we described in Chapter 3, according to which, we split σ_q into two different values; for the main bulk of the rate production that was estimated we use $\sigma_{q1} = 0.20$, while for the few measured rates at the end, we assign $\sigma_{q1} = 0.01$. In that way, we reflect our confidence for the actual measurements, and we favour them over the rest of the rate history. σ_{p_0} is fixed to 10, while σ_p is treated as a free parameter.

One of the main issues with the aforementioned configuration, is that some of the rate values in Figure 6.20 are very small, and based on the variance that we assigned on them, the corresponding posterior values could be negative, which is not desirable. To tackle this issue, we abandon the marginal model from Subsection 3.4.1, and instead, we sample from the joint posterior distribution of $\boldsymbol{\phi}$, σ_p, $\tilde{\boldsymbol{q}}$ and $\tilde{p}_0$. In that way, we can assign a uniform prior for the rates with a non negative lower boundary, and guarantee that the rates do not change signs. We choose $\tilde{q} \sim \mathrm{U}(0, 10)$. The

initial pressure is also assigned a uniform prior: $\tilde{p}_0 \sim \mathrm{U}(0, 5000)$ For the response parameters, we use the prior specification for gas reservoirs that we established in Chapter 3. That is: $T_M \sim \mathrm{N}(0.5, 0.15^2)$, $P_M \sim \mathrm{N}(-2.5, 0.15^2)$, $C_D e^{2S} \sim \mathrm{Ga}(1, 0.2)$, $R_D \sim \mathrm{N}(3, 2^2)$, $M \sim \mathrm{N}(0, 2^2)$ and $\eta \sim \mathrm{N}(0, 2^2)$. Finally, σ_p is assigned a much larger range prior, in order to account for the change of units to pseudo-pressure: $\sigma_p \sim \mathrm{U}(0, 50)$.

We performed MCMC using three parallel chains and the DEzs algorithm, for 666,667 iterations each, which, after applying thinning of 50 and discarding the first 10,000 iterations as burn-in, leave 10,002 samples for inference.

It is worth noting that we performed tests with models that comprised more than one transitions, but the results were similar, and the model selection techniques chose the one transition model, which is why it is the only one presented here.

6.2.1 MCMC output

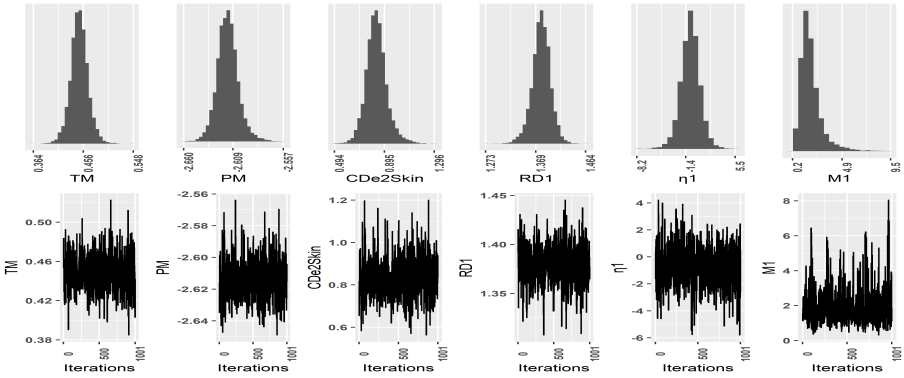


Figure 6.21: Posterior histograms and traces for the response parameters.

The histograms in Figure 6.21 and the summary statistics in Table 6.6 indicate that the early time and radius parameters have substantially smaller variances than their respective priors, indicating that these quantities can be precisely identified from the data. In contrast, the marginal posteriors for M and η show substantially larger variance, indicating an inability to identify specific values for these transition parameters.

Parameter	min	25%	mean	75%	max	sd
T_M	0.36	0.44	0.45	0.46	0.55	0.02
P_M	-2.66	-2.62	-2.61	-2.61	-2.56	0.01
$C_D e^{2S}$	0.49	0.78	0.83	0.88	1.30	0.08
R_{D1}	1.27	1.37	1.38	1.39	1.46	0.02
η_1	-8.21	-1.58	-0.75	0.09	5.46	1.35
M_1	0.24	1.26	1.91	2.28	9.54	1.01

Table 6.6: Posterior summary statistics for the response parameters.

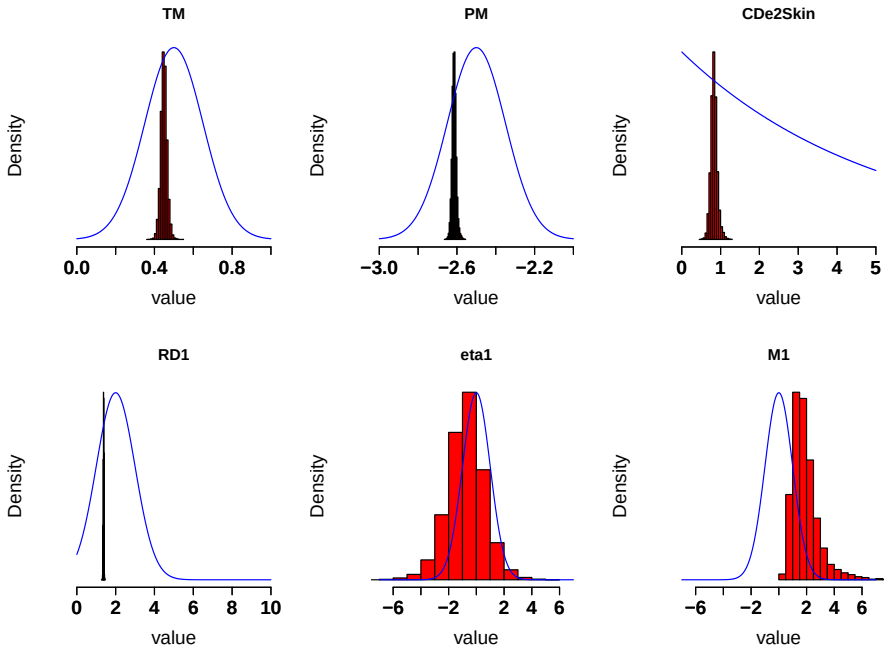


Figure 6.22: Posterior histograms and corresponding prior densities.

These results are more clear in Figure 6.22, where we see that both parameters show evidence of conflict between prior distribution and data. The form of the M distribution specifically indicates a late time response slope of 1, and possibly a closed system. In those situations, the M parameter is attempting to reach infinitely large values and the only thing restricting it is the prior distribution, which creates the conflict. The posterior of η is dictated by the M, due to the large correlation of the two, as mentioned in the analysis of Figure 5.5.

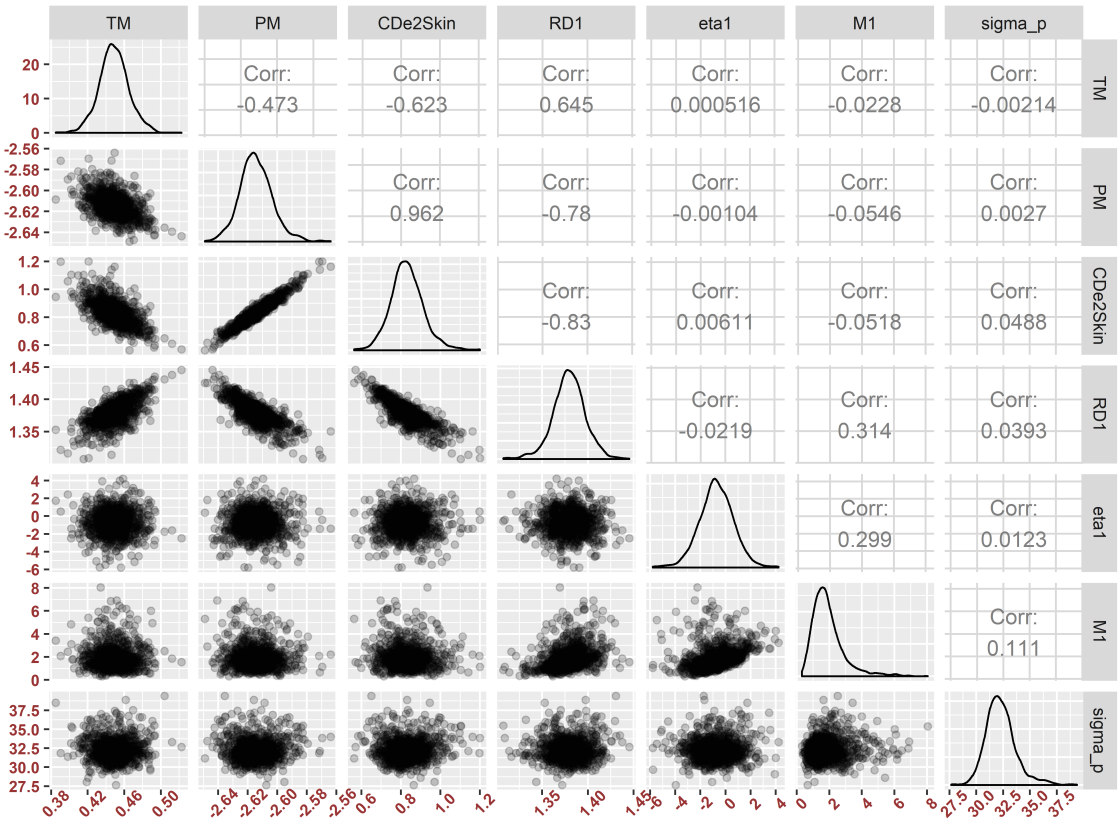


Figure 6.23: Posterior joint marginal scatterplots and Pearson correlations.

The scatterplots in Figure 6.23 show the usual strong correlation between P_M and $C_D e^{2S}$, but also some interesting correlations between the early time parameters and the transition radius. Specifically, T_M and R_{D1} are positively correlated, while the latter is also negatively correlated with P_M and $C_D e^{2S}$, which suggests a connection between the early time behaviour of the response and the moment that the transition takes place. For example, the downwards shift of the response curve due to larger

P_M values seems to be compensated by an earlier appearance of the effect of the transition.

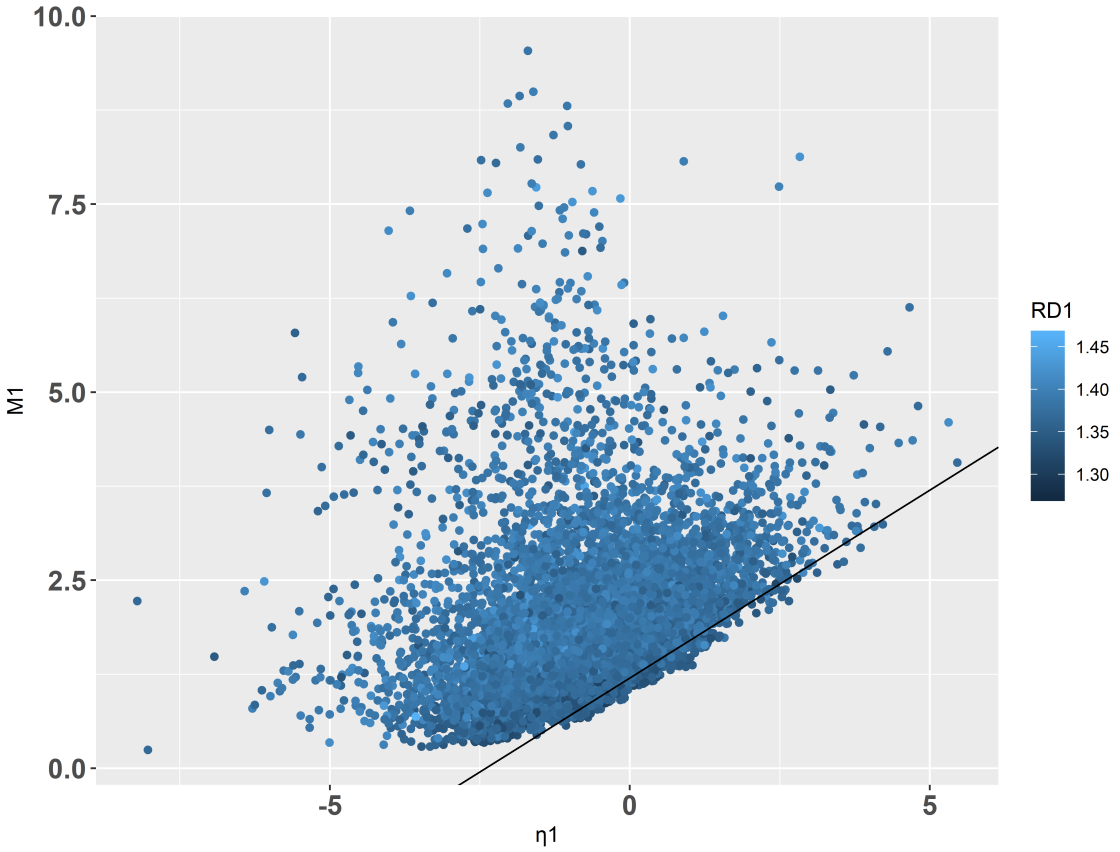
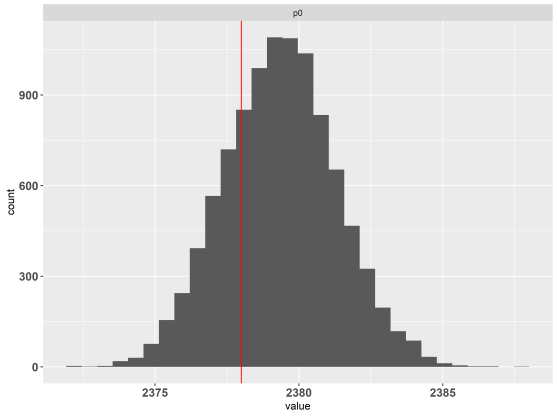
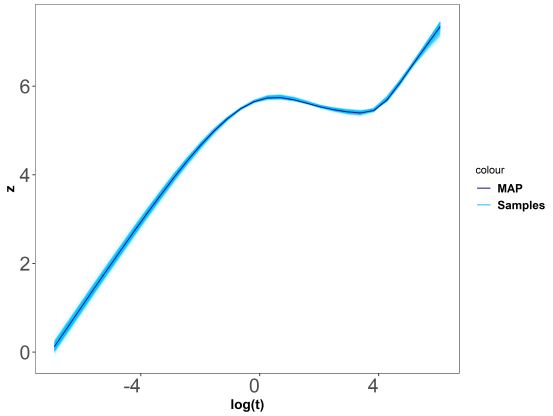


Figure 6.24: Posterior scatterplot for the late time parameters of the gas data set, coloured based on R_D.

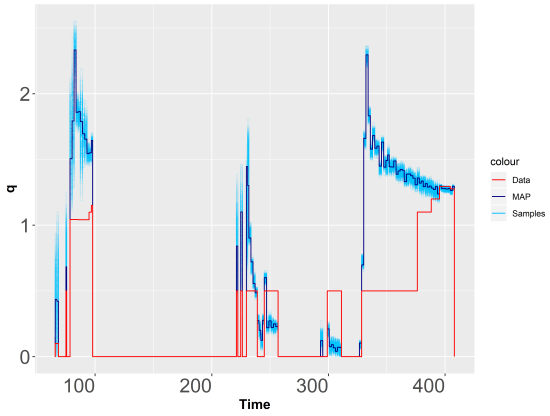
Focusing on the late parameters in Figure 6.24, we observe that the pairs of M and η lie closer to the half-unit slope line that passes from the point $(0, 1.2)$. According to the results of the analysis of Figure 5.7, this line was associated with the pseudo-steady state flow regime, which generally indicates a closed reservoir. Values of M on the upper subspace dictated by this line, as mentioned in the analysis of Figure 6.22, are actually an artefact of the radial composite model, and they also produce a unit slope. Specifically, if it weren't for the priors on those values, a much larger part of the subspace would be covered by the samples, and they would all produce the same late time unit-slope response behaviour.



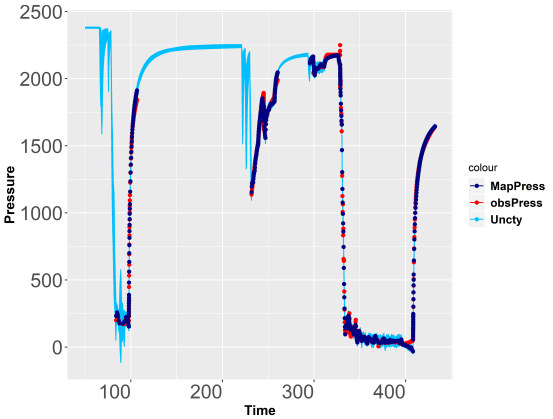
(a) Initial pressure.



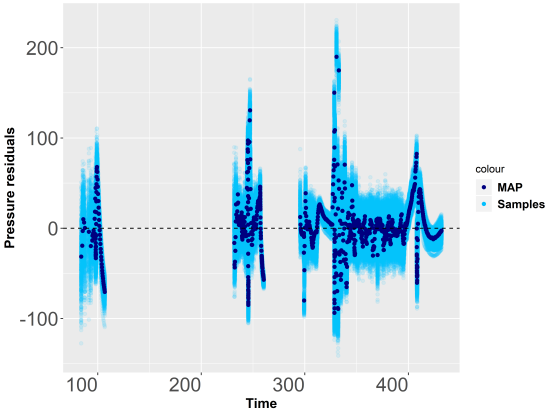
(b) Response.



(c) Rate.



(d) Pressure.



(e) Pressure residuals.

Figure 6.25: Uncertainty plots for the gas data set.

Uncertainty plots

In Figure 6.25 we present the uncertainty plots for the respective quantities, from the posterior samples of the MCMC.

The posterior marginal of the true initial pressure in Figure 6.25a comprises a large variance, and the corresponding histogram has a range of more than 10 psi. This is not very surprising, since the initial pressure was not measured, and had to be estimated.

The uncertainty plot in Figure 6.25b illustrates that, once again, the large uncertainty of the late time parameters will not necessarily be reflected in the response curve. Instead, the samples are centralised around the MAP. The late time behaviour is that of a slope, as the response parameter histograms suggested.

The plot of the rate posterior samples in Figure 6.25c, shows the effect of splitting the σ_q for measured and estimated values; for the former, where a low variance was assigned, the posterior values change slightly from the measurements. For the latter, instead, the posterior values can be more than four times larger than the estimated rates. It is worth noting that, thanks to the effect of the prior, even though some posterior rates come close to zero, they do not change their sign.

The pressure uncertainty plot in Figure 6.25d, suggests that the true pressures are significantly different from the measured ones, and even in the measurement scale, there are noticeable deviations. That is also reflected in Figure 6.25e, where some residuals are significantly large. There are two factors that contribute to this results: first, it is an artefact of the necessary change of units to pseudo-pressure. Second, the time intervals of the rate estimated values do not always agree with the rapid changes in the pressure history, adding to the complexity of the problem.

6.2.2 Principal component analysis

The PCA screeplot in Figure 6.26 suggests that there is one main component that accounts for almost 80% of the total system variation. The loading plot indicates

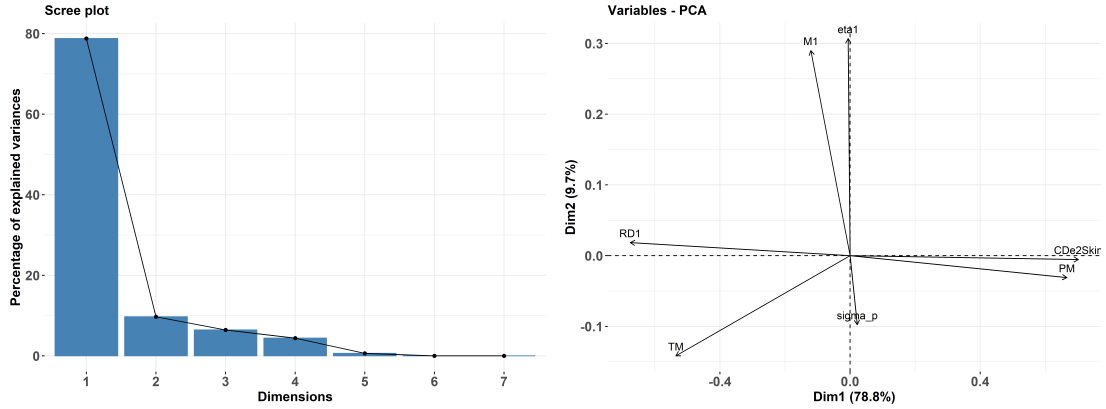


Figure 6.26: PCA results (scree plots on the left and loading plots on the right) for the gas data set.

that the parameters that contribute mostly to this component are the T_M, P_M, $C_D e^{2S}$ and R_D, while the late time ratio parameters' effect is mostly visible in the second component and explains about 10% of the system variation. Those results imply that the main source of variation come from a combination of early and middle times, while the late time slope explains a significantly smaller amount of the model's variation.

6.2.3 Step plots

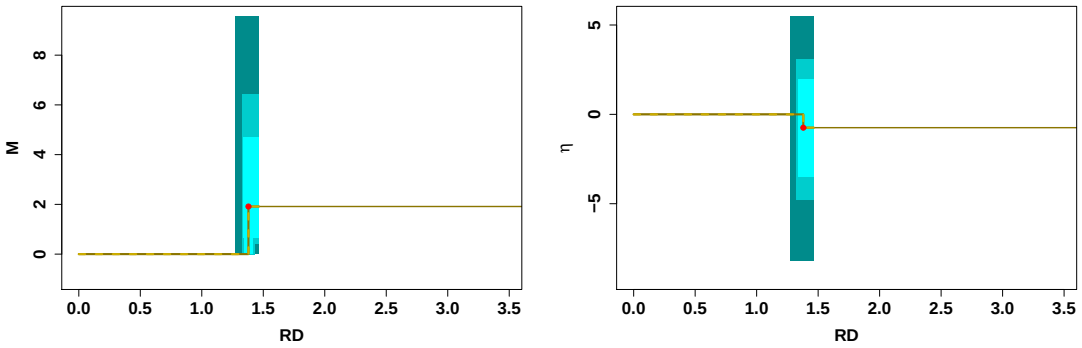


Figure 6.27: Step plots of the late parameters for the gas data set.

The step plots in Figure 6.27 suggest that the model is confident about the distance from the wellbore (radius - ρ) that the transition takes place, while the uncertainty for m and η is significantly large, a result that, as we saw before, is associated with late time slopes.

6.2.4 Buildup derivative

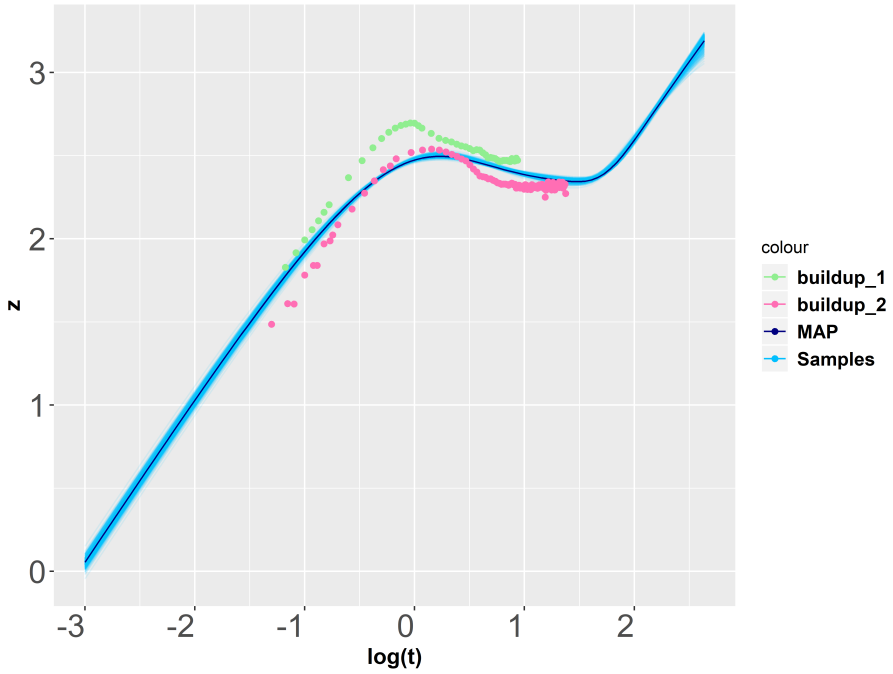


Figure 6.28: Buildup derivatives for the gas data and the response uncertainty plot.

As mentioned in Subsection 1.3.2 and Subsection 6.1.7, there are some limitations to the buildup derivative, coming from its derivation using a numerical derivative, the fact that depending on the flow rate history we might observe the derivatives to appear at different levels, and the limited information they provide for the late time flow behaviour. It appears that all those issues are evident in the case for the gas data buildup derivatives in Figure 6.28, where, the buildup derivatives show differences for the early times (e.g. the skin 'hump'), they stabilise at different levels, and they do not provide enough information about the late time behaviour. Therefore, the comparison between the two methods in this case is limited to the existence of the radial flow regime.

It is still worth noting that the response's radial flow level from the posterior samples lies within the two derivatives, in a way averaging their effect.

6.3 Comparison of MCMC algorithms

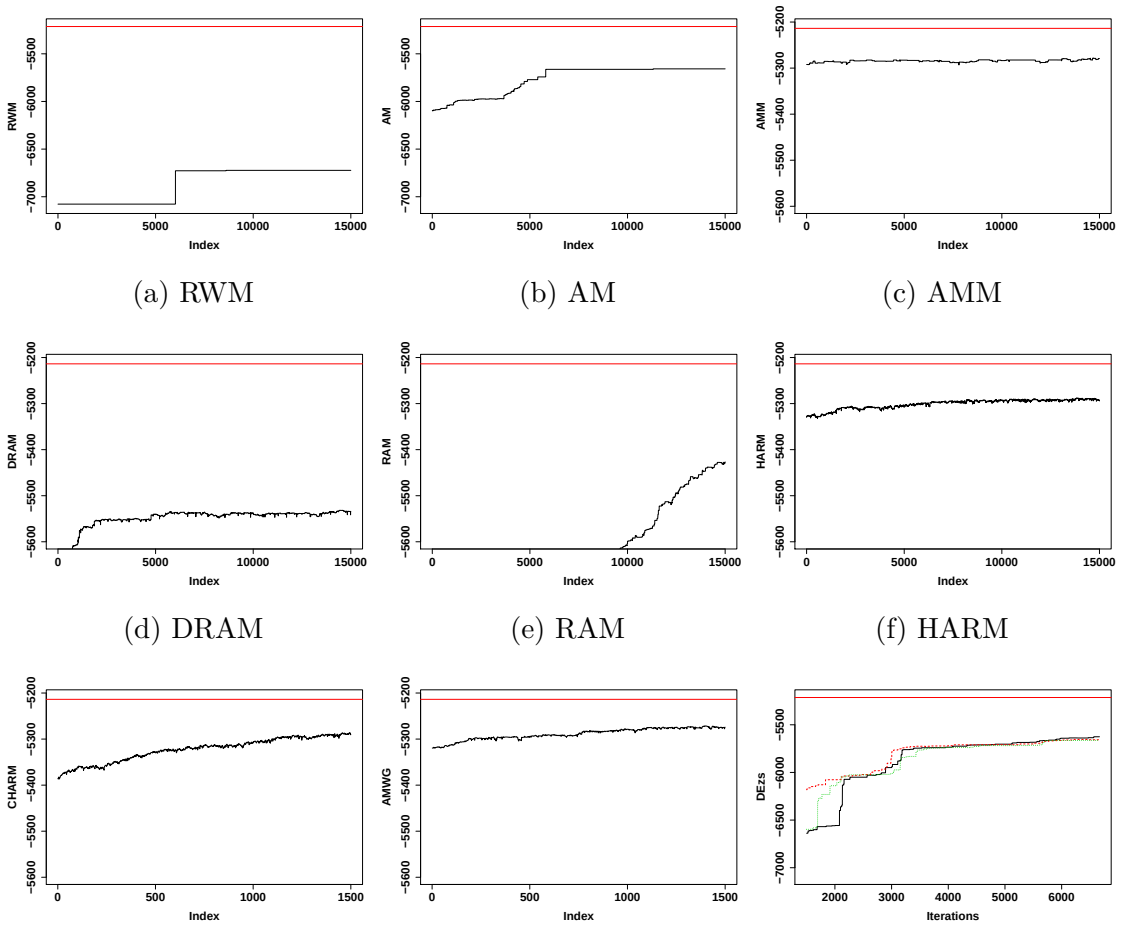
The MCMC algorithms we first presented in Section 4.2, show different levels of performance when we apply them to our model. There are, for example, significant differences regarding their efficiency, or how quick the adaptation is. In addition, when performing a two-phase simulation, we usually focus on different aspects of the algorithms; for the ones used in place of optimisation, we require an algorithm that can quickly scan the whole prior space and find the posterior, while for the second phase, we need algorithms that can capture the correlation structure of the posterior distribution.

In the next two subsections, we compare a variety of different algorithms for those two phases.

6.3.1 Optimisation phase

When dealing with a problem that comprises real data, one of the first challenges is to find the area where the posterior distribution lies. A common way to tackle this, is by using an optimisation algorithm. Unfortunately, as explained in Chapter 4, this is not a compelling option in our case. A popular alternative is incorporating the optimisation issue within the MCMC run. This is also associated with the burn-in phase of an MCMC, where the main goal of the algorithm is to scan the whole prior density, favouring the parameter configurations that result to higher posterior values.

In Figure 6.29 we compare nine of the algorithms described in Section 4.2. We use the Bayesian model from Chapter 3 on the oil dataset from Section 6.1. The algorithms that are based on a Metropolis step (RWM, AM, AMM, DRAM, RAM and HARM) run for 20,000 iterations, while the Gibbs-type algorithms (CHARM and AMWG) were permitted 2,000 iterations. Finally, we run the DEzs algorithm with 3 chains and 6,667 iterations for each chain. The different lengths of the chains aim to provide results that take approximately the same amount of time. We plot the posterior trace for the last 3/4 of the chains for scale purposes.



(a) RWM (b) AM (c) AMM

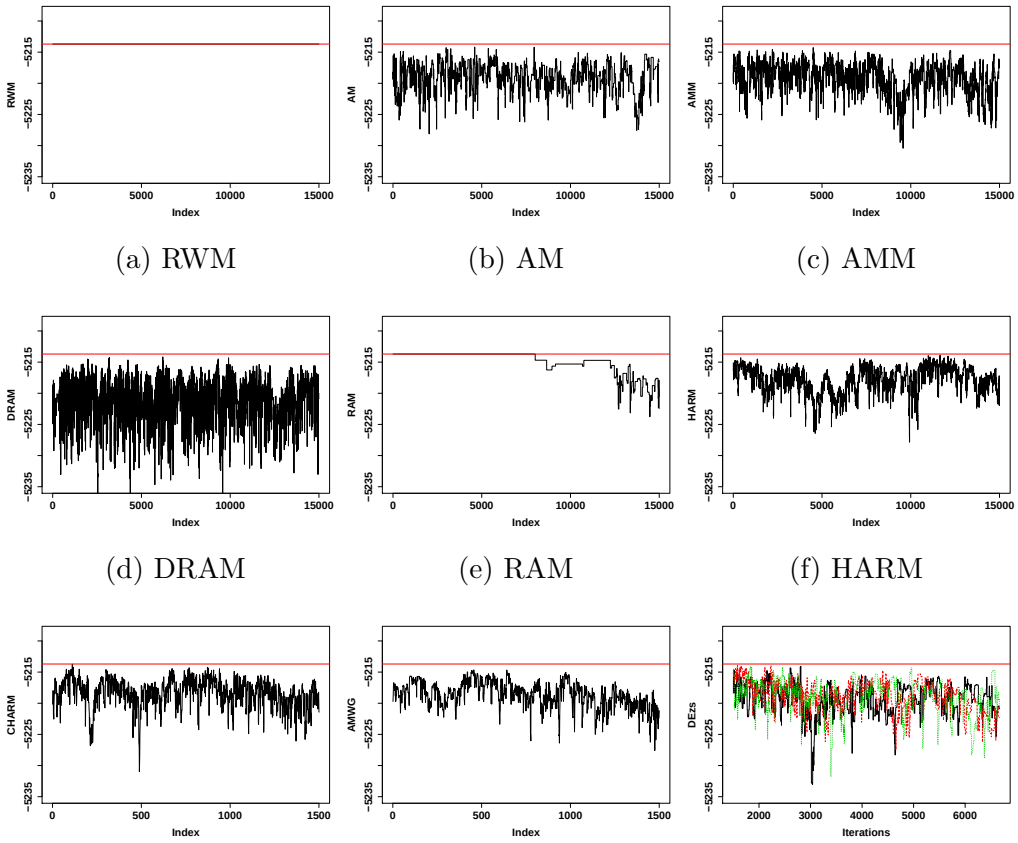
(d) DRAM (e) RAM (f) HARM

(g) CHARM (h) AMWG (i) DEzs

Figure 6.29: Posterior value trace for various algorithms, starting from a random initial point. The goal is to compare how quickly each algorithm reaches the MAP (red boundary line).

All algorithms start from the same random initial point, sampled from the prior. Note that the RWM is the only non-adaptive algorithm and is included for reference, while the red line on top of the plots denotes the MAP.

Firstly, we observe that RWM, AM and DEzs perform poorly compared to the rest of the algorithms, with the first two being unable to make any transition after the first 12,000 iterations. DRAM starts well, but gets stuck in a local maximum. RAM takes a long time to get started, overcoming the threshold of -5,600 only after 15,000 iterations. AMM and HARM performed well at the beginning, but get stuck in a local maximum before they reach the MAP. Finally, CHARM and AMWG perform well, given that, although they take much time for sampling a single iteration, they



(a) RWM (b) AM (c) AMM

(d) DRAM (e) RAM (f) HARM

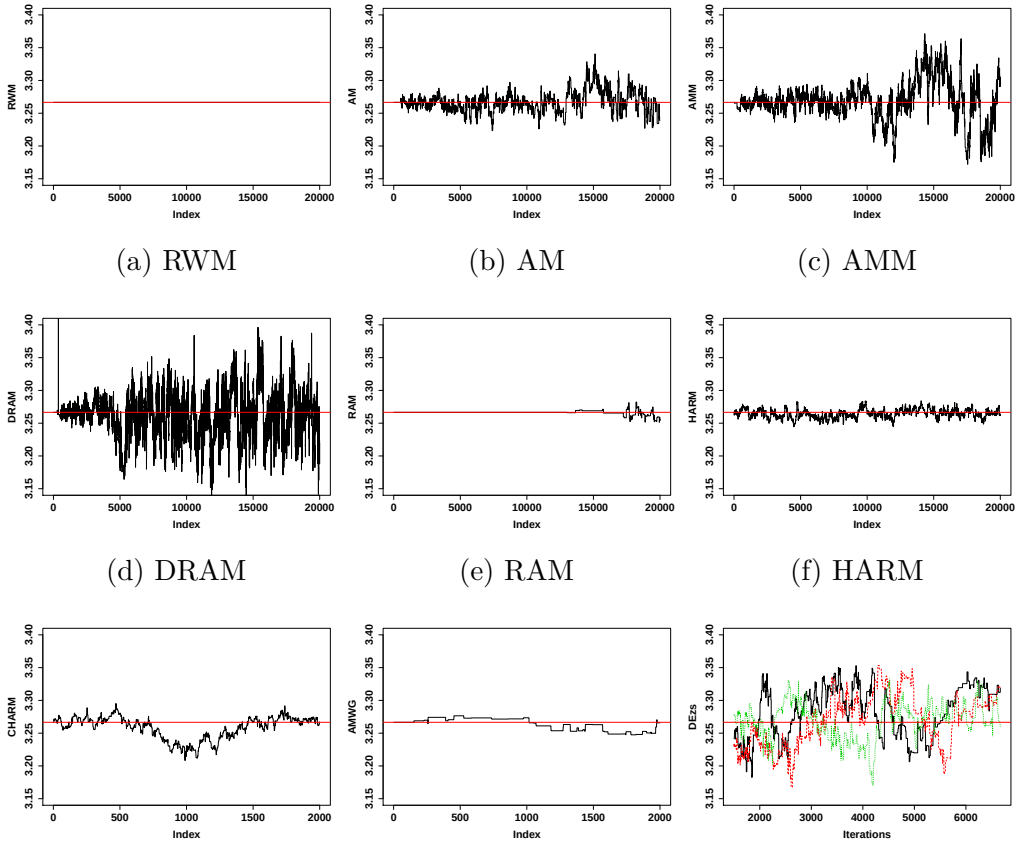
(g) CHARM (h) AMWG (i) DEzs

Figure 6.30: Posterior value trace for various algorithms, starting near the MAP (red boundary line).

only require one tenth of the iterations in order to match the results of the best Metropolis-type algorithms.

The conclusion we derived based on those results is that, even though there are some Metropolis-type algorithms that perform well, they can easily get stuck in a local maximum. Conversely, Gibbs-type algorithms, although slow, perform consistently better.

The algorithm we picked for this stage based on more extended results was the CHARM.



(a) RWM (b) AM (c) AMM

(d) DRAM (e) RAM (f) HARM

(g) CHARM (h) AMWG (i) DEzs

Figure 6.31: Trace of the T_M parameter, starting from the MAP (the red line is the T_M MAP value).

6.3.2 Comparison of MCMCs starting from the MAP

In contrast to the optimisation phase, we need a different approach when dealing with a situation where we already have a good approximation of the MAP, but need an efficient way to recover the shape of the posterior density.

In Figure 6.30 we examine the posterior trace of the nine MCMC algorithms, starting from the MAP, and in Figure 6.31 the corresponding trace of the first parameter (T_M). The running configuration is the same as in the previous section, but these plots show the whole chain, without omitting iterations from the beginning.

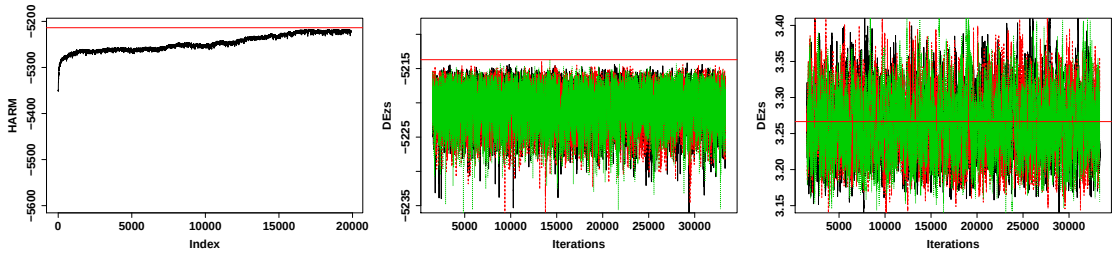
The algorithms that perform poorly are the RWM, which does not move from the MAP, and the RAM, which takes more than 12,000 iterations to start accepting candidates. All the other algorithms show similar posterior trace results with some,

like the AM and the AMWG indicating slower mixing.

Important additional information comes from the first parameter trace plot in Figure 6.31. HARM, CHARM and AMWG indicate that even though the posterior traces fluctuate close to the MAP, the traces of the first parameter T_M suggest poor mixing, since the associated variance has not increased, and all the accepted candidates are very close to the MAP. Furthermore, AM and AMM after 20,000 iterations are still in the process of 'learning' the covariance structure, and thus, the adaptation takes a lot of time to recover the variance of the first parameter. DRAM and DEzs indicate good mixing from the early stages of the chain.

For this stage, after more extended results, our choice was the DEzs algorithm. The results of DRAM are promising, but issues like multimodality, significantly affect this algorithm's performance.

In Figure 6.32, we indicatively show the results of the algorithms as chosen for a full two-phase simulation run: a first phase of 20,000 iterations for CHARM, where the componentwise nature of the algorithm finds quickly the region of high probability close to the MAP, and then a second phase of 5,000,000 iterations for DEzs, where the algorithm adapts the proposal and captures the correlation structure of the posterior, while tackling issues like multimodality.



(a) CHARM posterior value trace. (b) DEzs posterior value trace. (c) DEzs T_M marginal trace.

Figure 6.32: Results of the two-phase simulation.

6.4 Summary

In this chapter we presented the analysis of a real oil and a real gas reservoir, using our model. We examined the results of the MCMC and the convergence and efficiency diagnostics, we performed principal component analysis, and we used relevant visualisation tools that we introduced in the previous chapter, such as the uncertainty plots and the step plots. We also examined the effect of different priors, and we performed model selection. The analysis of the oil data set, with the original set of priors, suggested a three transition model with the presence of hemiradial flow, that experiences an early change of diffusion, possibly related with reservoir heterogeneities close to the wellbore, a big mobility change, which could be associated with the boundary structure of the reservoir, and a pseudo-steady state flow regime, which could indicate a closed reservoir. The alternative set of priors rules out the hemiradial flow regime and instead suggests two mobility changes, and again, the pseudo-steady state flow regime. The analysis of the gas data set indicates the immediate presence of pseudo-steady state flow, after the radial flow regime, which could be an indication that the transient pressure has reached all the reservoir boundaries within the radius of investigation. All those assessments were based on the structure of the posterior density and were possible because of the nature of the multi-region radial composite model of Section 2.5, and its ability to act as a surrogate model for almost every situation encountered in practice. Modelling all the different possibilities explicitly, would require a much more complicated model with many parameters, some of which is not straight forward how they would be combined (e.g. modelling all possible boundary conditions), as well as a way to perform model selection among them. Such a model is questionable whether it would be feasible and would probably not be practical. We also tested the Laplace method and we compared our results with the buildup derivative and the total least squares method. Finally, we compared a selection of different MCMC algorithms that we presented in Chapter 4, and concluded that, for our case, a two-phase simulation using CHARM

for the optimisation phase and DEzs for the main sampling phase, was the most efficient way to recover the posterior density.

In the next chapter, we conclude this thesis by summarising our methods and results, and we point to possible extensions of this work for the future.

Chapter 7

Conclusion

7.1 Summary

In this thesis, we presented a new method for the deconvolution of reservoir well test data, using parameterised, physical models for the response function, that provide a description of the flow behaviour in the reservoir, and we combined those models with a Bayesian statistical context, that can deal with observational error for all data.

We illustrated the use of the model by applying it on synthetic and field data sets, making inferences about the response function and other quantities of interest. We examined the effect of different numbers of regions on the same data set and the contribution of the parameters to the variation of the system.

Finally, we developed new forms of visualisation for the posterior parameters and their uncertainties.

7.2 Concluding remarks

The proposed method has a number of advantages over other methods, such as the buildup derivative and the total least squares and its derivatives.

First, the parameters can be linked to the flow behaviour of the reservoir, and thus, extract information about the physical model, which can either corroborate results produced by different well testing methods, or add information for the interpretation of the late-time response features. This becomes clear from tools such as the histograms and the scatterplots, that show the form of the marginal densities, and can provide insight into the correlation structure of the posterior, which can help interpreting the behaviour of the response; the step-plots, that can associate the transition parameters with specific response features; and the PCA, which shows how influential each parameter is to changes in the system.

Second, with help of the Bayesian context, and the use of full distributions for the parameters of interest, our method provides a principled way of evaluating the uncertainty of the model, which can be easily visualised with the use of the uncertainty plots for the response, the rates, the pressure and its residuals. It also allows us to account for the error in the independent variables, which is essential due to the nature of the problem. It gives us the ability to incorporate prior knowledge when it is available, and it also allows the marginalisation of parameters, either for efficiency, or because they might not be necessary for the interpretation of the system. Diagnostic tools such as the Geweke, Gelfand and autocorrelation plots can help us assess the convergence and efficiency of the algorithm, while information criteria and Bayes factors assist us to select the most appropriate model.

Finally, the method is based on a solution to the diffusion equation, and thus, it guarantees physical results, which is a potential issue for other methods, especially for noisy data sets, while at the same time, it provides enough flexibility to resemble a vast variety of plausible response shapes.

7.3 Future research

Further extensions and improvements can be implemented in order to advance this work.

Regarding the model optimisation there are a number of different approaches: first, there is ongoing research on faster methods for convolution on non-equispaced time [44], that can be easily implemented on the model and replace the current, relatively slow convolution method. The generation of a radial model, using an alternative parameterisation [50] that allows the treatment of the mobility and diffusivity as continuously varying functions of the distance from the wellbore can extend the flexibility of the model and possibly provide new methods of interpretation. In addition, the migration to a different programming language might improve the response generation, and also help with the efficiency of the MCMC, which is the main source of time consumption.

Alternative MCMC methods, such as the reversible jump MCMC in [32] can potentially facilitate model selection, in the sense that instead of running multiple chains and then using the information criteria or the Bayes factors in order to determine the best model, this can possibly be done 'on the fly' through an appropriate MCMC variation.

Finally, a possible next step is the extension to the multi-well case [20], where there are additional challenges, such as the fact that radial models at different points of the same reservoir are not apparently coherent, and the issue of modelling the interference effects between the wells.

Appendix A

Nomenclature

Table A.1: Nomenclature

Symbol	Description	Location
A	Marginalisation matrix	p.55
C	Wellbore storage coefficient	p.2
$C_D e^{2S}$	Curve match	p.32
C_t	t step covariance matrix	p.74
D	Deviance	p.86
E	Error function	p.17
I_a	Identity matrix	p.54
K	Prior knowledge	p.41
L	Likelihood	p.54
M	Mobility ratio	p.32
N	Number of rates	p.7
P	Pressure (general)	p.2
P_M	Pressure match	p.32
R_D	Dimensionless radius	p.32
S	Wellbore Skin	p.2
S_t	RAM lower-triagonal matrix	p.75
T_j	Time (rate)	p.7

Table A.1 – continued from previous page

Symbol	Description	Location
T_M	Time match	p.32
W	Laplace variable	p.25
X_0	Initial vector	p.66
X_t	Current state	p.66
Y	Proposal state	p.66
a	Acceptance ratio	p.66
a^*	Optimum acceptance ratio	p.77
a_q	Rate inverse gamma shape parameter	p.59
a_p	Pressure inverse gamma shape parameter	p.60
a_{p_0}	Initial pressure inverse gamma shape parameter	p.60
b	Marginalisation vector	p.55
b_q	Rate inverse gamma scale parameter	p.59
b_p	Pressure inverse gamma scale parameter	p.60
b_{p_0}	Initial pressure inverse gamma scale parameter	p.60
c	Marginalisation constant	p.55
$\boldsymbol{c}$	Vectorised convolution integral function	p.45
c_t	Total system compressibility	p.2
d	Parameter dimensions	p.70
e_t	HARM standardised variates	p.77
g	Response function	p.7
h	Formation thickness	p.2
k	Permeability	p.1
m	Number of pressures	p.7
p	Pressure (bottomhole)	p.2
$\tilde{p}$	True pressure	p.7
p_0	Observed initial pressure	p.7
$\tilde{p}_0$	True initial pressure	p.7

Table A.1 – continued from previous page

Symbol	Description	Location
q	Observed rate	p.2
$\tilde{q}$	True rate	p.7
q_n	Transition probability	p.66
r_i	Transition radius	p.30
r_n	HARM uniform sample	p.77
r_w	Wellbore radius	p.2
s	HARM distance	p.77
t	Time	p.2
t_p	Pseudo-production time	p.13
$\tilde{\boldsymbol{y}}$	True independent parameters	p.42
$x_i^{(j)}$	DEzs i state of the j chain	p.80
$\boldsymbol{x}$	Data vector	p.52
z	Response transformation	p.9
z_i	HARM variates	p.76
$\boldsymbol{\Sigma}$	Conditional variance matrix	p.56
$\boldsymbol{\Sigma}_p$	Pressure variance matrix	p.43
$\boldsymbol{\Sigma}_q$	Rate variance matrix	p.43
β	AMM weight	p.74
γ	Euler constant	p.25
γ_r	DEzs tuning parameter	p.79
ϵ	AM error term	p.74
ε	DEzs error term	p.79
ζ	Scalar pressure match weight	p.17
η	Diffusivity ratio	p.32
$\boldsymbol{\theta}$	Set of parameters	p.46
κ	Curvature penalty	p.17
λ	Scalar curvature penalty weight	p.17

Table A.1 – continued from previous page

Symbol	Description	Location
$\boldsymbol{\lambda}$	Data generation parameters	p.46
μ	Viscosity	p.1
μ_c	Conditional mean vector	p.56
ν	Scalar rate match weight	p.17
$\boldsymbol{\sigma}$	Variance parameters	p.46
σ_p^2	Pressure variance	p.43
σ_q^2	Rate variance	p.43
σ_{p0}	Initial pressure variance	p.44
τ	HARM transition parameter	p.77
ϕ	Porosity	p.1

Appendix B

Construction of the C matrix

Given the Equation 1.4.1, we can write the elements of $\boldsymbol{C}$ as:

$$\boldsymbol{C}_{ij}(z) = \int_{-\infty}^{\ln T} \theta_j(t_i - e^{\tau}) e^{z(\tau)} d\tau, \tag{B.0.1}$$

where $i = 1, \ldots, m$ and $j = 1, \ldots, N$, and

$$\theta_j(t) = \begin{cases} 1 & t \in [a_j, b_j] \\ 0 & \text{otherwise.} \end{cases}$$

$z(\tau)$ is a piecewise linear function, interpolated at the nodes $\tau_0, \ldots, \tau_1$. Thus, from Equation B.0.1:

$$\begin{aligned} \boldsymbol{C}_{ij}(z) &= \sum_{k=1}^{n} \int_{\tau_{k-1}}^{\tau_k} \theta_j(t_i - e^{\tau}) e^{\alpha_k + \beta_k \tau} d\tau \\ &= \sum_{k=1}^{n} e^{\alpha_k} \int_{\tau_{k-1}}^{\tau_k} \theta_j(t_i - e^{\tau}) e^{\beta_k \tau} d\tau \\ &= \sum_{k=1}^{n} e^{\alpha_k} C_{ijk}(z). \end{aligned}$$

and

$$C_{ijk}(z) = \int_{I_{ijk}} e^{\beta_k \tau} d\tau,$$

where each of the sets I_{ijk} is either empty, of the form $(-\infty, b]$, or $[a, b] = [\mu - \rho, \mu + \rho$, with midpoint μ and radius ρ, and accordingly, we obtain:

$$C_{ijk}(z) = \begin{cases} 0 & \text{if } I_{ijk} = \emptyset \\ \beta_k^{-1} e^{b\beta_k} & \text{if } \alpha = -\infty \\ 2\rho & \text{if } \alpha \neq -\infty, \beta_k = 0 \\ 2\beta_k^{-1} e^{\mu\beta_k} \sinh \rho\beta_k & \text{if } \alpha \neq -\infty, \beta_k \neq 0. \end{cases}$$

Bibliography

[1] LG Acosta, AK Ambastha et al. 'Thermal well test analysis using an analytical multi-region composite reservoir model'. In: *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers. 1994.

[2] Ram G Agarwal, Rafi Al-Hussainy, HJ Ramey Jr et al. 'An investigation of wellbore storage and skin effect in unsteady liquid flow: I. Analytical treatment'. In: *Society of Petroleum Engineers Journal* 10.03 (1970), pp. 279–290.

[3] Tarek Ahmed and Paul McKinney. *Advanced reservoir engineering*. Elsevier, 2011.

[4] Hirotogu Akaike. 'Information theory and an extension of the maximum likelihood principle'. In: *Selected papers of hirotugu akaike*. Springer, 1998, pp. 199–213.

[5] D Anderson and K Burnham. 'Model selection and multi-model inference'. In: *Second. NY: Springer-Verlag* 63 (2004).

[6] Ake Bjorck. *Numerical methods for least squares problems*. Vol. 51. Siam, 1996.

[7] Dominique Bourdet. *Well test analysis: the use of advanced interpretation models*. Vol. 3. Elsevier, 2002.

[8] Dominique Bourdet, JA Ayoub, YM Pirard et al. 'Use of pressure derivative in well test interpretation'. In: *SPE Formation Evaluation* 4.02 (1989), pp. 293–302.

[9] Dominique Bourdet, Alain C Gringarten et al. 'Determination of fissure volume and block size in fractured reservoirs by type-curve analysis'. In: *SPE annual technical conference and exhibition*. Society of Petroleum Engineers. 1980.

[10] Dominique Bourdet, TM Whittle, AA Douglas and YM Pirard. 'A new set of type curves simplifies well test analysis'. In: *World oil* 196.6 (1983), pp. 95–106.

[11] M Bourgeois, Roland N Horne et al. 'Well test model recognition using laplace space type curves'. In: *SPE formation evaluation* 8.01 (1993), pp. 17–25.

[12] Cajo JF ter Braak and Jasper A Vrugt. 'Differential evolution Markov chain with snooker updater and fewer chains'. In: *Statistics and Computing* 18.4 (2008), pp. 435–446.

[13] Steve Brooks, Andrew Gelman, Galin Jones and Xiao-Li Meng. *Handbook of markov chain monte carlo*. CRC press, 2011.

[14] Siddhartha Chib and Edward Greenberg. 'Understanding the metropolis-hastings algorithm'. In: *The american statistician* 49.4 (1995), pp. 327–335.

[15] Siddhartha Chib and Ivan Jeliazkov. 'Marginal likelihood from the Metropolis–Hastings output'. In: *Journal of the American Statistical Association* 96.453 (2001), pp. 270–281.

[16] J Andrés Christen, Bruno Sansó, Mario Santana-Cibrian and Jorge X Velasco-Hernández. 'Bayesian deconvolution of oil well test data using Gaussian processes'. In: *Journal of Applied Statistics* 43.4 (2016), pp. 721–737.

[17] KH Coats, LA Rapoport, JR McCord, WP Drews et al. 'Determination of aquifer influence functions from field data'. In: *Journal of Petroleum Technology* 16.12 (1964), pp. 1–417.

[18] Mary Kathryn Cowles and Bradley P Carlin. 'Markov chain Monte Carlo convergence diagnostics: a comparative review'. In: *Journal of the American Statistical Association* 91.434 (1996), pp. 883–904.

[19] JA Cumming, DA Wooff, T Whittle, RJ Crossman, AC Gringarten et al. 'Assessing the Non-Uniqueness of the Well Test Interpretation Model Using Deconvolution'. In: *EAGE Annual Conference & Exhibition incorporating SPE Europec*. Society of Petroleum Engineers. 2013.

[20] JA Cumming, DA Wooff, TM Whittle, AC Gringarten et al. 'Multiple well deconvolution'. In: *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers. 2013.

[21] Petros Dellaportas and David A Stephens. 'Bayesian analysis of errors-in-variables regression models'. In: *Biometrics* (1995), pp. 1085–1095.

[22] Sun He-Dong, Liu Yue-wua and Shi Yingc. 'A Well Test Model for Composite Reservoir with Resistance Force on Interface'. In: *Open Petroleum Engineering Journal* 6 (2013), pp. 43–48.

[23] Paul H Garthwaite, Yanan Fan and Scott A Sisson. 'Adaptive optimal scaling of Metropolis–Hastings algorithms using the Robbins–Monro process'. In: *Communications in Statistics-Theory and Methods* 45.17 (2016), pp. 5098–5111.

[24] Alan E Gelfand, Susan E Hills, Amy Racine-Poon and Adrian FM Smith. 'Illustration of Bayesian inference in normal data models using Gibbs sampling'. In: *Journal of the American Statistical Association* 85.412 (1990), pp. 972–985.

[25] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari and Donald B Rubin. *Bayesian data analysis*. Vol. 2. CRC press Boca Raton, FL, 2014.

[26] Andrew Gelman, Jessica Hwang and Aki Vehtari. 'Understanding predictive information criteria for Bayesian models'. In: *Statistics and computing* 24.6 (2014), pp. 997–1016.

[27] Stuart Geman and Donald Geman. 'Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images'. In: *IEEE Transactions on pattern analysis and machine intelligence* 6 (1984), pp. 721–741.

[28] John Geweke et al. *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Vol. 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, 1991.

[29] Wally R Gilks, NG Best and KKC Tan. 'Adaptive rejection Metropolis sampling within Gibbs sampling'. In: *Applied Statistics* (1995), pp. 455–472.

[30] Walter R Gilks, Sylvia Richardson and David Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC, 1995.

[31] Neil J Gordon, David J Salmond and Adrian FM Smith. 'Novel approach to nonlinear/non-Gaussian Bayesian state estimation'. In: *IEE proceedings F (radar and signal processing)*. Vol. 140. 2. IET. 1993, pp. 107–113.

[32] Peter J Green. 'Reversible jump Markov chain Monte Carlo computation and Bayesian model determination'. In: *Biometrika* 82.4 (1995), pp. 711–732.

[33] Alain C Gringarten. *Method for obtaining a dimensionless representation of well pressure data without the use of type-curves*. US Patent 4,607,524. Aug. 1986.

[34] Alain C Gringarten et al. 'From Straight Lines to Deconvolution: The Evolution of the State of the Art in Well Test Analysis'. In: *SPE Reservoir Evaluation & Engineering* 11.01 (2008), pp. 41–62.

[35] Alain C Gringarten et al. 'Practical use of well-test deconvolution'. In: *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers. 2010.

[36] Alain C Gringarten, Dominique P Bourdet, Pierre A Landel, Vladimir J Kniazeff et al. 'A comparison between different skin and wellbore storage type-curves for early-time transient analysis'. In: *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers. 1979.

[37] Alain C Gringarten, Henry J Ramey Jr et al. 'The use of source and Green's functions in solving unsteady-flow problems in reservoirs'. In: *Society of Petroleum Engineers Journal* 13.05 (1973), pp. 285–296.

[38] Heikki Haario, Marko Laine, Antonietta Mira and Eero Saksman. 'DRAM: efficient adaptive MCMC'. In: *Statistics and Computing* 16.4 (2006), pp. 339–354.

[39] Heikki Haario, Eero Saksman, Johanna Tamminen et al. 'An adaptive Metropolis algorithm'. In: *Bernoulli* 7.2 (2001), pp. 223–242.

[40] Florian Hartig, Francesco Minunno and Stefan Paul. *BayesianTools: General-Purpose MCMC and SMC Samplers and Tools for Bayesian Statistics*. R package version 0.1.6. 2019. `https://CRAN.R-project.org/package=BayesianTools`.

[41] DR Horner et al. 'Pressure build-up in wells'. In: *3rd world petroleum congress*. World Petroleum Congress. 1951.

[42] D Ilk, PP Valkó and TA Blasingame. 'Deconvolution of Variable-Rate Reservoir-Performance Data Using B-Splines'. In: (2006).

[43] Robert E Kass and Adrian E Raftery. 'Bayes factors'. In: *Journal of the american statistical association* 90.430 (1995), pp. 773–795.

[44] Jens Keiner, Stefan Kunis and Daniel Potts. 'Using NFFT 3—a software library for various nonequispaced fast Fourier transforms'. In: *ACM Transactions on Mathematical Software (TOMS)* 36.4 (2009), p. 19.

[45] James Kennedy. 'Particle swarm optimization'. In: *Encyclopedia of machine learning* (2010), pp. 760–766.

[46] Fikri J Kuchuk, Richard G Carter, Luis Ayestaran et al. 'Deconvolution of wellbore pressure and flow rate'. In: *SPE Formation Evaluation* 5.01 (1990), pp. 53–59.

[47] Pierre Simon Laplace. 'Memoir on the probability of the causes of events'. In: *Statistical Science* 1.3 (1986), pp. 364–378.

[48] Kenneth Levenberg. 'A method for the solution of certain non-linear problems in least squares'. In: *Quarterly of applied mathematics* 2.2 (1944), pp. 164–168.

[49] Michael M Levitan et al. 'Practical Application of Pressure-Rate Deconvolution to Analysis of Real Well Tests'. In: *SPE Reservoir Evaluation & Engineering* 8.02 (2005), pp. 113–121.

[50] MM Levitan, GE Crawford et al. 'General heterogeneous radial and linear models for well-test analysis'. In: *SPE Journal* 7.02 (2002), pp. 131–138.

[51] Donald W Marquardt. 'An algorithm for least-squares estimation of nonlinear parameters'. In: *Journal of the society for Industrial and Applied Mathematics* 11.2 (1963), pp. 431–441.

[52] Charles Sedwick Matthews and Donald G Russell. *Pressure buildup and flow tests in wells*. Vol. 1. Henry L. Doherty Memorial Fund of AIME, 1967.

[53] CS Matthews, F Brons, P Hazebroek et al. 'A method for determination of average pressure in a bounded reservoir'. In: (1954).

[54] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller and Edward Teller. 'Equation of state calculations by fast computing machines'. In: *The journal of chemical physics* 21.6 (1953), pp. 1087–1092.

[55] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[56] Mustafa Onur, Murat Cinar, Dilhan Ilk, Peter P Valko, Thomas A Blasingame, Peter S Hegeman et al. 'An investigation of recent deconvolution methods for well-test data analysis'. In: *SPE Journal* 13.02 (2008), pp. 226–247.

[57] Martyn Plummer, Nicky Best, Kate Cowles and Karen Vines. 'CODA: Convergence Diagnosis and Output Analysis for MCMC'. In: *R News* 6.1 (2006), pp. 7–11. `https://journal.r-project.org/archive/`.

[58] Kenneth Price, Rainer M Storn and Jouni A Lampinen. *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media, 2006.

[59] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2014. `http://www.R-project.org/`.

[60] HJ Ramey Jr et al. 'Short-time well test data interpretation in the presence of skin effect and wellbore storage'. In: *Journal of Petroleum Technology* 22.01 (1970), pp. 97–104.

[61] Lewis Fry Richardson and J Arthur Gaunt. 'VIII. The deferred approach to the limit'. In: *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character* 226.636-646 (1927), pp. 299–361.

[62] Christian Robert and George Casella. *Introducing Monte Carlo Methods with R*. Springer Science & Business Media, 2009.

[63] Gareth O Roberts, Andrew Gelman, Walter R Gilks et al. 'Weak convergence and optimal scaling of random walk Metropolis algorithms'. In: *The annals of applied probability* 7.1 (1997), pp. 110–120.

[64] Gareth O Roberts and Jeffrey S Rosenthal. 'Examples of adaptive MCMC'. In: *Journal of Computational and Graphical Statistics* 18.2 (2009), pp. 349–367.

[65] Gareth O Roberts, Jeffrey S Rosenthal et al. 'Optimal scaling for various Metropolis-Hastings algorithms'. In: *Statistical science* 16.4 (2001), pp. 351–367.

[66] A Roumboutsos, G Stewart et al. 'A direct deconvolution or convolution algorithm for well test analysis'. In: *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers. 1988.

[67] Abdurrahman Satman, Mauricio Eggenschwiler, Henry J Ramey Jr et al. 'Interpretation of injection well pressure transient data in thermal oil recovery'. In: *SPE California Regional Meeting*. Society of Petroleum Engineers. 1980.

[68] T von Schroeter, F Hollaender, AC Gringarten et al. 'Deconvolution of Well-Test Data as a Nonlinear Total Least-Squares Problem'. In: *SPE Journal* 9.04 (2004), pp. 375–390.

[69] Gideon Schwarz et al. 'Estimating the dimension of a model'. In: *The annals of statistics* 6.2 (1978), pp. 461–464.

[70] Robert L Smith. 'Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions'. In: *Operations Research* 32.6 (1984), pp. 1296–1308.

[71] Statisticat and LLC. *LaplacesDemon: Complete Environment for Bayesian Inference*. R package version 16.1.1. Bayesian-Inference.com, 2018. `https://web.archive.org/web/20150206004624%5C%5C/http://www.bayesian-inference.com/software`.

[72] Harald Stehfest. 'Algorithm 368: Numerical inversion of Laplace transforms [D5]'. In: *Communications of the ACM* 13.1 (1970), pp. 47–49.

[73] George Stewart. *Well test design & analysis*. PennWell Corporation, 2011.

[74] Rainer Storn and Kenneth Price. 'Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces'. In: *Journal of global optimization* 11.4 (1997), pp. 341–359.

[75] Cajo JF Ter Braak. 'A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces'. In: *Statistics and Computing* 16.3 (2006), pp. 239–249.

[76] LG Thompson, AC Reynolds et al. 'Analysis of variable-rate well-test pressure data using Duhamel's principle'. In: *SPE formation evaluation* 1.05 (1986), pp. 453–469.

[77] Djebbar Tiab. 'Analysis of pressure and pressure derivative without type-curve matching—Skin and wellbore storage'. In: *Journal of Petroleum Science and Engineering* 12.3 (1995), pp. 171–181.

[78] AF Van Everdingen, William Hurst et al. 'The application of the Laplace transformation to flow problems in reservoirs'. In: *Journal of Petroleum Technology* 1.12 (1949), pp. 305–324.

[79] Matti Vihola. 'Robust adaptive Metropolis algorithm with coerced acceptance rate'. In: *Statistics and Computing* 22.5 (2012), pp. 997–1008.

[80] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
`http://ggplot2.org`.

[81] Liehui Zhang, Jingjing Guo and Qiguo Liu. 'A well test model for stress-sensitive and heterogeneous reservoirs with non-uniform thicknesses'. In: *Petroleum Science* 7.4 (2010), pp. 524–529.