Modelling Discrepancy in Bayesian Calibration of Reservoir Models

Behzad Nezhad Karim Nobakht

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

Energy, Geoscience, Infrastructure and Society

Institute of Petroleum Engineering

December 2018

ABSTRACT

Simulation models of physical systems such as oil field reservoirs are subject to numerous uncertainties such as observation errors and inaccurate initial and boundary conditions. However, after accounting for these uncertainties, it is usually observed that the mismatch between the simulator output and the observations remains and the model is still inadequate. This incapability of computer models to reproduce the real-life processes is referred to as model inadequacy.

This thesis presents a comprehensive framework for modelling discrepancy in the Bayesian calibration and probabilistic forecasting of reservoir models. The framework efficiently implements data-driven approaches to handle uncertainty caused by ignoring the modelling discrepancy in reservoir predictions using two major hierarchical strategies, parametric and non-parametric hierarchical models.

The central focus of this thesis is on an appropriate way of modelling discrepancy and the importance of the model selection in controlling overfitting rather than different solutions to different noise models.

The thesis employs a model selection code to obtain the best candidate solutions to the form of non-parametric error models. This enables us to, first, interpolate the error in history period and, second, propagate it towards unseen data (i.e. error generalisation). The error models constructed by inferring parameters of selected models can predict the response variable (e.g. oil rate) at any point in input space (e.g. time) with corresponding generalisation uncertainty.

In the real field applications, the error models reliably track down the uncertainty regardless of the type of the sampling method and achieve a better model prediction score compared to the models that ignore discrepancy.

All the case studies confirm the enhancement of field variables prediction when the discrepancy is modelled. As for the model parameters, hierarchical error models render less global bias concerning the reference case. However, in the considered case studies, the evidence for better prediction of each of the model parameters by error modelling is inconclusive.

DEDICATION

To my Mother and Father

ACKNOWLEDGEMENT

I would like to thank the many people who have given me the opportunity to do my PhD study.

First and foremost, I would like to thank my supervisors Prof Mike Christie, and Prof Vasily Demyanov, for their immense help, valuable comments, and guidance. I appreciate all their contributions of continuous support of my PhD research and writing of this thesis.

Besides my supervisors, I wish to express my deep and sincere gratitude to the members of the Institute of Petroleum Engineering (IPE), particularly for supporting my PhD study funding (including Ali Danesh Scholarship and Adrian Todd Golden Key Award).

My sincere thanks also go to all my colleagues in Uncertainty Quantification group at Heriot-watt University who made my research an entertaining and memorable experience.

Edinburgh, September 2018

Behzad Nezhad Karim Nobakht

ACADEMIC REGISTRY

Research Thesis Submission

Name*:*	Behzad Nezhad Karim Nobakht		
School:	School of Energy, Geoscience, Infrastructure and Society		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought:	PhD in Petroleum Engineering

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.
6) I confirm that the thesis has been verified against plagiarism via an approved plagiarism detection application e.g. Turnitin.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate*:*		Date:	

Submission

Submitted By *(name in capitals):*	
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals):*			
Method of Submission *(Handed in to SSC; posted through internal/external mail):*			
E-thesis Submitted ***(mandatory for final theses)***			
Signature:		Date:	

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PUBLICATIONS BY THE CANDIDATE

Nobakht, B. N. K. & Christie, M. Model Prediction under Uncertainty Using Hierarchical Models. 79th EAGE Conference and Exhibition 2017, 2017.

Nezhad Karim Nobakht, B., Christie, M. & Demyanov, V. Model Selection for Error Generalization in History Matching. SPE Europec featured at 80th EAGE Conference and Exhibition, 2018. Society of Petroleum Engineers, SPE-190778-MS.

Nezhad Karim Nobakht, B., Christie, M. & Demyanov, V. Modelling Discrepancy for History Matching of Non-stationary Time-series, SPE Journal (in progress).

Chapter 1 – INTRODUCTION

The world economy is anticipated to approximately double in the next 20 years, with growth averaging 3.4% per year (BP, 2017). The growing world economy leads to higher energy consumption: the information from the International Energy Outlook 2016 (EIA, 2016) illustrates a significant growth in worldwide energy demand until 2040.

Among all the sources of energy, hydrocarbons and petroleum fuels remain the largest source of energy, even though their share of total world marketed energy consumption drops from 33% in 2012 to 30% in 2040 (EIA, 2016).

The world proved oil reserves have more than doubled since 1980 (Figure 1.1): for every barrel of oil produced more than two new barrels have been found (BP, 2017). However, ExxonMobil's analysis of the energy outlook expresses that the current production rate cannot meet the ever-growing demand for energy and new development/production programs are required (ExxonMobil, 2004).

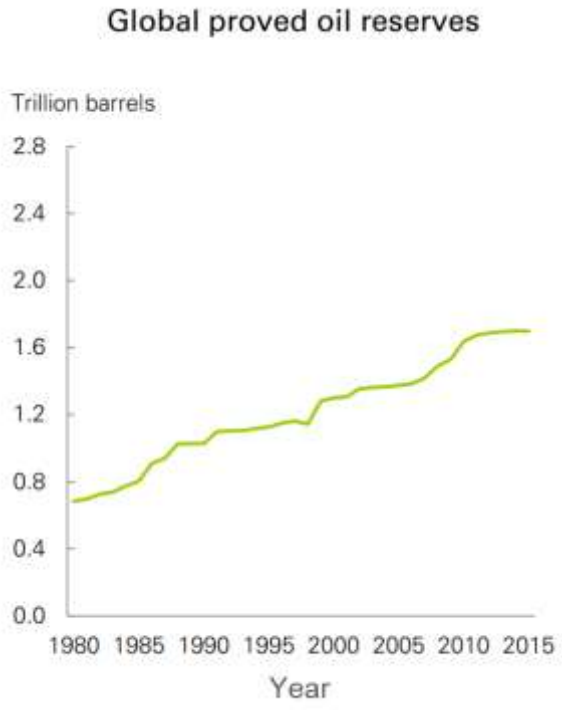

Figure 1.1 Global proved oil reserves from 1980 to 2015 (from BP, 2017)

The goal of oil reservoir development plan is to speed up oil and gas production with maximum recovery factors and at the lowest cost possible (Pacheco and Vellasco, 2009;

Lima et al., 2015). However, optimisation of production plans over possible alternatives requires a broad understanding of the oil reservoir performance prediction (Lima et al., 2015).

In reservoir engineering, the oil production optimisation means exploring production strategies that are economically more valuable. In this context, the application of intelligent systems and computer-based reservoir models is a critical factor in making strategic and operational decisions (Pacheco and Vellasco, 2009).

Computer-based modelling and simulation of a physical process provide a mathematical description of the real system behaviour based on physical principles (Hazelrigg, 1999). Computer models help engineers and scientists to understand and predict a given phenomenon. Although, models do not contain the full complexity of the true physical phenomenon (Ordaz-Hernandez et al., 2007). Instead, they provide a less complex (but valuable) abstraction in that simplifications usually alter the realism (Ordaz-Hernandez et al., 2007; Hazelrigg, 1999).

In the upstream sector of oil and gas industry, reservoir simulation models are broadly used for the field development, operational decision making, and further investments of the oil fields (Pacheco and Vellasco, 2009). These models are constructed in accordance with the analysis of many surface/subsurface physical and chemical measurements across the oil field.

Reservoir models are valuable tools for answering reservoir management questions and finding possible solutions to reservoir problems. The initial reservoir model requires some data representing fluid characteristics, multiphase flow features such as relative permeability, and well performance. Therefore, the selected reservoir model should provide a sufficient description of those parameters that dominate the fluid flow associated with the designed simulation study.

A reservoir model represents the significant geological characteristics such as faults, variation in reservoir properties, stratigraphy. Even though new reservoir models can handle more complicated studies, computer models are not perfect meaning that they can never recover the real physics of the oil reservoir (Al-Yahya, 2010). These models are designed to be appropriate approximations sufficient to inform the decision which the model is intended to support.

Because simulation models are not perfect, reservoir engineers try to adjust parameters of the simulation model to match the predicted output with the available production or seismic data from the field. This iterative model calibration process is known as history matching in the oil industry (Leo et al., 1986).

History matching is used to reduce the discrepancy between the observed data and the simulation output (Figure 1.2) and requires running many simulation models (Li et al., 2001; Gilman and Ozgen, 2013; Cancelliere et al., 2011).

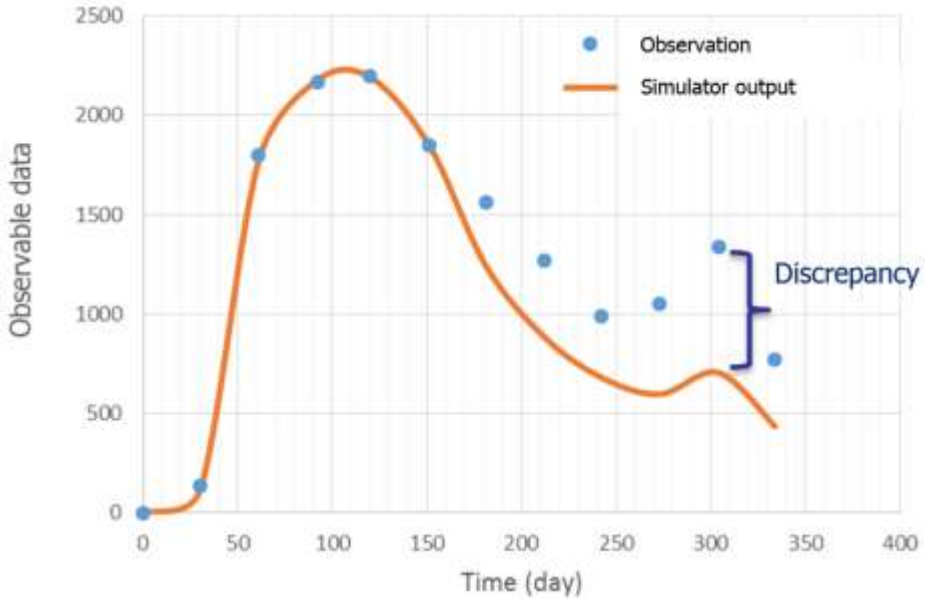

Figure 1.2 The discrepancy between a reservoir simulation model and observation data

The solutions to the history match process are later ranked based on their quality, such that the models with lower discrepancy/misfit values gain a higher rank (Bouzarkouna and Nobakht, 2015).

The misfit values show how well the simulated model can fit the measured data (Christie et al., 2002; Bouzarkouna et al., 2014; Bouzarkouna and Nobakht, 2015). Therefore, we need a misfit function (or objective function) that assigns a likelihood to each calibration model after comparing them to the observed data (Glimm and Sharp, 1999).

History match models should fit the available data, but their match quality does not necessarily equate to forecast value. If the matched reservoir and displacement processes

carry predictive information for the forecast (i.e. continuity), and the model case is based on the correct interpretation, then the history matched models have high predictive capability.

1.1 Statement of the problem

Theoretically, the likelihood function consists of the accessible data provided by a sample of the model contained in the observed data (Ratmann et al., 2009; Sargsyan et al., 2018). The likelihood of the calibration model m is the probability that the model m fits the observed data.

Furthermore, the definition of the likelihood function is in line with the assumptions about the errors including experimental errors of the observation data (O'sullivan and Christie, 2005a). For instance, if the measurement errors are random, then each pair of them are uncorrelated. The assumption of uncorrelated measurement noise is often implicitly used in the likelihood function, where a diagonal error covariance matrix represents the measurement errors (Seiler et al., 2011).

Now, if we consider that the errors are independent and identically distributed (i.i.d.) and follow Gaussian statistics, then the likelihood of observing data O given a model m is the exponential of the negative misfit:

$$\mathcal{L} = P(O|m) = \exp(-misfit) \tag{1-1}$$

Next, if there exists a known measure of uncertainty σ for the observation collected throughout the time steps j=1,…,m, then the misfit is proportional to the discrepancy between the reservoir simulator output S and observation as

$$misfit = \frac{1}{2}\sum_{j}^{m}\left(\frac{S_j - O_j}{\sigma}\right)^2 \tag{1-2}$$

In the above equation, errors follow uncorrelated Gaussian statistics with mean zero. The equation gives a straightforward definition of the mismatch known as standard linear least-squares (LSQ) from which we assess the quality of simulation models.

In the oil industry, like many other disciplines, measurement errors are considered as the dominant source of uncertainty in simulations with the discrepancies being the difference between observation and simulated data (Yusuf et al., 2018; Jones and Mitchell, 1978). It is often assumed that the measurement errors are independent and identically distributed (i.i.d) for all field variables (Nicotra et al., 2005; Rotondi et al., 2006; Erbas and Christie, 2007). Therefore, the standard least-squares (LSQ) misfit given in Eq. (1-2) is commonly used as the misfit/objective function in the history matching (Kuznetsova, 2017).

1.1.1 Sources of discrepancy in reservoir modelling

Simulation models of physical systems are subject to numerous uncertainties such as observation errors and inaccurate initial and boundary conditions (Glimm and Sharp, 1999). However, after accounting for these uncertainties, it is usually observed that the mismatch between the simulator output and the observations remain and the model is still inadequate (Christie et al., 2005).

Despite being biased, the inadequate model may be the best that is available since adjusting the model to remove the discrepancy, is often infeasible (Pernot and Cailliez, 2017; Rabosky and Goldberg, 2015; Jones and Mitchell, 1978). Therefore, the existing inadequate model is employed to make predictions of unmeasured quantities with the discrepancy often being ignored (Pernot and Cailliez, 2017).

Several reasons may cause the discrepancy between the simulators output and the real reservoir behaviour (see Figure 1.3).

There are 3 different ways of handling discrepancy in the calibration of reservoir models to data.

First, we can improve our reservoir model which requires a better understanding of physics, higher computation time and more expensive technology (Ling et al., 2014; Christie et al., 2005).

Second, we can rely on the simple assumptions about discrepancy (e.g. Gaussian white noise) which yields an overconfident prediction of the future model behaviour (O'Sullivan and Christie, 2006).

The third choice, as highlighted in Figure 1.3, is to account for the discrepancy by exploring possible correlation structures of errors using all information at hand and carry it forward to the misfit formulation (Pernot and Cailliez, 2017; Morrison et al., 2018). In this study, we do a full investigation of all types of information that can reliably account for the modelling discrepancy in simple and complex case studies.

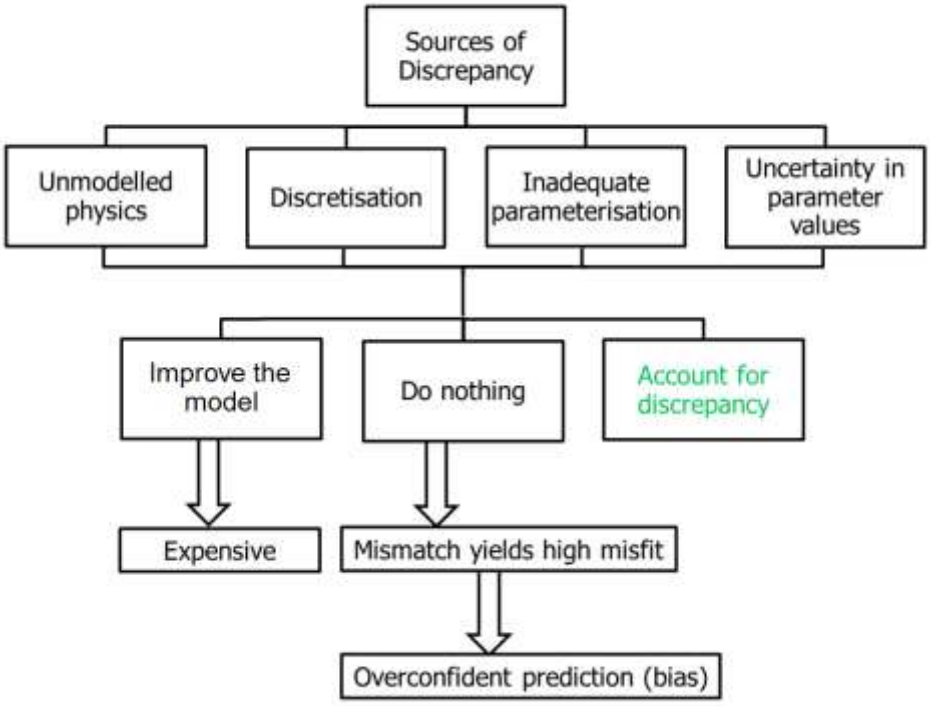

Figure 1.3 Sources of discrepancy between reservoir models and real-life reservoirs

In computer modelling of physical processes, the simulation model is unable to produce a perfect match, even when the actual quantities of the physical parameters are known (Kennedy and O'Hagan, 2001) due to the inherent inadequacy of the computer models (i.e. model error). This incapability of computer models to reproduce the real-life processes is referred to as model inadequacy (Kennedy and O'Hagan, 2001), model error (Sargsyan et al., 2018), model bias (Del Giudice et al., 2015), or model discrepancy (Arendt et al., 2012).

Both measurement errors and model errors determine the degree of accuracy that we require in matching our models to the field data (Del Giudice et al., 2015). If we ignore the model error, the estimation of uncertainty becomes biased, because the probability distribution of some parameters is far from the truth (Stephen et al., 2007; Vink et al., 2015). To avoid this bias in the prediction of field variables, we must account for the modelling discrepancy.

Modelling errors also play a pivotal role in controlling the accuracy of estimation (Christie et al., 2005). Typically, computer models rarely calculate the exact quantity that is measured. For example, in measuring the bottom hole pressure of a well, there will be a gauge somewhere close to the perforations which are measuring fluid pressure in flow that may be two-phase and will need correcting to the exact location of the perforations. Meanwhile, the simulator calculates grid-block average pressure and then assumes homogeneous sub-grid properties and radial flow in a well model to estimate the bottom hole pressure.

Figure 1.4-a, b shows how the LSQ model fails to reliably estimate the production profiles of a reservoir field variable and its corresponding uncertainty when the model error is ignored. The history match adjusts the model parameters to capture the unknown truth about the field variable in the history match period and then predict the unseen data (the shaded grey area in Figure 1.4). However, the simulation model (the blue line) is subject to mismatch, and the estimated uncertainty cannot capture most of the true response of the field variable. The reason is that the estimated measurement errors in Figure 1.4-b seem to have underestimated the uncertainty and yield narrow (overconfident) prediction.

Now, if we find a way to reliably estimate the model discrepancy and add it to the simulator output, then we arrive at the green line in Figure 1.4-c which has less mismatch to the truth. In addition, the error model assigns higher values for standard deviations (the black error bars) which are more significant than the measurement uncertainty. The reason is that the standard deviations estimated by error models include both measurement and model errors.

Adding the estimated model bias to the simulator output will lead to a more reliable prediction of the field variable, where the estimated confidence interval is likely to cover the true reservoir response (see Figure 1.4-d).

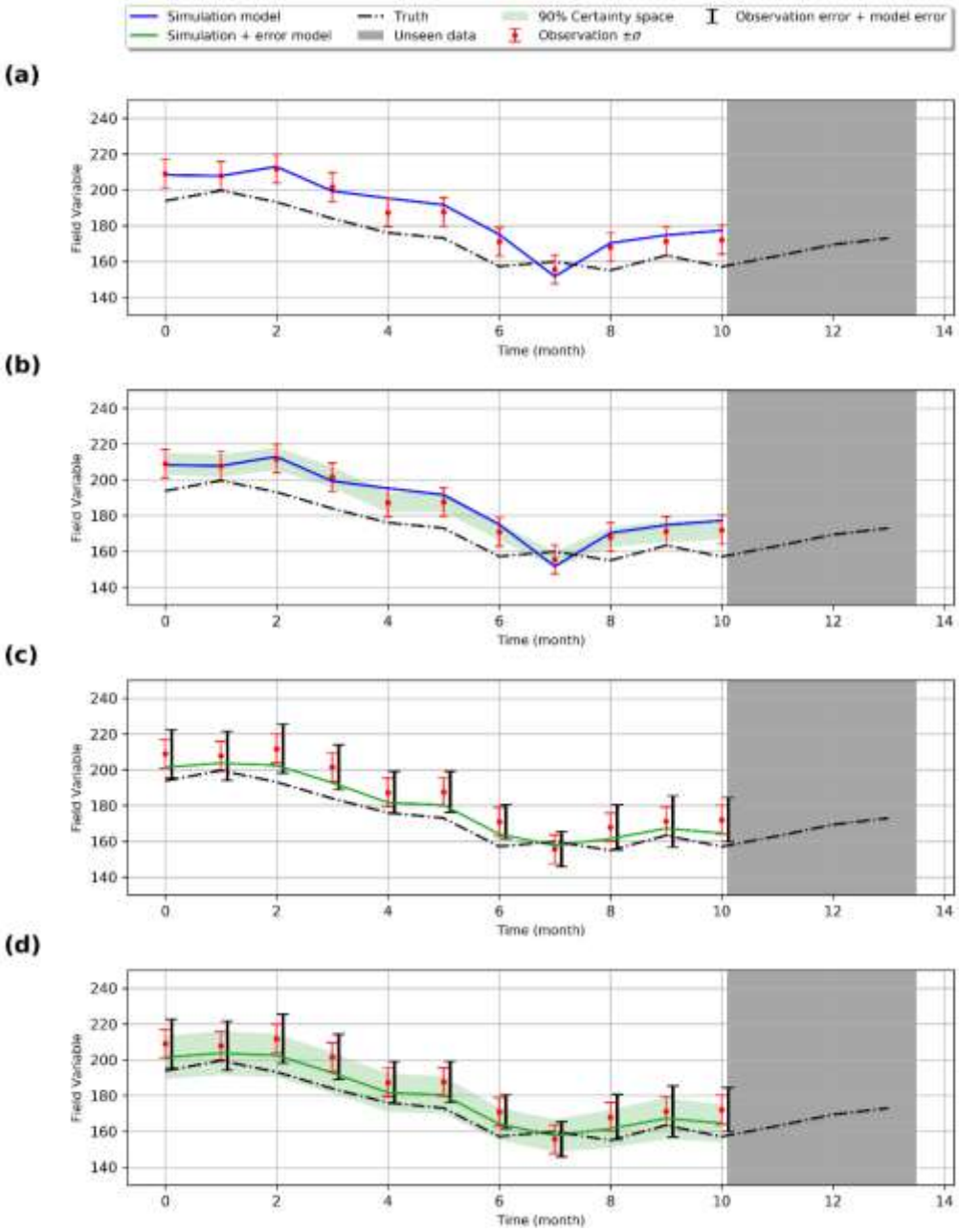

Figure 1.4 Modelling discrepancy in history matching: **a)** The history match runs simulations to match the observation data with known uncertainty $\pm\sigma$. **b)** The estimated 90% confidence interval fails to cover the truth. **c)** An error model estimates new σ (black error bars) and adds the estimated modelling discrepancy to the simulator output which reduces the discrepancy. **d)** Prediction under new estimated uncertainty leads to a wider confidence intervals covering the entire true response.

1.2 The predictive domain in oil reservoirs prediction

Oil companies spend significant time and effort forecasting for the long-term objectives, where estimations of hydrocarbons reserves and production are required. However, many other business intentions need estimates of production profiles to be made over shorter time domains (e.g. short-term forecasting of pressure decline). Such a short-term predictive domain requires no difference in procedure, in principle, to a long-term prediction. However, the critical features influencing the short-term prediction may be different from those of a long-term forecast.

The initial reservoir model is created by interpreting essentially static data, such as surfaces induced from seismic data, well logs analysis, core analysis, stratigraphic studies, geostatistical and geological information. Depending on the size of the field, this task may require plenty of efforts and time. By the time the static model is constructed, an agreement between the correct workflow process and the most uncertain and relevant parameters and their respective ranges of variability must be specified (Landa, 2001).

The principal sources of discrepancy are the inability of the model to represent heterogeneity and limitations in parameterisation. It is highly questionable that statistics, no matter how sophisticated, can predict events or features which are not represented in the reservoir model nor for which there is information in the observed data.

Also, in the calibration process, the reservoir models will be matched to imperfect data. Then, the question is whether there is some form of continuity between the model and imperfect observations, such that the discrepancy can be described statistically and extrapolated.

The underlying problem concerns the size of the predictive domain, within which it can be assumed there is some form of continuity of the represented drainage volumes and displacement processes, e.g. forecasting under the continuation of current well production and displacement/depletion process. In other words, we need to balance between the size of training (history) and forecast intervals. For instance, if there are five years of history available for calibration, then the predictive domain of 3 years can be considered for the forecast.

The general assumption in this thesis is that the modelling process is data led in that interpretation of data leads to model. However, in reservoir modelling, it is well

recognised that there is data insufficiency. Thus the modelling process must be interpretation led such that interpretation leads to a model conditioned by data. This is a critical difference in that the role of data is secondary to the interpretation. For example, in a reservoir model, expectations of the reservoir extent, quality and connectivity are primarily related to the depositional system (e.g. shoreface, fluvial deposits, carbonate platform).

Figure 1.5 shows the prediction strength of a statistical approach in short-term and long-term predictive domains for two adjacent wells. When the amount of the observed data is limited to 161 observation points (Figure 1.5-a), the long-term forecast becomes unreliable as the statistical model has not trained any point in the production decline phase. However, for short-term forecasting, the predictive domain suggests that the statistical model is reliable for only the first 32 points in the forecasting period. Once more data become available (Figure 1.5-b, c), the predictive domain can cover more points in the forecast period.

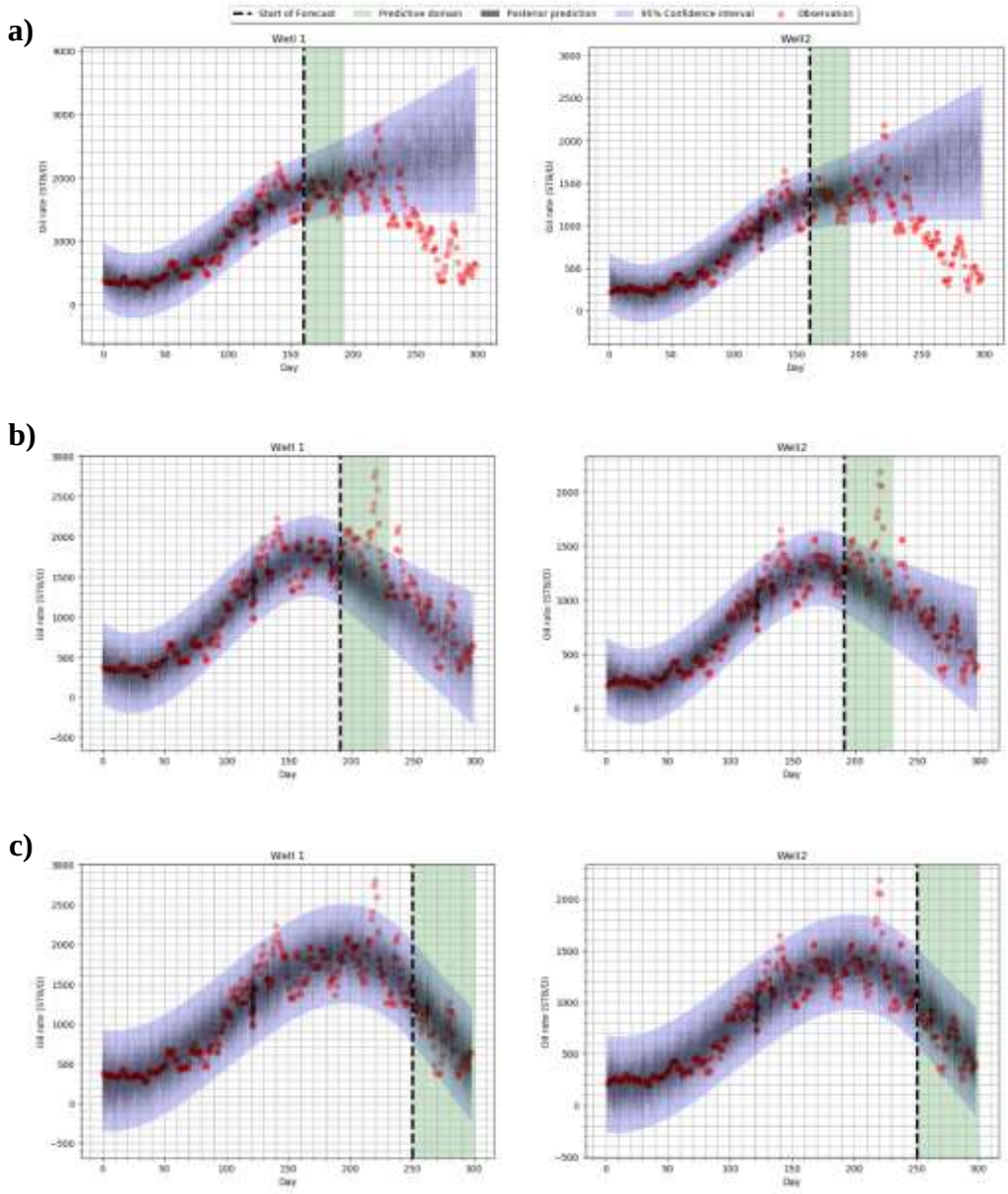

Figure 1.5 Schematic illustration of prediction strength of a statistical approach in short-term and long-term predictive domains for two adjacent wells: **a)** the history match phase is too small causing unreliable long-term forecasting and small predictive domain; **b and c)** when more data become available for history match, the predictive domain enlarges while the statistical approach rigorously predicts the future data.

Short-term forecasting is typically limited to existing wells, relatively short period and current flowing conditions, in which conditions it is reasonable to assume continuity of

information from the recent historical data. Consequently, short-term forecasting is considered in Chapter 6 where our statistical approach predicts the future data.

To sum up, error modelling by itself does not circumvent the limitations of long-term predictions. It is an approach to quantify the level of confidence in a computer-based prediction from a rigorous analysis of the source and extent of errors concerning the prediction. Hence, the metric of success in error modelling is the confidence it provides that the errors are of a particular magnitude—not necessarily that they are small (Christie et al., 2005).

1.3 Summary of chapters

This thesis is set out as follows:

Chapter 2 presents a literature review of reservoir modelling, simulation, calibration and prediction. It also discusses different types of data collected throughout the reservoir life to build reservoir models with a focus on production data and their associated errors. Next, the importance of match quality function in history matching and probabilistic forecasting of reservoir models are discussed. Finally, the chapter reviews some stochastic optimisation techniques employed in the thesis to generate an ensemble of simulation models. The simulation models are then used to approximate posterior probabilities of field production profiles by use of standard statistical approaches.

Chapter 3 gives a literature review of sources of discrepancy in computer modelling of real-life systems. This chapter demonstrates why ignoring modelling discrepancy can lead to bias prediction, and how accounting for discrepancy can improve the predictive performance of the system. The chapter also provides mathematical frameworks (error models) that plug modelling discrepancy into the match quality function. To do so, a parametric statistical model is built above the simulation model to establish a parametric hierarchical model. Our parametric model, however, suffers from its incapability to generalise into unseen data.

Chapter 4 explores two significant aspects of error modelling throughout the history matching of oil reservoirs: the non-parametric hierarchical modelling of discrepancy and the model selection problem. The non-parametric models place flexible priors on functions that are generalisable throughout the entire input space. This enables the error

models to generalise towards the forecast period which is of high importance for reservoir engineering problems. Finally, Chapter 4 introduces a workflow for implementation of error models within history matching and forecast.

Chapter 5 tests the parametric hierarchical error models on three experimental cases of the Teal South reservoir model and compares them to standard linear least-square models (LSQ) that ignore the model discrepancy. This chapter also quantifies model prediction improvement gained by using error modelling in history matching of reservoir models.

Chapter 6 applies non-parametric hierarchical models to the real case study, the Zagadka oilfield. Different solutions to non-parametric hierarchical models are provided to predict the future behaviour of the field variables. This chapter also examines the predictive performance of hierarchical modelling of discrepancy for estimates of model parameters and production profiles.

Chapter 7 concludes the thesis with a review of the chapter's results, significant contributions, and key findings. It also presents some recommendations for future research work on error modelling.

Chapter 2 – Reservoir modelling, simulation and history matching

2.1 Importance of reservoir modelling

Decision making in oil industry investigates the influence of multiple decisions that can govern the direction of billions of dollars. With rising business complexity in the oil and gas sector, making proper and informed decisions are becoming essential to improving the operating business performance. The decision-makers use some decision-analysis tools that combine information from different sources. Many of the influential factors may come from economic, technical and political uncertainties associated with the petroleum industry (Garb, 1988).

Reservoir modelling answers the question of how future performance of a hydrocarbon reservoir varies under different field development scenarios. A reservoir model is designed to be an appropriate representation of the subsurface sufficient to describe the range of uncertainty of outcomes of a development plan, to evaluate and support the development decision.

Reservoir models should provide a sufficient description of those parameters that dominate the fluid flow associated with the designed simulation study. In the oil and gas industry, a reservoir model refers to a computer model encompassing all the characteristics of the oil field reservoir. The selected reservoir model simulates the behaviour of the fluids flowing through the reservoir under different conditions and helps engineers find solutions to maximise the production.

Reservoir models set out a mathematical representation of the static and the dynamic description of petroleum reservoirs under study. The static model requires geologists and geophysicists to build a numerical equivalent of a 3D portrait of physical quantities in the rock (i.e. geological model). Then, reservoir engineers use the static model to construct a dynamic model that resolves changes in reservoir pressure, fluid saturations and flow properties (i.e. simulation model). The results of a simulation model are finally used for enhancing estimation of reserves, decision making under different development plans, well placement optimisation and, more generally, reservoir management policies. The reservoir study must address the following subjects:

- General investigation and modelling of rock characteristics, fluid flow properties and reservoir structure
- Estimation of recoverable hydrocarbon and hydrocarbon in place

- Exploring best development plans and their associated uncertainty
- Identification of the underlying drive mechanism of the petroleum reservoir
- management strategies and risk analysis

2.2 Data collection and integration

The primary step for the establishment of a static reservoir model is a database creation of all available data. The data mainly stem from seismic data, sedimentological/petrographic data, pressure and production test data, field/ well observation data, well logs and core data. In oil and gas reservoir modelling, a combination of dynamic data along with static data improves the quality of the reservoir models produced and provides the practitioners with a better idea for reservoir management. Consequently, the uncertainty of simulated scenarios is reduced, providing an unbiased economic evaluation of reservoir (Cunha, 2003; Riani et al., 2012; Sharifi et al., 2014).

One of the most significant obstacles in reservoir modelling is the reasonable combination of dynamic and static data (Sharifi et al., 2014). This is especially true where the modelling is done in a step-wise fashion, dealing with static and dynamic data matching in different tasks (Sætrom et al., (2016b)).

Several studies demonstrate that data integration can remarkably minimise the uncertainty in predictive reservoir performance and enhance the reservoir model as long as critical elements are not missing from the model. (Betz, 2015; Rushing and Newsham, 2001; Ehigie, 2010). Moreover, inefficient integration of the static and dynamic data in the process of facies modelling can drastically limit the predictability of the simulated reservoir models (Sætrom et al., (2016a), Sætrom et al., (2016b), Perrone et al., 2017).

Reasonably, this can restrict the interaction between various subsurface disciplines throughout the model specification and dynamic data matching. Thus, the output models might excellently honour the current dynamic data observations but ultimately fail to match the static data and the geology of the oilfield.

Sætrom et al. (2016a) addressed the integration problem through an ensemble-based method plugged into an adaptive pluri-Gaussian facies modelling scheme. They tested their method on a medium size case study with 15 years of production data. Meanwhile

the dynamic data conditioning, clear trends were learned in the facies model throughout the reservoir, which presented significant evidence of the expected facies distribution and associated connectivity. Consequently, with an enhanced description of the reservoir and subscale physics through the consistent combination of different types of data, we improve reservoir management and decisions under uncertainty (Perrone et al., 2017; Sharifi et al., 2014).

Because there may exist many plausible reservoir models matching the data (non-uniqueness nature of inverse modelling), not all of them describe the underlying geology. Complex reservoir features such as channels and fractures enormously influence reservoir production while any miscalculation in the integrity of such features causes unrealistic reservoir performance prediction.

Due to the sparseness of data, integration of secondary data in reservoir models is a vital and challenging task. Secondary data such as seismic data often give indirect knowledge about spatial variation of reservoir properties.

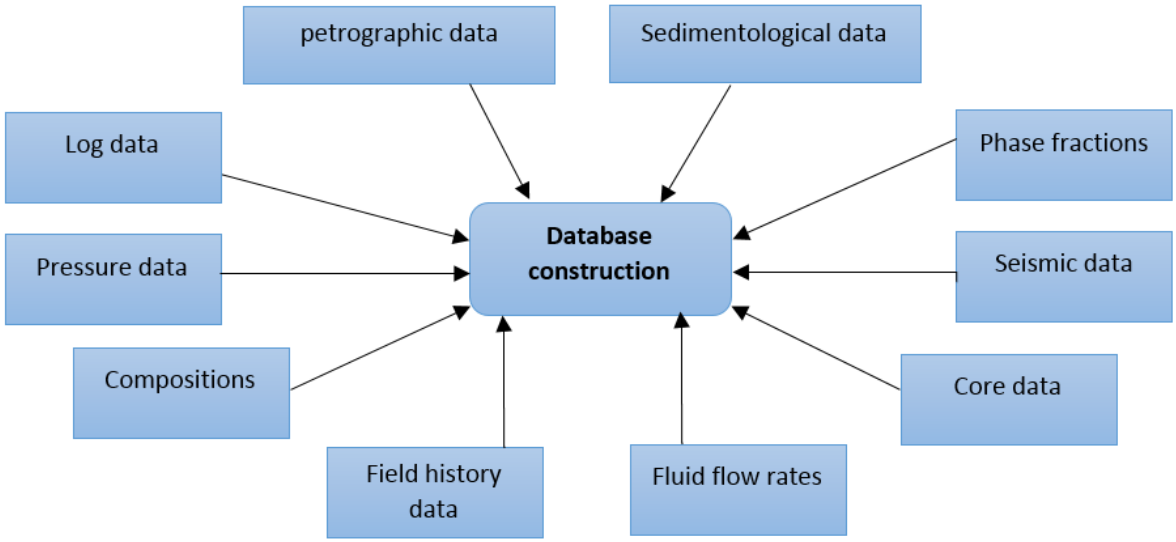

Figure 2.1. Conventional sources of data in oil and gas reservoir studies

As companies exert several integration solutions to handle specific needs, the number of individual tools optimised to handle specific kinds of data rises. Integration of data produced by these tools causes problems for businesses as data integration turns to be fragmented and complicated while lacking coherence throughout the modelling.

Not having a data integration standard across the company makes it difficult for businesses to store and manage information because using many tools can become costly, ineffective and risky (Mu and Kuang, 2010). Also, certain data types call for special tools, thereby driving businesses to implement integration with multiple vendors. With a multitude of integration solutions across the work, there is a lack of uniformity and simplicity which results in overall loss and inefficiency.

2.3 Types of data

2.3.1 Seismic data

Seismic data deliver a time-picture of the subsurface structure using 2D/3D refraction, reflection and shear wave data. For instance, in the 3D seismic reflection, practitioners designate many lines of receivers over the surface and employ lines of source points set out orthogonally to the receivers. The quality of the sub-surface image obtained may pertain to the statistical variation of the information gathered for each cell of sub-surface coverage (i.e. bin). The number of observations collected from the echoes at a particular area increases the chance of rebuilding the subsurface geological shape.

In general, seismic data analysis implies tracing and correlating along connected reflectors throughout the 2D or 3D dataset and utilising these as the evidence for the geological interpretation. The geological interpretation then produces structural maps that reflect the spatial variation in depth of specific geological layers. Using these maps makes an essential contribution to identifying hydrocarbons in place and, do volume calculations, and create the models of the subsurface.

Reflection seismic, for instance, allows for image variations in the subsurface geology by prompting an acoustic wave from near the surface and receiving the echoes from more profound stratigraphic boundaries. The sound moves towards the subsurface in the form of a spherical wavefront where boundaries between various types of rocks will reflect or transmit the waves. Then, at the surface, the geophones observe the returning signals. At the final stage, the signals distinguished by the geophones will be recorded and delivered to data processors.

Seismic interpretation is always subject to uncertainty as a particular dataset may have multiple solutions that match the data. In such a condition, analysis requires more data to

restrict the solution. Therefore, geologists and geophysicists can make use of additional borehole logging and seismic acquisition for gaining a clearer picture of the study.

2.3.2 Well Log and core data

Study of the residual fluid content with supplementary test data gives insight into the uncertain response to well treatment, future reservoir performance, and downhole log understanding. Core data analysis is the standard approach to directly measure the earth's subsurface by examining samples, or cores. Conventional core analysis provides the most basic data required regarding either presence/type of hydrocarbons or lithology of the rock that might be undetectable through downhole logging measurements alone.

To do core analysis, engineers take samples from the formation using specific coring tools. These tools collect the cores and pull them back to the surface. Laboratories then use the cores for calibration of well logs and detection of changes in reservoir characteristics. Analysts then use cores to characterise pore systems in the rock and model reservoir behaviour to optimise production based on the interpretation of core permeability, porosity, grain density, fluid saturation, lithology and texture (Keelan, 1972; Bergosh et al., 1985)

Conditioning the log and core data for computations of the different petrophysical parameters involves adjustments from surface-to-reservoir conditions, normalisation, and environmental-correction factors. Because engineers and geologists cannot evaluate the rock formations in their original place, they lower specific tools called "logging tools" down into the borehole. These tools measure the subsurface properties as a series of observation at different depths known as well logs.

The data are recorded either in a real-time mode at the surface or along the hole to an electronic data format. We usually perform Well log operations either during the process of drilling to supply real-time data about the formations influenced by the borehole or once the well has reached the total depth (Gearhart et al., 1981).

Well Logging tools can address different logs such as resistivity logs for formation/ mud resistivity, gamma ray for correlation log, sonic logs for formation interval transit time, neutron porosity log for log porosity and density, and borehole imaging for detailed reservoir description. Borehole images give additional insight into sedimentological information of the rock. However, information quality, tool resolution, and borehole

coverage constrain the reliability of the study (Varhaug, 2016). As far as the equipment collect the data digitally, we can quickly learn thickness distribution and statistical analysis to enhance hydrocarbon recovery eventually. By using different well logging tools and comparing many logs from different wells, engineers can develop promising production plans for the oilfield.

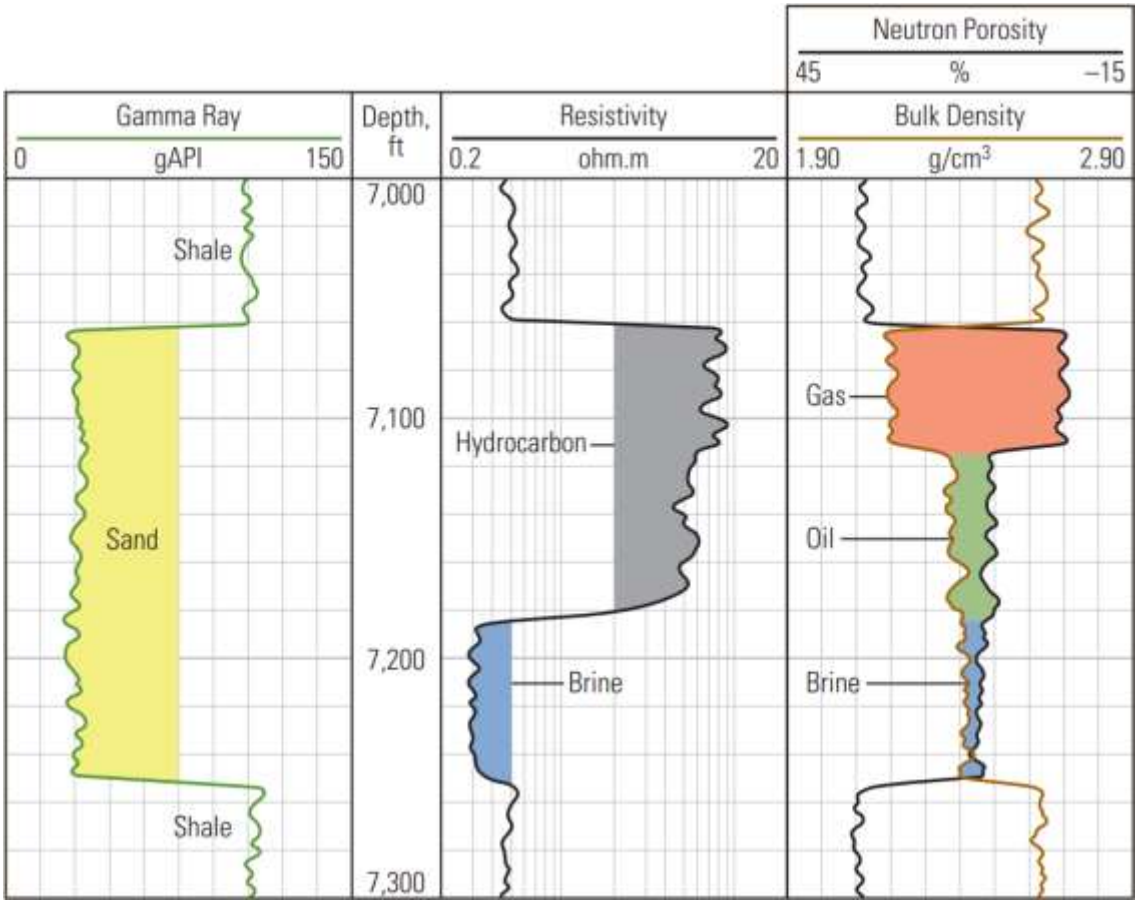

Figure 2.2 Well logs interpretation. A typical combination of log data comprises gamma ray, resistivity log, and neutron and density curves. Gamma-ray gives a high value to shale and low value to sand (Track 1). In the second track, the resistivity response is high in the presence of hydrocarbons, while low resistivity pertains to brines. The third track distinguishes the type of hydrocarbons encountered by neutron porosity and bulk density logs, both of which have the higher response to the gas than to the oil (from Varhaug, 2016).

Numerical simulation based on log and core analysis help us do production forecasting. However, extrapolation, whether statistically or deterministically, is highly uncertain regardless of the scale of grid sampling.

Simple curve estimation might be useful for this task although its accuracy is usually questionable (Nguyen and Chan, 2005). Therefore, surrogate models established by machine learning techniques such as Artificial Neural Network (ANN) can be viewed as an alternative approach.

The advantages of neural networks involve their generation features, applicability to the non-linear problems, computational efficiency, and ability to handle high-dimensional data. Nguyen and Chan (2005) used both ANN and curve fitting methodologies and supplied users with a range of likely solutions in the range of total production and length of production of a producing well. By interpreting the range of potential solutions, the data analyser can determine the best production forecast.

2.3.3 Production data

One of the principal challenges for practitioners in oil and gas studies is the quality and the reliability of oil field production data. The reliability of the model prediction and estimated production profiles highly depends on the authenticity of the reservoir measured production data (Kabir and Young, 2004; Valjak, 2008). Reliable production data, including fluid flow rates, temperatures, compositions, pressures and phase fractions, are of great value in a productive industrial process.

Kabir and Young (2004) showed that the operational problem influences the real production data. In their study, the use of data analysis tools such as type curves of fluid flow in the reservoir and standalone simulation of the fluid performance in the tubing system demonstrated how history matching to those recorded measurements causes unreliable prediction. They showed that the measured water cut was biased because the needle valves were too close to the wellhead and led to unstable flow.

When traditional well testing cannot speculate the complex and heterogeneous reservoir model, Pressure transient tests and production data provide further information about the characteristics of the fault and fracture network within a reservoir (Li et al., 2011; Zheng et al., 2000; Doublet et al., 1996).

Pressure transient data is broadly applied in history matching due to its accessibility and fast response at the well (Li and King, 2016). Common pressure inference tests include injection or production of fluid from one well while the pressure is observed in other

observational wells. The well pressure response can be influenced by both the geometry and the flow characteristics of the reservoir.

Transient tests provide the simplest way to estimate reservoir parameters by analysis of pressure changes (Hamdi and Sousa, 2016; Li et al., 2011). Although the transient can be used to calibrate each of the different models, it will not necessarily distinguish between them (i.e. if each of the model calibrations remains within plausible uncertainty ranges of their respective parameters).

In case more careful characterisation of reservoir heterogeneity is needed, a numerical inversion method can be used to integrate the observational data into reservoir models (Li and King, 2016). Overall, research engineers are still exploring innovative automated data analysis tools that correlate the state/quality of the production data, well history, and water-oil ratios which in turn are the basis of a successful oil field study.

Likewise many industrial processes, oil field operations review the quality of production data as one of the fundamental factors before inducing a model from them. Due to the growing application of computers in the industry and their need for additional information in field development studies, process optimisation and control require many raw measured data. However, there is no guarantee that raw measurements are accurate enough because for any data obtained in a real-life reservoir model, there exists a degree of uncertainty (Kabir and Young, 2004; Valjak, 2008).

The reservoir simulator makes use of numerical simulation of flow equations to reproduce the measured data. Measurement errors can originate from random errors and systematic errors that are an inevitable part of operations (Paffenholz et al., 1994).

In most of the experiments, random errors are well quantified (and often follow a normal distribution). However, the lack of repeatability in reservoir measurements and tests may confound the quantification (Bu and Damsleth, 1996; Paffenholz et al., 1994). While measurement precision in most cases can be quantified, systematic errors cannot be accounted for before they are known (and when they are known they can usually be corrected).

Random errors (Figure 2.3-a) come from the fluctuations that are perceived by making multiple measures of a given experiment (Bu and Damsleth, 1996; Adams and Markus, 2013). This is to say that measuring the same experiment in several trials may have

different outcomes (reproducibility). Some common causes of such random uncertainties in industrial experiments would reasonably be:

-unpredictable fluctuations in initial conditions in the observations

-Limitations derived by the precision of measuring equipment and the uncertainty in interpolating among the smallest pieces

-Influence of an uncontrolled variable on the measured quantity (e.g. the weight of the object)

-changes in instruments or the environmental conditions

As opposed to random errors, systematic errors systematically affect the measurements (see Figure 2.3-b). They may shift the mean and the variance and introduce skewness not inherent in the rock variability distribution (Adams and Markus, 2013; Bu and Damsleth, 1996; Bajkowski et al., 2013). These errors usually come from sensor drift, calibration inaccuracies, instrument failures and leaks. As a result, the measurements fail to represent energy/material balances or other constraints in the model precisely (Valjak, 2008; Baker et al., 2003).

Below are some of the examples of systematic errors in the oil industry:

-A consequence of measuring parameters at low pressure rather than under reservoir conditions. The lack of representativeness of core measures may produce significant systematic errors, particularly in the measured capillary pressures and relative permeabilities (Baker et al., 2003).

-The logging condition, type of instrument, logging time, and operation environment all can produce plenty of systematic errors in well-logging data. The use of logging data contaminated by systematic errors for reservoir modelling will cause unrealistic prediction. Hence, logging data must be standardised. The purpose of standardisation is to exclude systematic errors among various well logging data (Bhushan et al., 2009; Cardone et al., 1980).

-In measuring the bottom hole pressure of a well, there will be a gauge somewhere adjacent to the perforations which is measuring fluid pressure in flow that might be two-phase and will require adjusting to the exact location of the perforations. Meanwhile, the reservoir simulator computes the grid-block average pressure and

then assumes homogeneous sub-grid properties and radial flow in a well model to estimate the bottom hole pressure (Nobakht and Christie, 2017).

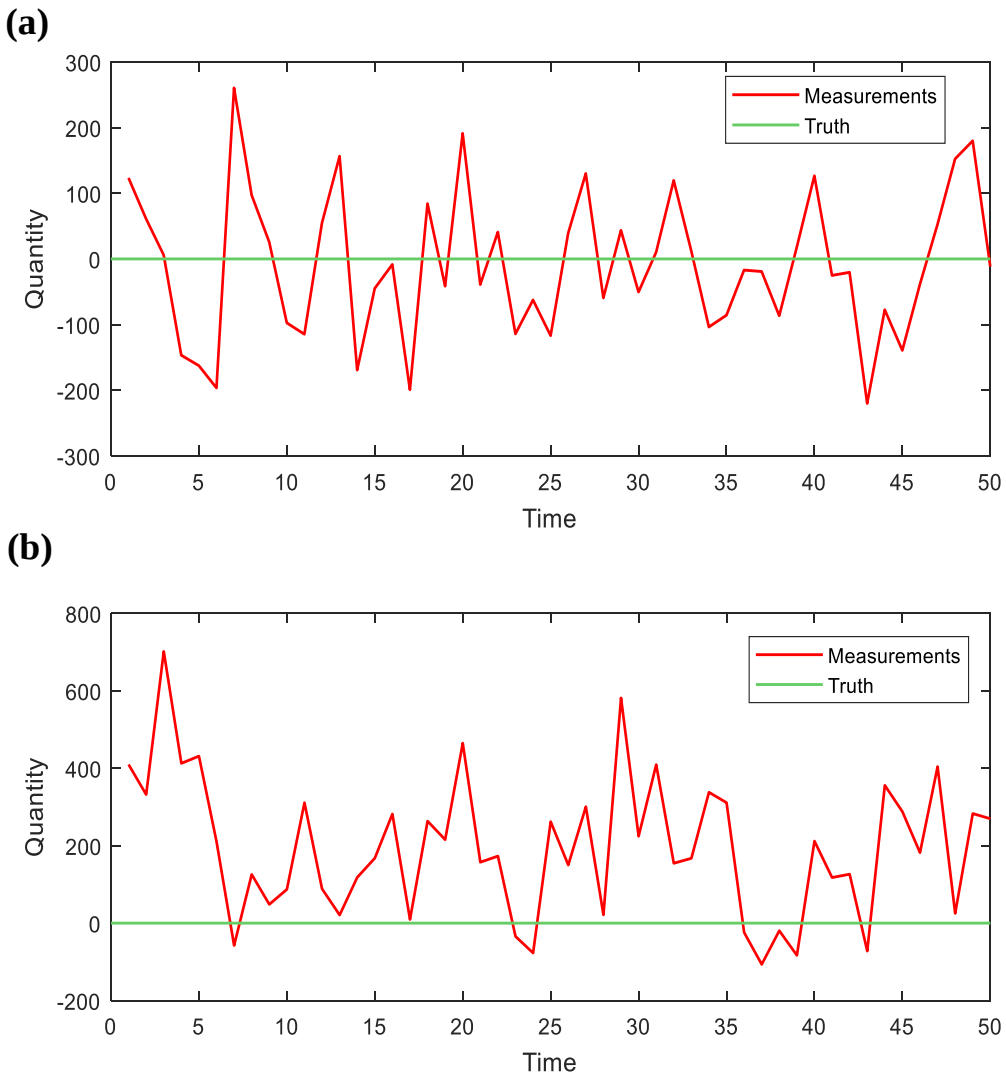

Figure 2.3 **a)** Random errors. **b)** Systematic errors (tend to be consistently positive in this example)

Operating decisions based on such failure can affect reserves predictions. The conventional approach to overcome these issues by combining best practices and measurement devices is not adequate any longer (Bybee, 2008).

In general, there are several commonly used methods for production data analysis, including conventional decline type curve analysis, Data correlation check, real-time and pressure-time plots, data viability assessment, model-based analysis and advanced decline curve analysis (Ilk et al., 2010; Bybee, 2008). Mattar and McNeil (1998)

Presented a per-well basis analysis of production data by merging Pseudo-steady state flow and material balance calculations. They provided a practical tool for estimation of the reserves with reasonable certainty, before being confirmed by production decline curve analysis.

Baker et al. (2003) studied the accuracy of the material balance computations influenced by errors in PVT data. Systematic and random errors were deliberately imposed on reservoir properties such as fluids formation volume factors, bubble point pressure, solution gas-oil ratio, and API gravity.

Figure 2.4 shows how systematic errors were introduced into the reservoir PVT properties by sets of values 10% above and below the actual values of the gas oil ratio. The systematic errors introduced to the PVT data eventually resulted in a range of errors in estimated OOIP and water influx values.

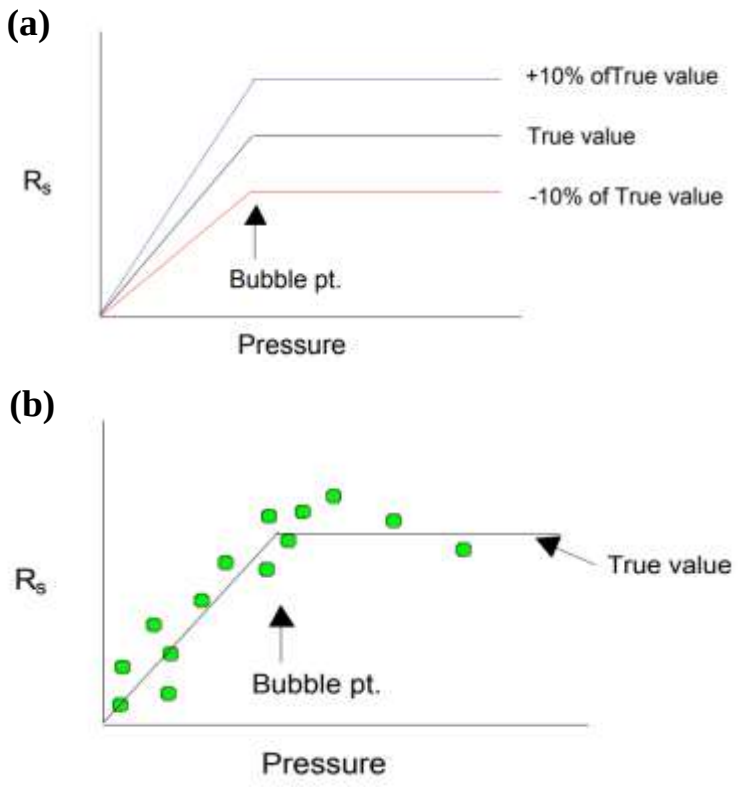

Figure 2.4 Schematic of a) systematic errors and b) random errors imposed on the True Gas oil Ratio R_S (from Baker et al., 2003)

2.4 Data validation and reconciliation

Data interpretation and their use form the basis of reservoir engineering study to determine hydrocarbons in place and their most efficient recovery. Data review and validation is canonical to acquire the precision and limitations of the data and should always be the beginning step in reliable reservoir simulation (Pederson et al., 1997; Lasrado, 2009).

Reservoir prediction results based on inaccurate production data may be more detrimental to the economics of reservoir development than not using a predictive reservoir model at all (Pederson et al., 1997).

To reduce the effect of random and gross errors in data, we need a specialised technique that infers the relationships between the variables of the system automatically. The Data Validation and Reconciliation (DVR) is a standard technique that has been introduced to capture the measured data random errors and thereby boost the accuracy of estimated variables. The DVR process adjusts measurement outputs by process model constraints, e.g. thermodynamic constraints to collect the interrelationship between variables one wishes to estimate (see Figure 2.5). Van der Geest et al. (2001) proposed a flexible simulator of an oil well that applies to different forms of online, real-time reconciliation of measured data from equipment. They tested their technology on two wells where data reconciliation constitutes an excessive level of instrumentation. This level of instrumentation made the existing measurement device more reliable by verifying the validity of recorded data. For DVR to be efficient, no gross error should exist either in the measured data or the constraints, since they may bias the robustness of the reconciliation results. The DVR depends highly upon a reliable estimate for the covariance/correlation matrix which is extremely vulnerable to the results of measurement and the presence of outliers.

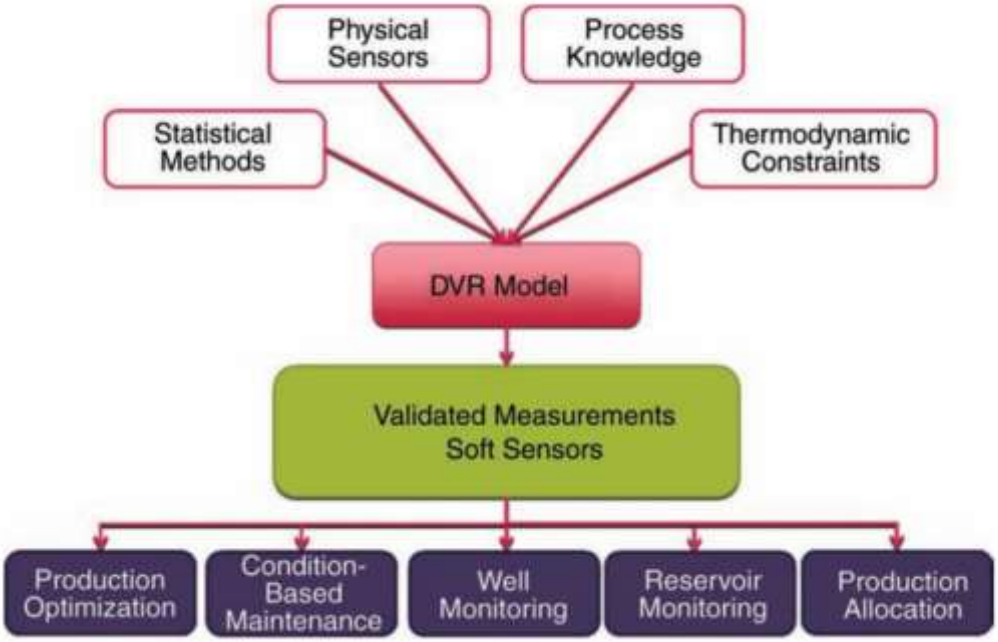

Figure 2.5. Data Validation and Reconciliation (DVR) procedure (from Bybee, 2008)

The DVR also extracts sound and reliable information about the state of calibration processes from raw measurement data and generates a single valid set of data describing the most likely states of the process (Wising et al., 2009). To do so, the DVR adjusts the contradictions among the measured data and their constraints, estimates the real values of measurements. Next, it implements gross errors remediation and generates a consistent set of validated and reconciled process data with higher accuracy (Heyen et al., 1996). The DVR can also be treated as the first step in the real-time optimisation used in downstream, followed by reservoir model update, model-based optimisation, optimiser evaluation, and final decision (Wood and Mokhatab, 2007). The availability of the real-time production (RTO) data collected from intelligent completions can facilitate short-term development plans and data-driven reservoir optimisation. Hence, the downstream industry uses the RTO for automated optimisation of plant control settings, either in closed-loop or open-loop to enhance profitability and efficiency. Employing such forward-looking technologies in line with other sources of field information build a more reliable and coherent database which in turn lead to sounder decisions.

2.5 Reservoir modelling

Reservoir models are valuable tools for answering reservoir management questions and finding possible solutions to reservoir problems. According to the *Occam's razor principle*, from a family of otherwise equal models of a given system, the best candidate model is the simplest one (Domingos, 1999). Similarly, reservoir engineers aim to distinguish among candidate models appropriately, and if other things are equal, select a model with the fewest assumptions (Jefferys and Berger, 1992). Despite the fact that absolute results might not be precise, simpler models have more powerful generalisation capability than complex ones and, therefore, are amenable to empirical testing (Jefferys and Berger, 1992).

One consideration in the reservoir model selection is determining the primary forces such as capillary forces, viscosity and gravity that have the highest impact on reservoir description (Singh et al., 2013; Gilman and Ozgen, 2013). The initial reservoir model requires some data representing fluid characteristics, multiphase flow features such as relative permeability, and well performance (Mattax and Dalton, 1990). Therefore, the selected reservoir model should provide a sufficient description of those parameters that dominate the fluid flow associated with the designed simulation study (Mattax and Dalton, 1990; Slater and Durrer, 1971).

As described previously, each type of data gives specific information about the reservoir and partially contributes to the understanding of the model. The available data should be used for calibration of the reservoir model's geometry and continuity.

On the other hand, stratigraphic interpretation presents nearly deterministic constraints on main stratigraphic surfaces and faults. Geostatistical techniques then posit correlated errors on the deterministic aspects of seismic data by adjusting the structural model of seismic. This is mainly because seismic data are not able to visualise the reservoir characteristics at the scale of interest and therefore, some degree of uncertainty appears due to the physical constraint of the reservoir properties being modelled (Deutsch, 2000). By the time we successfully plug all geological/structural features into the static reservoir model, we start building a reservoir simulation model populated by petrophysical properties.

2.5.1 Static reservoir model

A static reservoir model is the one that develops an integrated 3D reservoir model combining all the significant geological features such as structural, stratigraphic, lithological and petrophysical properties of the reservoir in different locations (cells). The static model captures the distribution of reservoir facies before adding the petrophysical properties of cells to the model.

In structural modelling, as depicted in Figure 2.6, reservoir-scale structure merges with all available information (e.g. seismic data) before building the simulation model. The structural model illustrates faults, horizons, reservoir boundaries and layering by the use of different well log data and borehole images.

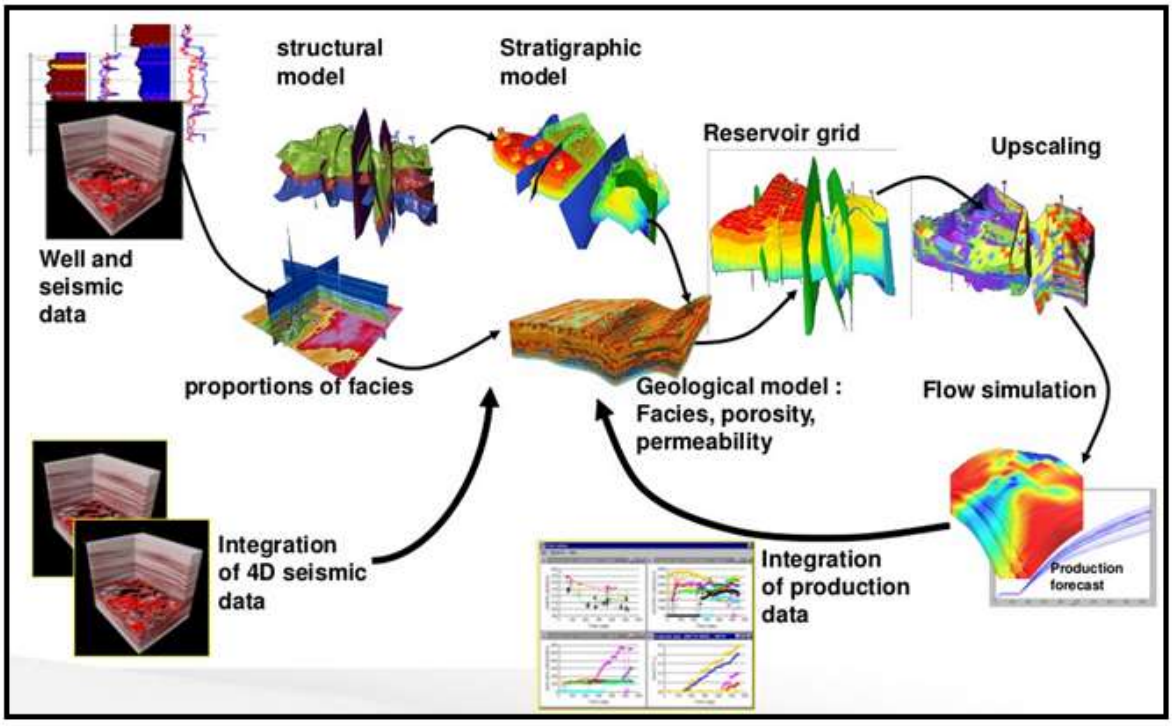

Figure 2.6. Schematic representation of reservoir modelling and simulation (from Estublier et al., 2014)

It is important to recognise there may be aspects of the reservoir not seen in the seismic or wells that are nevertheless likely in the context of the geological interpretation, e.g.

fractures and fault damage zones in carbonates are typically not visible in seismic and are under sampled by vertical appraisal wells.

The conventional grid-based approach partition the reservoir space into three-dimensional grids cells each representing a unique value for reservoir parameters and fluid flow properties.

Jackson et al. (2013) stated that if there exists heterogeneity controlling the spatial distribution of petrophysical characteristics in the reservoir, then surface-based reservoir models can explicitly model the heterogeneity without reference to a grid. In this case, underlying grids do not limit the models created from surfaces.

In the surface-based approach, the top and the base surfaces define the volume of the reservoir model until the additional information about faults reshapes the form of surfaces (Jackson et al., 2015). The fault surfaces then partition the reservoir into different geological zones based on seismic interpretation. Besides, reservoir volume divides into different grid cells (each having a unique porosity, permeability and facies code) within each geological zone. Conventional surface-based reservoir modelling assigns an equal grid resolution throughout the entire zone. However, in most of the reservoir simulation case studies, some zones have larger gradients of interest and, hence, require higher grid resolution. For example, water flooding needs a high resolution adjacent to the waterfront since the saturation gradient is considerably large. For those grids far from the waterfront, however, the saturation changes gradually and high resolution is unnecessary (Jackson et al., 2015; Deutsch, 2000; Al-Busafi et al., 2005).

Before setting up a history matching process, it is essential to evaluate how well the constructed reservoir model represents the true reservoir behaviour. The initial reservoir model will need to describe the best available static and dynamic representation of the reservoir since history matching, in essence, is non-unique. Moreover, production forecasts are highly dependent upon the initial geological model and its assumption. Therefore, to reliably predict the future performance of the reservoir, the initial reservoir model should ensure that the simulation model is consistent with the geological/structural features of the reservoir (Gilman and Ozgen, 2013).

2.5.2 Dynamic modelling and model validation

Reservoir simulation needs to predict multiphase flow in the reservoir. Since the geological models often have fine-grid cells, simulators cannot efficiently run the geological model. This means that the geological model created by geoscientists at the static model preparation stage is used as an input for the reservoir simulation model. Therefore, upscaling the geological model to a coarser scale simulation model is inevitable following the construction of 3D-reservoir simulation grid.

The static model needs to sample the structure, stratigraphy, facies and property variations of depositional units (bodies). The dynamic model normally needs to sample the pressure and saturation (sometimes also other dynamic responses such as composition and temperature). It is usual for the dynamic variations to occur at a coarser grid scales than the static sampling, but some dynamic phenomena may require a finer scale sampling, e.g. coning.

Coarsening and upscaling are conventional practices in reservoir simulation for managing the model size. They replace a heterogeneous grid property with an equivalent homogeneous one to improve computational efficiency and elevate reservoir model performance (Qu et al., 2015).

Reservoir simulation grid constructs a model based on the significant geological characteristics such as faults, variation in reservoir properties, stratigraphy. The simulation grid needs a simplified version of the fault geometry, regular cell grid geometry, and general homogenisation of rock properties. Moreover, well completions must also be resolved in the reservoir simulation grid (Qu et al., 2015).

Christie and Clifford (1998) introduced a streamlined technique that produces upscaled compositional fluxes almost equivalent to those taken from post-processing conventional compositional model runs in a shorter time. Their technique enhance the speed for the loss of a minor amount of accuracy. Hu et al. (2007) developed a new integrated model based on wellbore model and reservoir model that exchanges flow and pressure data at the sand face. The reservoir model computes the flow rate of each phase while the wellbore model constrains pressure to the reservoir model.

Upscaling uses the geological models to build simulation models that reduce computational time and yet maintain the reservoir characteristics to a reasonable degree.

Coarsening, used as reduced-order modelling for a large number of simulation models, also decrease the size of the model and computation time. Once the upscaling of simulation model established, we perform the simulation run to evaluate its match quality.

At present geologists and reservoir engineers share their knowledge and information with a different perspective of the subscale physics. The share of knowledge, either offer the reservoir engineer a better understanding of the subsurface or give the geoscientist the ability to validate the geological model. For example, grid coarsening can exploit information from high-resolution geological models to apply to both structured and unstructured grids.

In the literature, flow simulation by the use of a coarsened model is shown to provide a reliable approximation to high-resolution computations performed in the original geological model. King et al. (2006) Introduced a constrained optimisation strategy to the coarsening process of 3D reservoir models for fluid flow simulation. They assumed constraints from stratigraphy, reservoir fluids, well locations, and large-scale reservoir structure.

Aarnes et al. (2007) developed a non-uniform coarsening methodology for modelling subsurface flow properties. They observed that using an equal number of cells, the non-uniform model offers more consistent results compared with the uniformly coarsened grids model because the nonuniform coarse grid accurately determines the flow velocity.

Based on the match quality of the initial reservoir model against the observation, skilled interpreters investigate the best candidate solutions to integrate some new type of data produced by production log, the downhole pressure build up and falloff tests. The integration of new data provides further validation of the static model against the history of wells and reservoirs and constrains the initial reservoir model to more reasonable simulation (Bouska et al., 1999).

In model validation, we evaluate match quality of a representative sample of the model realisations from the geostatistical inversion. If the assumed properties in the model are close to the truth, the simulated results of response variables such as well bottom hole pressure improve.

Model verification, known as an extension of the model selection problem, affirms that a model and its computational performance reproduce those variables that the model

explicitly computes. Model validation should illustrate that other relevant characteristics of the generated flow simulation are in equilibrium with those of the measured flows.

2.6 History matching and prediction of oil reservoirs

Reservoir model calibration, known as history matching in petroleum engineering, is a canonical step in the oil industry.

The primary goal of model calibration in reservoir management is to gain some degree of belief to the solution of the true response of reservoir, validate the reservoir simulation model, make a realistic assessment of uncertainty, and predict the future response of reservoir (Christie et al., 2005). In this regard, model calibration explores some solutions to the theoretical model from which our prior believes about reservoir parameters will be revised.

History matching is used to reduce the discrepancy between the observed data and the simulation output and requires running many simulation models. In history matching, we try to adjust the parameters of the simulation model to match model's output with the available production or seismic data from the field.

Likewise many engineering processes, history matching is an inverse problem with non-unique solutions. The adjustment of those spatially varying parameters from noisy production data is referred to as an ill-posed problem. The ill-posed problem prompt history match to non-unique solutions and different realisations of the reservoir may give equally good models (Leo et al., 1986; Bouzarkouna and Nobakht, 2015). In other words, many configurations of different model parameters, e.g. rock porosity saturation and permeability, can afford a proper fit to the data.

In a forward problem, we begin with the causes of the problem and then compute the results, whereas, in an inverse problem, we start with the observed results of a system and then look for the causes. To do so, history match explores likely solutions in multi-dimensional parameter space throughout a tuning process. Each calibration model, which reproduces the available production measurements may feature various geological and petrophysical properties (Ibrahimov, 2015).

However, designing a simulation model involves subsurface uncertainties which can considerably affect prediction results. Quantifying such uncertainties for a field under

development makes history matching a challenging task. In many engineering simulation-based optimisation problems, the number of function outputs is a limiting factor affected by time or cost.

In general, the history matching process uses the following well/reservoir measurement data (Mattax and Dalton, 1990; Satter and Thakur, 1994):

- General data from core and log analyses: base maps and well pressure tests contribute to the specification of the grid structure, dimensions, initial contacts between hydrocarbons, reservoir pressure and layering.
- Geological data: well productivity tests, well logging, core analysis are the basis for the understanding of gross/net thickness, permeabilities, anisotropy ratio and initial fluid saturations.
- Production/injection well data including flow saturation profiles, shut-in pressures during well test, well production/injection rates, RFT pressures (measured pressure along the depth), productivity index, Gas oil ratio and water-cut-rate, skin factor, future production/injection plan for each well.
- Rock and fluid data including transmissibility barriers, capillary pressure and PVT data.

Because of the high number of uncertain parameters in most of the oil and gas reservoir models, it is very challenging to deal with the reservoir management manually. Hence, researchers have tried to introduce a range of Assisted (automatic) History Methods (AHM) to speed up the history match. A complete workflow of automatic history matching includes a selection of reservoir variables that require adjustment and parameter updating schemes, data analysis, and a combination of the matched models to obtain an ensemble of best reservoir models.

In addition, reservoir production performance remarkably outlines the economic feasibility of hydrocarbon recovery and also the future of production operations. Thus, for efficient reservoir management, a thorough analysis of past, present and future reservoir performance is required.

2.6.1 Match Quality Standard

Reservoir modellers employ history matching for conditioning of the reservoir properties to the production performance data. For that reason, a history match study tunes the reservoir properties until it finds a good match to the observed well pressure and flow properties (Watson et al., 1984). The solutions to this iterative process are later ranked based on their quality, such that the models with lower misfit values gain a higher rank and likelihood.

The standard least-squares (LSQ) misfit given in Eq. (1-2) is commonly used as the misfit/objective function in the history matching (Kuznetsova, 2017). However, the mismatch is usually more significant than would be expected from timing and picking errors alone (Shearer, 2009). Moreover, because many combinations of the reservoir model parameters, e.g. rock porosity and permeability can provide an excellent match to the data, finding a match quality standard is still a matter of debate. Nevertheless, despite many years of model calibration in the history of oil and gas study, the petroleum industry has been less willing to review the quality of history matching models.

One consideration about calibration of reservoir models may be that if a model gives a fair match to the observed values, then it should be able to predict accurately within the historically tested drainage area and production mechanism. However, there are several issues with this assumption since having a good match will not necessarily guarantee an appropriate prediction (Tavassoli et al., 2004). The accuracy and precision of estimates of reservoir properties can restrain the benefit of reservoir simulation process in predicting future performance of reservoir. This becomes evident when different sets of parameter values generate a nearly identical match to the observation data (Seinfeld and Kravaris, 1982; Jahns, 1966).

Because the number of physical parameters (e.g. permeability and porosity in oil reservoirs) to be estimated in a calibration process is usually large, it is significantly important to specify which parameters can have precise inference (i.e. identifiability). A non-identifiable model invariably has two or more parameterisation generate equiprobable observations (Watson et al., 1984; Seinfeld and Kravaris, 1982; Jahns, 1966). For instance, the identifiability of porosity should answer this question: how

confidently can we estimate spatially variable porosity given pressure and production data?

Concerning the problem of non-uniqueness in history match, new technologies such as time-lapse 4D-seismic surveys can be employed. The goal of 4D seismic is to observe and compare the changes in the reservoir behaviour as a result of oil/gas production or water/gas injection into the reservoir (Al-Busafi et al., 2005). Then, the history matching process updates the initial reservoir description as dynamic information of field response, and time-lapse seismic results become available.

The time-lapse seismic data add more constraints to the fluid flow modelling by mapping fluid movements in the vertical and lateral space of the reservoir. The additional faults in the new model can add to the total uncertainty within history matching framework, forcing history matching to provide more evidence.

2.6.2 Bayesian calibration of simulation models

Estimates for the model's parameters are often determined by Bayesian analysis throughout the process of history matching. In the Bayesian framework, questions about uncertainties of flow parameters estimates are addressed via a posteriori probability density (Nezhad Karim Nobakht et al., 2018). If this probability is simple (e.g., Gaussian), this analysis is easily predictable and unchallenging. Otherwise, more elaborate procedures such as Monte Carlo sampling may be required (Bazargan et al., 2013). Bayesian analysis uses a statistical model to relate the observations to the model output (data) via iterative progress. Roughly speaking, Bayesian imposes probability densities on the models themselves. These probabilities, which represent measures of degrees of belief, are coupled with the data misfit function into a final (a posteriori) probability density on the parameter space.

$$posterior \propto prior \times likelihood \qquad (2\text{-}1)$$

In the Bayesian framework, the analytical solution to the posterior probability needs the integration of the likelihood function over the all possible values within the entire parameter space. Even though the Bayesian statistics provides the optimal means for making the statistical inference, the exact use of those tools is difficult which makes the analytical evaluation of the posterior impossible. Considering that the exact Bayesian inference is impossible, the approximate solution to the real posterior probability densities

can be obtained by algorithms such as Neighbourhood algorithm- Bayes (Sambridge, 1999a).

2.6.3 History matching implementation

History match begins with matching average pressure and flow rates to address the material balance for the reservoir. Then, to better understand the transmissibility barriers and compartmentalisation, petroleum engineers match the RFT pressure for each well. The last step of fitting the input data includes matching gas to oil ratio GOR, well water-cut rate and well pressure response to shut-in/build up.

The problem of reservoir history matching is typically investigated through two different approaches (Bouzarkouna and Nobakht, 2015): the manual history match and the assisted history match (AHM).

The manual history match workflow involves a sequential method that starts with the matching of field variables and follows with the tuning of corresponding flow layers/unit. In this case, reservoir engineers can subjectively enhance the match quality for a given model based on their judgment and experience (Christie et al., 2002). The reservoir model should provide a sufficient description of those parameters that dominate the fluid flow associated with the designed simulation study.

The disadvantage of manual history match is mainly due to the tedious trial-and-error tuning procedure in a single history match model (Bouzarkouna and Nobakht, 2015). Then, the manual approaches suffer either from long computation times or from the need to rebuild a physical simulation model for each reservoir.

The main advantage of assisted history matching is to automate the manual adjustments of the reservoir simulation model or comparison of filed measurements and reservoir simulation output. Setting reasonable parameter range limits helps automated history match to explore those solutions that are physically valid (Christie et al., 2002). However, assisted history matching can only automate adjustments within the defined parameter space.

Assisted history matching benefits from appropriate optimisation techniques to accelerate convergence, find the optimum solution(s) and perform a better search in the model parameter space. The optimisation process bounded by prior model constraints needs an

objective function to build the solution space for history match problem (Cancelliere et al., 2011; Mattax and Dalton, 1990). A suitable optimisation algorithm should have a robust exploration capability throughout the entire search space without being trapped in local minima in the solution space. Optimisation then generates a set of uncertain reservoir parameters drawn from prior information and enhances the quality of the calibration models during the history match study

Besides, advanced optimisation algorithms with a fast and parallel computation capability are still ongoing research in assisted history matching. Automatic history matching is based on algorithms written to evaluate an objective function explicitly using several realisations. It requires the building of a mathematical model, setting up an objective function, and performing a minimisation algorithm to the defined objective function (Cancelliere et al., 2011). The objective function evaluates the quality of simulator output using a mismatch/misfit formulae. The algorithm endeavour to reduce the misfit value and thus to find the model that best approximates the fluid flow rates and well/reservoir pressure data collected during the reservoir life (Christie et al., 2002; Cosentino, 2001).

A generalised framework for history matching is schematically described in the following workflow diagram (Christie et al., 2006).

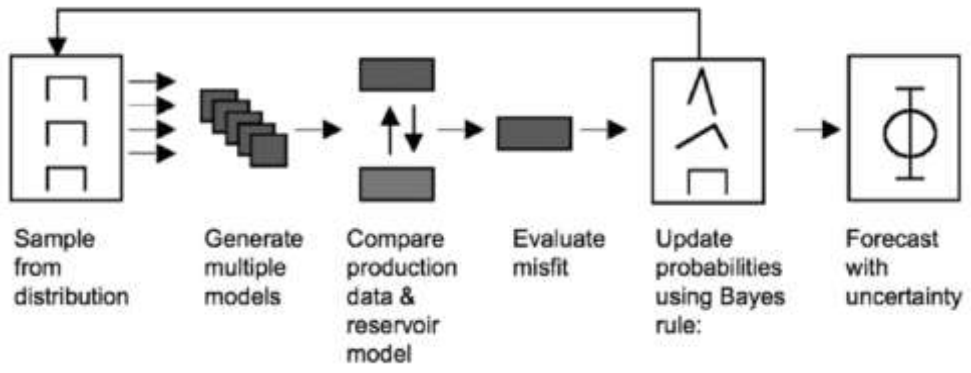

Figure 2.7. Schematic of reservoir history match and prediction under uncertainty (from Christie et al., 2006)

At the beginning of a history matching process, some sample models are drawn from the prior distribution of parameters. Then, the history match generates and runs multiple simulation models. The simulator output is further compared with the observation field

and seismic data. The resulting misfit firstly defines the likelihood of each realisation, and secondly, updates the probability density function for each parameter based on Bayes' rule. In an iterative process of the optimisation method, history matching finds some candidate simulation models with an improved match quality. At the final step, prediction runs are generated from the posterior probability distribution.

2.7 A literature review of assisted history matching

Recent development in computational capacities enables automation of the optimisation algorithms in history matching problem. Many optimisation methodologies introduced in the literature address the problem of field development scenarios (Li et al., 2001; Thomas et al., 1972). Slater and Durrer (1971) proposed a balanced error-weighted gradient technique that systematically decreases the discrepancy between observed and measured reservoir field data. Their approach quantitatively determines the connections among the individual well regions by adjusting the parameter of each region, while keeping other regional parameters constant.

The dominant optimisation strategies in assisted history matching are data assimilation methods, gradient-based algorithms and stochastic methods.

Data assimilation integrates several kinds of data such that the observed quantities are accumulated into the model state to improve knowledge of the past, present, or future states (Evensen et al., 2007; Aanonsen et al., 2009). In data assimilation techniques such as the ensemble Kalman filter (EnKF), the new estimate is a function of the previous estimates, and thereby it updates the ensembles to match the observed data (Jung et al., 2018).

The EnKF techniques were initially applied to weather forecasting and have recently been adapted to the optimisation of reservoir history matching (Gao et al., 2004). These methods are highly parallelisable, reliable for handling systems with a large number of parameters and easily adjustable to different types of simulations (Oliver and Chen, 2011; Evensen et al., 2007; Aanonsen et al., 2009). However, the different-scaled data, underestimation of uncertainty (unless there is an additional perturbation), high nonlinearity, and limited available data produce difficulties for more reliable data assimilations with mathematical clarity, and also for enhanced predictions of unknown properties (Jung et al., 2018; Oliver and Chen, 2011).

In gradient-based optimisations, the partial derivatives define the change in production data because of a small variation in reservoir parameters. Also, gradient-based methods such as the Gauss-Newton and Levenberg-Marquardt are helpful due to their faster convergence rates (Bissell et al., 1992).

The optimisation tools based on Newton-Raphson terminology, progressively improve approximations of roots of a univariate real-valued function (Polyak, 2007). These techniques compute the derivatives of the specified objective function for the calibration process to uncertain reservoir parameters. Thomas et al. (1972) deployed a nonlinear optimisation method that automatically tunes the reservoir parameters based on the standard Gauss-Newton least-squares procedure. They assigned some limits to the range of each parameter to manage highly nonlinear cases with particular provisions.

For constrained problems, once the procedure calculated the derivatives, methods such as sequential quadratic programming can find the minimum (Gill et al., 2005). On the contrary, for an unconstrained problem, constraints are placed as a penalising factor to the objective function. Moreover, quasi-Newton methods are computationally efficient for unconstrained problems, usually in association with a line search procedure (Dennis Jr and Schnabel, 1996). In either case, statisticians anticipate the computation time of the optimisation to be approximately proportional to the number of parameters.

Chen et al. (1974) established history-matching study for a set of optimal control problems in gradient-based optimisation settings to challenge the assumption of constant-zone gradient optimisation. They assumed that reservoir characteristics belong to continuous functions properties of position instead of a uniform distribution. Moreover, streamline derived sensitivity methods proved to be a great potential (Datta-Gupta and King, 2007; Vasco and Datta-Gupta, 1999) as they only need a single forward simulation model to produce the sensitivities analytically.

Recent gradient-base approaches (see Li et al., 2001) considered adjoint equations for three-dimensional, three-phase flow to compute the sensitivity of production field data to permeability and skin factors. Even if there is a lack of interest in a fully automated history matching, sensitivity analyses are beneficial for understanding the underlying physics that control three-dimensional multiphase flow.

Since the primary objective of the gradient-based methods is to obtain a single best solution to the history match, it is difficult for engineers to assess the associated uncertainty. Gradient-based techniques are not reliable when the reservoir model has a complex geological structure with many unknowns (Zingg et al., 2008; Asadollahi and Naevdal, 2009). If the starting point is selected far from the truth, then the Gradient-based techniques may fail to reach a reasonable solution. As a result, the algorithms quantify the uncertainty in light of either an improper solution or a unique local optimal solution. Moreover, by performing different trials of history match, a gradient-based optimisation will not necessarily find the global optimum (Zingg et al., 2008).

In many applications, information about the gradients of the objective function is unreliable or impossible to evaluate. For instance, finding the derivative of a non-smooth function subject to noisy data are of little value.

One of the sources of noisy data in reservoir simulations is linear, and non-linear solver convergence which causes different errors at each time step (Mishev et al., 2008). In such a situation, mathematicians exert gradient-free (derivative-free) stochastic optimisations. Gradient-free methods do not require derivative information in the classical sense to obtain an optimal matched model.

In the global optimisation problem, many powerful techniques originate from stochastic strategies as they can reliably explore the search space without being trapped in local optima. Some of the classes of stochastic global optimisation methods comprise evolutionary strategies (ES), Bayesian optimisation, particle swarm optimisation, and Differential Evolution.

The stochastic optimisations, in general, assume that the production data pertain to a realisation of a stochastic process (Zingg et al., 2008). As opposed to gradient-based techniques, the stochastic sampling methods do not require the sensitivity coefficients to converge to the global minimum and can better assess the uncertainty. However, in practice, stochastic optimisations have essential need for many simulation runs to secure convergence. Hence they are computationally expensive, especially for a significant field-scale application.

In recent years, researchers devoted a fair amount of work to build hybrid versions of stochastic population algorithms, mainly to improve the performance of algorithms and

to reduce the number of control parameters. This consideration improves practical applications of stochastic algorithms regarding convergence speed, exploration, exploitation. Gao et al. (2016) introduced a parallelised and Hybrid data-integration algorithm that ensures the convergence of the objective function minimisation, even in case the objective function is nonsmooth.

To sum up, the fair use of gradient information can remarkably enhance the convergence speed. Whereas, in general, gradient-free techniques converge very slowly, particularly around an optimum. Another benefit of gradient-based methods is that they follow a clear convergence criterion while reaching at least a local optimum. On the contrary, a specification of a termination basis is not an easy task for gradient-free optimisations.

Disadvantages of Gradient-based methods include a local rather than a global solution, inconsistency of dealing with noisy objective function solution spaces, bias computation of the derivatives, and topology optimisation (Zingg et al., 2008). On the other hand, development cost is minimal in a gradient-free optimisation such as genetic algorithms since they handle the function estimations through a "black box" process that finds a global optimum. Also, gradient-free techniques can better handle the noise in the objective function and have no trouble with topology changes. Some of the optimisation methods implemented in history matching of oil and gas reservoirs are:

1. Evolutionary strategies: Romero et al., 2000, Schulze-Riegert et al., 2002

2. Genetic algorithms: Tokuda et al., 2004, Sanghyun and Stephen, 2018, Castellini et al., 2008

3. Neighbourhood algorithm: Nicotra et al., 2005, Ahmadi et al., 2013

4. Bayesian optimisation algorithm (BOA): Abdollahzadeh et al., 2012, Abdollahzadeh et al., 2013

5. Simulated Annealing: Ouenes et al., 1992, Ouenes and Saad, 1993, Panda and Lake, 1993

6. Ant colony optimisation (ACO): Hajizadeh et al., 2009

7. Differential Evolution (DE): Nghiem et al., 2013, Hajizadeh et al., 2010

8. Particle swarm optimisation (PSO): Mohamed et al., 2011

9. Data assimilation: Skjervheim et al., 2005, Evensen et al., 2007

10. Hybrid data-integration: Gao et al., 2016.

2.7.1 Bayesian optimisation algorithm

The Bayesian optimisation algorithm (Pelikan et al., 1999) produces a group of likely solutions to a given problem by sampling Bayesian networks. BOA sampling randomly initialises a population of strings assuming a uniform distribution over all possibilities. Then, the optimisation updates the population for some iterations each following four steps. First, Genetic algorithm selection method selects the likely solutions from the current population. The second stage contains the construction of a Bayesian network that fits the population of candidate solutions. Third, the optimisation samples the constructed Bayesian network to find new candidate solutions. Fourth, the new candidate solutions amend the original population.

A Bayesian network is a directed acyclic graph with the nodes associating to the positions in the solution strings. From a statistical perspective, a Bayesian network encompasses a joint probability distribution computed by:

$$P(X) = \prod_{i=0}^{n-1} p(X_i | \prod_i) \qquad (2\text{-}2)$$

where $X = (X_0, X_1, \dots, X_{n-1})$ is a set of all variables; $\prod_i$ is a set of nodes of X_i; and $p(X_i|\prod_i)$ is the conditional probability of X_i given $\prod_i$.

To better understand the algorithm process, it is essential to illustrate the construction of Bayesian networks and the way it proposes new solutions in the search space. As stated in Abdollahzadeh et al. (2013), for a five-bit problem, the solution representation comprises five bits, and the Bayesian network includes a node for each bit in the solution and edges that depict connections between the bits. In other words, the probabilities of setting associated bit in different nodes depend on the probabilities of setting bits in parent nodes. For instance, the flowchart in Figure 2.8 implies that the probability of setting the bit 3 is conditional on whether the bit 2 has been set or not. If now suppose that node 1 has been observed, and that bit 1 is set to zero. Then, the value of bit 0 becomes 0.9 conditional on the already visited node 1.

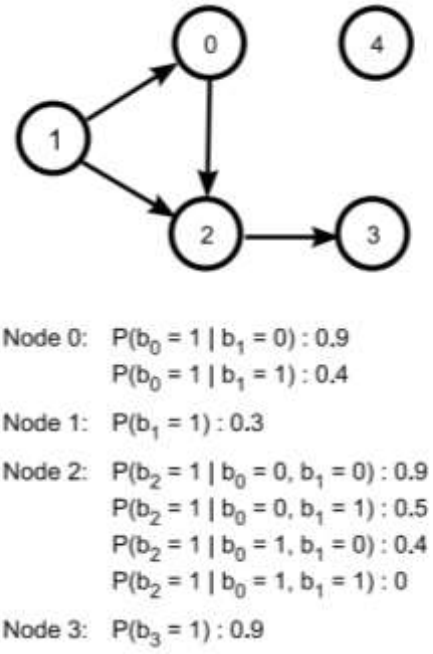

Figure 2.8. v A simple Bayesian network for a five-bit problem (from Abdollahzadeh et al., 2013)

The assumption of the acyclic network implies a constraint on the Bayesian network. Additional constraints can be considered for controlling the complexity of the network such as placing an upper bound on the number of required parent nodes for each node.

Specification of network structure requires an optimisation process that improves the fitness of the structures. Two approaches to evaluate the quality of fitness are Bayesian metrics and minimum description length metrics. Bayesian metrics quantify the uncertainty of network parameters and structures by assuming prior information for them within the Bayesian framework. Minimum description length, on the other hand, assumes that the best hypothesis for a structure and its parameters is the one with the best compression of data allowed by the model. The optimisation begins with adding edges to the network at a time in that it maximises the quality of network while meeting the constraints.

Different methods can build the network from the set of picked solutions. All methods have two necessary components: a scoring metric which discriminates the networks according to their quality and the search algorithm which searches over the networks to find the one with the best scoring metric value.

2.7.2 Particle Swarm Optimisation

Some scientists have produced a computer simulation of various descriptions of the action of organisms in a bird flock patterns or fish school. Reynolds (1987) and Heppner and Grenander (1990) performed simulations of bird flocking by the aesthetics of bird flocking choreography that enables large numbers of birds to flock synchronously, change direction abruptly, and scatter. Both models relied heavily on the direction of inter-individual distances where synchrony of flocking behaviour was assumed to be a function of a bird's attempts to preserve an optimum distance between themselves and their neighbour fellows.

Throughout the simulation of a simplified model, Eberhart and Kennedy (1995) discovered a population-based optimisation strategy for continuous nonlinear functions called Particle Swarm Optimisation (PSO). Particle swarm optimisation has close ties with Artificial Intelligence (AI), fish schooling, bird flocking and swarming theory in particular.

In PSO an ensemble of simple objects, the particles, are located in the search space of some functions, and each assesses the objective function at its place (Eberhart and Kennedy, 1995; Kennedy, 2011). The starting points for the model parameters in the parameter space are drawn randomly. Each particle then decides its direction through the search space by evaluating some aspect of the trace of its current and best states compared with those of other members of the swarm. The subsequent iteration begins after each particle has been moved forward. Ultimately the swarm as a group of particles, like a flock of birds jointly searching for food, is likely to move adjacent to an optimum of the fitness function.

The dynamic process for PSO has two significant steps, first updating the velocity of each particle at each iteration, and second updating the location of each particle in parameter space. Even though the swarm is designed to walk in 2-dimensional space in principle, the PSO can prolong to multi-dimensional space as an optimisation framework. As depicted in Figure 2.9, each particle carries four information: the location of the i-th particle after k-th iteration in n-th dimensional space $X_i = (X_{i1}, X_{i2}, \dots, X_{in})$ where $i = 1, \dots, \mathrm{m}$ indicates the size of swarm and $k = 1, \dots, K$ specifies the number of iteration; the

velocity of the i-th particle $V_i = (V_{i1}V_{i2}, \ldots, V_{in})$; the best location of i-th particle seen so far pbest; the best location of the whole group visited so far, called gbest.

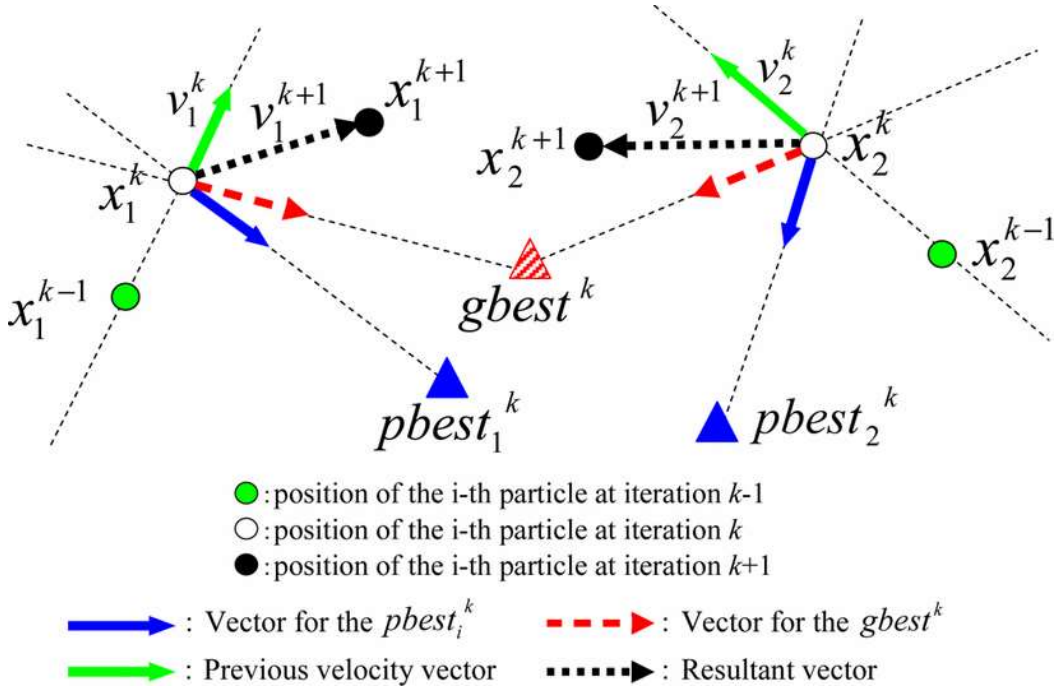

Figure 2.9. Working principles of Particle swarm optimisation (from Choi et al., 2010)

The locations of particles and their velocities will be updated as follows:

$$X_i^{k+1} = X_i + V_i^{k+1} \tag{2-3}$$

$$V_i^{k+1} = \omega \times V_i^k + c_1 \times rand()_1\left(p_{best}{}^i - X_i^k\right) + c_2 \times rand()_2\left(g_{best} - X_i^k\right) \tag{2-4}$$

where $rand()$ is a random number of a uniform distribution [0 1] and ω, c_1, c_2 are the parameters for controlling the impact of the previous velocity, p_{best}, and g_{best}.

PSO demands only primitive mathematical operations and is computationally cheap regarding both memory usage and convergence speed (Kennedy, 2011). Because PSO is gradient-free, the derivative of the function in the space is not required. It uses a heuristic algorithm meaning that the particles learn from each other throughout their generation. It is a robust optimisation algorithm for different applications in history matching because it is capable of successfully establishing parallel computations (Mohamed et al., 2010).

The main disadvantage of using the PSO is the high risk of being trapped into local minima because all agents become nearly identical while converging around the best solution (Choi et al., 2010).

2.7.3 Neighbourhood Algorithm

The Neighbourhood Algorithm (NA) is a derivative-free stochastic optimisation strategy designed and introduced by Sambridge (1999a). It was initially applied to a seismic waveform problem where linearised inversion methods suffer from extreme dependence on the primary solution.

The search algorithm in neighbourhood algorithm exert the geometrical partitions known as Voronoi cells (named after Georgy Voronoi) to run a proper search in parameter space where the nearest neighbours regions are established under a proper distance norm (Liebling and Pournin, 2012).

Voronoi cells are commonly the nearest-neighbour zones around each point in model space, as represented by a distance norm. To find suitable regions of the search space, Voronoi cells break down the multi-dimensional search space into separate areas by centring around the generated points. Thus commencing from randomly generating n_{si} samples, the search space will be split into n_{si} cells nearby n_{si} initial points, in that the interfaces of each Voronoi cell are the equidistant lines between neighbour points (Hajizadeh, 2011; Ahmadi, 2012).

The NA algorithm then chooses the n_r best-fitting models with the lowest misfits and produces n_s new samples using a random walk search in the Voronoi cells of selected models. Consequently, at each iteration of the algorithm, n_s/n_r new models are formed at each cell. The geometry of the Voronoi cells reshapes once new models found their places in the parameter space.

For instance, as can be seen in Figure *2.10*, the initial number of random models is n_{si}=10 followed by n_r=2 selected best-fitting models. Then, NA generates 4 new models and ranks all models based on their objective function value. Next, the algorithm selects the best cells and carries them forward to further improvement. Finally, new n_s models are produced by a uniform random walk inside Voronoi cells of those selected n_r cells. Note

that the NA approximates the misfit values within each Voronoi cell by assigning a constant misfit to samples within each cell. The process will be continuing until a user-defined termination criterion such as a maximum number of realisations is met.

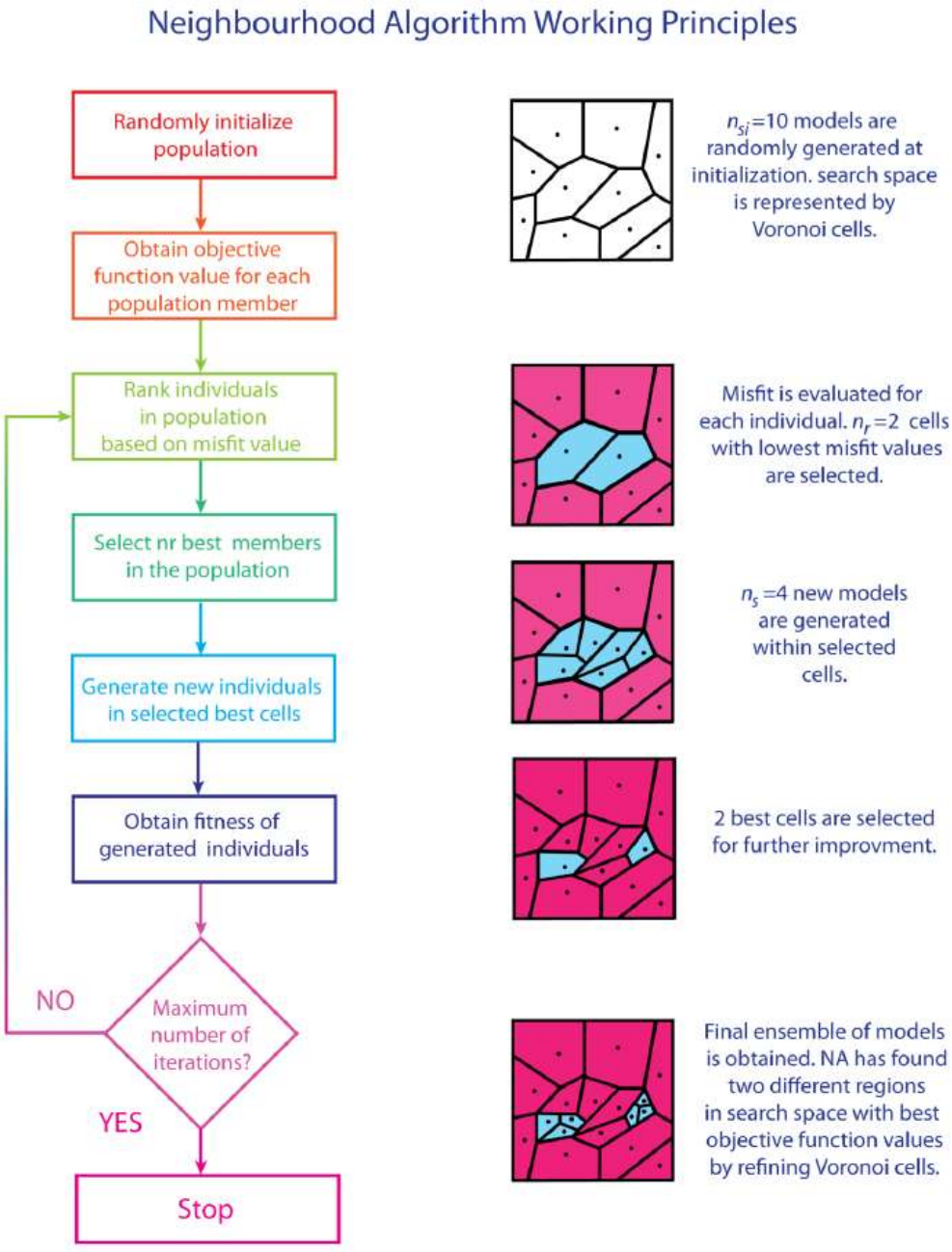

Figure 2.10. Schematic of Neighbourhood Algorithm(from Hajizadeh, 2011)

The neighbourhood algorithm procedure is a two-stage statistical approach for non-linear inverse problems applicable to different scientific disciplines. Initially, the NA uses a

gradient-free search algorithm in multidimensional parameter space. The purpose of the search stage is to find model configurations in the parameter space with the sound quality of a user-defined objective function. The appealing feature is that all available samples entirely and equally influence the extent and the form of the neighbourhoods of each sample including the regions of better and worse data fit. The NA, then, uses a direct search routine for global optimisation of computer models.

The algorithm is conceptually straightforward and needs merely two tuning parameters. Instead of using numerical value of the objective function, NA employs the rank of a data fit criterion. In this manner, all challenges regarding the scaling of a misfit function are circumvented. Voronoi cells can be adapted to improve any existing direct search algorithm, by periodically substituting the forward modelling computations with nearest neighbour approach.

The NA-surface presents a straightforward process of implementing non-smooth interpolation of an unusual distribution of points in multivariate parameter space. With regards to the distribution and sampling density, the Voronoi cells are always space filling, unique and have size dependent on the sampling density (see Figure 2.10). The NA-surface will include short-range variations in misfit only where they exist in the original samples and longer-range variations in case of sparse sampling.

Compared with NA, on average, PSO minimises the misfit (reach local minima) in each generation much quicker. Mohamed (2011) carried out a comparison between NA and PSO in history matching of oil reservoirs. He showed that it is much likely that PSO yields a good history match in a less number of samples than the Neighbourhood algorithm, and this behaviour is consistent with varying the initial random starting points. PSO has more tendency to shift new samples towards the low misfit areas than NA. NA and PSO need a separate calculation to go from sampled models to forecasts of uncertainty.

There may be some cases with weak prior information, numerical limitations, or complicated physical interrelationships between parameters which restrict the parameter space by complex boundaries. In such cases, classical search algorithms are unable to handle irregular limits efficiently. Wathelet (2008) expanded the NA formulation to such parameter spaces affected by non-uniqueness wherein the ideal solution involves the

ensemble of all models that equally match the data and prior information. Employing the properties of the Voronoi cells, he demonstrated how a dynamic scaling of the parameters throughout the optimisation of the solutions could considerably improve the exploration.

2.8 Uncertainty Quantification (UQ) in reservoir modelling

The Bayesian framework has become a standard tool for reliable quantification of uncertainty in post-production fields where the observation data is used to estimate the likelihood (Elsheikh et al., 2015). Once the models history matched to data using an optimisation method described in the previous section, the model probability is updated according to Baye's rule (Eq. (2-1)). The updated probability (posterior) then shrink down the estimated range of uncertainty. The ideal approach for statisticians is to implement Bayesian inference using Markov Chain Monte Carlo (MCMC). Convergence of MCMC methods, however, needs a remarkably high number of simulation models and, hence, is often very costly regarding runtime (Arnold et al., 2016).

On the other hand, adaptive stochastic optimisation algorithms designed to speed up the convergence by reducing the number of required realisations cannot estimate the posterior accurately from these algorithms (Arnold et al., 2016). Because these algorithms intrinsically bias the sampling towards some good fitting regions of parameter space and hence influence the estimate of the posterior. This bias must be eliminated to make reliable estimates of the posterior from the results of stochastic optimisation.

History matching makes use of Optimisation algorithms to generate an ensemble of simulation models by exploring the search space. The question now arises: how can one make a reliable prediction of reservoir response variables future while estimating reservoir model parameters. We answer this question by referring back to Bayes' rule where inferences are drawn from the collected ensemble to generate the posterior probability densities (PPD) of parameters.

The process of evaluating PPD requires multiple calibration models, usually referred to as an ensemble of reservoir models. The analytical solution to posterior probabilities within a Bayesian framework needs integrating the likelihood function over the entire parameter space. To avoid such heavy computations, we must find a statistical approach to approximate the integration discretely.

There are several ways to evaluate the PPD of model parameters (Hutahaean, 2017) including:

1) a local characterisation of PPD around Maximum Likelihood (ML) or Maximum *a Posteriori* (MAP) solutions.
2) Solutions that pick only a subset of the ensemble of the models such as Randomised Maximum Likelihood (RML) method
3) Sampling from the complete PPD based on acceptance criteria including Markov chain Monte Carlo (MCMC) and Rejection Sampling (RS)

2.8.1 Model prediction by Neighbourhood Bayes Algorithm

To adjust the unknown sampling distribution from a stochastic algorithm, we exert the appraisal stage of the Neighbourhood Algorithm (referred to as NA-Bayes). NA-Bayes is a Markov chain Monte Carlo (MCMC) stochastic sampling algorithm designed by Sambridge (1999b) as the complementary stage of the NA-sampler algorithm. The neighbourhood Bayesian inference algorithm is based on the idea that even poor-fitting models carry valuable information about the underlying system.

NA-Bayes facilitates error analysis of the ensemble of simulation models produced in the search stage. Instead of being limited to draw inferences about the probabilities from the single best-fitting model, NA-Bayes offers the option of utilising the suite of all models in making inferences regarding the system (Arnold et al., 2016; Sambridge, 1999b).

NA-Bayes partitions the parameter space into different Voronoi cells in that the misfit remains constant in each cell. Voronoi cells function as a surrogate strategy to estimate unknown misfit values. The algorithm then uses Voronoi cells to interpolate the posterior probabilities of unknown instances in the search space. This interpolation of the misfit surface provides posterior inference without extensive new solving of the forward problem.

The central computational challenge in the new algorithm is the creation of the resampled ensemble. The NA-Bayes approximates PPD by employing a standard statistical technique recognised as Gibbs sampler (Geman and Geman, 1984). The figure below illustrates how the Gibbs-sampler takes a random walk over the Voronoi partitioned parameter space throughout the NA-Bayes algorithm. Each proposed point of the random

walk is accepted/rejected proportional to the probabilities of the cells such that the larger the cell, the higher the chance of acceptance.

The computational costs and memory usage of the method are carefully analysed by Sambridge (1999b). It is shown that the resulting numerical algorithm lends itself easily to a parallel implementation.

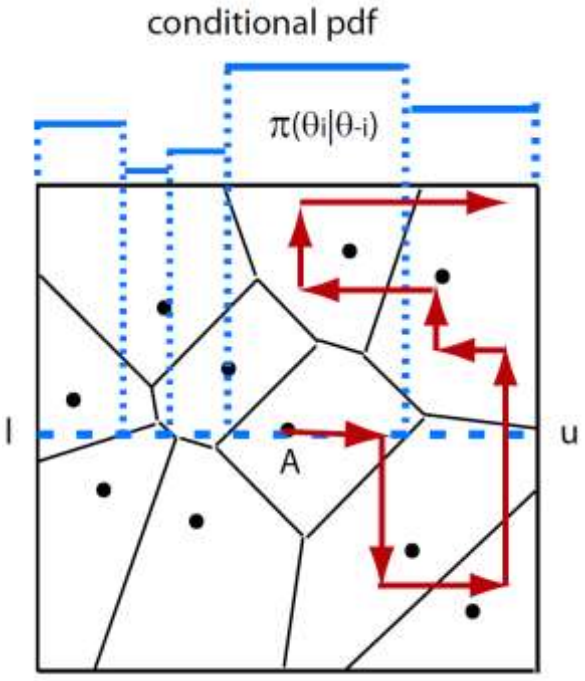

Figure 2.11 Working principles of Gibbs sampler algorithm (from Erbas, 2007)

The interpolation of scattered data in two or three-dimensional space is not a matter of concern. However, Gibbs sampler implementation may cause some geometrical issues in case of multidimensional Voronoi cells. To rectify this problem, NA-Bayes uses a prior model covariance matrix C_M encompassing model parameters to define Voronoi cells in terms of L2-norm. Because L2-norm removes the dimensionality of the parameters by controlling the impact of each parameter on the shape of Voronoi cells. The L-2 norm between points x_a and x_b in the model space is computed as follows:

$$\| x_a - x_b \| = ((x_a - x_b)^T C_M^{-1} (x_a - x_b))^{1/2} \qquad (2\text{-}5)$$

Similar to MCMC, Rejection Sampling (RS) technique uses the whole ensemble of history matching models to approximate posterior probability distribution (Casella et al., 2004). In Rejection Sampling (RS), samples are generated independently from an initial model distribution. Then the samples are accepted or rejected based on a decision function (Casella et al., 2004).

Liu et al. (2001) stated that the RS and the MCMC methods as the reliable optimisation methods for PPD approximation. However, the acceptance rule for both RS and MCMC is described in line with the likelihood of the models. Consequently, the PPD is evaluated as a result of a long chain, which often requires a considerable number of iterations (Hutahaean, 2017; Liu et al., 2001).

Chapter 3 – Modelling discrepancy in history matching of reservoir models

3.1 Introduction

In order to calibrate data in oil reservoirs, a significant amount of numerical-based simulations are generally required. The degree of accuracy of numerical solution procedures is a critical term in simulation-based prediction (Rausch et al., 2012). Such simulations deal with models that involve highly nonlinear degrees of freedom (Thomas et al., 1972) followed by equations that comprise multiple physical properties over multiple scales of time, depth, etc.

Regarding oilfield reservoirs, an accurate simulation of fluid flow through the reservoirs is complicated because, even though the fluid properties can be measured with fair accuracy, the dynamic behaviour of fluids through the reservoir is under control of the rock properties. The rock properties such as permeability and porosity can be computed by taking some samples at wells. However, these samples are representative of only a small section of the total reservoir volume and, therefore, are not sufficient or may not be even representative. Consequently, the inaccurate estimation of rock properties leads to high uncertainty in fluid flow predictions (Christie et al., 2005).

Also, reservoir simulation itself is extremely affected by the underlying geological (static) model that is inadequate (Singh et al., 2013; Oberkampf et al., 2002). The inadequate static model mainly arises from the inability to capture subscale details. For instance, Faults, permeability barriers, and other geological properties of the reservoir are not always reasonably interpreted from seismic and production data, and their exact location may be off by hundreds of meters (Bouska et al., 1999). These limitations directly influence the reservoir static model reliability.

For solid reservoir management, a careful and fast analysis of past, present and future reservoir performance is required (Watts, 1997). As stated in Chapter 2, assisted history matching is a potent tool to perform such analysis. Assisted history matching tools are based on algorithms written to correctly evaluate an objective function using multiple iterations to obtain acceptable history matched models and simulation model's parameters from a set of acceptance/rejection logic (Bhark and Dehghani, 2014). The solutions to this iterative process are ranked based on their quality, such that the models with lower

misfit values gain a higher rank (Christie et al., 2013; Subbey et al., 2004; Jones and Mitchell, 1978).

To better understand the real reservoir behaviour, reservoir engineers make sure that the reservoir simulation model fits the data appropriately. The question of how well a model matches the measured data is described by the match quality function that includes some assumptions about the errors. From a statistical perspective, incorrect assumptions about the errors may lead to the wrong prediction of the future response of reservoir models (Nobakht and Christie, 2017). Consequently, an accurate estimation of different sources of errors is of high value in evaluating the true values of reservoir model parameters which are recovered by the inverse solver (Nobakht and Christie, 2017; Morrison et al., 2018; Rabosky and Goldberg, 2015; Jones and Mitchell, 1978).

In the oil industry, like many other disciplines, measurement errors are assumed as the only source of discrepancy in simulations with the discrepancy being the difference between observation and simulated data (Yusuf et al., 2018; Jones and Mitchell, 1978). However, experienced modellers are acutely aware of bias, ranging from shortcomings of poor sampling/representativeness, anchoring on perceived best interpretations, influence from external factors to human error and organisational behaviours.

It is often assumed that the measurement errors are independent and identically distributed (*i.i.d*) for all field variables (Nicotra et al., 2005; Rotondi et al., 2006; Erbas and Christie, 2007).

Standard Least-squares misfit definition (Eq. (1-2)) is widely used for solving misfit functions in oil and gas industry as long as it analytically offers point-by-point Euclidean distances between the observed data and the calibrated data dealing with the discrepancies in minimisation progress (Bouzarkouna et al., 2014). It is the most popular approach by simply addressing the Gaussian statistics as it assumes that the theoretical relation between input and output of the solver is linear or weakly nonlinear (Gouveia, 1996).

One of the sources of noise in reservoir simulations, however, is linear, and non-linear solver convergence which causes different errors at each time step (Mishev et al., 2008). Consequently, for different solver settings, the Gaussian statistics may not be valid. On the other hand, when buildup tests are performed for reservoirs without accurate

knowledge of the initial reservoir condition, significant errors are expected to happen (Mohammad et al., 2014; Hategan and Hawkes, 2007).

In many science settings, there may exist a mutual dependence between different data points. This attribute implies that the assumption of independent observations no longer holds.

Interestingly, on the condition that the actual quantities of the physical parameters are known, the simulation model is still unable to produce a perfect match (Kennedy and O'Hagan, 2001) due to inherent inadequacy of the computer models (i.e. model error). Therefore, there always exist some uncertainty bars on the simulator output due to the model error.

If we ignore the model error, the estimation of uncertainty becomes biased, because the probability distribution of some parameters is far from the truth (Stephen et al., 2007; Vink et al., 2015). To avoid this bias in the prediction of field variables, we must account for the modelling discrepancy.

Model error is usually ignored for the following reasons: First, engineers rely on simplifications and abstractions which do not correctly represent the modelled system. The simplifying assumption here is that the mismatch is small enough that any discrepancy between the simulator output and the real value of a quantity is dominated by observation error (Morrison et al., 2018). Second, any misleading prior belief about the discrepancy will impose an additional uncertainty to the predicted quantities and must be prohibited (Brynjarsdóttir and O'Hagan, 2014). Also, the question remains of how to account for the model discrepancy.

Conventionally, a history match with standard least-squares regression model makes use of measurement uncertainty (the grey dashed-line in Figure 3.1) solely and ignores the modelling discrepancy (the blue dashed line in Figure 3.1). On the other hand, History match with an error model posits uncertainty bars on both the simulation model output and observation.

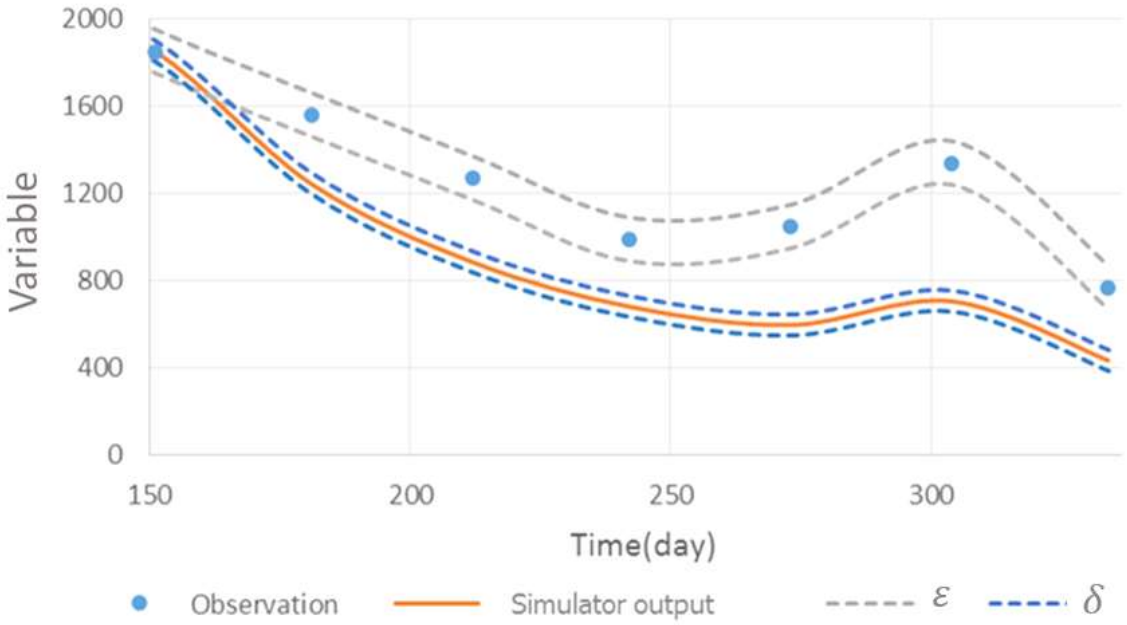

Figure 3.1 An Error model aims at placing uncertainty bars δ on the simulator output in addition to the observation uncertainty ε

In this chapter, we endeavour to corroborate modelling discrepancy to the misfit function. To do so, we construct an error model which makes use of both measurement and model errors in the misfit function to calibrate our models (see Figure 3.1). The error model downweight the misfit reasonably by placing higher uncertainty on the observable variable (Stephen et al., 2006).

We will then discuss how the new misfit/likelihood representation ultimately influences our belief about each calibration model and the uncertainty quantification of production profiles.

The estimated production profiles gained by history matching with modelled discrepancy should differ from those of standard least-squares which ignored modelling discrepancy. The difference comes from the change of representative uncertainty in the likelihood/misfit function.

We present two methods of lumping uncertainties caused by modelling discrepancy into the misfit function. First, the unknown covariance matrix of modelling discrepancy can be estimated based on some ensemble of history matched models (i.e. single-level

modelling) based on a single deterministic case study. Second, we can parametrise the total uncertainty with some unknown parameters that represent the sum of both measurement and model errors. Then, we tune these new parameters along with model parameters to perform a probabilistic forecast. The second method is referred to as hierarchical/multilevel modelling (Kang et al., 2015) because the statistical model that accounts for the uncertainty is established above the simulation model.

3.2 Errors in computer modelling of reservoir models

As discussed in the introduction, ignoring any source of errors can lead to biased parameters estimation in a calibration process which may cause further poor decisions at the top-level problem. However, assessing uncertainty in the model predictions can be a challenging task. It requires the models to match the known data adequately, and then an appropriate calculation of probabilities to establish the likely ranges of behaviour (Rausch et al., 2012).

On the other hand, realistic simulations for multiscale and complex problems will cause a significant degree of uncertainty in solution accuracy to make one believe estimates of solution error in these procedures are of high importance (O'Sullivan and Christie, 2005b). For instance, compressible fluid flow equations usually have a high degree of nonlinearity. These equations are linearised by the reservoir simulator to obtain the pressure distribution in the reservoir.

Tavassoli et al. (2004) showed that a good match in productions close to those of the "true" case might not be a good representative, nor does it always result in a reliable forecast. Even if we know all the true inputs and properties required to do the forecast of the process being modelled, where no uncertainty applies, the predicted value will not thoroughly match the real value of the process. This discrepancy between the experiment and the mathematical model (computer model) is referred to as model inadequacy or modelling discrepancy (Rabosky and Goldberg, 2015; Morrison et al., 2018).

O'Sullivan and Christie (2005b) stated that there always exists uncertainty in the problem specification due to observation and simulation errors in that the likelihood will be defined by assigning probabilities to the solution and experimental errors of different sizes. Thus, the required probability models for both types of errors must be supplied by independent analysis.

Watkins and Modine (1992) claimed the existence of a considerable error in the information supplied underlines the possibility of using stochastic methods for uncertainty quantification. In another study, Glimm et al. (2004) tried to sort out the sources of uncertainty and error by identifying separate contributions to the total standard deviation σ.

Kennedy and O'Hagan (2001) discussed the different sources of uncertainty in computer modelling besides uncertainty about context-specific inputs. They classified these errors as parameter uncertainty, parametric variability, residual variability, code uncertainty observation error, and modelling error. Since the last two sources of errors were discussed in the introduction of this Chapter, we give a brief explanation to the rest as follows:

- Parameter uncertainty stems from uncertain parameters of the model, which are expected to have a value over a presumed range of contexts. However, the exact values of the parameters remain unknown to practitioners and cannot be controlled in physical tests.
- Parametric variability comes from the use cases where some of the inputs are deliberately uncontrolled or unspecified. Therefore they add further uncertainty to the model output (Saracco et al., 2014).
- The residual variability is defined as the variability that happens in the process output yet under replicated experimental conditions. Stochastic characteristics of the experimental setup can cause the residual variability (Kalyanaraman et al., 2016).
- Code uncertainty is attributed to the uncertainty in the unknown output of computer code before running simulation by a different configuration of inputs.

With regard to the calibration of physical systems, Christie et al. (2005) specified three main categories of simulation errors: (*i*) inaccurate input data, (*ii*) inaccurate physics models and (*iii*) the limited accuracy of the solutions. A perfect physics model with exact input data will produce biased solutions if the underlying physical equations are solved poorly. Besides, a perfect solution of the incorrect equations will also generate unreliable answers.

O'Sullivan and Christie (2005b) attempted to estimate the viscosity of an injected gas displacing oil in a porous medium by introducing the model bias into the misfit function

using a solution error model. The model bias was considered to be the discrepancy between the coarse grid and fine grid simulations.

They showed that the use of the traditional least-squares misfit in a biased model yielded a biased estimate of parameters to a very great extent and proposed the solution error model to replace a single variance with a time-dependent covariance. However, only one parameter, the viscosity of the gas, was estimated in their work to prove the efficiency of error modelling.

In our study, we use multivariate reservoir models and account for the model bias within the history matching process. Figure 3.2 demonstrates the importance of modelling discrepancy in the prediction of production profiles of a reservoir field variable. The history match aims to capture the unknown truth about the field variable in the calibration period and then forecast into the future. However, the simulation model is subject to discrepancy (Figure 3.2-c), and the prediction interval from an ensemble of matched models cannot capture most of the true response of the field variable (Figure 3.2-d). The reason is that the estimated measurement errors seem to have underestimated the uncertainty and yield narrow (overconfident) confidence interval.

Next, accounting for model bias results in higher values for standard deviations. The total uncertainty estimated by an error model (the black error bars in Figure 3.2-e) are more significant than the measurement uncertainty. Adding the estimated model bias to the simulator output eventually reduces the mismatch between the predicted field variable and the truth. This is followed by a reasonable forecast of the field variable where the estimated confidence interval covers the entire truth.

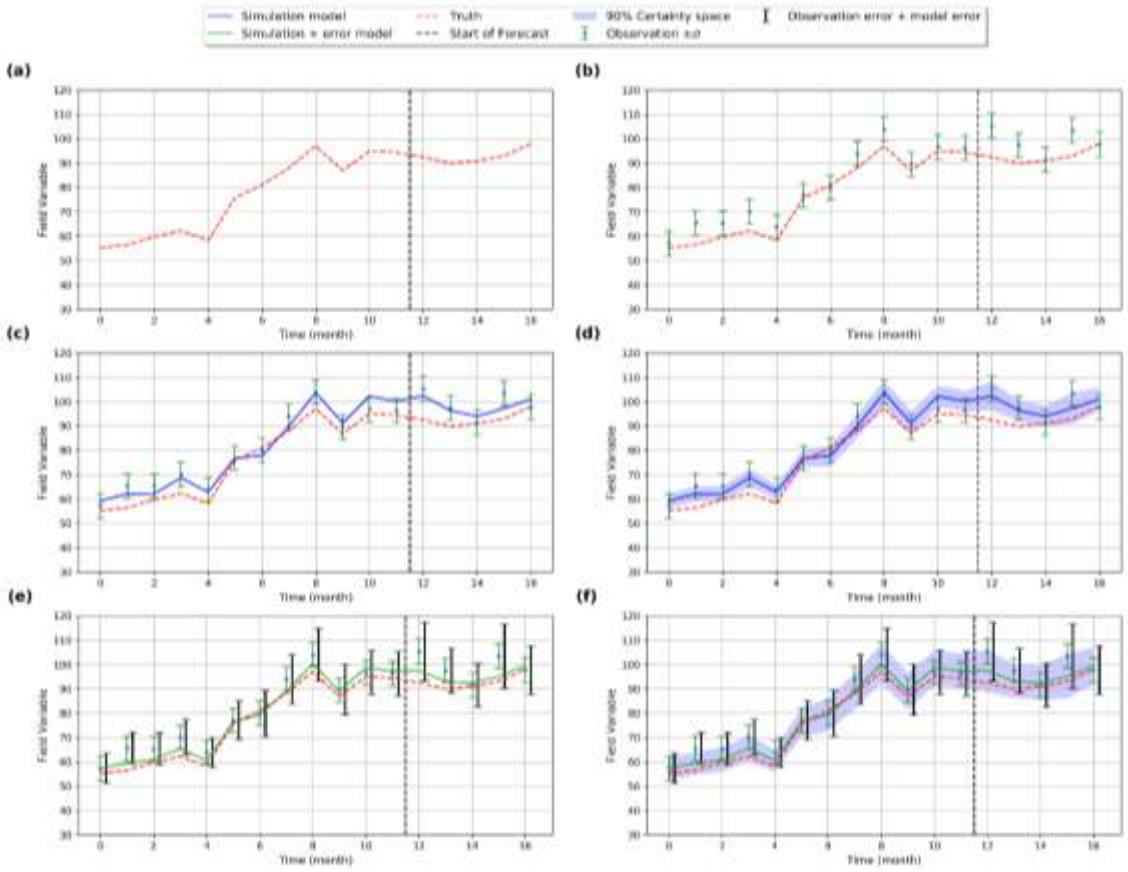

Figure 3.2 Building an error model for a field variable: **a)** The True response of the field variable we aim to recover by the simulator; the time steps 12 through 16 are held for validation of forecasts **b)** The experimental data collected throughout the life of reservoir with some uncertainty $\pm\sigma$. **c)** The simulation model history matched to the first 11 points has some degree of discrepancy with respect to the truth; the discrepancy becomes more significant in the forecast period. **d)** The estimated 90% confidence interval for the simulator response fails to cover the truth; the narrow uncertainty implies overconfident prediction of the field variable. **e)** The error model takes account of model bias with associated uncertainty (black error bars) and adds it to the simulator output which reduces the discrepancy; the estimated uncertainty by error model are larger than the measurement uncertainty (green error bars) confirming the fact that uncertainty had been underestimated. **f)** Prediction under new estimated uncertainty leads to a wider confidence intervals covering the entire true response.

3.3 Bayesian statistics

Estimates for the model parameters are often determined by Bayesian analysis throughout the process of history matching (Okano, 2013). Bayesian analysis uses a statistical model to relate the observations to the model output (data) via iterative progress.

In the Bayesian framework, questions about uncertainties of flow parameters estimates are addressed via a posteriori probability density (Morrison et al., 2018). If this probability is simple (e.g., Gaussian), this analysis is easily predictable and unchallenging. Because, if the sample size is sufficiently large, then the posterior densities are Gaussian or near-Gaussian due to the Bayesian central limit theorem (Miroshnikov et al., 2015; Van der Waart, 1998; Le Cam and Yang, 2012). Otherwise, more elaborate procedures such as Monte Carlo sampling may be required (Okano, 2013; Miroshnikov et al., 2015).

Bayesian imposes probability densities on the models themselves. These probabilities, which represent measures of degrees of belief, are coupled with the data misfit function into a final (a posteriori) probability density on the parameter space.

Now, let $\theta = \{\theta_1, \dots, \theta_l\}$ be the model parameters in parameter space Θ; and D be the observed data, typically obtained as input-output pairs $D = (x_D, y_D)$. x_D are control variable corresponding to instances of observation (e.g. time and location), whereas y_D are the observed variables (or response variables) at those instances. Then in a Bayesian framework, given the observation D the prior probability of the model denoted as $p(\theta)$ is updated by sampling from the likelihood $p(D|\theta)$ to generate posterior probability density function $p(\theta|D)$ which yields the following equation (see Pernot and Cailliez, 2017) :

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta} \tag{3-1}$$

The denominator in the above equation affirms that all posterior densities integrate to 1. Note that the term $p(D|\theta)$ is the likelihood of the data under the assumption that the model is correct.

Probability densities in Eq. (3-1) can describe the results of measurements of observable data, the prior knowledge of model parameters and the real correlation between model parameters and observable data. A major consideration is when a priori information can reasonably be represented probabilistically.

Now suppose a code or function $S(x_D, \theta)$ is used to reproduce the observations for a defined model configuration θ. In other words, $S(x_D, \theta)$ corresponds to the relationship between model inputs and outputs. Then, the mismatch between simulator output $S(x_D, \theta)$ and the observations at each instance i is:

$$mismatch = y_{D,i} - S(x_D, \theta)_i \tag{3-2}$$

The mismatch can originate from different sources such as uncertainty in model parameters, discretisation error, numerical approximation, inadequate parameterisation or unmodelled physics. Practitioners can deal with the mismatch in 3 different ways. First, one can add more physics to the model which requires a more extensive understanding of the model and higher computation time (Christie et al., 2005). Second, one shall rely on the simple assumptions about discrepancy (e.g. Gaussian white noise) which yields an overconfident prediction of the future model behaviour (O'Sullivan and Christie, 2006). Third, a comprehensive analysis can be done to explore possible correlation structures of discrepancy using all information at hand and carry it forward to the misfit formulation (Pernot and Cailliez, 2017; Morrison et al., 2018).

In this study, we compare the predictive capability of the second (ignoring model bias), and the third (accounting for bias) approaches. We then do a full investigation of all types of information that can precisely analyse the uncertainty in simple and complex calibration processes.

Theoretically, the likelihood function or likelihood consists of the accessible data provided by a sample of the model contained in the observation data. As discussed earlier, the definition of the likelihood function relies on assumptions about modelling errors and experimental errors of the observation data.

The question is: to what extent do we believe in the imposed likelihood for each model? One problem may be that if the likelihood function is centred on a biased estimation of parameters, then the posterior probability is no more credible (Curran, 2005; Nandram and Xu, 2011). This later yields inaccuracies in oil prediction along with a misleading assessment of reservoir properties.

A simple assumption about the noise model is that the errors are independent and identically distributed (*i.i.d.*) and are drawn from a fixed probabilistic model (Gouveia, 1996).

If we assume that errors are *i.i.d* around zero mean with a known measurement uncertainty σ (i.e. Gaussian white noise model), then the likelihood of matching each point i is

$$\mathrm{p}_i = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(- \frac{\left(y_{D,i} - S(x_D, \theta)_{\,i}\right)^2}{2\sigma^{\,2}} \right) \tag{3-3}$$

If the output constitutes a set of random variables having joint multivariate distribution, then the probability of matching all the points is the product of matching each point individually, and so is

$$p = \prod \mathrm{p}_i = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n exp\left(- \sum_{i=1}^{n} \frac{\left(y_{D,i} - S(x_D, \theta)_{\,i}\right)^2}{2\sigma^{\,2}} \right) \tag{3-4}$$

The above joint density is a product of marginal densities, which means that matching points are independent of each other. Moreover, if we take logs of Bayes' rule, then we obtain

$$log\, p(\theta|D) = log\, p(D|\theta) + log\, p(\theta) \tag{3-5}$$

Multiplying both sides by -1 and substituting for the likelihood, we obtain an expression for the standard least squares misfit formulae:

$$LSQ = -\log p(\theta|D) = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{(y_{D,i} - S(x_D,\theta)_i)^2}{\sigma^2}\right) - \log p(\theta) \qquad (3\text{-}6)$$

Eq. (3-6) is a very simple definition of the misfit based on least-squares regression model which comes from the logarithm of the likelihood. However, the misfit is normally higher than would be anticipated from significant errors.

Significant errors are assigned a higher weight, because of the squaring, than are minor errors. Thus, least-squares favours many medium-sized errors over several large errors which result in consistently over-predicting the mean value (O'Sullivan and Christie, 2005).

The standard deviation may be subject to further inaccuracies in parameter estimation causing overconfident probability distribution for the estimated parameter. It also may be that the fixed value for variance is chosen too large, leading to an under-confident parameter estimate (Figure 3.3-a). For example, the simulated solution in a coarse grid may evaluate a late breakthrough time such that the difference between the observed and predicted data is higher than at other instances of time.

On the other hand, choosing too small values for variance results in an unwantedly too low likelihood for most of the parameter space and too overconfident estimates for the rest of the space (Figure 3.3-b). In addition, the noise characteristic of field variables such as injected gas is not always constant in which cases time-varying noise statistics should be considered (Jahanshahi et al., 2008). For such reasons, a single value for a variance, as in Eq. (3-6) is yet to be unrealistic.

Moreover, a standard least-squares misfit can cause highly biased parameter predictions when building an approximate reservoir model, such as an up-scaled model (O'Sullivan and Christie, 2006).

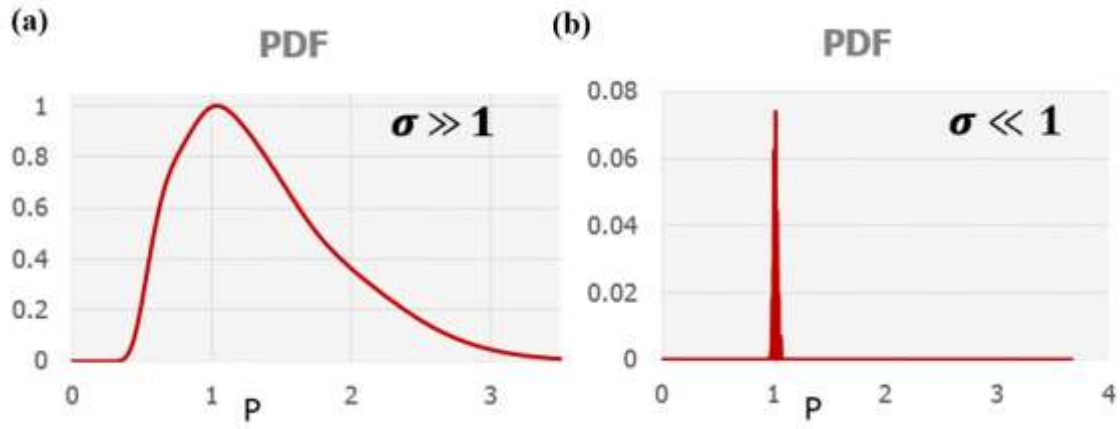

Figure 3.3 Posterior estimates of a physical model parameter P using standard least-squares misfit with **a)** large standard deviation and **b)** small standard deviation.

Eq. (3-6) assumes a constant σ and ignores the orrelation between different instances of measured quantities. For more generality, a full covariance matrix of measurement uncertainty C_y can give a formulae for the LSQ misfit function as

$$LSQ = \frac{1}{2}\left|(y_D - S(x_D,\theta))^T {C_y}^{-1}(y_D - S(x_D,\theta))\right| - log\, p(\theta) \qquad (3\text{-}7)$$

Note that if $C_y = \sigma^2 I$, then Eq. (3-7) becomes equal to Eq. (3-6).

During the process of depletion, reservoir fluids flow through the production wells to the field surface, because the pressure at the bottom of the well is greater than that caused by the hydrostatic pressure of the column of oil in the well. The depletion occurs naturally until the oil flow rate drops over time since the reservoir pressure declines. At this point, pumping might be employed to keep the oil rate at economic levels (Muggeridge et al., 2014).

If the reservoir pressure drops below oil bubble point pressure, then the gas in the oil will release from the solution. Since the separated gas has a considerably lower viscosity, it will flow up front to the production well. Subsequently, the viscosity of the remaining oil in the well rises. This will decrease the oil rate and causes a great deal of uncertainty in

the oil rate predictions, necessitating higher error bars on the observed oil rate (Muggeridge et al., 2014; Mohammad et al., 2014; Jahanshahi et al., 2008).

Figure 3.4 shows how a time-varying noise model can reliably estimate the uncertainty in the time domain while the Gaussian white noise model assumption no longer holds.

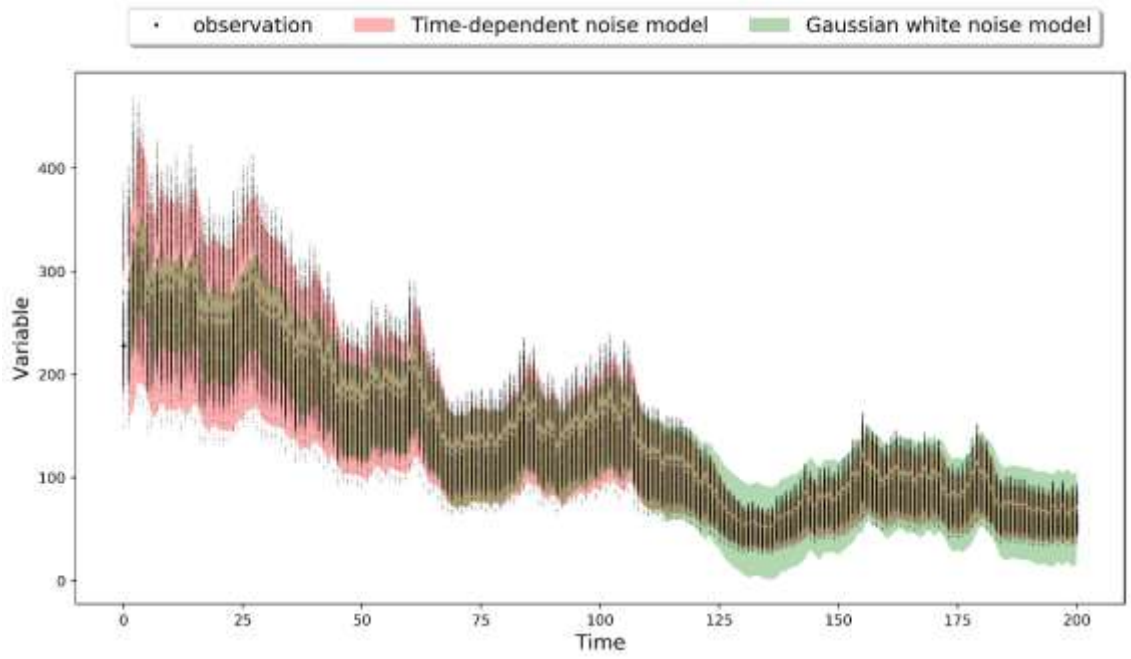

Figure 3.4 Noisy observation from a time-series variable: a time-varying noise model can estimate large and small variations in time. A Gaussian white noise overestimates (and underestimates) certain parts of the time domain.

In other studies on misfit definition, an attempt has been made to add the term of shape matching into the formulation of the misfit. Bouzarkouna and Nobakht (2015) proposed a new misfit definition to track the similarities in sets of realisations using directed, quantile and average Hausdorff distances. They claimed that the new misfit definition could be a better alternative objective function in handling the curve similarities along with the Euclidean distances between the history and the simulated models at the same time.

Bertolini and Schiozer (2011) assessed the impact of 8 different objective functions on the history matching. They concluded that the normalised and weighted functions do not

outperform the square error objective function in their case studies. Also, Bertolini and Schiozer (2011) do not test the influence of objective function on estimates of model parameters.

In this work, we evolve as much information as available into the misfit function and examine the predictive performance of our models in the estimation of model parameters and production profiles. However, before doing so, an extensive analysis of sources of uncertainties with underlying assumptions shall be done in the following sections.

3.4 Modelling discrepancy

Recent studies indicate that the mismatch may not follow a Gaussian distribution, especially when extensive approximations of the physical processes are considered (Muggeridge et al., 2014; Jahanshahi et al., 2008). In previous sections, we discussed the causes of model bias but did not address how it can change the likelihood function.

Christie et al. (2005) showed that neglecting mean errors in the likelihood function (mean zero assumption) undoubtedly misleads the prediction when extrapolating discrepancy beyond the range of experimental data and eventually causes a fair amount of uncertainty. They claimed to have achieved a high accuracy as of a fine grid model by merging a coarse grid model with error data to avoid high computation costs of fine grid models. As a result, the maximum likelihood estimate produces a more realistic parameter estimation compared to the actual values.

Kennedy and O'Hagan (2001) presented a comprehensive analysis of all errors, whether those associated with the input variables or the model discrepancy in the context of model calibration. However, they have not answered the question of why these errors occur.

They picked the best input approach and assumed that there is a single 'best-fitting' value of input parameters $\hat{\theta}$ such that the model most precisely represents the system given the imposed error structure. However, their approach takes account of predictions in the regions where data exist (i.e. interpolation) and the situation may be subject to complexity where extrapolation is of interest.

Consequently, posterior density converges around the best-fitting value and $\hat{\theta}$ lies in some arbitrary tight neighbourhood which is, in general, biased and unable to be estimated correctly. Hence, Kennedy and O'Hagan's terminology suffers from a weak prior

information and cannot provide good learning about physical model parameters. The other issue is that if model inadequacy is modelled with an uninformative prior assumptions, interpolations are unbiased, but extrapolations outside the range of the observations as well as parameters estimate will be biased (Brynjarsdóttir and O'Hagan, 2014). In the context of reservoir models, that the historical data only conditions the models for the explored drainage areas and displacement processes, the model forecast becomes less reliable as the step out range increases.

In the case of equifinality where θ is not identifiable in the model, there will not be a unique $\hat{\theta}$ but a set of such values. Brynjarsdóttir and O'Hagan (2014) modelled the model discrepancy as an additive correlated error term. They constrained the model discrepancy function to achieve a more realistic prior information about the model discrepancy function and physical parameters. However, the issue of formulating prior knowledge about model inadequacy is still a new area of research in realistic calibration to take accounts of what simulator is missing throughout the calibration (Morrison et al., 2018; Ling et al., 2014; Parish and Duraisamy, 2016).

3.4.1 Single-level error modelling

Computational models are built to make predictions, by which we intend to predict values of model response for unseen quantities. Such predictions are fundamental to engineering design and development plans.

If we make use of an inadequate computer model that fails to capture the true physical process, as is often the case, it is crucial to characterise the uncertainty in the prediction due to modelling discrepancy (Morrison et al., 2018; Jones and Mitchell, 1978). For simulation models of complex physical processes, modelling discrepancy is usually a substantial contributor to the predictive uncertainty (Sargsyan et al., 2018). Consequently, a reliable representation of the modelling discrepancy is required (Sargsyan et al., 2018; Rabosky and Goldberg, 2015).

A single-level error model can be carried out to evaluate the correlation between observed errors in different instances of time. The term single-level modelling of uncertainty suggests that the analysis is done at one analytical level without parameterization. Therefore, the covariance matrix of uncertainties is estimated analytically based upon observed errors.

However, this interpretation would not support any inferences to be made about individual-level correlations, such as the individual level relationship that may exist between a physical model parameter and observed errors. In that regards, it is obvious that we must estimate model bias using inadequate data to determine correlations/covariances. However, we should expect such estimates to be erroneous (Lucas, 2014). Consequently, some researches seem to support the use of the multilevel models on such samples (Hox, 1995; Lucas, 2014).

Despite giving a straightforward representation of the distribution of errors, the single-level error models ignore that error can be grouped within a time-series domain.

Therefore, a standard approach is to model and calibrate a fully statistical description of the discrepancy (often referred to as a bias function) between the simulator output and the actual quantity of that output (Pernot and Cailliez, 2017; Morrison et al., 2018). These functions, often known as emulators/surrogate models, can generalise towards entire data space. In Chapter 4 we give a concise explanation of these emulators.

To describe the theory of a physical system like oil reservoirs accurately, some experiments must be tested. This can be done by comparing the observed data with the predicted output of a mathematical model that represent the whole physics. The theoretical notation for a given model m can be denoted:

$$m \mapsto\ y_D = S(x_D, \theta) \tag{3-8}$$

where $D = (x_D, y_D) = \{x_{D,1}, y_{\mathrm{D},1}, x_{D,2}, y_{\mathrm{D},2}, \dots, x_{D,m}, y_{D,m}\}$ comprises input-output pairs of observed data in a finite-dimensional data space $\mathfrak{D}$; $\theta = \{\theta_1, \theta_2, \dots, \theta_l\}$ is a set of parameters for a finite-dimensional model space Θ; and $S(\,.\,)$ is a linear/nonlinear operator that maps model space into data space to predict y_D.

If uncertainties can be neglected (i.e. idealised model), a functional relationship $y_D = S(x_D, \theta)$ recovers the exact values of the model parameters θ, and the predicted data values equal y_D.

However, the predicted values for a physical system by the operator $S(.)$ can never recover the observed values due to inherent model inadequacy and experimental errors previously described as sources of errors.

Now, let $\mathbb{j}$ be a joint space manifold for parameter space Θ and data space $\mathfrak{D}$ which can be described as $\mathbb{j} = \Theta \times \mathfrak{D}$. Therefore, the points in the new manifold are of the form $j = \{\theta, D\}$. Now, assuming that the joint probability density $p(\theta, D)$ over $\mathbb{j}$ is normalised, then the marginal probability densities are:

$$p(\theta) = \int_{\mathfrak{D}} dD\ p(\theta, D) \quad , \quad \int_{\mathfrak{D}} dD\ \ p(D) = 1 \tag{3-9}$$

$$p(D) = \int_{\Theta} d\theta\ p(\theta, D) \quad , \quad \int_{\Theta} d\theta\ p(\theta) = 1 \tag{3-10}$$

The intuitive interpretation of the marginal probability equations above is clear, as the projection of the joint probability density $p(\theta, D)$ over $\mathfrak{D}$ and Θ.

Now, Let us simply admit here that our problem is mildly nonlinear (or it is linear), and the coordinates θ and y are not too distant from being 'Cartesian coordinates' over nearly linear manifolds (Tarantola, 2005). Then, assuming homogeneous marginal probability density for model parameters $p(\theta)$, the conditional probabilities can be captured by the division of the joint probability density $p(\theta, D)$ and the marginal probabilities. Under these restrictive conditions, we arrive at

$$p\,(D|\theta) = \frac{p(\theta, D)}{p(\theta)} \tag{3-11}$$

$$p(\theta|D) = \frac{p(\theta, D)}{p(D)} \tag{3-12}$$

and merging the two conditional probability equations, we arrive at the Bayes' rule:

$$p(\theta|D) = \frac{p(D|\theta) \times\ p(\theta)}{p(D)} \tag{3-13}$$

Further, assume to have independent knowledge of measurement uncertainties for y_D represented as covariance C_y. Then, the conditional probability $p(D|\theta)$ can be rewritten in terms of least-squares criterion which yields the following equation:

$$p(\theta|D) = (2\pi)^{-\frac{nm}{2}} \left|C_y\right|^{-\frac{n}{2}} \exp(-\frac{1}{2} \left|(y_D - S(x_D, \theta) - \bar{e})^T {C_y}^{-1} (y_D - S(x_D, \theta) - \bar{e})\right|) \times p(\theta) \tag{3-14}$$

where n is the number of realisations, m is the length of the history (or error) points and $\bar{e}$ is the prior assumption we make about the mean error . The logarithm of Eq. (3-14) equals the LSQ misfit in Eq. (3-7) if the mean error is assumed zero. It accounts for measurement uncertainties in form of a full covariance C_y, whereas the uncertainty caused by model bias is ignored.

Now, let us impose some uncertainties on the simulator output regarding the model inadequacy regarding a covariance matrix C_m which is independent of C_y. Then, under the Gaussian assumption, the uncertainty in observable data can be represented by the addition of the covariances of measurement errors and model inadequacy (Tarantola, 2005). Then, the new misfit function we call as Full Error Model (FEM) describing the error model can be computed by taking $log\ p(\theta|D)$ as:

$$FEM = \frac{1}{2}\left|(y_D - S(x_D,\theta) - \bar{e})^T (C_m + C_y)^{-1}(y_D - S(x_D,\theta) - \bar{e})\right| - log\, p(\theta) \quad (3\text{-}15)$$

This result is remarkable because it explains that, under the Gaussian assumption, measurement uncertainties and model inadequacy readily combine by addition of their respective covariance operators, even when the solver is nonlinear (Tarantola, 2005). However, in real-world applications, the C_m is not known before running calibrations. Therefore, we can forget there are two different elements of uncertainties and find the unknown *total uncertainty* $\Sigma = C_m + C_y$. In the next section, we treat the total uncertainty as an unknown term and give a different solution to Σ.

Eq. (3-15) is a general formula for correlated/uncorrelated uncertainty while information about measurements and Gaussian modelling error are available. This can also be developed when initially a naive calibration is done to gain a joint *a priori* state of information on the model parameters. For a maximum generality, let us define the covariance C_θ accounting for the prior information about model parameters which is entirely independent of the results of measurements. The latter should find a physical correlation between input and output of a fixed solver and add it into the uncertainty analysis. In simulations, this may be achieved by doing a first quick history match with a standard-least squares misfit function when there is no knowledge of C_m and/or C_θ and start the process over using new misfit definition in a separate analysis. The new misfit shall combine model inadequacy, measurements errors and *a priori* information on a Gaussian assumption as follows:

$$FEM = \frac{1}{2}\left[\left|(y_D - S(x_D,\theta) - \bar{e})^T \left(C_m + C_y\right)^{-1}(y_D - S(x_D,\theta) - \bar{e})\right| + \left|(\theta - \bar{\theta})^T C_\theta^{-1}(\theta - \bar{\theta})\right|\right] - log\, p(\theta) \quad (3\text{-}16)$$

where θ refers to the model being calibrated in second calibration progress and $\bar{\theta}$ provides prior information about parameters. Unless having a reliable *a priori* information about

model parameters, Eq. (3-16) is prohibitive and causes further biased prediction. Since, no such information is provided in our case studies, the Eq. (3-15) is preferred over Eq. (3-16) in both of history match and forecast.

The reality is that these models can never be adjusted to perfectly neither by LSQ (Eq. (3-7)) nor by FEM. Even if they are matched very well, there is no certainty that they will produce a reasonable forecast. Even with the best configuration of the parameters, there will always be some residual between the data and the model, due to either model misspecification, measurement error, or both (Vaart et al., 2018).

Brynjarsdóttir and O'Hagan (2014) showed that less informative priors about the model bias appeared to make better inference about unknowns when extrapolating untrained data. That being said, in all of our computations, we avoid making a prior assumption about the mean error and set it to be zero.

Figure 3.5 schematically compares the LSQ model (Eq. (3-7)) with the FEM model (Eq. (3-15)). As can be seen from Figure 3.5-a, the LSQ regression model only accounts for measurement uncertainty C_y to estimate the posterior probability of θ given data $p(\theta|D)$. Then, a single-level error model is constructed (Figure 3.5-b) which makes use of a covariance of model inadequacy C_m and update $p(\theta|D)$.

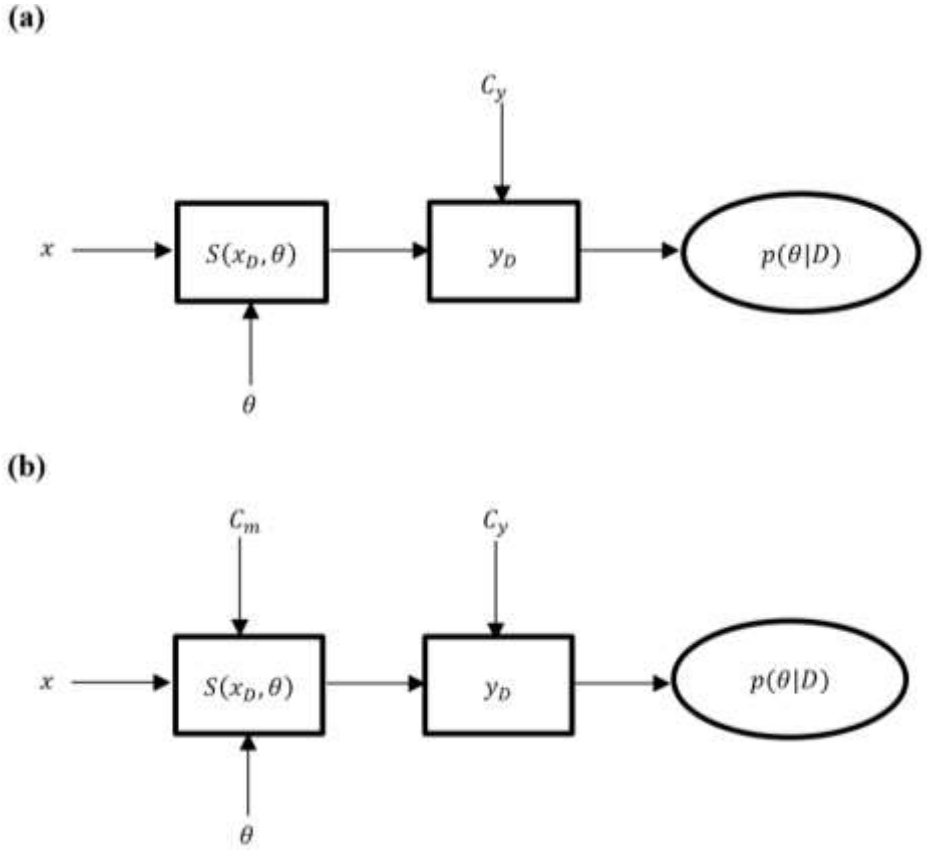

Figure 3.5 Schematic of **a)** LSQ model (Eq. (3-7)) and **b)** Single-level FEM error model (Eq. (3-15)) for posterior $p(\theta|D)$ estimate. LSQ model places uncertainty only on the measured output as C_y covariance whereas the error model also places uncertainty bars on simulator output in terms of C_m.

3.4.2 Parametric hierarchical modelling of discrepancy

If there is no knowledge of measurement uncertainty, the standard least squares framework can no longer find an optimal solution. Tarantola (2005) gives a concise statement about unknown uncertainty bars on data:

"If hesitation exists in choosing the a priori uncertainty bars, it is of course best to be over conservative and to choose them very large. A conservative choice for correlations is to neglect them". It may be that the calibration variable has a different scale of changes throughout the calibration and then σ varies at different periods. This implies that there should be a proper regression model exploring the possibility of σ variations (for instance,

the time-varying noise model in Figure 3.4). The model should also consist of a penalising factor which hinders overestimation of σ (Nobakht and Christie, 2017).

The notation of exploring different regression models above the reservoir model is referred to as a hierarchical or multi-level model. In a simple hierarchical model, the data can be classified into different groups, and two types of parameters jointly assess probability distribution over model outcome: model parameters θ shared among all groups and tuning parameters φ shared across each group. This enables a hierarchical model to learn about physical model parameters θ and a set of tuning parameters (e.g. time-varying σ) which are not modelled in the simulator. Tuning parameters, applied as a simplified surrogate in a different discipline, can contribute to fit best to reality and enhance the predictive performance of the model; are not dependent on the context of simulator's application; and nor they are of high scientific interest (Brynjarsdóttir and O'Hagan, 2014).

Apart from a rather great inferential imprecision, the advantage to the multi-level approaches is that in case of a considerable number of data points, grouping will ascertain a better match to data rather than pool analysis (Jackman, 2009).

A hierarchical model that is based on a finite set of tuning parameters (hyper-parameters) is referred to as a parametric hierarchical model. As an example of a parametric regression model, observations can fall into different groups, each with a constant unknown σ. Then history matching is able to automatically select the group-specific standard deviation. If we assume that the measurement errors are uncorrelated Gaussian scattered around a non-zero mean $\overline{e_i}$ with a standard deviation of the measurement error σ_i varying along the input space (e.g. location, time) then we can rewrite the likelihood of matching all points as

$$p = \prod \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\left(y_{D,i} - S(x_{D,i}, \theta) - \bar{e}\right)^2 / 2{\sigma_i}^2\right) \qquad (3\text{-}17)$$

Following the Bayes' rule, the hierarchical FEM misfit formulae for input-dependent standard deviation becomes

$$FEM = -\log p(\theta|D) = \sum_{i=1}^{n} \frac{\left(y_{D,i} - S(x_{D,i},\theta) - \bar{e}\right)^2}{2{\sigma_i}^2} + \sum_{i=1}^{n} \log \sigma_i \quad - \log p(\theta) \tag{3-18}$$

Note that the unknown σ_is are the tuning parameters or hyper-parameters of our parametric hierarchical model. The $\sum_{i=1}^{n} \log \sigma_i$ term penalises large standard deviations, and allows the history matching to obtain appropriate standard deviations. Likewise, in case there exists a full covariance matrix carrying information about correlated uncertainty, the logarithmic term is replaced by the log determinant of the covariance matrix Σ and the misfit formulae turns to

$$FEM = \frac{1}{2} \sum_{i=1}^{n} (y_{D,i} - S(x_{D,i},\theta) - \bar{e})^T \, \Sigma^{-1} \, (y_{D,i} - s(\theta)_i - \bar{e}) + \frac{n}{2} \log|\Sigma| - \log p(\theta) \tag{3-19}$$

The main difference between the hierarchical FEM model (Eq. (3-18) and Eq. (3-19)) and the LSQ model (Eq. (3-7)) is the FEM model treats the uncertainty as unknown Σ or σ_i's, which needs to be acquired from data, whereas the LSQ model already has this knowledge.

In conventional history match, $\bar{e}$ is set to zero; $\Sigma = \sigma^2 I$ is known for each field parameter before running simulations. One reason is that assuming a correlation between elements of Σ needs too high computational efforts in search of all elements. The limitation arises as n increases (i.e. the curse of dimensionality). Note that the covariance Σ must always hold positive-semi-definiteness (i.e. valid covariance) to ensure invertibility.

Due to the simplicity, when using the Gaussian assumption for uncertainty, we can neglect that there are two different sources of errors in the data space. All happens as if

the model was perfect and the measurement uncertainties were those represented by the covariance matrix $\Sigma = C_m + C_y$. Another consideration is to assume a diagonal matrix (with all off-diagonal elements being zero) to reduce the dimension of the regression model. Consequently, a parametric model can parametrise the unknown standard deviation of the diagonal matrix Σ.

Now, suppose $\varphi = \{\sigma_1, \sigma_2, \dots, \sigma_m\} \in \Phi$ is a set of hyper-parameters we intend to tune along with real model parameters θ with a prior distribution of θ and φ defined in advance. If the objective is to evaluate the posterior estimates of both the model parameters and the tuning parameters, then all unknown parameters must be tuned at the same time. Consequently, the posterior estimates are affected jointly by θ and φ. However, some of the parameters, often with no precise physical meaning, are not important in terms of estimation while their estimates can immensely affect the inference on the parameter of interest. In the literature, these parameters are said to be the nuisance parameters (Kuss, 2006; Garbuno-Inigo et al., 2016).

The nuisance parameters are usually the parameters of a flexible model that may better interpret data at hand, but their inference is not of high priority (Spall and Garner, 1990). For instance, in the adaptive design of a clinical trial (Pritchett et al., 2015), the variance of a continuous variable or the control group event rate are examples of the nuisance parameters. Ideally, the objective would be to integrate nuisance parameters out and evaluate posterior estimates for the parameter of interest only (Liseo, 2005).

The choice of the hyper-parameters is generally done by using Maximum Likelihood (ML) estimates, Maximum a Posteriori estimates (MAP), or marginalising over nuisance parameters in a fully Bayesian manner (Garbuno-Inigo et al., 2016; Rasmussen, 2004).

Marginalising over nuisance parameters is a standard Bayesian approach (Spall and Garner, 1990). Since the joint probability distribution depends on the value of the unknown standard deviations, one can carry out a probability-weighted average over the unknowns to get a posterior that no longer depends on the nuisance parameters. To do so, we must place a prior on the tuning parameters φ and calculate Bayesian posterior of the model parameters by integrating out the φ in a Fully Bayesian Hierarchical (FBH) manner:

$$p(\theta|D) = \int p(\varphi, \theta|D)p(\varphi)d\varphi = \int p(\theta|\varphi, D)p(\varphi|D)p(\varphi)d\varphi \qquad (3\text{-}20)$$

An approximation to a Full Bayesian solution is Maximum Likelihood (ML) solution wherein the hyper-parameters are evaluated after data are observed instead of being marginalised out. As implied by the name, Maximum Likelihood solution is concerned with searching the entire hyper-parameter space Φ and finding a set of hyper-parameter values $\hat{\varphi}$ that maximises the likelihood of observing the data D given those parameters $p(D|\varphi)$:

$$\hat{\varphi} = \left\{ \varphi \in \Phi \ : \ \arg\max_{\varphi} p(D|\varphi) = \arg\max_{\varphi} \int p(\varphi, \theta|D)d\theta \right\} \qquad (3\text{-}21)$$

In the above equation, ML returns point estimates (i.e. modal estimates) of standard deviations. An alternative Bayesian solution for ML is Maximum *a Posteriori* estimates (MAP). The MAP solution maximises the posterior probability of nuisance parameters $p(\varphi|D)$ instead of the likelihood as follows:

$$\hat{\varphi} = \left\{ \varphi \in \Phi \ : \ \arg\max_{\varphi} p(\varphi|D) = \arg\max_{\varphi} \frac{p(\varphi)}{p(D)} \int p(\varphi, \theta|D)d\theta \right\} \qquad (3\text{-}22)$$

Comparing the ML (Eq. (3-21)) and the MAP (Eq. (3-22)), the only difference is the inclusion of prior $p(\varphi)$ in the MAP. If, therefore, a uniform prior is assigned for the hyper-parameters φ, then the Eq. (3-21) and Eq. (3-22) and MAP solutions become identical.

In case the distribution is multimodal (with different global maxima), one can establish some, but not a single, ML settings for $\hat{\varphi}$, and launch the calibration with all different settings. This is an indispensable consideration in obtaining more reliable estimates of physical model parameters (i.e. lower level of hierarchy) and unbiased prediction for calibration parameters. The ML and MAP approaches, however, fail to rigorously assess the uncertainty about the true values of hyper-parameters, throwing away valuable information (e.g. full posterior distribution) about φ (Nezhad Karim Nobakht et al., 2018).

Likewise ML, MAP solution cannot rigorously estimate the parameters when the likelihood is multimodal or is a non-convex function with the possibility of optimisation being trapped in local optima (Garbuno-Inigo et al., 2016).

Figure 3.6 shows the parametric hierarchical error model we constructed in our study. As opposed to the error model presented in Figure 3.5-b, the model inadequacy is not estimated analytically but rather inferred through Bayesian statistics. The hierarchical model places a statistical model above the simulation model to produce forecast under uncertainty.

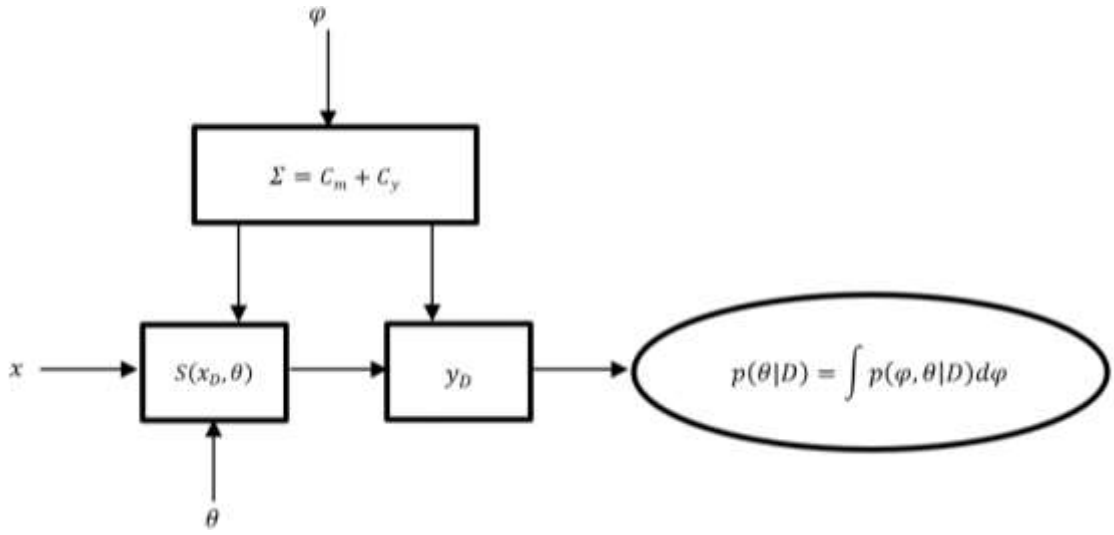

Figure 3.6 Parametric hierarchical error model for posterior $p(\theta|D)$ estimate. The FEM error model (Eq. (3-19)) parametrise the total uncertainty $\Sigma = C_m + C_y$ where both the measurement and the model errors are assumed unknown. Prior distribution must be placed for both model parameters θ and hyper-parameters of error model φ.

3.5 Conclusion

In this chapter, we reviewed sources of errors in computer modelling of physical systems, and in particular, oilfield reservoir simulation. We discussed the importance of modelling discrepancy in the likelihood function, Bayesian calibration and posterior estimates of model parameters.

Two methods were presented to construct the error models. The first error model (Eq. (3-15)) renders an analytical solution to the unknown covariance matrix of modelling error based on some ensemble of history matched models (i.e. single-level modelling). The second error model, parametric hierarchical model (Eq. (3-19)), parametrises the total uncertainty with some hyper-parameters that represents the sum of both measurement and model errors.

As for the hierarchical model, we discussed three different standard solutions to the hyper-parameters: Maximum Likelihood (ML) estimate, Maximum a posteriori (MAP) estimate, and Full Bayesian Hierarchical (FBH) solution enabling posterior estimates of model parameters independent of nuisance tuning parameters.

Although our parametric hierarchical model is able to predict the trained data, it fails to generalise since the parameters are only defined for particular instances of the time domain. On the other hand, assuming unknown standard deviation for each point of large datasets is prohibitive, because the number of parameters to be identified rises (i.e. curse of dimensionality).

Therefore, if the objective is to find the error model at any instance, we must deploy non-parametric emulators to generalise towards the entire input space. In the next chapter, we discuss these types of hierarchical models in detail.

Chapter 4 – Modelling discrepancy using non- parametric hierarchical models

4.1 Introduction

The purpose of regression is to interpolate a continuous function from a set of observed data consisting of input-output pairs. In the context of reservoir modelling, interpolation implies for datasets that are within the explored drainage area, or for time-series datasets collected throughout the life of the reservoir.

Parametric regression methods, such as the (Eq. (3-19)) we presented in Chapter 3, parametrise the function with a finite number of parameters and learn these parameters from the data.

In Chapter 3 we described why the variation of standard deviations of a time-series variable throughout the progress of time cannot always be neglected. Then, a parametric error model was used to estimate the total uncertainty of production data. Our parametric model, however, suffered from its incapability to generalise into unseen data.

Most of the parametric models are applied where measurements are homogeneous. In homogenous data structures, the statistical properties such as variances of various subsets of observations are equal.

In numerous real-world problems, however, measurements are usually not homogeneous (i.e. data come from more than one distribution) but often structured in different ways. In these problems, therefore, we must consider the possibility of mutual dependence between different measurements. This attribute implies that the assumption of independent observations no longer holds. Subsequently, observations are broken down into different groups each sharing some properties. As a result, the statistical model fitting the data must capture interdependence across the groups.

On the other hand, the information density (i.e. the amount of information per unit (Jaeger and Levy, 2007) in data is not uniform, even if all the data points have similar measurement error. For example, one key event like water breakthrough may carry much more information than many other rate measurements.

A successful practice, to handle complex non-homogeneous data structures, is the application of non-parametric models (Johnson and Willsky, 2013; Cao and Van Keilegom, 2006; Orbanz and Teh, 2011). Non-parametric models have become increasingly popular due to their flexibility, adaptability to the nature of problems and high predictive capability (McIntire et al., 2016). These models make use of a subset of parameters to explain a set of data such that the complexity of the model adapts to the data (Gershman and Blei, 2012; Johnson and Willsky, 2013).

In this chapter, we try to model the discrepancy using non-parametric models, which enables us to predict the response variable at any point of interest. Non-parametric approaches can firstly fit a model in the range of measurement data, and secondly, predict the target variable with associated uncertainty. These models place flexible priors on functions that are generalisable throughout the entire input space.

To do so, we use Gaussian Process (GP) models (Rasmussen and Williams, 2006) as emulators in the context of nonparametric model selection. A Gaussian process can take the shape of a full predictive distribution at the entire space, on the condition that a positive-definite covariance function and a mean function are known. The parameters of the GP model, known as hyper-parameters, such as the kernel length scale and the kernel variance, are learned from the data (Rasmussen and Williams, 2006).

As for covariance functions, there might be a set of candidate functions with a different number of hyper-parameters and different degree of complexity. We choose covariance functions through an error model selection framework prior to doing error modelling in the history match process.

In this work, we interpret the model selection problem for error modelling in data-driven settings that enables us to, first, interpolate the error in history period and, second, propagate it towards unseen data (i.e. error generalisation). The error models constructed by inferring parameters of selected models can predict the response variable (e.g. oil rate) at any point in input space (e.g. time) with corresponding generalisation uncertainty.

4.2 Non-parametric models

In many science settings, there may exist a mutual dependence between different variables of a real-life process. It may be that two randomly picked individuals from the same cluster/group seem to be more alike than two individuals chosen from a different

cluster. For instance, students' ability to learn at school depends on the characteristics of their class, including the quality of their teacher and the ability of other students in the class. Considering the influence of classes, we would expect exam results for students in the same class to be more similar than results for students from other classes.

By a similar argument, experimental data recorded on the same process at neighbouring instances of time turn up to be more correlated than those of far instances. Hence, dependencies can be expected to arise, and we need models that can represent correlations and variations of response variables at different time steps.

In time-series data analysis, it might be that the time instant at which some statistical characteristic such as mean, or variance varies abruptly. Therefore the time domain can be divided into some sections each representing the same statistical properties. Figure 4.1-a shows a noisy observed data whose mean and variance considerably change over time in different partitions. Each of these partitions can be assumed to share the same standard deviations (Figure 4.1-b).

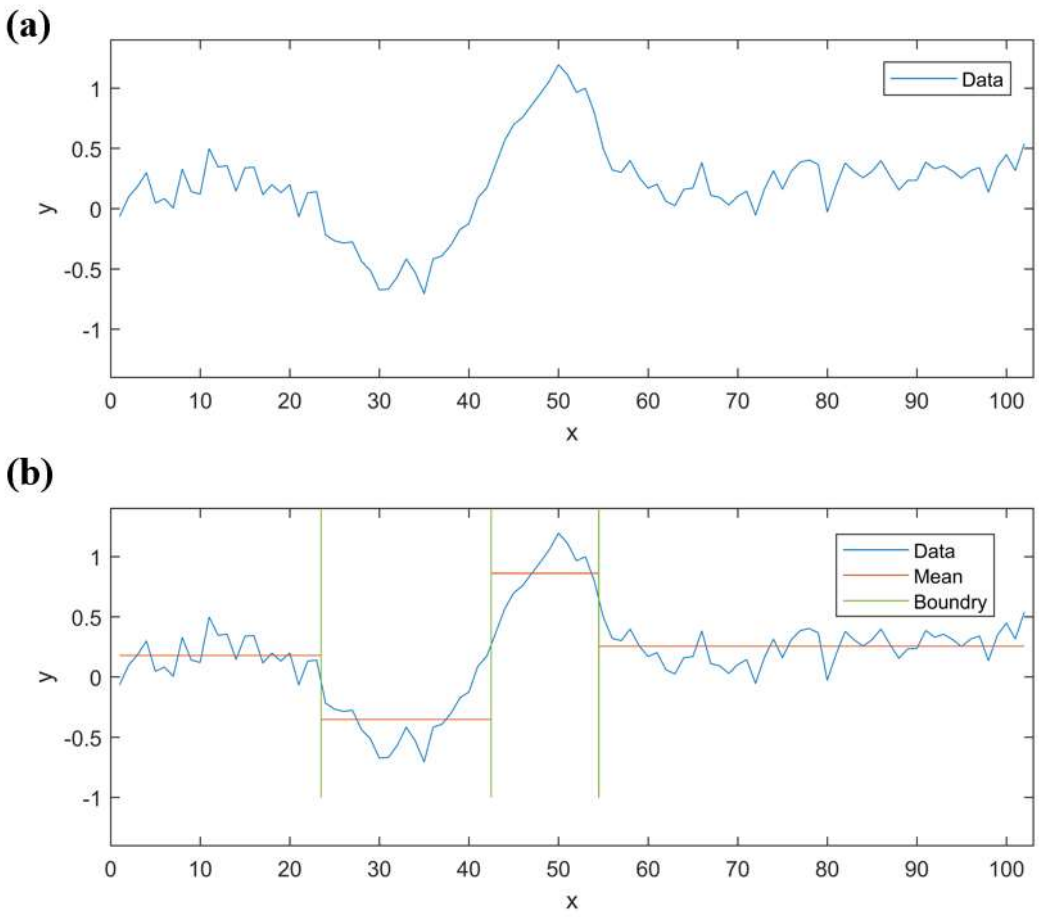

Figure 4.1 A set of noisy data partitioned into different sections: The sample standard deviation experience abrupt changes in the division points

For the above example, a parametric regression model assumes a known relationship between input x and output y at each section. However, if the size of data increases, then the parametric model is unable to adapt to the data complexity (Orbanz and Teh, 2011).

In a parametric model, we make a strong assumption about the form of regression model fitting the data. For instance, a regression model such as

$$y_i = \beta_0 + \beta_1 x_i + e_i \tag{4-1}$$

assumes a linear relation between the input x_i and the output y_i where the regression parameters β_0 and β_1 are constant numbers.

Therefore, parametric models such as Eq. (4-1) are limited to particular shapes, which may not always be appropriate; for example, modelling a multimodal distribution model with a single, unimodal model is not a reasonable practice.

Now, suppose we make no prior assumption about the shape of regression models. Instead, the shape of the regression model can be treated as an unknown function f which is to be learned from the data as follows:

$$y_i = f(x_i) + e_i \tag{4-2}$$

Furthermore, we can consider probability distributions over the function space f, such that the practice of modelling, explaining and generalising data is accomplished by treating these distributions until they give a reasonable match to data. As these functions, unlike parametric models, are not defined in the context of explicit sets of parameters, this approach is referred to as non-parametric modelling.

Non-parametric models enable us to choose from a broader class of functions f whilst doing the probabilistic inference of the model parameters (Gershman and Blei, 2012). These are flexible because of their strength to scale in model complexity with the observed data, especially where parametric models become challenging (McIntire et al., 2016).

There are numerous situations in which we have inadequate, or no, prior knowledge concerning a proper regression model. However, in case the data are visible samples from a known process which is continuous, smooth and changes in the observation take place over characteristic time-scales with a typical amplitude (Gershman and Blei, 2012). These characteristics (i.e. latent variables) can be controlled by functions that can generalise to the entire parameter space.

As the inference of the parameters of non-parametric models is usually done by Bayesian statistics, non-parameteric models may also be referred to as Bayesian non-parametric models (Orbanz and Teh, 2011).

4.2.1 Gaussian Process (GP) models

In order to fit complex mappings among inputs and outputs, a proper non-parametric hierarchical structure is needed (Roberts et al., 2013). Non-parametric surrogate models (i.e. emulators) provide approximations to the output of a simulation model and enable efficient exploration of the input space (Garbuno-Inigo et al., 2016). A standard approach is to choose a surrogate model that accounts for the uncertainty within a Bayesian framework (Bukkapatnam and Cheng, 2010; Garbuno-Inigo et al., 2016).

Bayesian approaches provide the posterior distribution of the model parameters wherein a prior probability of each parameter is known. To retain the probabilistic treatment, we must use a non-parametric regression model that enables us to choose from a large class of functions (Kuss, 2006; Roberts et al., 2013). In this context, Gaussian processes (GPs) are popular non-parametric emulators to build statistical models (Svensson et al., 2015; Roberts et al., 2013).

Gaussian processes have seen increasing popularity in the past two decades due to their flexibility and adaptability with complex data structures (Svensson et al., 2015; Gershman and Blei, 2012). A Gaussian Process model is a hierarchical regression technique for probabilistic modelling of functions (Rasmussen and Williams, 2006). Prior information for a Gaussian process model is described by some tuning parameters or hyper-parameters, which directly influence the posterior structure of the GP model (Kuss, 2006, Svensson et al., 2015).

In turn, Gaussian Process models use all previously trained data along with estimated quantities to perform predictions of untrained data, whereas, parametric models do not need the trained data that has been used to estimate the parameters (coefficients) of the underlying regression model (Roberts et al., 2013).

Gaussian Processes place a prior on function values directly (the coloured lines in Figure 4.2-a) rather than assuming a particular parametric structure of function f (Murray-Smith and Girard, 2001; Kuss, 2006). Since there might be a large class of functions belonging to the prior, the uncertainty is very high before observing data (see the large confidence interval in Figure 4.2-a). Consequently, the GP model is free to explore a different form of function covering the entire input space x and output space y.

Now, when some observations become available, the GP model runs some samples and compute the likelihood of each through Bayesian inference. Then, the posterior samples are estimated with lower uncertainty (Figure 4.2-b). By the time more data are observed, the estimated uncertainty interval becomes smaller while the GP model adapts to the new data (Figure 4.2-c).

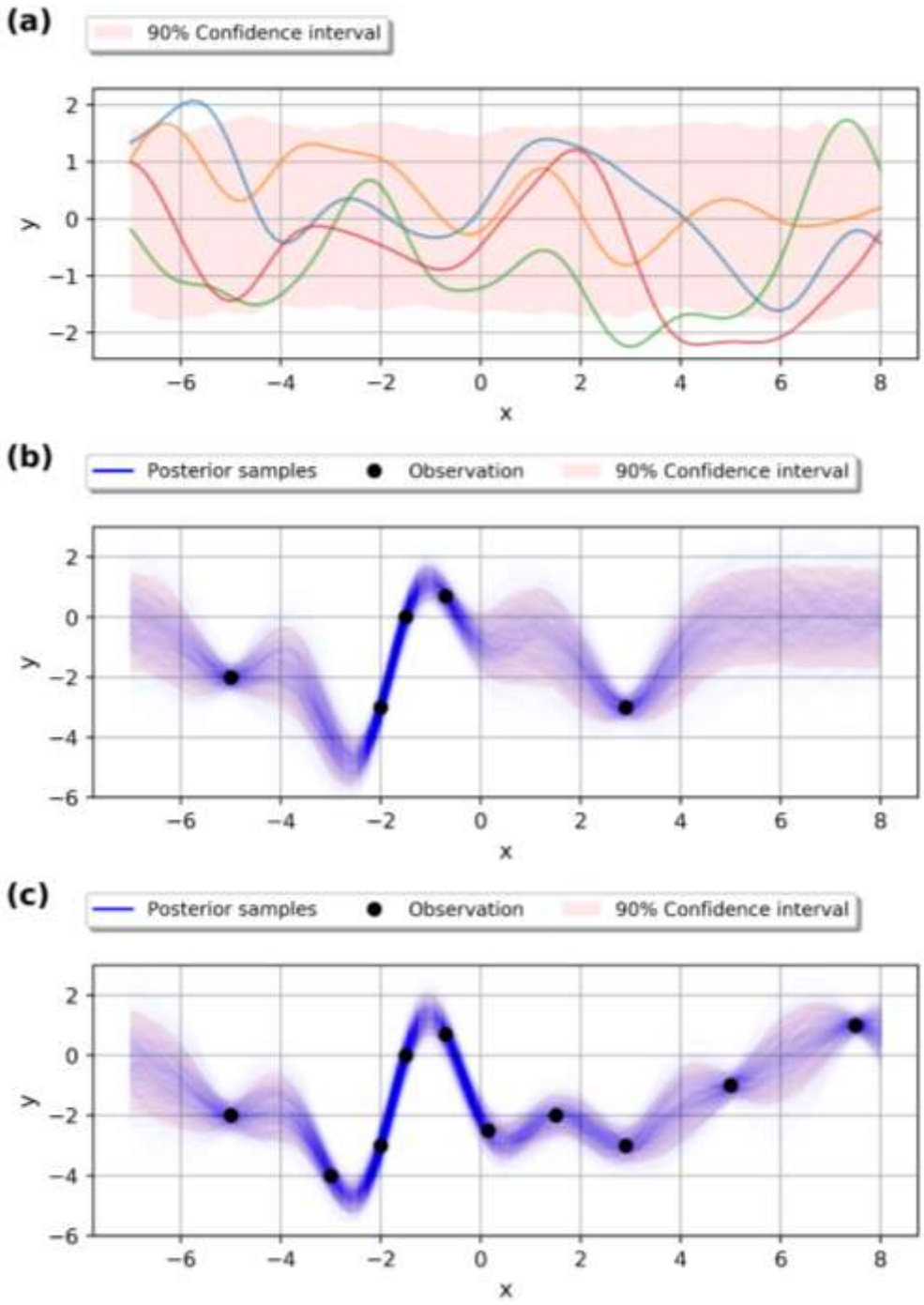

Figure 4.2 Schematic of a Gaussian process model: **a)** Prior distribution over functions shown in colours before observing data; **b)** The posterior GP samples after observing 5 data points; **c)** The Posterior GP samples when more data are available. The estimated confidence interval for the function f is wide before evaluating the likelihood of each samples. The interval becomes narrower as the number of observed data increases.

The GPs aim to model an unknown function $f(x)$ from measured input data x and output data $y = f(x)$ and assume that the values of $f(x)$ are a set of random variables of a joint n-dimensional Gaussian distribution (Murray-Smith and Girard, 2001). To construct a GP model, no explicit form of f is needed, but some assumptions about the form of f are encoded through the GP prior (e.g. mean zero prior).

Then, the GP finds the posterior state of $f(x)$ by a mean function $m(x)$, a covariance function $k(x, x')$ between x and x' instances, and their hyper-parameters. In mathematical terms, the GP models the $f(x)$ as

$$f(x) \sim GP(m(x), k(x, x')) \tag{4-3}$$

Now, we have a prior distribution for the output/response variable y that is a multivariate Gaussian:

$$p(y|x) = (2\pi)^{-\frac{n}{2}} |\Sigma_{NN}|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}(y-m)^T {\Sigma_{NN}}^{-1}(y-m)\right) \times p(y) \tag{4-4}$$

where n is the number, Σ_{NN} is the covariance, and m is the mean of the training set. Next, suppose there is a set of unseen data as $\{x^T, y^T\}$ we wish to predict by the emulator $f(.)$. Then, the joint distribution for the training and the testing set is Gaussian as follows:

$$\begin{bmatrix} y \\ y^T \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m \\ m^T \end{bmatrix}, \begin{bmatrix} \Sigma_{NN} & \Sigma_{NT} \\ \Sigma_{TN} & \Sigma_{TT} \end{bmatrix}\right) \tag{4-5}$$

where Σ_{TT} is the covariance of testing set, m^T is the prior mean of testing set, Σ_{NT} is the covariance between training and testing set, and Σ_{TN} is the transpose of Σ_{NT}.

The posterior distribution over y^T conditional on training data y yields the posterior mean m^* with the variance Σ^*

$$m^* = m^T + \Sigma_{TN} {\Sigma_{NN}}^{-1}(y - m) \tag{4-6}$$

$$\Sigma^* = \Sigma_{TT} - \Sigma_{TN} {\Sigma_{NN}}^{-1}\Sigma_{NT} \tag{4-7}$$

The resulting posterior m^*, Σ^* at the test points are multivariate normal distributions whose statistical properties not only provide point estimates but a full predictive distribution of the entire input space x (Bukkapatnam and Cheng, 2010).

4.3 Non-parametric modelling of discrepancy

As stated in section 4.2, the main advantage of non-parametric models is their flexibility in the model selection enabling practitioners to capture broader aspects of data especially for prediction of future unlabelled data (Roberts et al., 2013).

In Chapter 3, we established a hierarchical parametric model to handle the discrepancy within the training set. However our parametric model was unable to generalise to the untrained data, nor did it estimate the uncertainty outside the training domain.

In modelling discrepancy within history matching, we are interested in propagating uncertainty towards future input space where the amount of information about the error propagation rises as more data become available. This is to say that the error distribution cannot be defined in such a finite set of parameters, but an infinite dimensional function which makes them more flexible (Seeger, 2004).

In this section, we use Gaussian process models that enable us to predict the errors at any point of interest along with their associated uncertainty. We firstly fit a model in the range of observed errors, and secondly, forecast the error into the future data.

In advance of building our non-parametric model to capture discrepancy, we exert Kennedy and O'Hagan (2001) notation (KOH) for modelling discrepancy within the

Bayesian framework. According to KOH, the simulator response, albeit at the true input value, can never trace the true response of a physical process.

To enhance our understanding of KOH approach, let $S(.)$ denote a nonlinear function mapping input space $\mathbb{x}$ to output space $\mathbb{Y}$ such that $x = \{x_1, \dots, x_n\} \subset \mathbb{x}$ underlies the trained data $D = \{x, y\}$; x_i denotes a particular instance of our system, e.g. time; and y_i is the observation associated with the input x_i. Then the model discrepancy function $\delta(x)$ is related to simulator response as follows:

$$y_i = S(x_i, \theta) + \delta(x_i) + \varepsilon_i \tag{4-8}$$

In a reservoir model, the true value of the parameters can never be learned with certainty. Hence, following KOH modelling error strategy, we assume that there exists a true $\hat{\theta}$ estimated by a best-fitting model with the highest marginal likelihood. In our experiments, measurement errors ε_is follow Gaussian white noise distribution $N(0, {\sigma_\varepsilon}^2)$ in that the errors are independent and identically distributed (*i.i.d*). The δ is assumed unknown and independent of ε. The total estimated discrepancy at i^{th} instance e_i can be recognised as the sum of model discrepancy and measurement error

$$e_i = \delta(x_i) + \varepsilon_i \tag{4-9}$$

where e_is are equivalent to the residuals of the observed values. It is important to acknowledge that whereas the ε_is are assumed independent, this will not be true of the e_is as independence cannot always hold for $\delta(x_i)$s.

Our primary goal is to address the following question: how can we assess the uncertainty related to modelling discrepancy by training some data and propagate it towards areas where no data exist (i.e. extrapolation) without overfitting the data? But before doing so, one should have an appropriate prior belief about whether the functional form and the parameters of model inadequacy. Brynjarsdóttir and O'Hagan, (2014) discussed the challenges associated with a prior assumption about $\delta(.)$ and model parameters θ

especially in case of extrapolation. Their work ascertains that only with a realistic, informative priors does the posterior certainty space cover the true value of θ. Interestingly, however, less informative priors appeared to make better inference about unknowns when extrapolating untrained data. Subsequently, one shall be able to assign a non-complex flexible a priori distribution over $\delta(.)$ which returns a realistic estimation of structural uncertainty at any instance of interest. For simplicity, we assume that the model discrepancy pertains to a random variable of Gaussian distribution at any instance x_i. If further assume that any finite subset of those random variables follow statistics of a joint Gaussian distribution, then Gaussian processes (GPs) can be employed to construct the non-parametric model. A Gaussian process posits a prior distribution $P(\delta)$ over the function δ and enables fitting it into the Bayesian regression conditional on the trained data:

$$P(\delta, \theta|D) = \frac{P(D|\delta, \theta)P(\theta|\delta)P(\delta)}{\int P(D|\delta, \theta)P(\theta|\delta)P(\delta)d\delta} \tag{4-10}$$

As opposed to Eq. (3-1), the Eq. (4-10) plugs the model discrepancy term δ into the Bayesian inference from which the posterior probability of δ and θ are learned from data. Consequently, the probability of parameters of interest θ can be obtained by marginalising over the nuisance parameters of δ:

$$p(\theta|D) = \int p(\delta, \theta|D)d\delta \tag{4-11}$$

The GP model defines a function characterized by the mean m, returning the expectation of δ, and covariance function $k(x, x')$ of the process. Note that the covariance functions must always hold positive-semi-definiteness (i.e. valid covariance function) to ensure invertibility. The GP model is able to relate the inputs to the outputs as follows:

$$\delta(x) \sim GP(m(x), k(x, x')\ ) \tag{4-12}$$

There are many correlation functions in the literature which provide valid covariance functions. Firstly, we wish to introduce a covariance function which has a constant mean; is invariant to shift in time (i.e. wide sense stationarity (WSS)); and is invariant under rotations about the origin (i.e. isotropic). The most-widely used kernel within kernel machines field is the Squared Exponential (SE) covariance which has mean square derivatives of all orders (i.e. infinitely differentiable):

$$k(x, x') = {\sigma_f}^2\ exp\left(\frac{-(x - x')^2}{2\ {\sigma_l}^2}\right), \qquad \sigma_{\ l} > 0 \tag{4-13}$$

where σ_l is characteristic length scale and σ_f is kernel standard deviation. The SE is a smooth covariance function that relaxes the assumption of WSS as $k(x, x') = k(x - x')$, and the assumption of isotropy as the correlation structure is a function of the Euclidean distance between locations in one direction. Table A- 1 illustrates several stationary kernel functions that we used to carry out GP regression.

4.3.1 Model Selection for non-parametric error models

In the previous section, we readily modelled the model discrepancy based upon a known structure of kernel/mean function. However, there might be a set of candidate models with a different number of parameters or carrying more complex structures. These candidate models can be treated as unknown functional forms hidden within data, prior to doing a history match process.

The foregoing problem of providing equilibrium between the goodness of fitting models and the simplicity of regression model give rise to a major area of interest known as model selection problem (Ling et al., 2014; Stine, 2004).

Information theory addresses a coherent view of the model selection problem by converting the model into a code. Akaike (1974) introduced a model selection framework

centred on maximising the likelihood function for each candidate model and then penalising the model complexity. Going further, Schwarz (1978) assigned a more severe penalty for complex models in terms of a penalized likelihood known as Bayesian Information Criterion (BIC).

In a groundbreaking study by Rissanen (1989), the model selection has been investigated in terms of Minimum Description Length (MDL) principle. It holds that the best explanation for a set of observation is the one that enables the maximum compression of the data. The MDL techniques are particularly well-founded for identifying suitable candidate models in situations where the models under consideration hold a complex nature, and overfitting the data is a serious concern (Rissanen, 1989; Stine, 2004).

Our study sets out the model selection for model discrepancy δ through BIC framework as it often do well in resolving the overfitting of the models (Stine, 2004). In addition, compared with the MDL, the BIC is less mathematically demanding (i.e. computationally efficient) and yet approximately equivalent to MDL once the number of sample size n is larger than the number of free parameters p. Then, the BIC is related to maximum likelihood $\mathcal{L}_{max}$ of the fitting kernel function as follows:

$$BIC = -2\ log\ \mathcal{L}_{max} + p\ log\ n \tag{4-14}$$

and the best error model $\hat{\delta}$ is the one that minimises the BIC:

$$\hat{\delta} = \arg\max_{\delta}\ BIC \tag{4-15}$$

4.3.2 Model prediction

In order for GP to account for model inadequacy, a training phase for the GP is essential for estimation of hyper-parameters $\varphi = \{\sigma_l\ , \sigma_f, \dots\}$. This can be done by computing the maximum likelihood (ML) estimates, maximum *a posteriori* estimates (MAP), and a Full Bayesian Hierarchical model (FBH).

ML returns point estimates of hyper-parameters $\hat{\varphi}$ such that the joint probability density (or the likelihood) is maximised. In practice, ML represents an approximation to a Full Bayesian multi-level model wherein the hyper-parameters are evaluated after data are observed instead of being marginalised out.

In case the distribution is multimodal, one can establish some, but not a single, ML settings for $\hat{\varphi}$, and launch the calibration with all different settings. This is an indispensable consideration in obtaining more reliable estimates of physical model parameters (i.e. lower level of hierarchy) and unbiased prediction for calibration parameters. The ML and MAP approaches, however, fail to rigorously assess the uncertainty about the true values of hyper-parameters, throwing away valuable information about φ. The ML and MAP for our model can be computed maximizing Eq. (4-16) and Eq. (4-17) respectively:

$$log\, \mathcal{L}(\varphi) = log\, p(\delta|x,\varphi) = -1/2(\, \delta^T k_N\, \delta \;\; + log|k_N| + n\, log(2\pi)) \qquad (4\text{-}16)$$

$$p(\varphi|D) = p(\delta|x,\varphi\,)p(\varphi) \qquad (4\text{-}17)$$

where the subscript N refers to the training data. Likewise ML, MAP estimation lacks realistic estimation when the likelihood is multimodal or is a non-convex function with the possibility of optimisation being trapped in local optima. Hierarchical models employed along with Bayesian framework, are concerned not merely with returning modal estimates for hyper-parameters, but rather with inferring a joint posterior probability density for the model parameters (Garbuno-Inigo et al., 2016; Svensson et al., 2015).

Ideally, we place a prior and calculate a Bayesian posterior on hyper-parameters in a fully Bayesian manner (FBH). This would give rise to a probability-weighted average over the hyper-parameters to generate a forecast that no longer pertains to the nuisance parameters (i.e. marginalisation). FBH enables marginalising over hyper-parameters conditional on

the trained data to make prediction δ^* for a non-observed input configuration x^* as follows:

$$p(\delta^* | x^*, D) = \int_{\varphi} p(\delta^* | x^*, D, \varphi) \, p(\varphi | D) \, d\varphi \tag{4-18}$$

This integrated predictive distribution can be approximated by means of Markov Chain Monte Carlo (MCMC) as follows (Garbuno-Inigo et al., 2016):

$$p(\delta^* | x^*, D) \approx \sum_{j=1}^{m} \omega^{(j)} \, p\left(\delta^* \middle| x^*, D, \varphi^{(j)}\right) \tag{4-19}$$

where the weights $\omega^{(j)}$ are normalised for m weighted samples $\{\omega^{(j)}, \varphi^{(j)}\}_{j=1}^{m}$ of the distribution $p(\varphi|D)$.

4.3.3 Non-stationarity and segmentation

The hypothesis of data stationarity has been the most common framework in time series analysis. However, in real-world processes, the relationship between a contemporary response variable and its own past may follow a non-stationary stochastic process.

On the other hand, modelling a non-stationary process with stationary techniques to capture the correlation structure of data may lead to crude approximation. Consequently, prediction of a non-stationary class of data using a stationary model is high-risk (Korkas and Fryzlewicz, 2017) as it is unable to realistically quantify main sources of non-stationarity (i.e. changes in mean, trend and standard deviation).

The kernel functions presented in Table A- 1 are all stationary implying that the statistical properties of the error are all constant within the time domain. This necessitates initial correlation analysis to ensure stationarity of errors before fitting a kernel to the errors.

Figure 4.3 compares non-stationary and stationary time-series by autocorrelation plots. In this example, autocorrelations for non-stationary data have an obvious downward trend, 11 of which outside the confidence interval. On the contrary, almost all autocorrelations for stationary data remains within confidence intervals, and no specific trend can be observed.

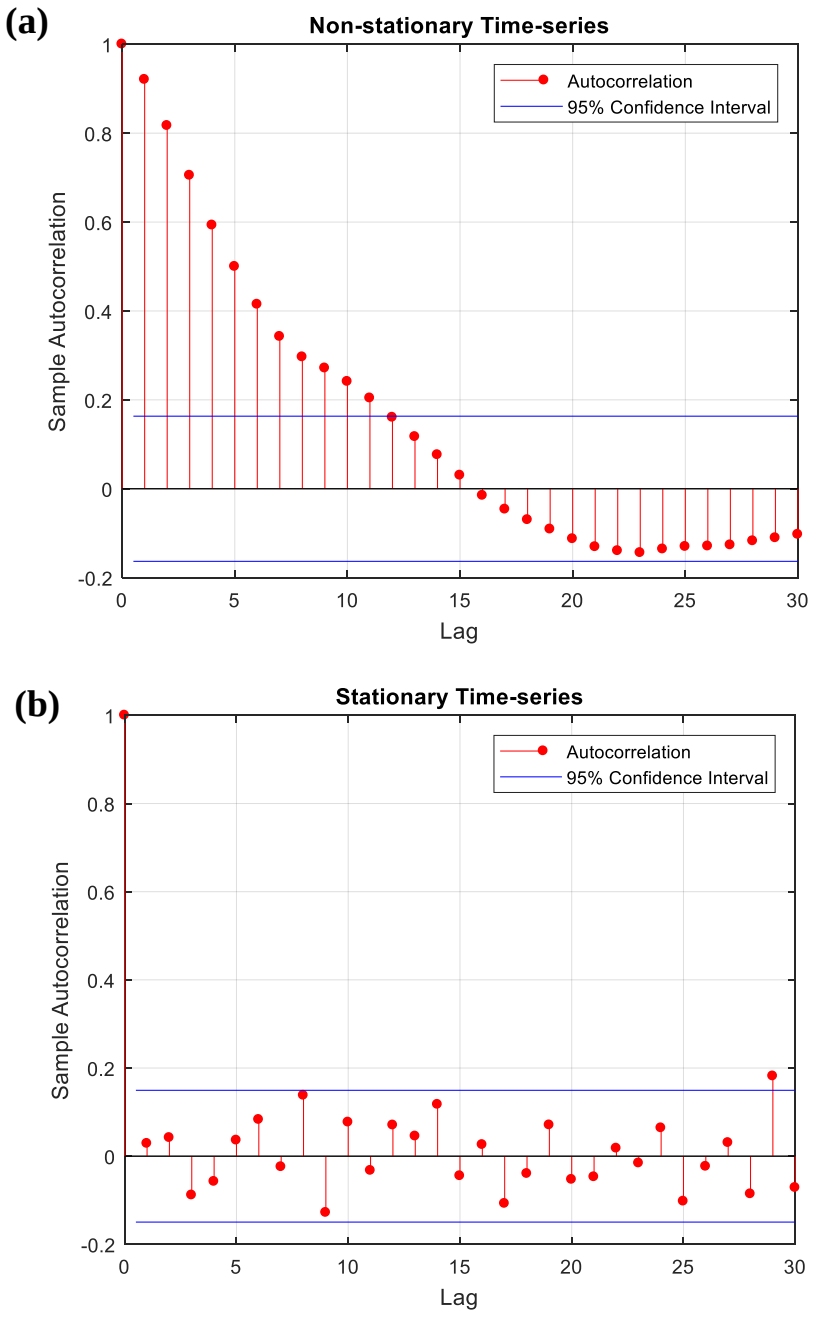

Figure 4.3 Autocorrelation plot for two sets of sample: (a) non-stationary data (b) stationary data

One common solution to the problem of modelling a non-stationary time-series data is the division of the series into segments of approximately stationary processes in that non-stationary time-series can be seen as a concatenation of stationary partitions.

The segmentation traces the periods of stability, optimal moments of change (change-point detection) and the pattern in the non-stationary time-series. The best segmentation is described as the optimiser of a cost function that maximised the posterior distribution of the observation (Davis et al., 2006).

In this work, we make use of a penalised contrast for the detection of changes in mean and slope proposed by (Lavielle, 2005). His approach provides an automatic choice of location of stationary segments on the basis of the specified statistics to solve the following problem. Given observation $\{x_1, \dots, x_n\} \in \mathbb{R}^D$ and a specified number of change-points $m \in \mathbb{N}$, the objective is to find integers $0 < d_1 < \cdots < d_m < n$ that minimise the sum of the within-segment least-squares loss

$$e_d = \sum_{j=1}^{m+1} Loss(x_{d_{j-1}} + 1, \dots, x_{d_j}) \tag{4-20}$$

where $d_0 = 0$, $d_{m+1} = n$ and m is addressed by penalised contrast.

4.4 History Match, modelling discrepancy and forecast workflow

The autocorrelation and segmentation helped us better understand the non-stationarity of the residuals and partitioned the data into different stationary segments. Now, we can employ the error modelling in the history matching process and generalise the response to the forecast.

Figure 4.4 illustrates our workflow for incorporating the modelling discrepancy into the history match process of reservoir models. The history match begins with sampling multiple models from prior distribution of model parameters θ. Then, the reservoir simulator reproduces the production data for all generated models. Furthermore, the Least-squares misfit (LSQ) evaluates the quality of each model by comparing simulator

output with observation data. The best fitting model estimates the true value of model parameters $\hat{\theta}$ and the model discrepancy δ.

Error model selection is done throughout the BIC minimisation process wherein kernel functions that best match to the estimated δ are selected. By the time the BIC code evaluated the proper form of the emulator $\hat{\delta}$, the GP regression is able to predict the model bias and associated uncertainty at any instance of input space (e.g. forecast). We can also generate the posterior probability density of model parameters θ by marginalising over $\hat{\delta}$.

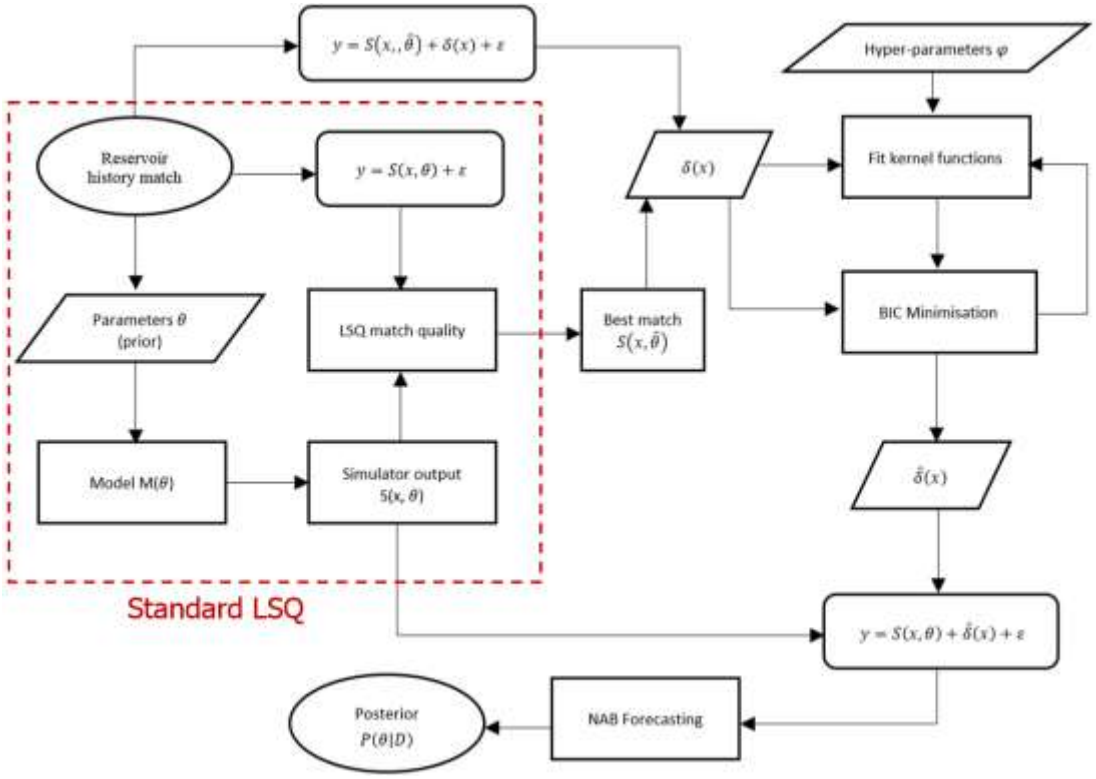

Figure 4.4 Flowchart of history match and forecast with modelling discrepancy: the top and the bottom eclipses show the start and the end of the flowchart. The red dashed-box corresponds to standard LSQ approach while the rest of the flowchart models the error through FEM.

4.5 Conclusion

In this chapter, we examined two significant aspects of error modelling throughout the history matching of oil reservoirs: the non-parametric hierarchical modelling of model inadequacy, and the model selection problem.

In contrast to parametric methods, nonparametric methods on function estimation do not assume any parametric form of the function other than certain smoothness assumption. This enabled our non-parametric Gaussian process models to generalise towards the forecast which is of high importance for reservoir engineering problems (Nezhad Karim Nobakht et al., 2018).

The Gaussian process models we fitted to data make use of stationary kernel functions which are valid only for the stationary time-series. Therefore, a primary correlation analysis should be performed on the field variables' residuals to detect the likely non-stationary in time-series errors.

When the non-stationarity of data confirmed by the correlation analysis, Lavielle, (2005) segmentation technique partitions the non-stationary dataset into stationary segments.

Then, the BIC code obtains the best solution to the form of GP kernels from which we predict the error for the unseen future data. This allows us to select those models that best fit the data without overfitting.

After finding the best correlation structure of the GP kernels, we perform history matching and forecasting of the reservoir model while accounting for the modelling discrepancy. The resulting posterior probability distribution found by the error model (Eq. (4-11)) is different from that of the LSQ model (Eq. (3-1)) which ignores modelling discrepancy.

Consequently, we expect the error model's posterior estimates of model parameters and field/response variables to be different from LSQ models. These changes are compared in history match and forecast of a real reservoir model in Chapter 6.

Chapter 5 – Hierarchical error modelling: Application to the Teal south

5.1 Introduction

In comparison with single case studies, comparative case studies often provide more generalisable knowledge around causal questions – why and how a particular model specification may impact the outcome. Moreover, researchers investigate the influence of the methodology of interest on comparative case studies to draw more reliable conclusions. The selection of case studies, however, needs careful thought as the decisions made at this step have implications for how well we achieved the objectives of our research.

In Chapter 3, we presented a hierarchical parametric approach for evaluating reservoir history matching models and compared it to the history matching models in which the model discrepancy is ignored. We then discussed how ignoring model error might lead to biased estimation of uncertainty. Now, it is time to test our statistical model in real-world cases and scrutinise the impact of error modelling on different scenarios. With regards to regression models, the Full Error Model (FEM) and the Linear Least-Squares (LSQ) described in Chapter 3 are used to perform history matching with and without modelling discrepancy, respectively. To be able to do this well, the precise characteristics of each case is explained in details at the start of the study.

In the parametric modelling of discrepancy, the size of the training dataset, the spread of errors, and the choice of input variables can directly influence the results of modelling strategies. For instance, one cannot assume unknown standard deviation for each point of large datasets, because the number of parameters to be identified rises (i.e. curse of dimensionality). The selection of the input parameters is another critical factor to take into account when evaluating the performance of modelling approaches since the selected set of model parameters, and their associated correlations can affect model performance (Dupin et al., 2011).

We apply our statistical framework for modelling discrepancy to the history matching of a real oilfield case study, the Teal South reservoir, and uncertainty quantification of production variables. We firstly manipulate and set up different case studies of the Teal South to address a different aspect of error modelling and its comparison to unmodelled discrepancy scenarios.

The central objectives of this chapter can be summarised as follows:

In the first case study, we implement our error model in the Bayesian calibration of an idealised model (or 'inverse crime' in the language of Kaipio and Somersalo, 2007) which has no systematic nor random error. The idealised model is generated by simulation of the Teal South reservoir model at the designated 'correct' set. We anticipate that both FEM and LSQ recover the real value of input variables since the simulator can correctly recover the data.

In the second case study, we impose Gaussian white noise on the first case study (idealised model) and track down changes in estimated parameters while the model configuration remains unchanged.

In the last case study, we use the original observation data from the Teal South reservoir model and tune 6 unknown model parameters. This case study contains structural uncertainty since the equation used in the simulator approximates the real underlying physics of the reservoir. Both FEM and LSQ obtain a probability distribution for model parameters and estimate uncertainty for the output variable. These estimates are compared statistically to examine the predictive performance of each modelling strategy. In all of the 3 case studies mentioned above, we use six early production data points and hold the remainder for model evaluations.

5.2 The Teal South oil reservoir

To scrutinise the applicability of error modelling throughout the process of history matching and to realise how accounting for modelling discrepancy differs from ignoring it we make use of a simple proof-of-concept reservoir, the Teal South reservoir model. The Teal South reservoir, located in Eugene island (in the central Gulf of Mexico), was developed by Mobil Oil in the mid 80's and is currently being operated by Apache. Energy Resources Clearing House (ERCH) in Houston provided the production data for this oilfield on a monthly basis. Production began in November 1996 from a single horizontal well (Pickup et al., 2008). The monthly data consist of field oil production rate FOPR, field water production rate FWPR and field gas production rate FGPR.

There are several studies on the history matching of the Teal South model with different configurations, geological parameterizations and unknown variables (Hajizadeh et al., 2009; Christie et al., 2002; Pickup et al., 2008). In this work, we set up the simulation

model on an 11×11×5 corner point grid in conjunction with five geological layers in the model with uniform properties. Porosity is assumed to be fixed at 28% in the reservoir.

The simulation model has 6 unknown model parameters each with a uniform prior described in Table 5.1. The history matching uncertain parameters are permeabilities in different layers, anisotropy ratio, rock compressibility and aquifer strength. As for modelling discrepancy scenarios, the FEM models also have a set of hyper-parameters s1, s2, … s6 all with a uniform distribution U(1, 300). This allows history match to tune the hyper-parameters within their most likely range along with the model parameters.

A short number of reservoir data comprising of some PVT data is available for the Teal South reservoir. For instance, reservoir pressure is measured at two stages of the reservoir life (Christie et al., 2002): the initial reservoir pressure (Pi=3096 psi) and the reservoir pressure measured after 540 days of production (2458 psi). Figure 5.1 illustrates 11×11×5 simulation grid in 4500-ft sand structure map of the Teal South oilfield. A single well drills the 4500-ft sand restricted on three sides by geological faults and surrounded by a dip to the north (Christie et al., 2002). As for the simulation aspect of the reservoir model, the Eclipse reservoir simulator was used to produce the Teal South model's response variables. In the case of modelled discrepancy scenarios, each iteration of the history match process is followed by a post process which computes the misfit value based on the assumed error model structure. The sampling algorithm process benefits from a parallel computation scheme which enables them to be run on multiple compute nodes.

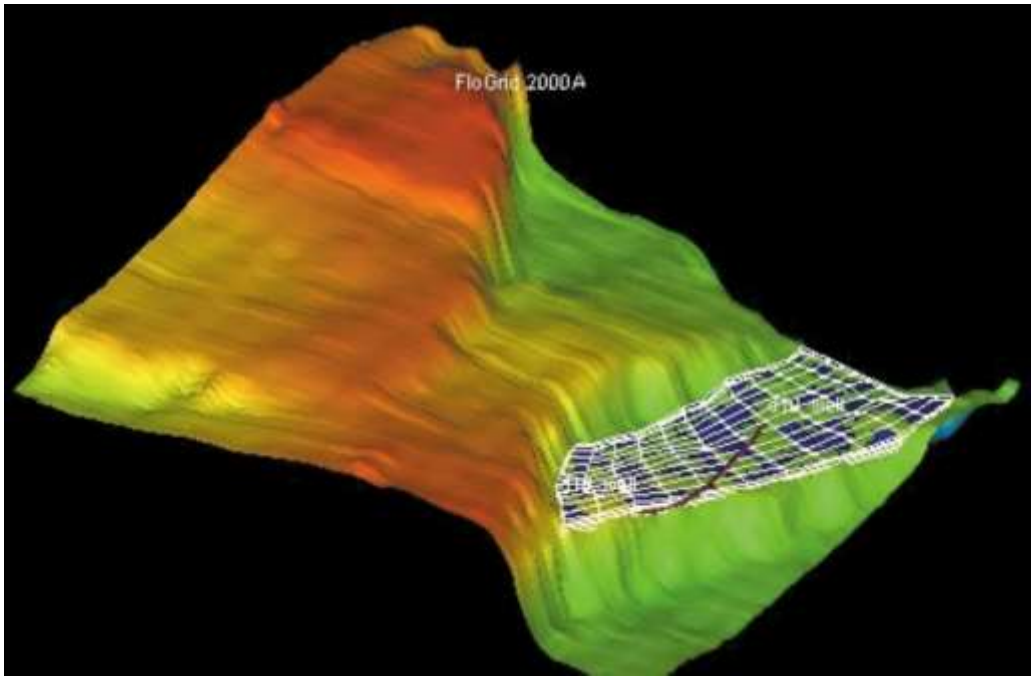

Figure 5.1 Simulation grid and structure map of the Teal south oilfield reservoir

Table 5.1: Model parameters distribution in the Teal south Case Study

Model Parameters	***Symbol***	***Unit***	***Range***
Log aquifer strength	aq_str	MMSTB	7 – 10
Log anisotropy ratio	logkvkh	-	-4 – -1
Log permeability multiplier at layer 1	P1	mD	0 – 4
Log permeability multiplier at layer 2	P2	mD	0 – 4
Log permeability multiplier at layer 3	P3	mD	0 – 4
Rock compressibility	rock_cr	psi^{-1}	0.000005 – 0.0001

Our primary focus is examining LSQ and FEM method with as few field parameters as possible. That is why we have chosen one objective parameter, FOPR. The production data of the reservoir comprises field oil, field gas and field water production rates for 1247 days of history. A considerably high production rate flowing through the small size

of the reservoir causes a quick depletion. Hence, as can be seen in the filed oil/gas production rates (Figure 5.2), there is a sharp slump in rates right after 180 days of production.

Because the choice of sampling method affects the exploration of parameter space and estimates of uncertainty (Erbas, 2007), we run history matching studies with different sampling methods. Our goal is to compare history matching and uncertainty quantification of the Teal South using the two sampling algorithms: Particle Swarm Optimisation (PSO) and Bayesian Optimisation Algorithm (BOA). Another consideration is to run an equal number of realisations for each study. The reason is to make sure our probability estimation and forecasting analysis are not affected by the type of optimisation algorithm. For example, it may be that for a certain number of iterations PSO converges considerably well for a calibration model while BOA fails to reach adequate ensemble of history matched model or gets trapped in local minima.

Apart from our original dataset, synthetic data are also produced to fulfil particular requirements or several conditions that might not be inferred from the original datasets. Synthetic data are deployed by practitioners to objectively test the exact or approximate mathematical models which may not apply to the original data (Lafferty et al., 2001; Wirgin, 2004). Such types of data are often made up to portray the authentic data and provide a foundation to assess the impact of particular solutions to a simplified version of a full model (Weiss, 1977; SUE, 1987; Lafferty et al., 2001). This can be worthy when interpreting any system response because the synthetic data are employed as a simulation or as a theoretical value, condition, etc. Consequently, the benefits we gain from synthetic data helps us to better understand the unexpected results of a theoretical/statistical model when it is not in agreement with our real physical model (SUE, 1987). In case the results prove to be unreliable, we will then be able to rebuild our statistical model and look for alternative solutions based on the response of our model to the synthetic data.

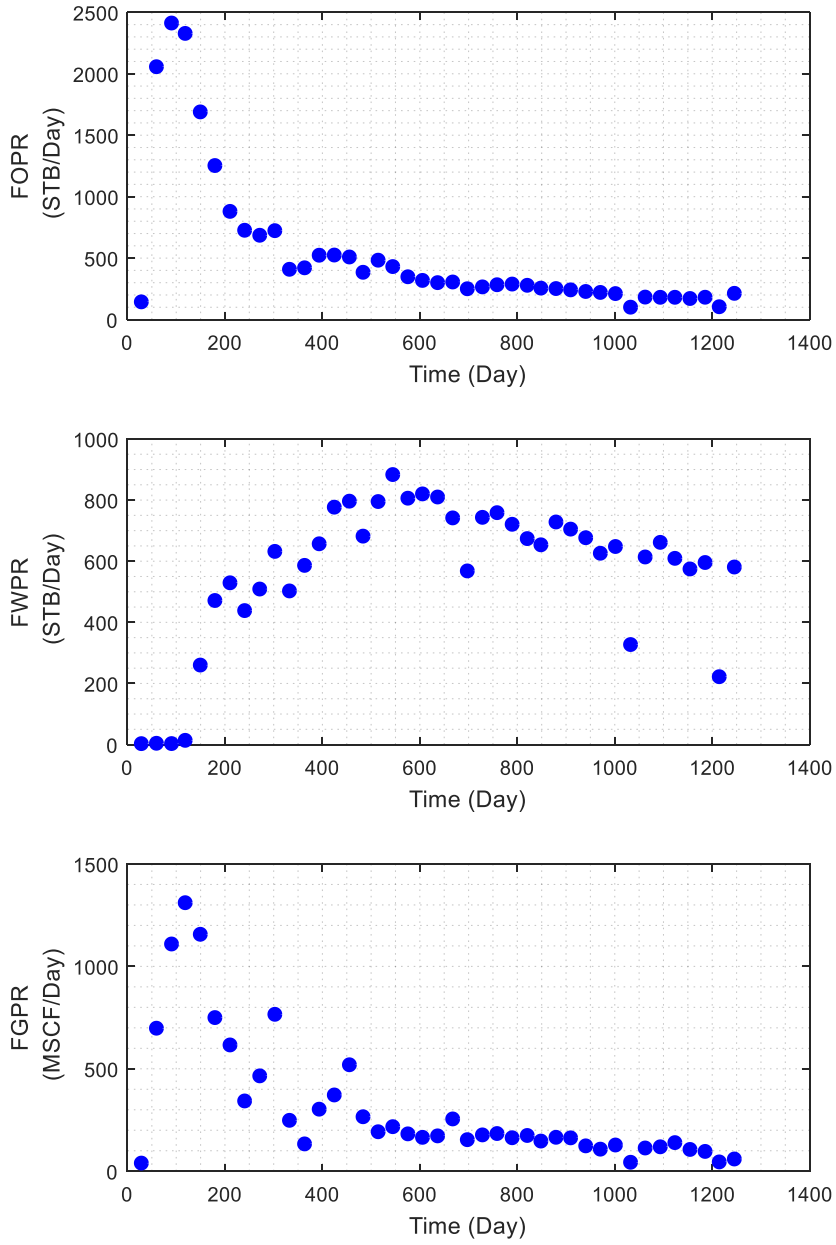

Figure 5.2: Production historical data collated for the Teal south oilfield

Now, we aim to do a preliminary analysis to examine the influence of assumptions about the distribution of errors in modelling discrepancy. To do so, we make use of synthetic data from which the deterministic aspect of modelling discrepancy is inferred.

As stated previously, the synthetic data explores a system reaction to certain circumstances or criteria. We produce synthetic data to assist in developing a baseline for forthcoming studies and analysis. The results of such analysis allow us to recognize these situations and make decisions correspondingly. In this section, we examine the FEM and LSQ response to synthetic data while removing structural uncertainty from original data.

As mentioned in previous chapters, the primary goal of FEM is not to better match the models in calibration progress. FEM is designed to reduce biased parameter estimation imposed by LSQ and therefore improve forecasting of field parameters. In the following, we aim to prove the fact that FEM can effectively improve the estimation of multiple parameters in simulations examining our theory with the following objective test cases:

a) Synthetic data: idealised model ($\delta=0$)

b) Synthetic data: imposed randomised Gaussian noise

c) Authentic data: full reservoir model

In these case studies, history matching runs are repeated with different random starting points to avoid biased estimation of model parameters and assure the credibility of our results.

5.3 Case study 1: Synthetic data from idealised model ($\delta=0$)

As mentioned in the introduction of this chapter, the data recorded through the history of a reservoir encompass structural uncertainty because reservoir simulators use equations and simplifying assumptions to approximate the underlying physics of the reservoirs. Such equations may assume homogeneous sub-grid properties in the reservoir or use average parameter values in the grid-blocks. Hence, regardless of the type of regression model used to calibrate simulation models, the inverse solver can never recover the actual value of uncertain parameters. In consequence, we need a reference case study to firstly present a model with a known true value of the parameters; secondly, remove structural uncertainty; and thirdly, compare the capability of LSQ and FEM.

In inversion scheme, if one aims to replace the original measurements of data with the one created by the theoretical model, there no more exist structural uncertainty in the model, and the model is called an "idealised model". In other words, when the observation represents information within model and data space, a mathematical model can thoroughly recover the "truth" and reach to the likelihood exactly equal to 1. This intentional removal of modelling incapability is referred to as "inverse crime" in Kaipio and Somersalo (2007) with solutions being always trivial.

Committing the so-called inverse crime will reduce model inadequacy described as C_m to zero. For simplicity, we make use of one field variable, FOPR, and two model parameters: P1 and P2 (see Table 5.1). The rest of the parameters are set to their mean values in that they have no longer impact on the history matching uncertainty.

We initially simulate our idealised model to gain the predicted oil rate (FOPR) corresponding to $\mathrm{p}_1 = \mathrm{p}_2 = 1$ and therefore generate synthetic data for oil rate FOPR. The synthetic data will be replaced with the observation in the history matching where p_1 and p_2 are treated to be unknowns with uniform distributions $U(0,4)$ and $U(0,2)$ respectively.

Although our synthetic data represent exact values without uncertainty, for the history match purpose, we need to assume a standard deviation for synthetic data. This comes from the fact that misfit functions whatsoever require standard deviations to evaluate the likelihood of each model. Hence, we assume a small standard deviation ($\sigma = 0.1$) for oil rate FOPR. At this stage, we expect both LSQ and FEM to recover parameters and observable data exactly. Figure 5.3 represents the process of constructing the truth case by removing structural uncertainty from original observations.

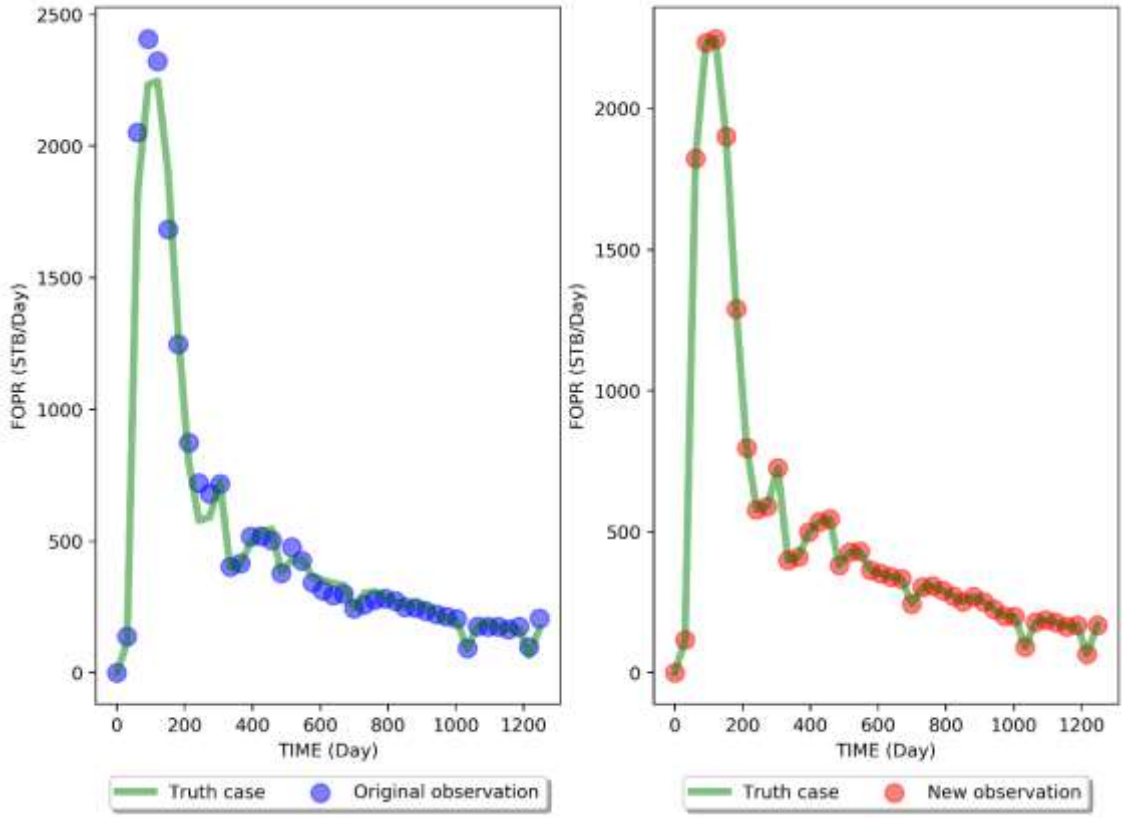

Figure 5.3 Structural uncertainty removal from the Teal South field variable FOPR: The original history points (blue dots) are replaced with the "truth" case (the idealised model simulated with $p_1 = p_2 = 1$). The new data (red dots) are expected to be recovered by both LSQ and FEM since they carry no structural uncertainty.

The simulated synthetic data of Figure 5.3 can now be used as new observations for LSQ and FEM history matching runs. Figure 5.4 portrays the progression of LSQ and FEM history matching studies through 200 iterations. Not surprisingly, both of the regression models tend to the exact value of the parameters p_1 and p_2 at some point around the 100th iteration. This can be confirmed by Figure 5.5 where all the posterior Probability Density Function (PDF) converge to 1. Therefore, the theoretical model FEM proves to be valid for our idealised model case study when there is no uncertainty in the model. In the following section, we investigate how changes in the idealised model influence the FEM and LSQ models.

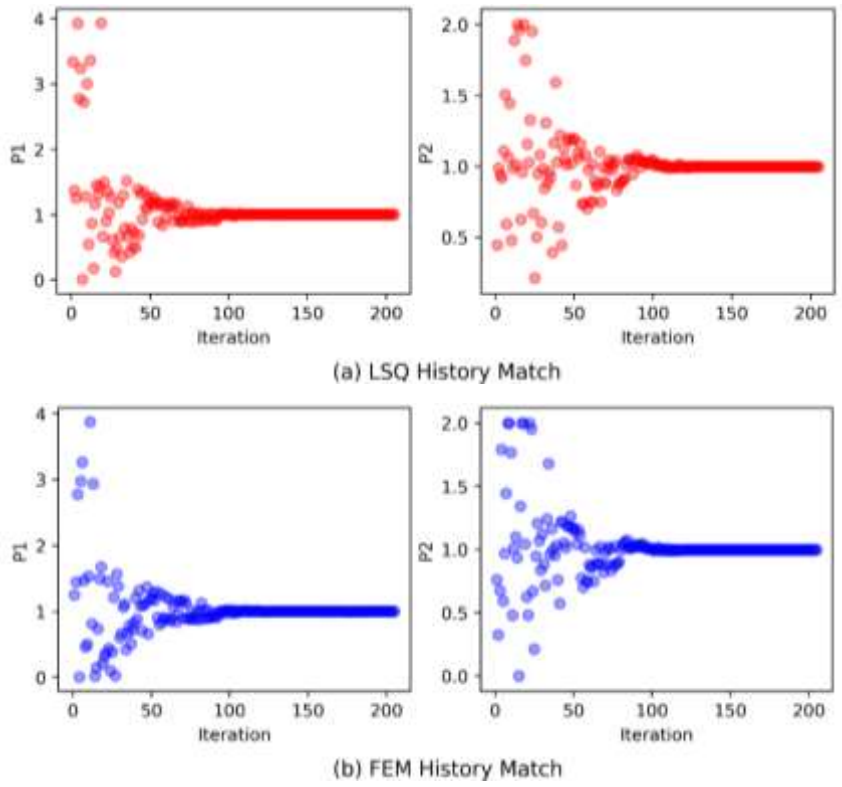

Figure 5.4 Value of model parameters against 200 iterations for a) LSQ and b) FEM

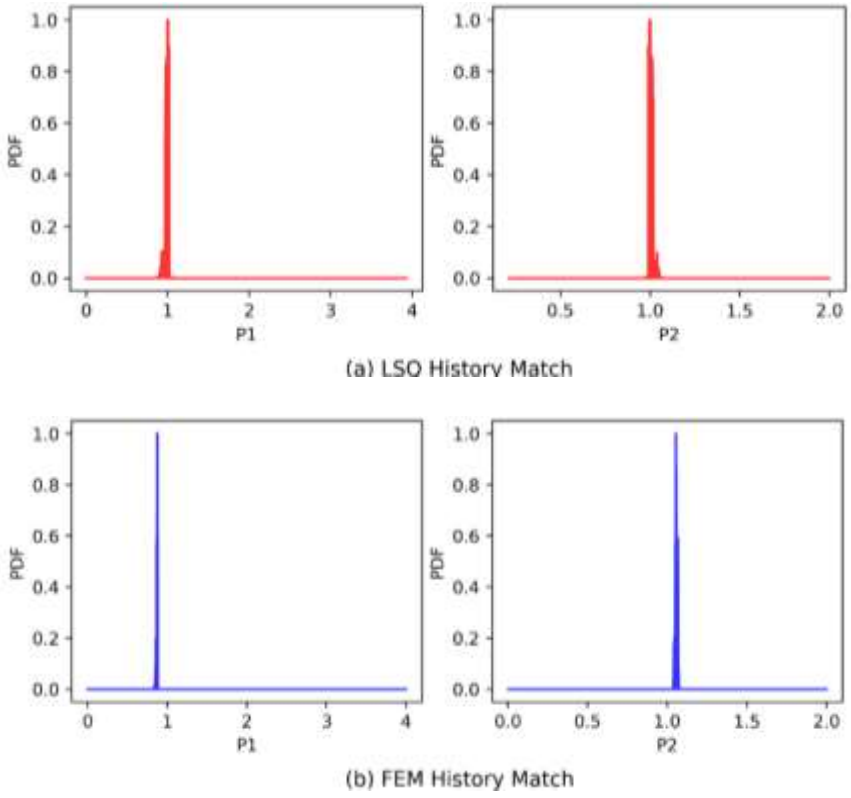

Figure 5.5 Estimated PDFs: Both the LSQ (in red) and the FEM (in blue) can predict the exact parameter values $p_1 = p_2 = 1$

5.4 Case study 2: Synthetic data from imposed randomised Gaussian noise

In this test, we add a set of randomised Gaussian noise with a known standard deviation ($\sigma = 10$) into the data from the previous case study (see Figure 5.6). The imposed noise will place Gaussian noise on the idealised model data and construct new observation. As a result, the posterior distribution of model parameters reacts to the changes in the calibration process.

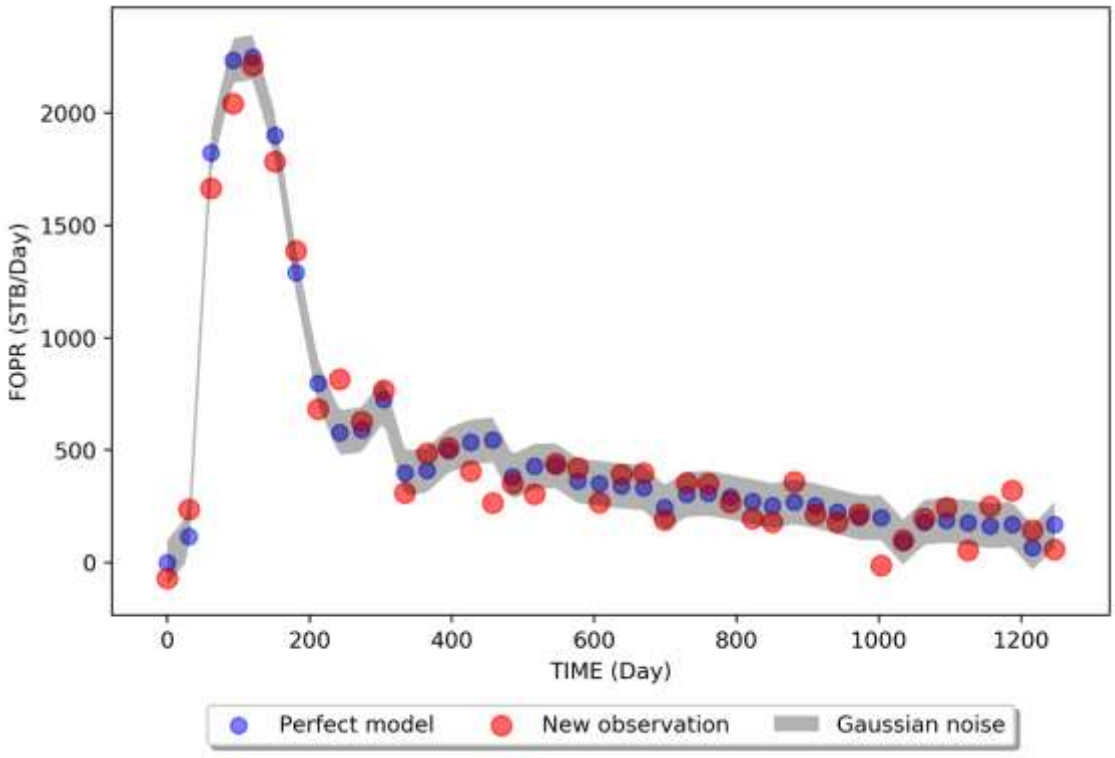

Figure 5.6 Synthetic data generated by imposing Gaussian white noise to the idealised model case study

In addition, we run 3 history match trials, each set up with various standard deviations (σ =2, 5 and 10) for the measurement errors. This results in different misfit values and therefore different value of likelihood for each model. Our expectation is that posteriors for both least-square and error model take the form of a Gaussian distribution scattered around the truth ($p_1 = p_2 = 1$). For a history match run with low sigma values ($\sigma = 2, C_y = \sigma^2 I$), LSQ becomes overconfident because it assigns low likelihood to the models outside the uncertainty bars. On the contrary, since FEM makes use of two covariances

($C_m + C_y$), it makes up the underestimated uncertainty by allowing C_m to add up to the total uncertainty. This enables FEM to be more flexible to changes in assumptions about the measurement noise and to obtain more consistent results (see Figure *5.7*).

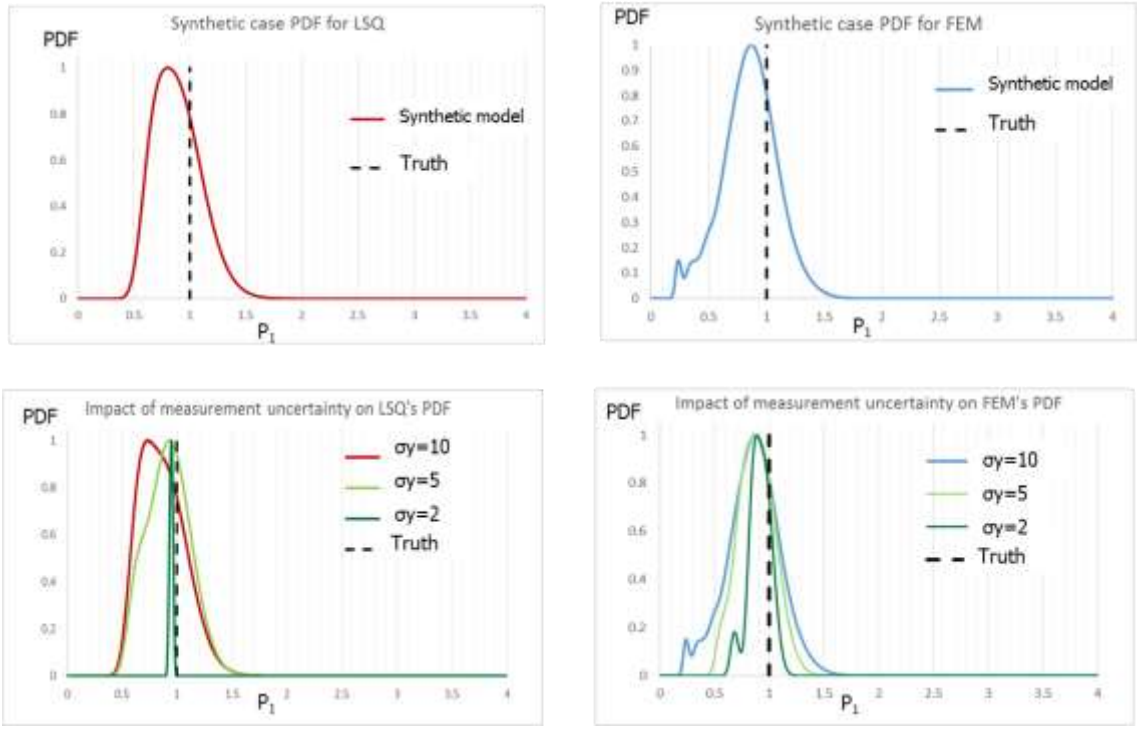

Figure 5.7 Posterior PDF of the model parameter p_1: in the upper graphs, FEM model (in blue) and the LSQ model (in red) follow Gaussian distribution around the truth ($p_1 = 1$). In the bottom left graph, LSQ is very sensitive to the changes in the standard deviation of the noise. In the bottom right graph, FEM is more consistent with changes to the assumed standard deviation and holds Gaussian.

The two upper graphs in Figure 5.7 indicate that both FEM and LSQ have met our expectation scattering around the truth $p_1 = 1$. Afterwards, in several experiments, we tried to monitor the impact of varying measurement uncertainty as in the two bottom graphs. Having decreased standard deviation from 10 to 2, LSQ has resulted in overconfident prediction of parameter p_1, while biasing towards a narrow range of parameter space. On the other hand, as can be seen in the bottom right graph, FEM is more consistent (in this case study) with changes in measurement uncertainty as it encompasses two sources of uncertainty ($C_m + C_y$). Following up, for the maximum

generality, we require examining more complicated models where structural uncertainty exists in the model.

5.5 Case study 3: Real data from the Teal South reservoir model

In the last case study, we make use of the real Teal South reservoir model without any change to the original data (Figure 5.2) nor to the model configuration (Table 5.1). The real data are deemed to have structural uncertainty which influences the estimation of model parameters and prediction of field variables. Therefore, we no longer expect a Gaussian distribution for all posterior estimates. As previous case studies, we calibrate only one field variable, Field Oil Production Rate FOPR, as the only response variable of the misfit.

To test the predictive capability of our models when generalising to unseen data, they are only calibrated to the first 6 points (up to the day 181) of history match period (training set in Figure 5.8). The remainder will be used to evaluate the predictive performance of our models (Testing set in Figure 5.8). The reservoir simulation model has 6 unknown model parameters each with a uniform prior designated in Table 5.1.

Because we used a small dataset to train our models, the estimated uncertainty rises shortly after the start of the forecast. Figure 5.8 depicts LSQ history matching and forecast of the oil rate while a constant sigma $\sigma = 100$ is assumed for measurement uncertainty. The calibration does not thoroughly match the observation data due to both the measurement noise and modelling discrepancy.

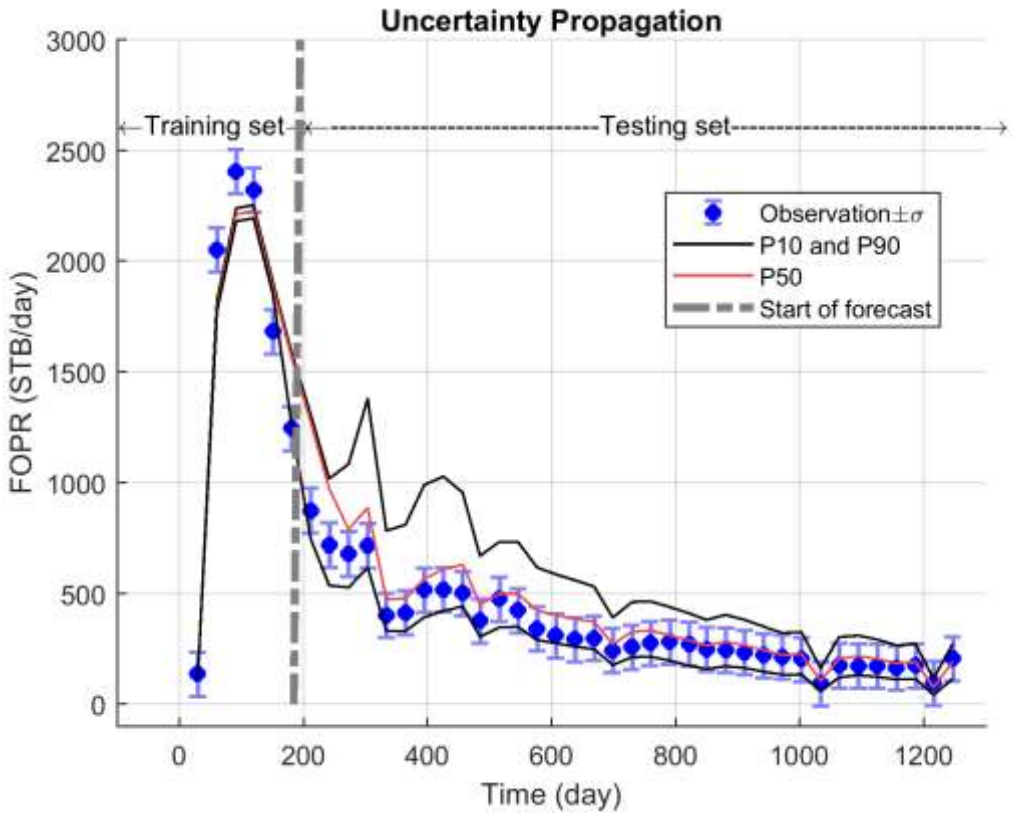

Figure 5.8 History matching of the Teal South reservoir model: LSQ history match is performed on the training set to forecast into the testing set.

Now, we treat measurement uncertainty and modelling error at 6 training points as unknowns which are to be learned from data. The parametric hierarchical model parameterises the total uncertainty ($C_m + C_y$) and solves for the 6 time-varying standard deviations. Prior information comprises a uniform U (1, 300) distribution for each of 6 unknown standard deviations. The solutions in Figure 5.9 include both the Maximum likelihood (ML) and the Full Bayesian Hierarchical (FBH) solutions to the standard deviations (s1, s2, …, s6 in Figure 5.9). We repeated this process 2 more times and compared the results of ML and FBH solutions in the figure below.

As expected, our analysis shown in Figure 5.9 depicts different modal estimates (ML1, ML2 and ML3) of standard deviations. The standard deviation s6 has the widest range of ML estimates which is indicative of objective function having multiple optimal solutions for s6.

The black lines then provide full posterior distributions of the hyper-parameters s1, s2, …, s6 (the unknown standard deviations for the first 6 data points). There is a clear consistency in the results of FBH solutions where multiple runs yield very similar posterior distributions.

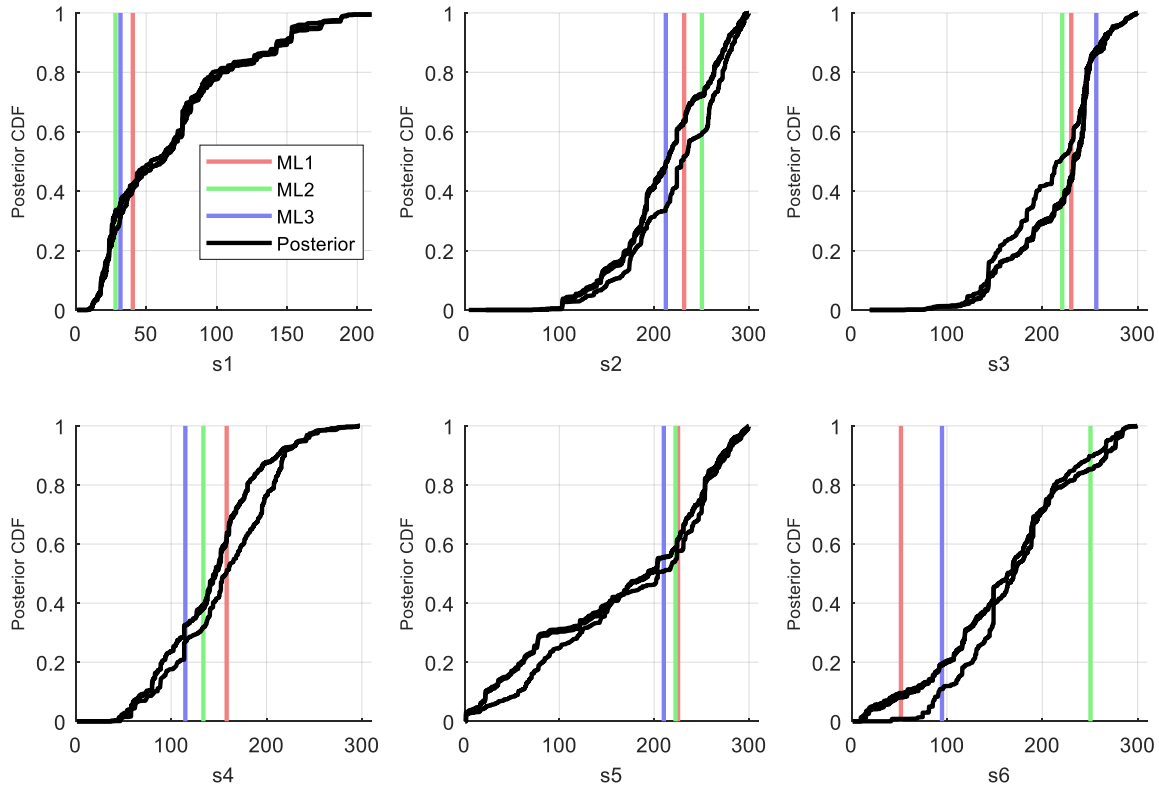

Figure 5.9 Estimation of hyper-parameters s1, s2, …, s6 of hierarchical modelling of discrepancy for the Teal South oilfield: Maximum likelihood (ML) estimates of standard deviations and full posteriors are collected through 3 history match runs with different starting points.

Once we estimated the standard deviations and evaluated the likelihood of each model, then, the posterior estimates of errors can be calculated. Posterior estimates of time-series errors can simply be calculated by the summation of the error of each calibrated model multiplied by the corresponding probability (see Figure 5.10). The maximum likelihood (ML) estimates, highlighted as coloured lines in Figure 5.10, correspond to point

estimates of errors having the highest likelihood, whereas black lines refer to full posterior estimates.

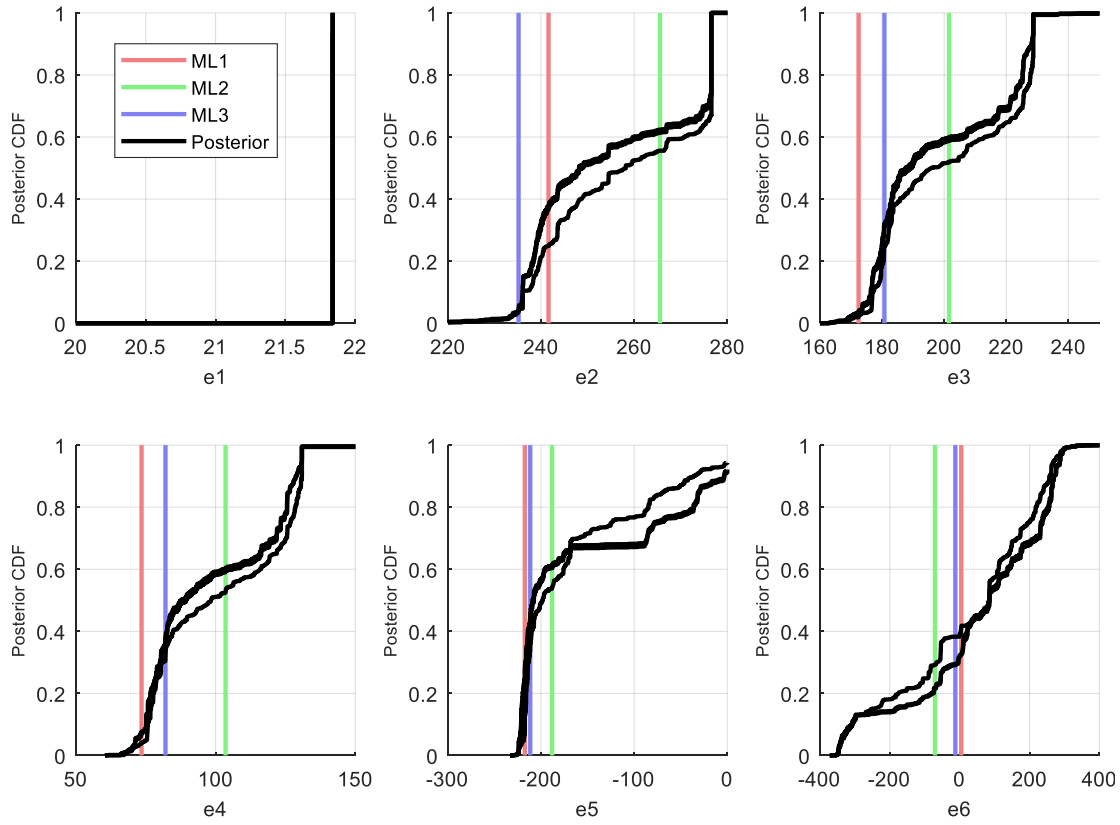

Figure 5.10 Posterior and modal estimates of the time-series errors (e1,…, e6) for the Teal South: coloured lines display Maximum likelihood (ML) estimates of errors, and black lines refer to the full posterior estimates

Now, we can observe the influence of the hierarchical approach on the prediction of the Teal South reservoir model. At the first step, we subtract the posterior mean errors estimated previously from the original observations to create new observations. The new observations shown as red dots in Figure 5.11, therefore, represents the mean-zero observations. The σ_is include both measurement uncertainty σ_ε and model discrepancy σ_δ in that $\sigma = \sigma_\delta + \sigma_\varepsilon$. At first view, the estimated σ_is seem to have traced the small and large gaps between P10 and P90 (e.g. at the last time step, high standard deviation allows for higher variation around the observation). Because, the FEM history match can quantify the degree of changes of the oil rate at different time steps.

The estimated s1, s2, …, s6 obtained from the FEM history matching of the Teal South reservoir model implies a higher standard deviation at the last time step, wherein the FEM allows for more significant variations. Interestingly, the high estimated sigma is in agreement with the credible uncertainty interval as the P10-P90 range becomes more extensive. On the contrary, the estimated time-varying standard deviation at the first history point is lower than the rest, implying that the calibration should penalise discrepancy at first point with higher intensity.

The posterior mean errors and standard deviations for the training dataset (interpolation) can change the uncertainty quantification since they account for the model discrepancy. This change becomes apparent where the best-fitting history match models lie within the 80% credible interval.

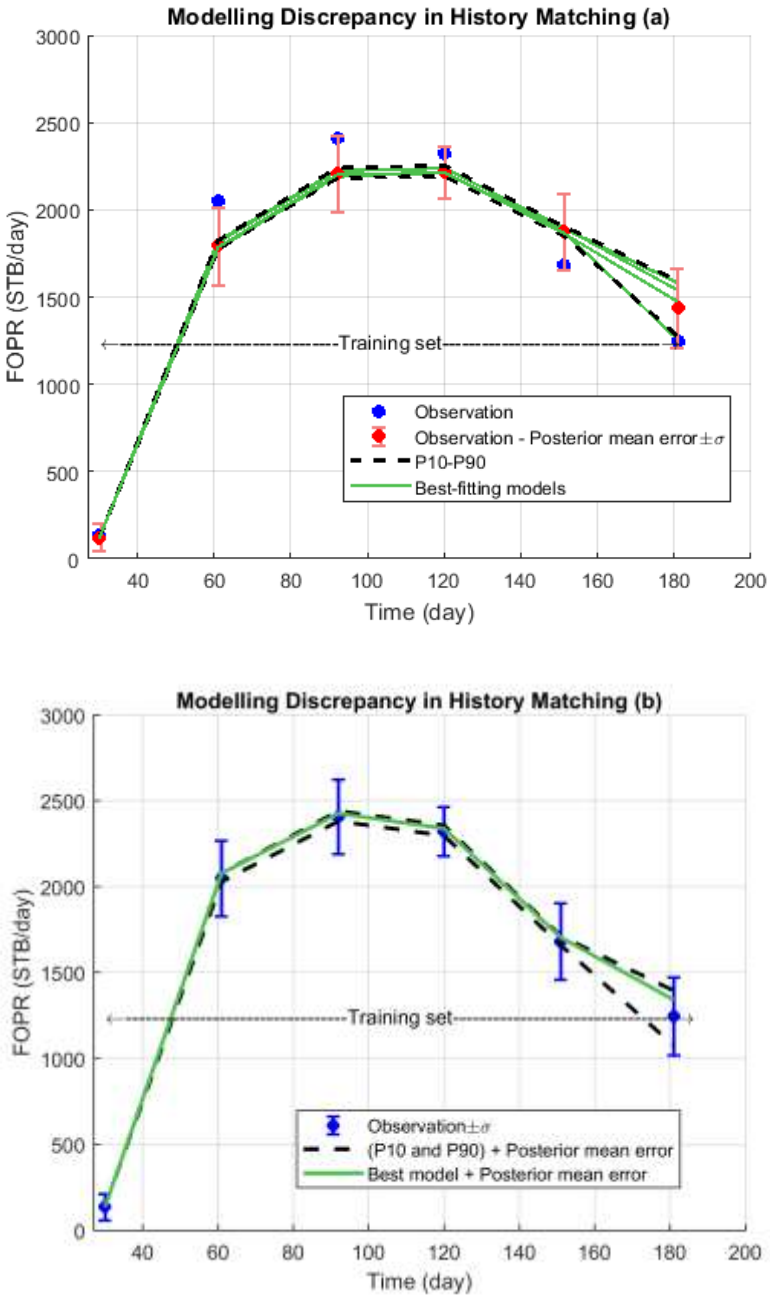

Figure 5.11 Posterior mean errors and standard deviations for the training data set estimated by a parametric error model: a) posterior mean error is subtracted from the original history to gain mean-zero observation which falls within the P10-P90 range. b) posterior mean error is added to the uncertainty intervals.

5.5.1 Estimation of the Teal South model parameters

Once we solved the unknown hyper-parameters of our hierarchical model, the posterior estimates of reservoir model parameters can be derived in terms of either the maximum likelihood estimates or the full posterior distribution.

To compare the unmodelled discrepancy scenario (LSQ) and modelled discrepancy scenario (FEM) we need a reference case as the true mean value of model parameters. To do so, we history match the whole observation (41 points instead of 6 early history period used by LSQ and FEM) and call it the reference. The accuracy of parameter estimation by LSQ and FEM can now be calculated using the Mean Absolute Percentage Deviation (MAPD) which computes the error of the estimated posterior mean P_k with respect to the true mean values $\overline{P_k}$:

$$MAPD = \frac{100}{p} \times \sum_{k=1}^{p} \left| \frac{P_k - \overline{P_k}}{\overline{P_k}} \right| \tag{5-1}$$

where p is the number of the model parameters. From history matching perspective, the sampling algorithm may also have the impact on our calibration/prediction results. Therefore, we run 3 restarts of history match using Particle Swarm Optimisation (PSO) and Bayesian Optimisation Algorithm (BOA). Then, the NAB computes the posterior distribution of each model parameter. The following tables summarizes the accuracy of estimation of parameters for different history match trials when compared to the reference scenario.

Table 5.2 Accuracy of posterior mean estimation of model parameters: 3 history match trials are run through PSO sampling for both LSQ and FEM scenarios

Modelling Strategy	***MAPD (1st Trial)***	***MAPD (2nd Trial)***	***MAPD (3rd Trial)***
LSQ-PSO	**22.6085**	**10.8895**	**16.0394**
FEM-PSO	**18.4951**	**20.122**	**14.1335**

Table 5.3 Accuracy of posterior mean estimation of model parameters: 3 history match trials are run through BOA sampling for both LSQ and FEM scenarios

Modelling Strategy	***MAPD (1st Trial)***	***MAPD (2nd Trial)***	***MAPD (3rd Trial)***
LSQ-BOA	**44.8289**	**36.0144**	**25.9021**
FEM-BOA	**33.6223**	**31.1214**	**26.0206**

From the tables above these conclusions can be drawn:

The choice of sampling algorithm obviously influences the accuracy of our estimation. For instance, from the data in Table 5.3, it is apparent that BOA fails to reach the accuracy below 25% error, whereas all history match trials with PSO have MAPD score between 10% and 23% (see Table 5.2). From Table 5.2, however, it is revealing that both of LSQ and FEM give a reliable estimate of posterior mean when sampling with PSO. On average, FEM is less influenced by the sampling algorithm, while LSQ yields unreliable posterior estimates in case sampling is carried out by BOA.

Figure 5.12 and Figure 5.13 depict the results of Table 5.2 and Table 5.3 regarding the posterior cumulative distribution function (CDF) where maximum likelihood estimates (ML1, ML2, ML3) and full posterior distributions are compared with the reference scenario. Not surprisingly, in all scenarios, the resulting ML solutions can be quite far from the reference value. Therefore, we interpret the estimated probabilities in the form of full posteriors and compare their variations.

Making a comparison between the posteriors of Figure 5.12-a and Figure 5.12-b, it is evident that FEM is less likely to be influenced by multiple restarts of history match runs. Hence, the FEM regression model is more consistent when different initial random starting points are concerned. Now, if we change the sampling algorithm to BOA, the same behaviour holds for posterior distributions in Figure 5.13. There exists a considerable degree of variations among multiple history match runs in Figure 5.13-a where LSQ fails to produce consistent posteriors. On the other hand, FEM improves this consistency to some extent and narrow down the uncertainty caused by different starting points.

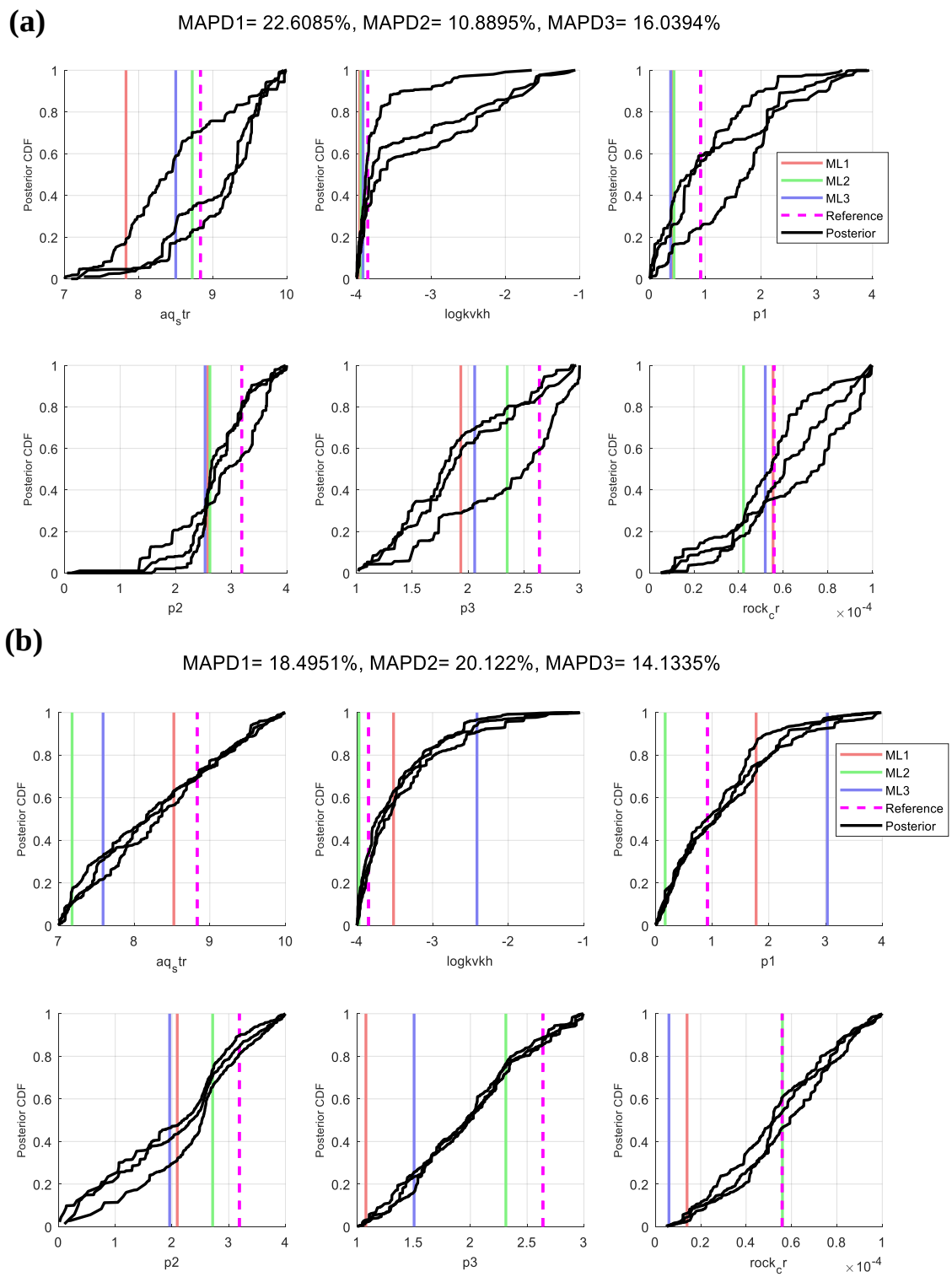

Figure 5.12 Posterior CDF of 6 model parameters history matched with PSO sampling: a) LSQ regression model and b) FEM regression model. 3 LSQ history match trials give the full posterior CDF shown as black lines and the maximum likelihood solutions (ML) shown as coloured lines.

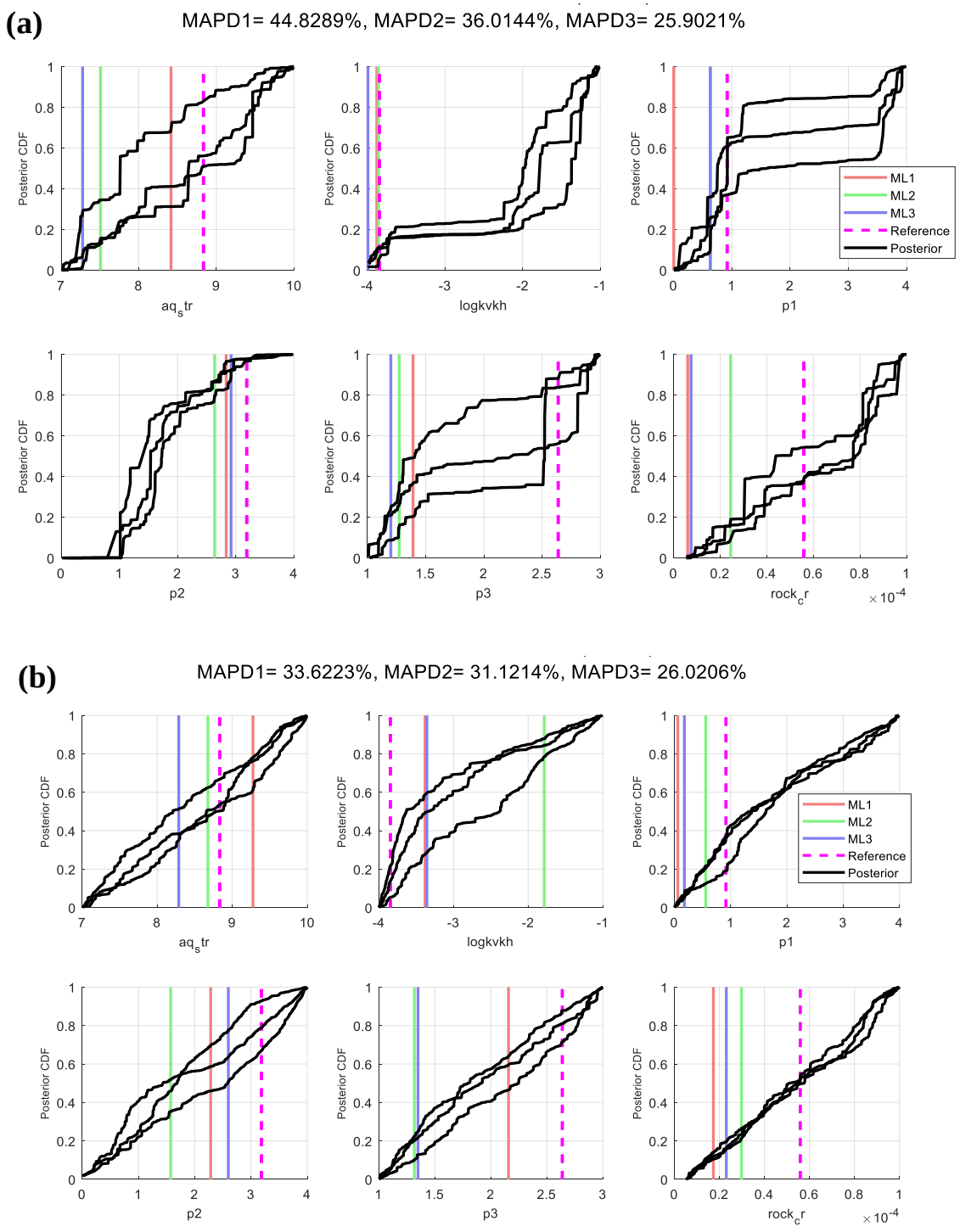

Figure 5.13 P osterior CDF of 6 model parameters history matched with BOA sampling: a) LSQ regression model and b) FEM regression model. 3 LSQ history match trials give the full posterior CDF shown as black lines and the maximum likelihood solutions (ML) shown as coloured lines.

5.5.2 Comparison of Uncertainty Quantification

In this section, the LSQ and FEM calibration models of the previous section are used to produce a probabilistic forecast of the oil rate. The quality of the forecasts was assessed by the Brier score (BS) – a quantity used to assess the reliability of weather forecasts such that the lower the Brier Score the better the quality of model prediction (Brier, 1950). In other words, the estimated BS quantifies what share of the observation data fall in uncertainty interval (P10-P90). The Brier Score, however simple, is a potent metric for quantifying the robustness of the forecast (Brier, 1950. For 80% certainty space (P10-P90), the BS is obtained by

$$\mathrm{BS} = \frac{1}{\mathrm{n}}\left(\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{f_i}^{\mathrm{p10}} - 0.9\right)^2 + \sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{f_i}^{\mathrm{p90}} - 0.9\right)^2\right) \tag{5-2}$$

where n is the number of forecast points ($\mathrm{n} = 35$ in our experiments), $\mathrm{f_i}^{\mathrm{p10}}$ and $\mathrm{f_i}^{\mathrm{p90}}$ are the binary value of 1 or zero, when predicted within or outside certainty space respectively.

Table 5.4 and Table 5.5 provide a comparison of oil rate prediction with and without modelled discrepancy based upon 3 trials of history matching sampled by PSO and BOA. Taking together, the results from Table 5.4 and Table 5.5 suggest that FEM is less likely to be influenced by sampling algorithms with the Brier Score not exceeding 0.043. However, this is not true of LSQ, because LSQ-PSO, has better predictive performance when compared with LSQ-BOA. It is also remarkable to note that both LSQ and FEM regression models of this study robustly predict the uncertainty when sampling is carried out by PSO.

Table 5.4 Accuracy of probabilistic prediction: 3 history match trials are run through PSO sampling for both LSQ and FEM scenarios

Modelling Strategy	***BS (1st Trial)***	***BS (2nd Trial)***	***BS (3rd Trial)***
LSQ-PSO	**0.0429**	**0.0200**	**0.0657**
FEM-PSO	**0.0200**	**0.0200**	**0.0429**

Table 5.5 Accuracy of probabilistic prediction: 3 history match trials are run through BOA sampling for both LSQ and FEM scenarios

Modelling Strategy	***BS (1st Trial)***	***BS (2nd Trial)***	***BS (3rd Trial)***
LSQ-BOA	**0.0657**	**0.1571**	**0.0429**
FEM-BOA	**0.0200**	**0.0200**	**0.0200**

To schematically understand the impact of modelling discrepancy on uncertainty quantification, and to better interpret the results of the tables above, we present the forecast results in

Figure 5.14 and Figure 5.15. As can be derived from the following figures, the average Brier Score evaluated over 3 trials (with different random starting points) for each modelling strategies reflect on strong predictive capability in both of LSQ and FEM models.

Overall, the most striking result to emerge from our uncertainty assessment is that for each sampling algorithm, modelling discrepancy significantly enlarges prediction interval of our calibration target FOPR. This enlargement is attributed to the high values of estimated standard deviations evaluated by the hierarchical approach which allows for higher variations of oil rate.

Looking at the models sampled by LSQ-PSO in

Figure 5.14, most of the observations fall within the certainty space where the average Brier Score is 0.043. Instead, FEM-PSO has slightly better predictive performance (BS= 0.0276) and offers a larger estimation of certainty space.

About models sampled by BOA, the distinction between LSQ and FEM becomes remarkably significant. From Figure 5.15, FEM (BS =0.02) offers not only a better predictive performance but also a wider credible interval when compared with LSQ (BS=0.0885). Therefore the sampling method is more likely to influence LSQ history matching models rather than FEM models.

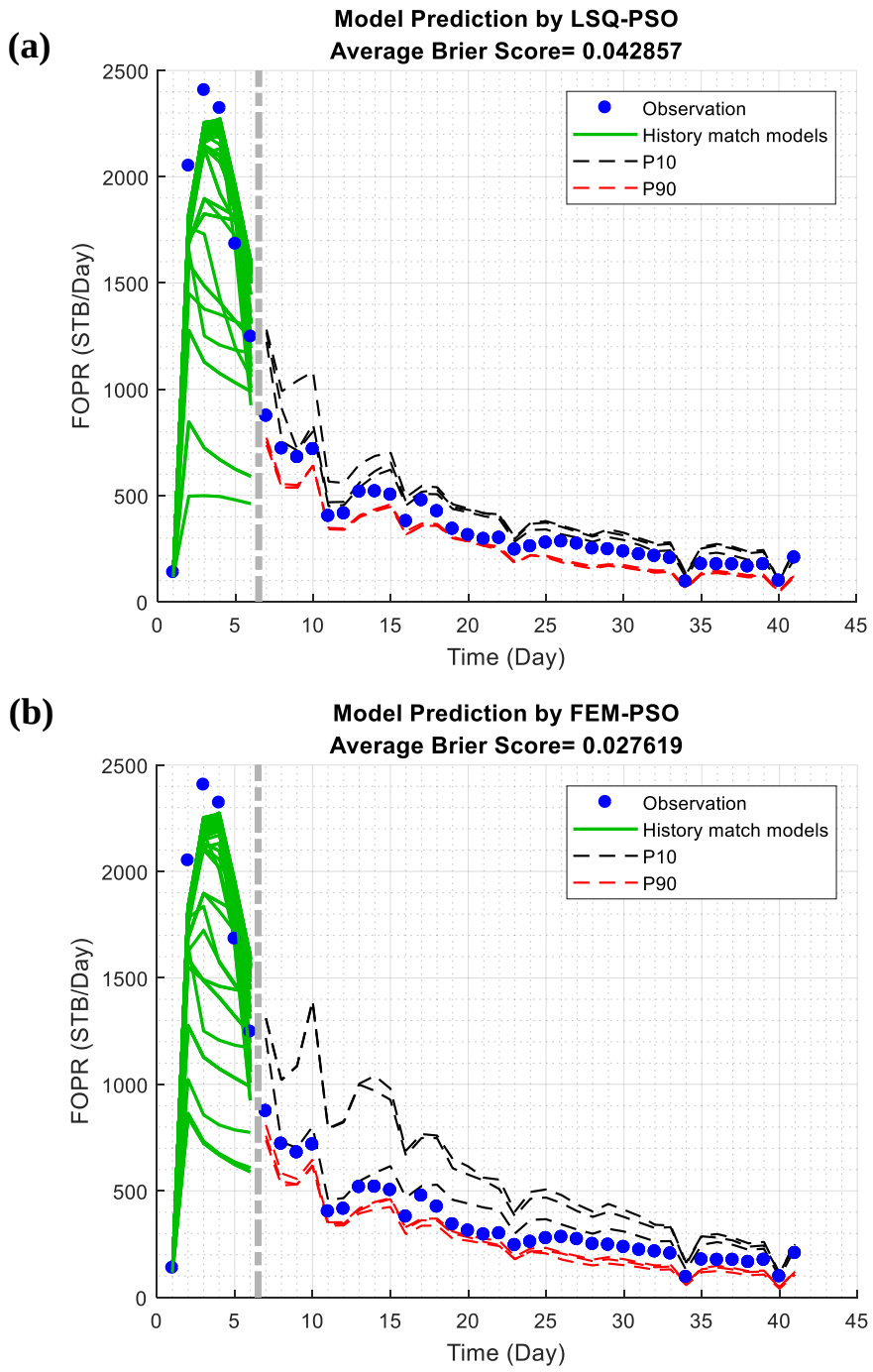

Figure 5.14 Model prediction by a) LSQ history matching; and b) FEM history matching. 3 trials of history matching with random starting points are sampled by PSO.

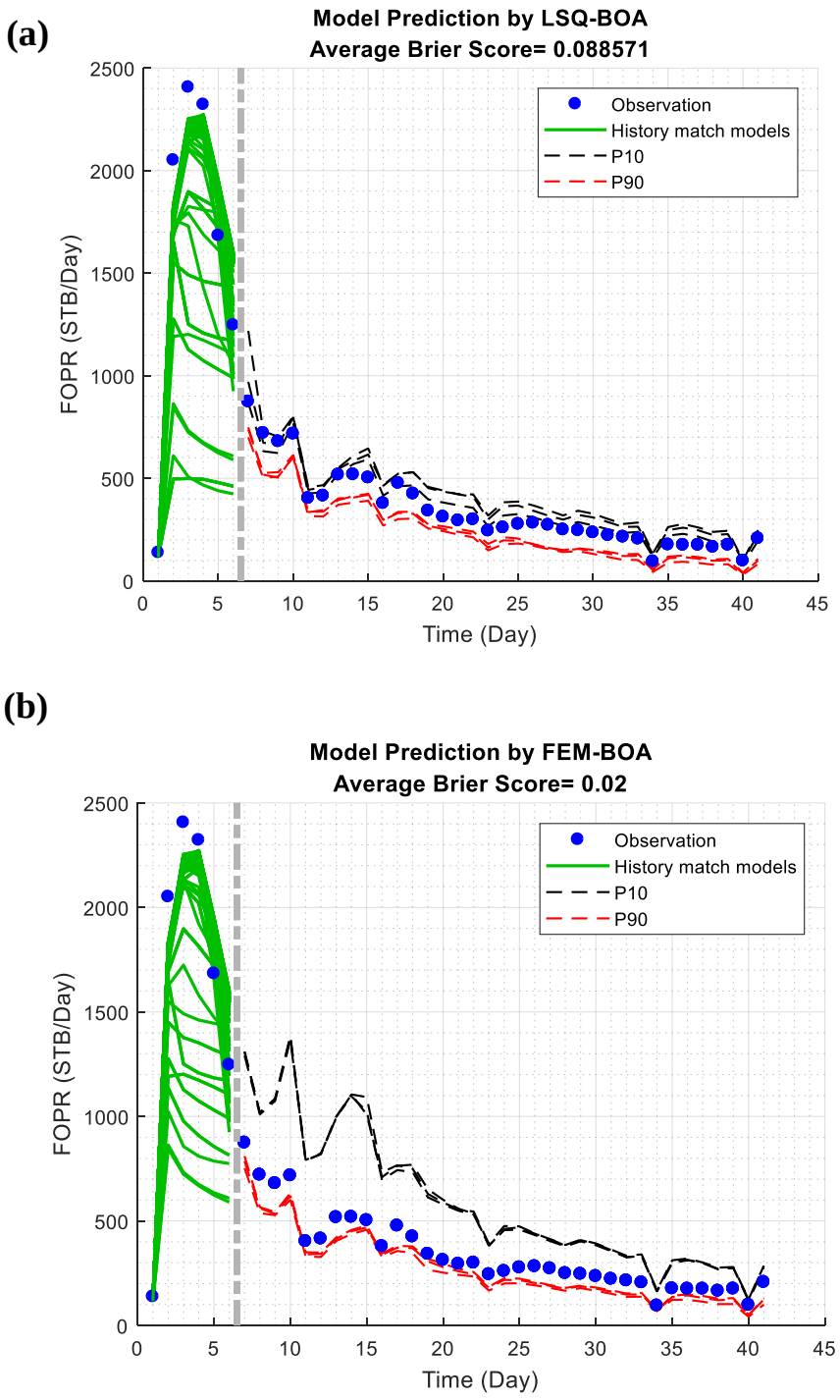

Figure 5.15 Model prediction by a) LSQ history matching; and b) FEM history matching. 3 trials of history matching with random starting points are sampled by BOA.

5.6 Conclusion

To investigate a comparative study of different error modelling strategies, we established 2 different modelling scenarios. In the first scenario, LSQ history match was used to match 6 (out of 41) history match points ignoring modelling discrepancy for the target FOPR while forecasting the remainder (35 observations). The second scenario, the Full Error Model (FEM), was also set to match 6 observations accounting for modelling discrepancy regarding a parametric hierarchical model.

Therefore, we made use of a standard least-squares regression (LSQ) with a constant measurement noise (σ_ε =100). The second scenario, the Full Error Model (FEM), utilises 500 samples yet to match 6 observations accounting for modelling discrepancy. Both scenarios are followed by two more restarts to monitor the impact of different starting points on the estimated posteriors.

We examined our modelling strategies with 3 case studies. In the first case study, we built an idealised model by simulation of the Teal South reservoir model at the actual value of 2 model parameters. Then, LSQ and FEM history match tuned the 2 parameters until finding their actual values. Both of FEM and LSQ recovered the real value of input variables with the same speed of convergence and the same distribution.

We then tried to examine the impact of varying standard deviations on posterior probabilities gained by each modelling strategies. Therefore, in the second case study, we added Gaussian white noise to the idealised model data while the model configuration remained unchanged. Our findings demonstrated that LSQ is very sensitive to the changes in the standard deviation of the noise, whereas FEM is more consistent with changes to the assumed standard deviation.

In the last case study, we utilised the original observation data from the Teal South reservoir model with all 6 model parameters. In addition to model parameters, FEM made use of 6 unknown time-varying standard deviations which tracked down the large and small gaps in the confidence interval. Regardless of the type of the sampling method, the FEM reached a better model prediction score compared to the LSQ models.

As for the estimation of model parameters, the choice of sampling algorithm changed the accuracy of our estimation. BOA failed to achieve the accuracy below 25% error, whereas all history match trials with PSO had MAPD score between 10% and 23%. On average,

FEM was less influenced by the sampling algorithm or by multiple restarts of history match runs. Hence, the FEM model was more consistent when the different initial random starting point is concerned.

Chapter 6 – Hierarchical error modelling: Application to the Zagadka

6.1 Introduction

In Chapter 5 we used a parametric hierarchical model to model the discrepancy in history matching of the Teal South. Although our parametric hierarchical model was able to predict the error within the history match period, it failed to generalise to the forecast period. Now, we aim to find the error model at any instance by use of non-parametric emulators that generalise towards the entire history match/forecast period.

In Chapter 4, we presented a non-parametric hierarchical approach for evaluating reservoir history matching models and compared it to the history matching in which the model discrepancy is ignored. We then discussed how neglecting model error could result in biased estimation of uncertainty. Now, it is time to test our statistical model in a more complex reservoir model while doing a comparative study on different scenarios.

This chapter also intends to examine two significant aspects of error modelling throughout the history matching of oil reservoirs: the non-parametric hierarchical modelling of model inadequacy, and the model selection problem. Therefore, a primary correlation analysis is performed on the field production variables to detect the likely non-stationary behaviours in time-series errors. Afterwards, the segmentation technique described in chapter 4 partitions the non-stationary dataset into stationary segments. This enables us to select those models that best fit the data without overfitting.

We interpret the model selection problem in the data-driven setup that enables us first to interpolate the error in the history period and then to propagate it towards unseen data (i.e. error generalisation). The error models constructed by inferring parameters of selected models (e.g. kernel sigma and length scale) can predict the response variable (e.g. oil rate) at any point in input space (e.g. time) with corresponding generalisation uncertainty. These models are inferred through the training set (history period) and further compared in terms of generalisation error on a test set (forecast).

Once proper kernels selected for each segment, the error models are ready to plug into Bayesian calibration and prediction. Note that our central focus is on an appropriate way of error modelling rather that different solutions to non-stationarity.

Similar to the previous chapter, the Full Bayesian Hierarchical (FBH) solutions, Maximum Likelihood (ML) solutions and the Linear Least-Squares (LSQ) are used to perform history matching with and without modelling discrepancy. We parametrise uncertainty within Kennedy O'Hagan (KOH) framework in that the modelling discrepancy adds up to the total uncertainty and compare it to linear least-squares LSQ model which ignores model discrepancy.

6.2 Application to the real case study: Zagadka oilfield

To examine the influence of non-parametric error modelling throughout the process of history matching and to realise how accounting for modelling discrepancy differs from ignoring it we make use of the Zagadka reservoir model (Christie et al., 2013). Zagadka is a medium sized oilfield under aquifer/water injectors pressure support, produced through a combination of waterflood and aquifer drive with over 100 wells drilled progressively during the reservoir life. The field is compartmentalised, with sealing faults creating 7 compartments.

The 108 × 265 ×7 reservoir model used by Christie et al. (2013) comprises around 135,000 active cells and 95 wells, most of which have over 10–15 years of history. All wells are divided into 9 main groups (G1 to G9 in Figure 6.1) based on the geological structure of the field, time of drilling and fault block in the model (Christie et al., 2013). As can be seen from Table 6.1, the Zagadka reservoir is parameterised by 12 model parameters, all with uniform prior distribution. Note that parameters such as vertical permeability multipliers are not in the parameter set. However, in other applications, such parameter combinations may be included in the reservoir model.

Table 6.1 Model parameters distribution in the Zagadka Case Study

Model Parameters	***Quantity***	***Range***
Fault transmissibility	1	0.0 – 1.0
Oil relative permeability	1	0.6 – 0.9
Water relative permeability	1	0.2–1.25
Capillary pressure 3	1	1.01 – 3.0
Aquifer support multiplier	2	2.5 – 4.5
K_h multiplier	6	1 – 30

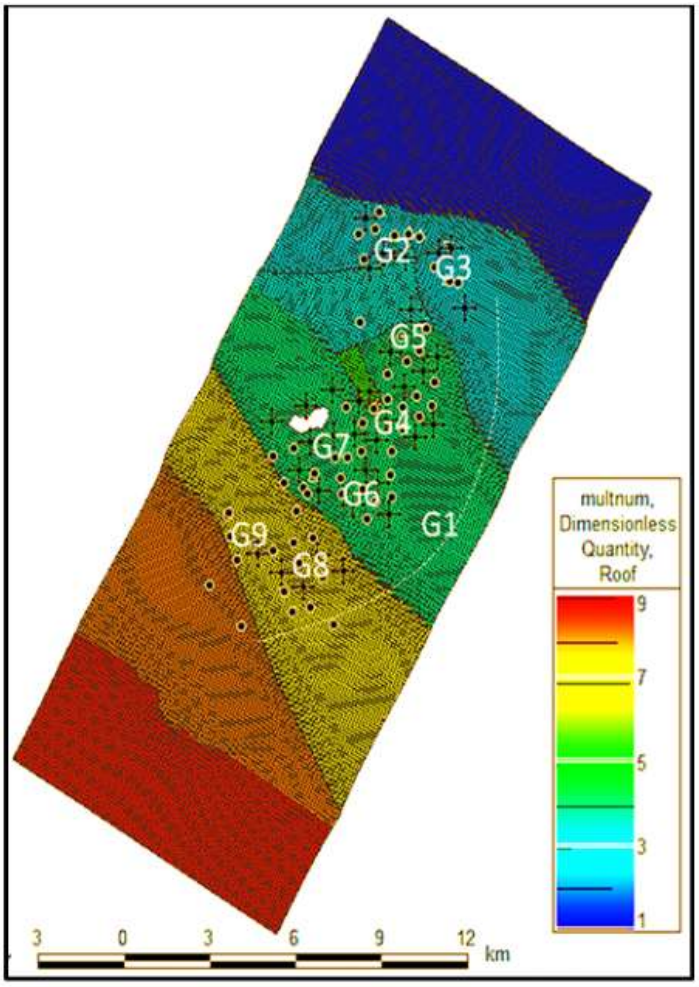

Figure 6.1 The Zagadka reservoir model

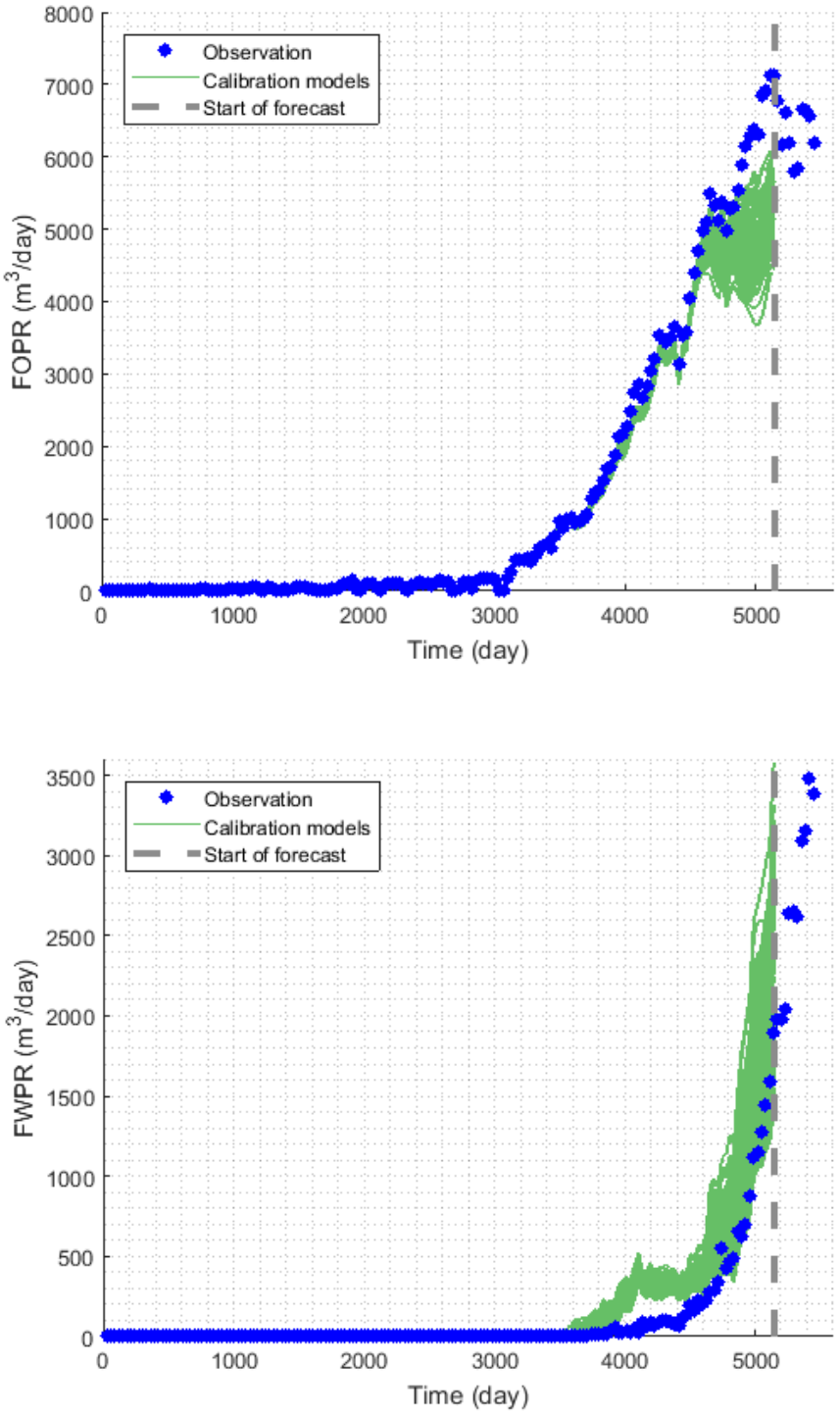

Figure 6.2 Production historical data collated for the Zagadka oilfield

We history match 2 field variables: Field Oil Production Rate FOPR and Field Water Production Rate FWPR. The sum of these misfit values computes the global objective

function. Moreover, the oil rate and the water rate (time-series variables) will be calibrated to 5144 days of history (see Figure 6.2); the remainder is to be used in the forecast and uncertainty quantification (similar to the Teal South case study).

The calibration does not give a perfect match due to the observation error and modelling discrepancy with early water breakthrough and under-production of oil as presented in Figure 6.2. Note that simulation results in Figure 6.2 may look different from previous studies of the Zagadka due to the removal of some parameters of the model (e.g. 19 parameters are reduced to 12). For each of the calibration targets, the residuals are recorded throughout history matching runs.

The autocorrelation is then computed to explore the likely non-stationarity of best-fitting model residuals, and if there is such non-stationary behaviour, find an optimal segmentation solution. We do correlation analysis for the residuals of the best-fitting model for both FOPR and FWPR. The autocorrelation plot depicted in Figure 6.3-a indicates a strong autocorrelation between residuals at early history period while, except some last points, an overall downward trend throughout the entire history is observed. To resolve this issue, we exert segmentation through optimal change point detection described in Chapter 4.

Segmentation replaces our non-stationary time series with 3 stationary segments ranging from x_{1-130}, $x_{131-155}$ and $x_{156-169}$ where x is the time, and the indices are the time steps. Figure 6.3-b shows optimal change point detection with the maximum number of change points being set to 2. The segmentation traces the optimal locations wherein the mean and the slope of the signal change most abruptly.

The bottom plots in Figure 6.3 refer to autocorrelation plots for 3 optimal segments found by the technique above. Except for the first time lags, the rest of the correlations remain within the confidence interval implying stationary processes. The same method is used to obtain stationary segments for FWPR. Afterwards, the model selection identifies the form of the covariance function of the modelling discrepancy. The error generalisation is finally performed using ML, and FBH approaches. The history match and the modelling discrepancy is then repeated with 2 more restarts to spot the variability of the error generalisation and estimation of model parameters.

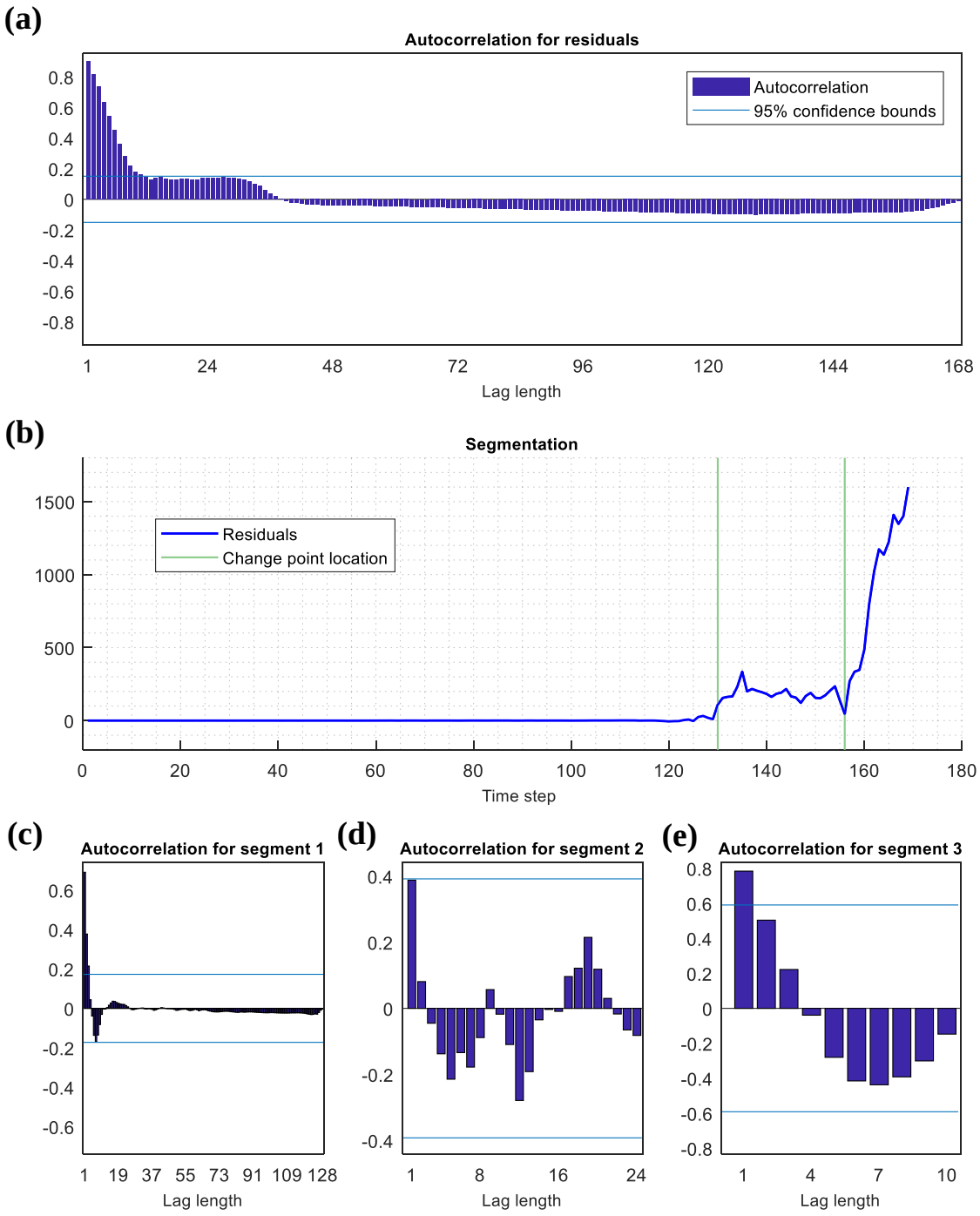

Figure 6.3 Autocorrelation plots for FOPR residuals: **a)** The plot shows autocorrelation for 169 residuals wrt the time lag. The first 11 lags have significant correlation outside confidence interval while correlation decays gradually for the rest. **b)** The Segmentation of residuals by optimal change point detection. **c, d, e)** The bottom plots refer to autocorrelation plot for 3 optimal segments after segmentation of 169 residuals.

6.3 BIC Minimisation

The autocorrelation and segmentation helped us better understand the non-stationarity of the residuals and partitioned the data into different stationary segments. Now, we can employ stationary kernels to fit each partition separately and generalise the response to the future data. However, a significant obstacle in this kind of application is how to make reliable assumptions about the correlation structures and mean errors.

In this part, we try to fit a zero mean Gaussian Process with unknown kernel function to the estimated modelling discrepancy δ at each segment in an optimisation process. The fitting models are evaluated through BIC model selection (Eq. (4-14)) which posits penalty on complex models. To do so, 10 types of randomly chosen kernel functions (50 experiments for each segment) are drawn without any constrained prior on the parameters of the regression models (see Table A- 1). The best regression model with the lowest BIC ($\hat{\delta}$) is then selected by Eq. (4-15) to address our prior knowledge about the model discrepancy. In all of our experiments, a constant measurement noise is assumed to be fairly known (σ_ε=200).

From Figure *6.4* we can see that, on average, models for segment 1 have larger BIC values compared to the other segments. This comes from the more significant number of points, and hence a higher code length for segment 1. The average estimated uncertainty bars due to the model discrepancy σ_δ for our best-fitting model in segments 1, 2, 3 and test set are 0.2870, 24.8095, 7.1990 and 92.9042 respectively (To see estimated total uncertainty bars at each point see the shaded green area in Figure *6.5*). The highest average σ_δ comes from extrapolation region (test set) as expected, whereas the lowest belongs to the first segment whose residuals have little variations from zero.

Overall, the estimated uncertainty reflects the extent to which the Gaussian process model is uncertain about its prediction. Therefore, testing points which are farther away from the training set will have higher uncertainty.

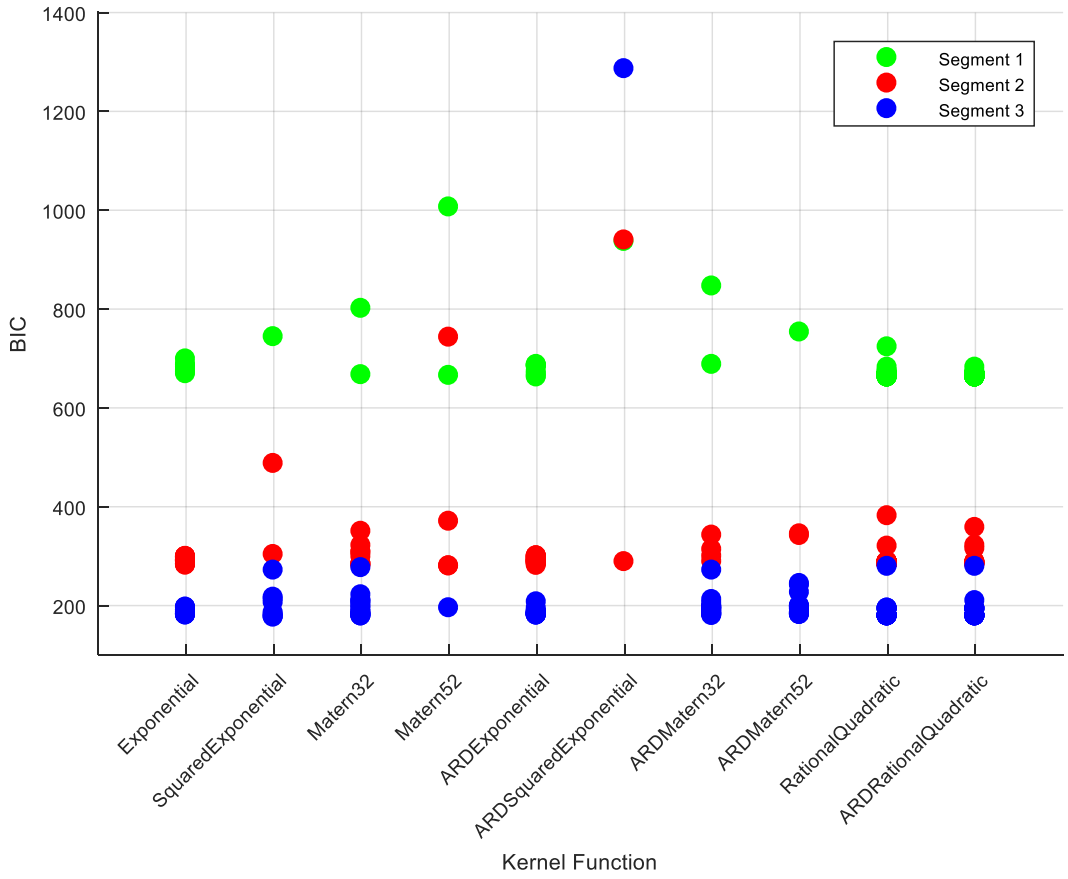

Figure 6.4 Optimisation of GP regression model fitting FOPR residuals by BIC minimisation: 10 stationary kernel functions with random prior means are used for each segment.

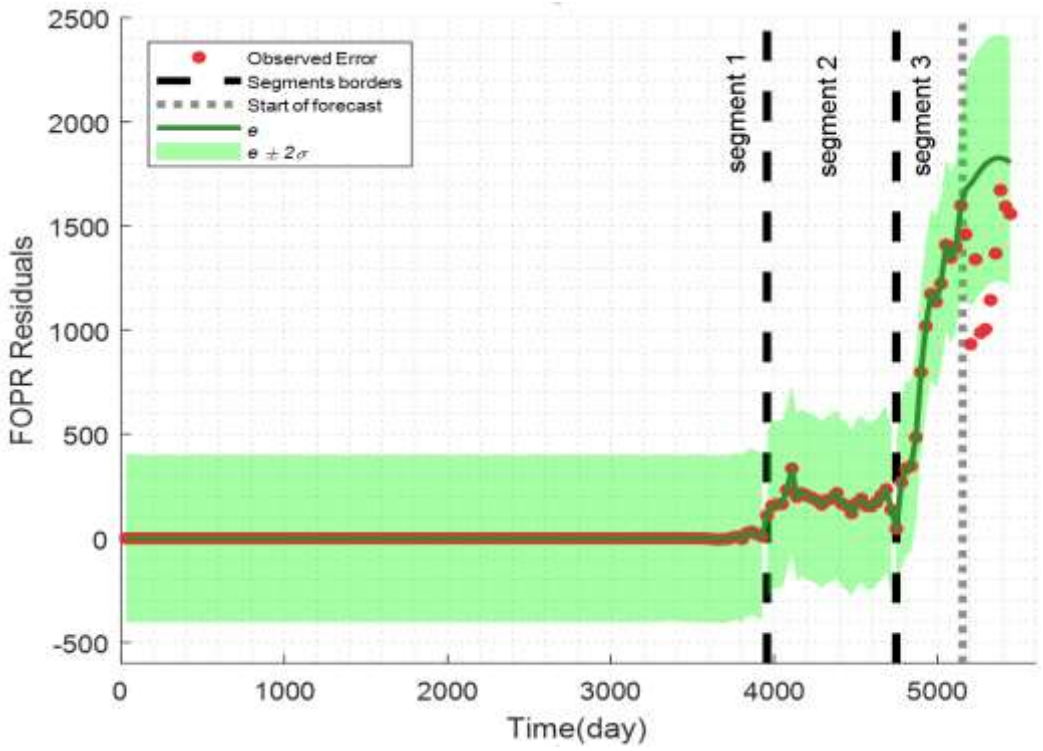

Figure 6.5 Best regression model fits the residuals with the lowest BIC at each segment. The last ten residuals are predicted with the associated uncertainty bars shown in shaded green ($\boldsymbol{\sigma = \sigma_{\delta} + \sigma_{\varepsilon}}$).

6.4 Comparison of Uncertainty Quantification (UQ)

Based on the knowledge gained with BIC minimisation, we set up the final regression model for each segment to carry out hierarchical modelling of the discrepancy. Besides, the selected Gaussian Process regression models estimate the total uncertainty of the field variables which are to be used in history matching. Now, the estimated uncertainties are plugged into the Bayesian framework to predict Bayesian confidence intervals.

To investigate a comparative study of different error modelling strategies, we establish 3 different modelling scenarios. In the first scenario, 400 samples are used to match 169 observations ignoring modelling discrepancy for each target FOPR and FWPR. Therefore, standard least-squares regression (LSQ) is used with a constant measurement noise (σ_{ε}=200). The second and third scenarios use 400 samples yet to match 169 observations accounting for modelling discrepancy by ML and FBH solutions respectively.

All these scenarios are followed by two more restarts to monitor the impact of different starting points on the estimated posteriors. As for the Automated History Matching, we made use of Bayesian Optimisation Algorithm (BOA).

The BOA (Pelikan et al., 1999) algorithm is deployed asynchronously throughout multiple cores, with each core performing a new simulation as soon as it has terminated the previous one.

Regarding the reservoir simulator, a parallel simulator is used to conduct the high efficient parallel optimisation with multiple concurrent simulations being computed throughout each history match run. The NA-Bayes (NAB) resampler (Sambridge, 1999a) is adopted for the posterior uncertainty analysis to approximate the value of various Bayesian integrals.

Because we use 3 different error modelling strategies (LSQ, ML and FBH), the posterior estimates of model parameters and the NAB forecasts will be different. Like the Teal South reservoir model, we compare predicted uncertainties for each modelling strategy using the Brier Score (5-2).

In this case study, we also compare the P50 prediction of our models against the observations since P50 offers the same probability to the lower/higher production rates. Once the model prediction is addressed by NAB, the result of each approach is evaluated through Normalized Root Mean Square Deviation (NRMSD) in that the lower the NRMSD, the better the prediction. The NRMSD computes P50 (the median) deviation from observation y_i to quantify the quality of predictions as follows:

$$NRMSD = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - P50_i)^2}}{Max(y_i) - Min(y_i)} \tag{6-1}$$

Table 6.2 provides a comparison of FOPR prediction with and without modelled discrepancy based upon 3 trials of history matching (with different random starting points). The estimated NRMSD quantifies the deviation of predicted P50 from observations in both interpolation (training phase) and extrapolation (testing phase).

Different trials of history matching with modelled discrepancy yield better posterior estimates of FOPR when compared to unmodelled discrepancy (LSQ) history matching models. It can be seen from NRMSD data in Table 6.2 that the modelled discrepancy significantly improves P50 prediction of our calibration target FOPR whereas there is little variation among ML and FBH approaches. Similarly, the uncertainty quantification (BS) gained by FBH and ML models remarkably outperforms those of LSQ models when estimating P10-P90 credible intervals.

On the other hand, modelled discrepancy offers a less statistically significant improvement for FWPR prediction as set out in Table 6.3. The NRMSD data collected through 3 trials of history matching indicate a slight improvement in P50 predictions by hierarchical models whereas the BS scores remain unchanged for each trial. Overall, the Brier Scores demonstrate that, regardless of the modelling strategy, FWPR gains a much better Brier Score when compared with FOPR.

Figure 6.6-Figure 6.11 plot the findings of Table 6.2 and Table 6.3 where the prediction improvement gained by modelling discrepancy becomes evident especially for oil rate. The GPs add an additive correlated uncertainty term to the ongoing history match process that is propagated towards extrapolation areas.

In Figure 6.6-Figure 6.11, the shaded green area becomes wider once the forecast begins for unseen data which is an inevitable part of GP regression. The resulting predictions demonstrate the importance of using error modelling in the prediction of response variables both regarding deviation from the median (NRMSD) or estimated credible interval (BS).

Table 6.2 Comparison of modelling strategies in terms of better prediction of the water rate FOPR (lower NRMSD)

Modelling Strategy	***1st Trial***		***2nd Trial***		***3rd Trial***	
	NRMSD	***BS***	***NRMSD***	***BS***	***NRMSD***	***BS***
LSQ	**0.3671**	**0.8200**	**0.3663**	**0.8200**	**0.3698**	**0.8200**
ML	**0.1074**	**0.5000**	**0.1166**	**0.5000**	**0.2029**	**0.7400**
FBH	**0.1011**	**0.5000**	**0.1211**	**0.5000**	**0.1861**	**0.7400**

Table 6.3 Comparison of modelling strategies in terms of better prediction of the water rate FWPR (lower NRMSD)

Modelling Strategy	***1st Trial***		***2nd Trial***		***3rd Trial***	
	NRMSD	***BS***	***NRMSD***	***BS***	***NRMSD***	***BS***
LSQ	**0.7878**	**0.0200**	**0.8525**	**0.1800**	**0.8684**	**0.1800**
ML	**0.7339**	**0.0200**	**0.7510**	**0.1800**	**0.7783**	**0.1800**
FBH	**0.7523**	**0.0200**	**0.7596**	**0.1800**	**0.7561**	**0.1800**

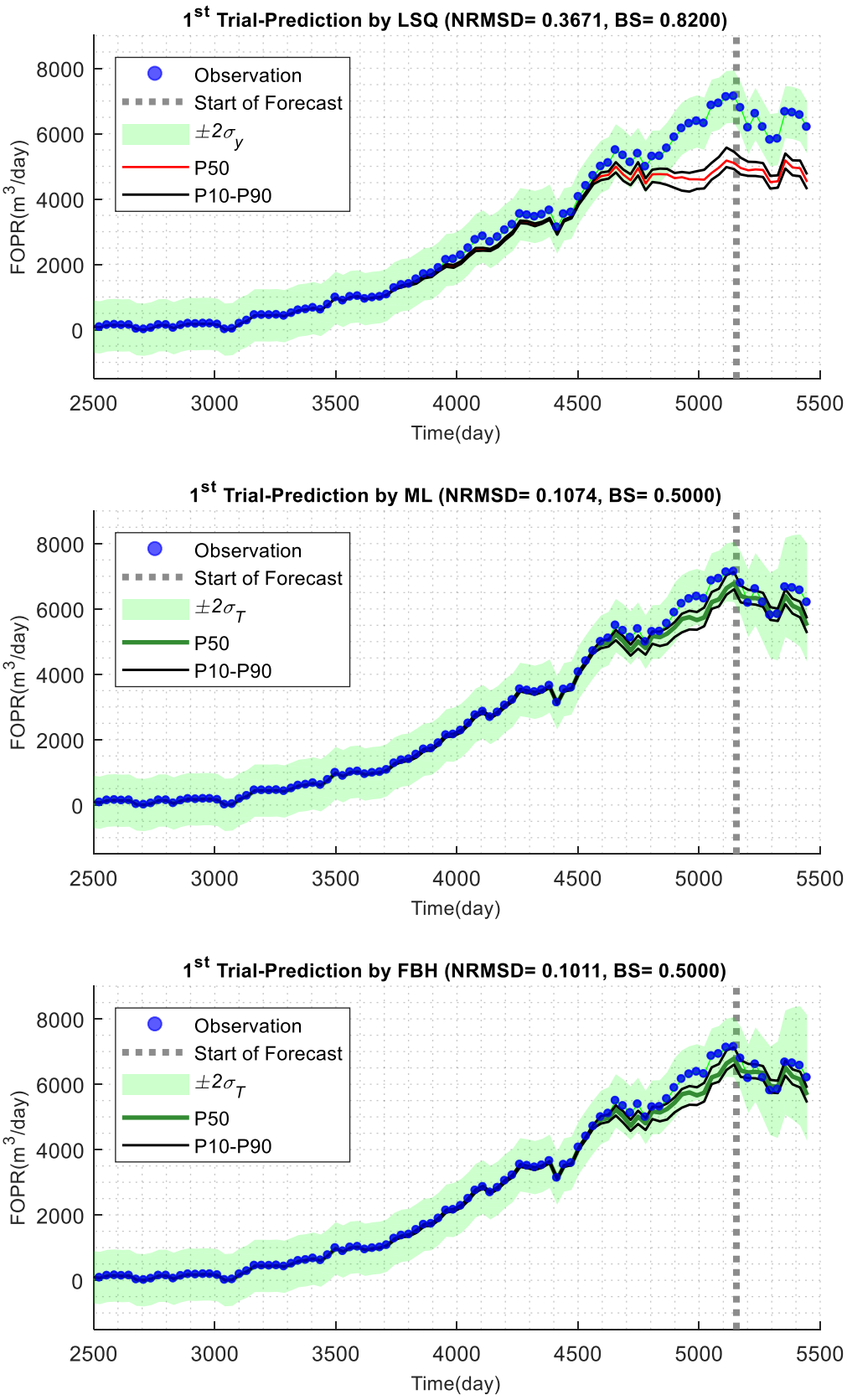

Figure 6.6 FOPR Prediction by LSQ, FBH and ML models (1st Trial)

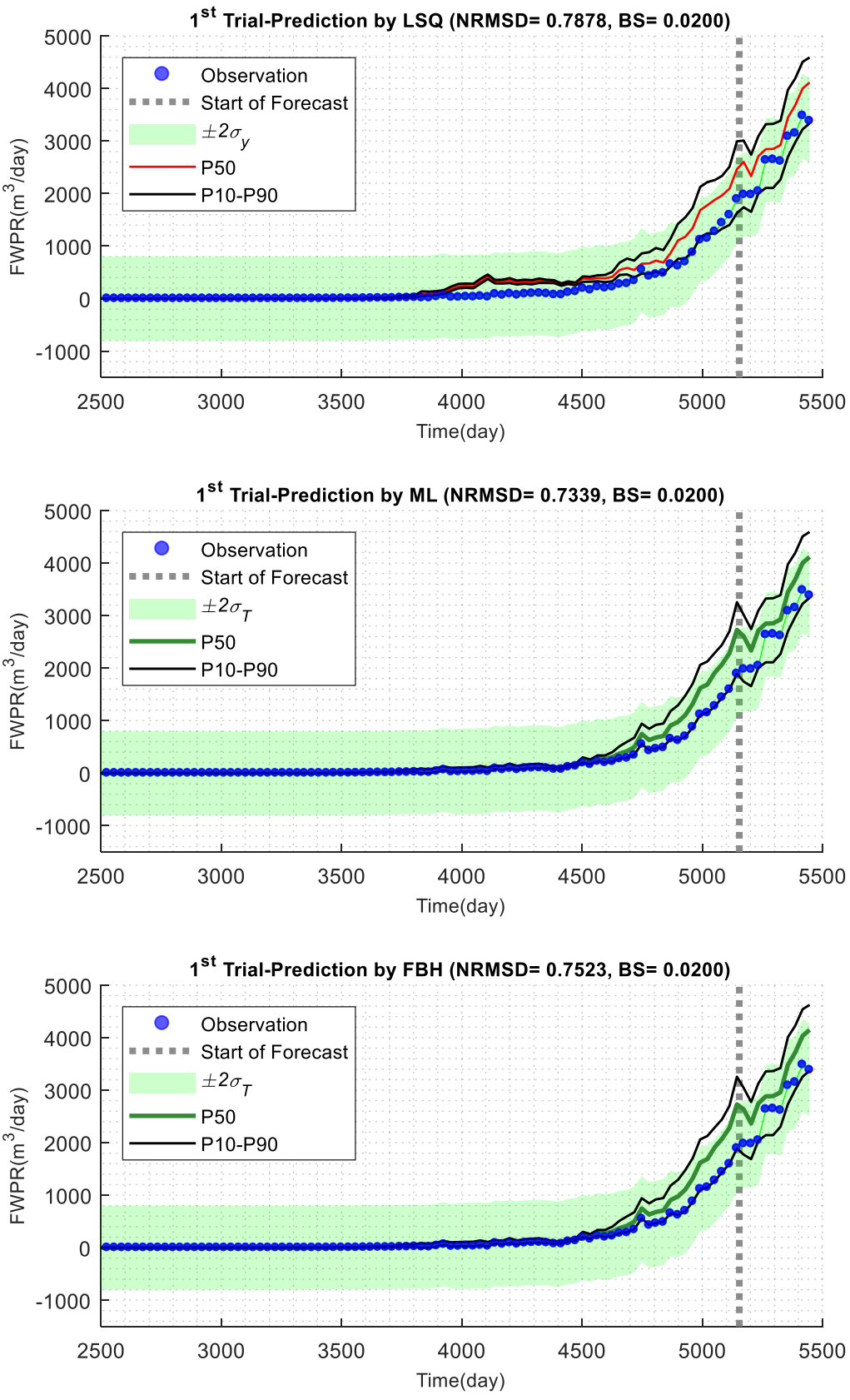

Figure 6.7 FWPR Prediction by LSQ, FBH and ML models (1st Trial)

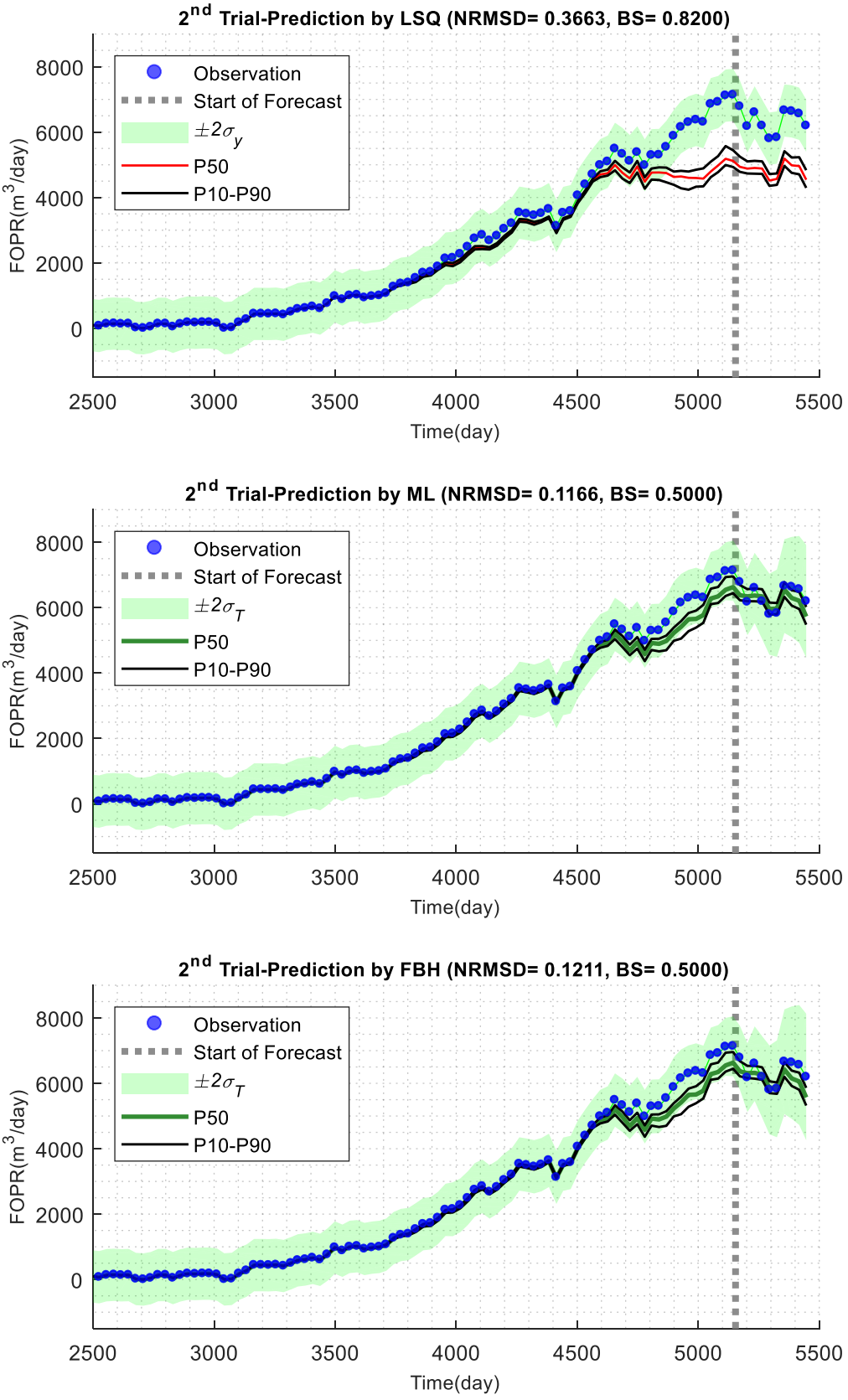

Figure 6.8 FOPR Prediction by LSQ, FBH and ML models (2nd Trial)

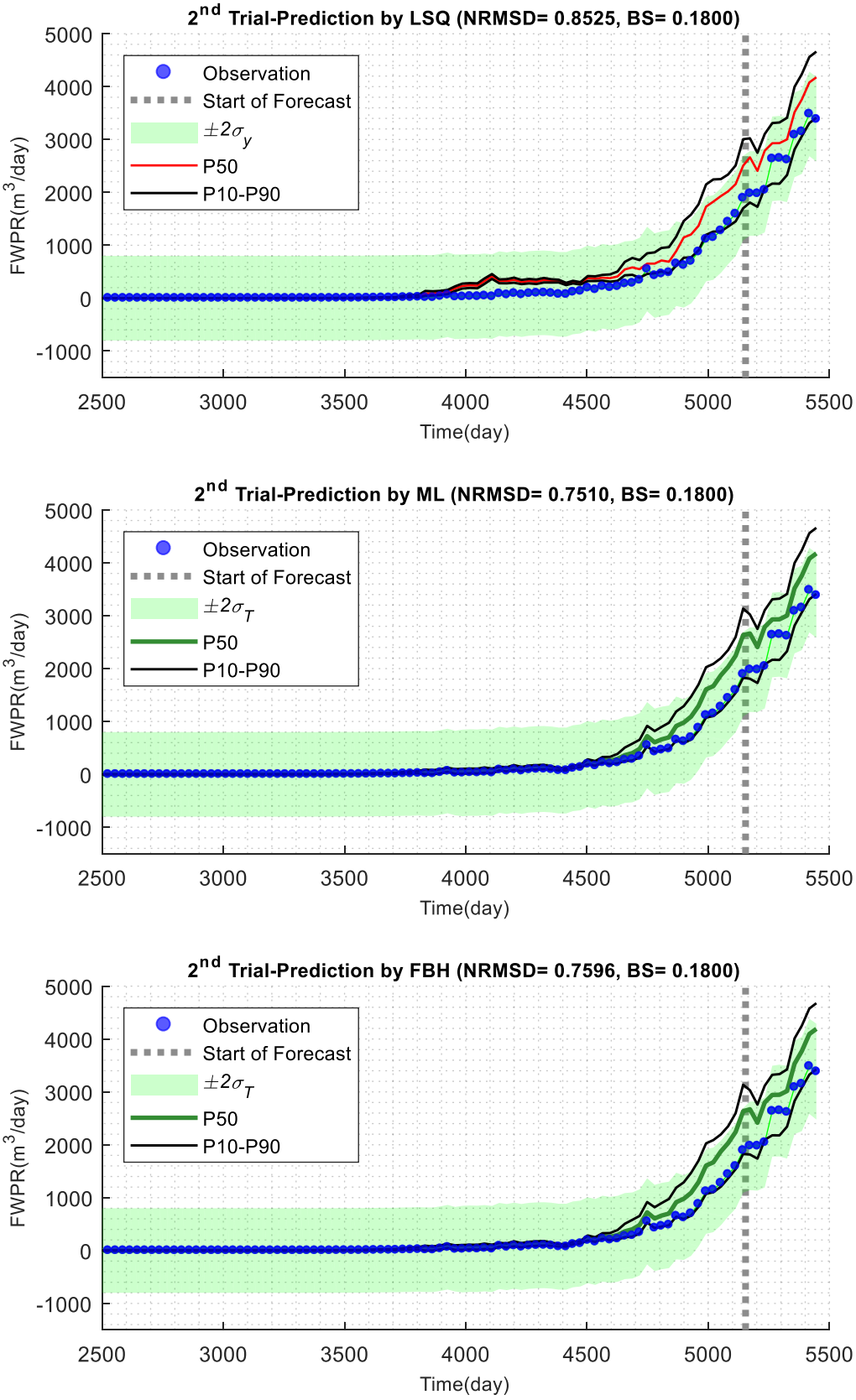

Figure 6.9 FWPR Prediction by LSQ, FBH and ML models (2nd Trial)

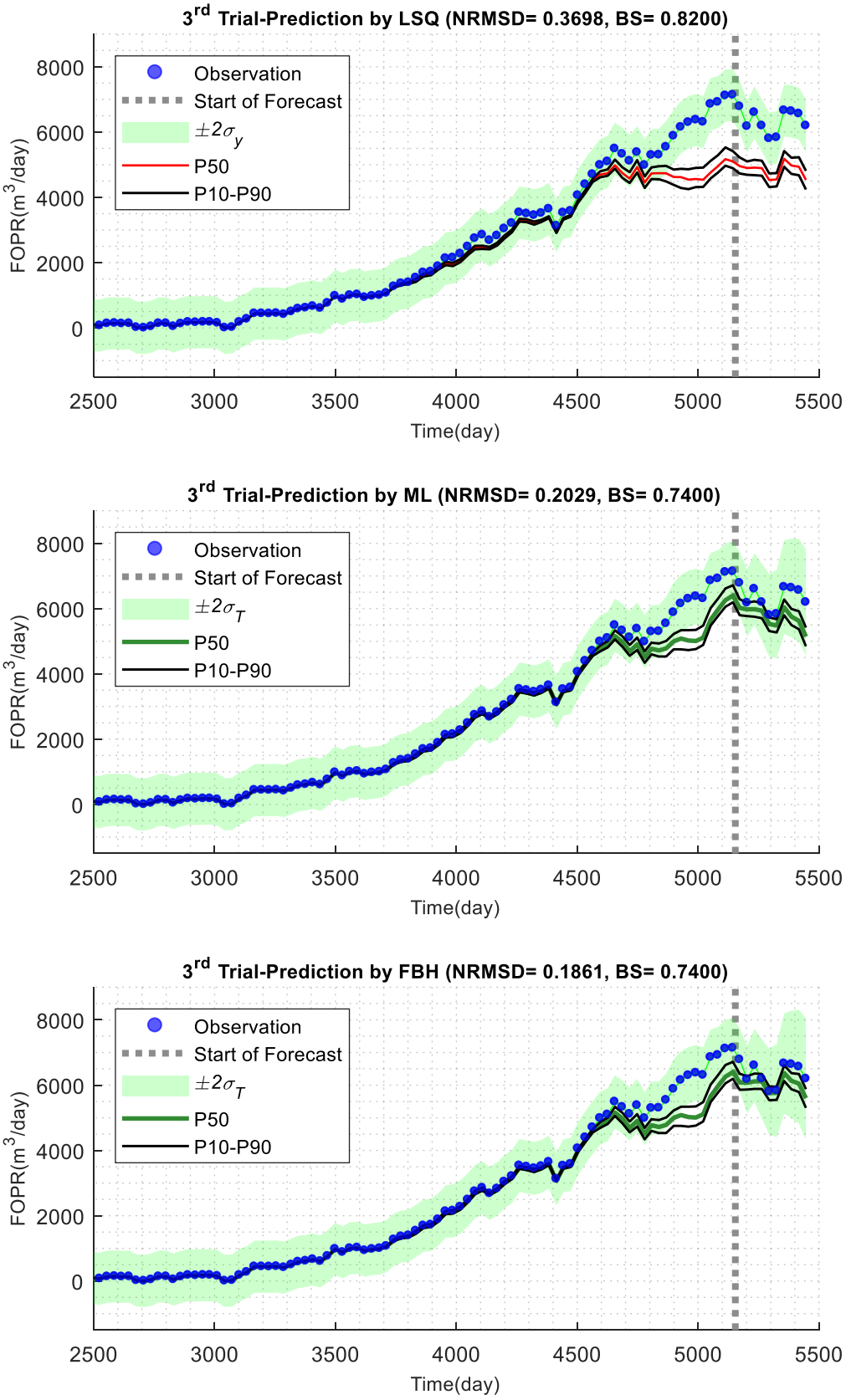

Figure 6.10 FOPR Prediction by LSQ, FBH and ML models (3rd Trial)

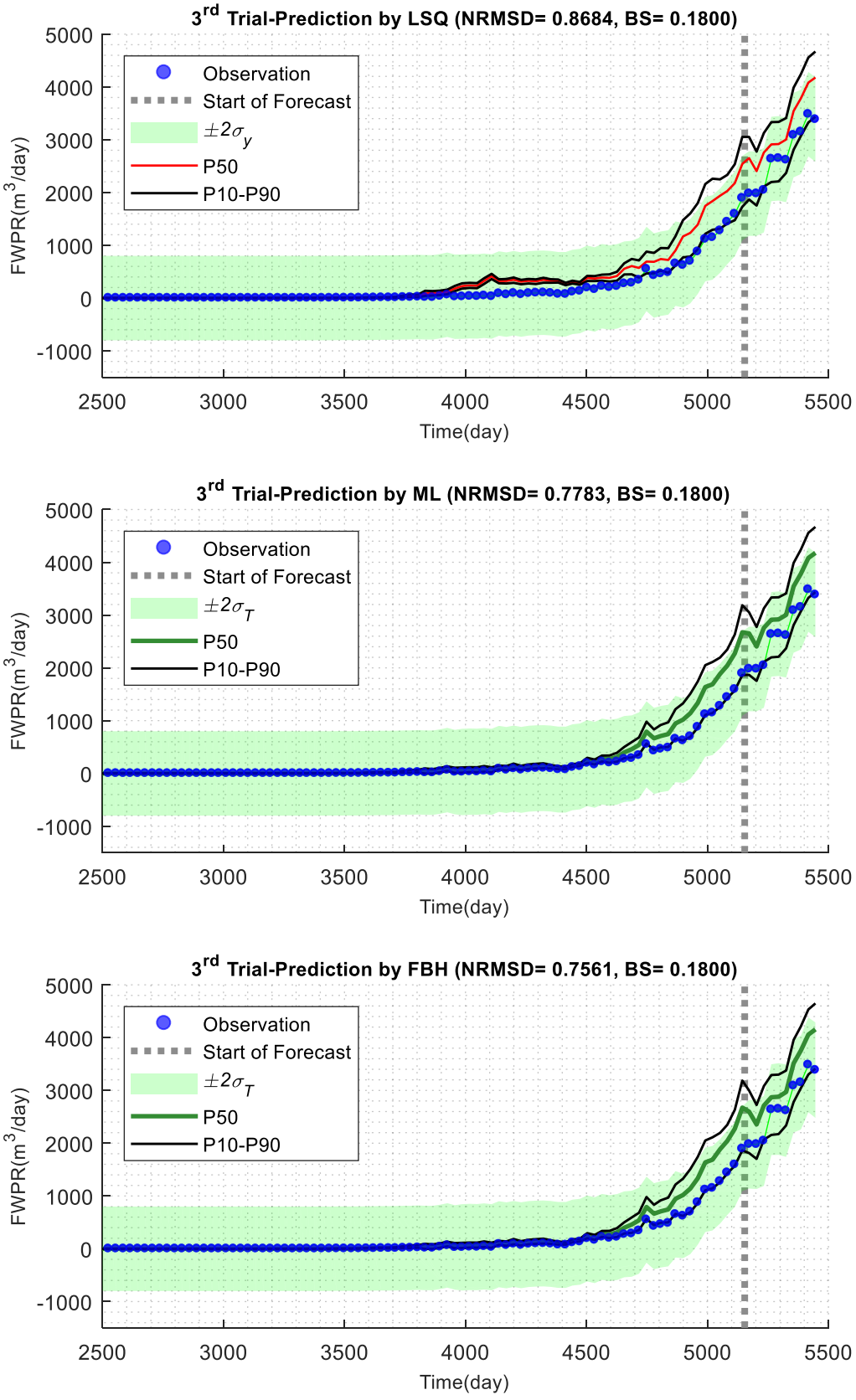

Figure 6.11 FWPR Prediction by LSQ, FBH and ML models (3rd Trial)

6.5 Comparison of posterior estimates of model parameters

We studied the outcome of modelling discrepancy in the improvement of response/field variables prediction and uncertainty quantification. Now, the question may arise as what can be the influence of ignoring/accounting for modelling discrepancy on model's parameters estimates. To address this question, we first need to define a reference model for distinguishing between the 3 scenarios mentioned above.

The reference case history matches to the whole observations (all available 180 observations are trained) and estimates the posterior distribution of model's parameters. Note that LSQ, ML and FBH scenarios were calibrated to only 169 observations. The posterior distribution of model's parameters is obtained by NAB for all cases. Figure 6.12 compares the posterior distribution of all the scenarios in the range of p10-p90 (squared lines). The results of LSQ and FBH are the average of all 3 trials of history matching whereas ML1, ML2 and ML3 are set up with different maximum likelihood point estimates.

As can be seen from the data in Figure 6.12, each of maximum likelihood solution finds diverse estimates of parameters (especially for the parameters aq2, krw and kxmult6). Therefore, despite being less biased (on average) compared to the LSQ model, the ML solutions to each parameter are less reliable when compared to the FBH solution.

Figure 6.13-a quantifies the deviation of estimated posterior means from the reference case mean in terms of Mean Absolute Percentage Deviation (MAPD, Eq. (5-1)). Among all models, the FBH model seems to have more reliable estimates since its worst estimate is 'kmult2' MAPD=101%.

From the average MAPDs (Figure 6.13-b) we can see that even though there is a slightly better estimation by all hierarchical models compared with the LSQ model (MAPD=51%), there is as yet too much deviation between hierarchical models and the reference case. The FBH, ML1, ML2 and ML3 have MAPD of approximately 34, 45, 40, and 35 per cent respectively. This also come from the fact that the most recent observations have a higher influence on the Bayesian inference of our reservoir model, and thus, the reference case, gives different results.

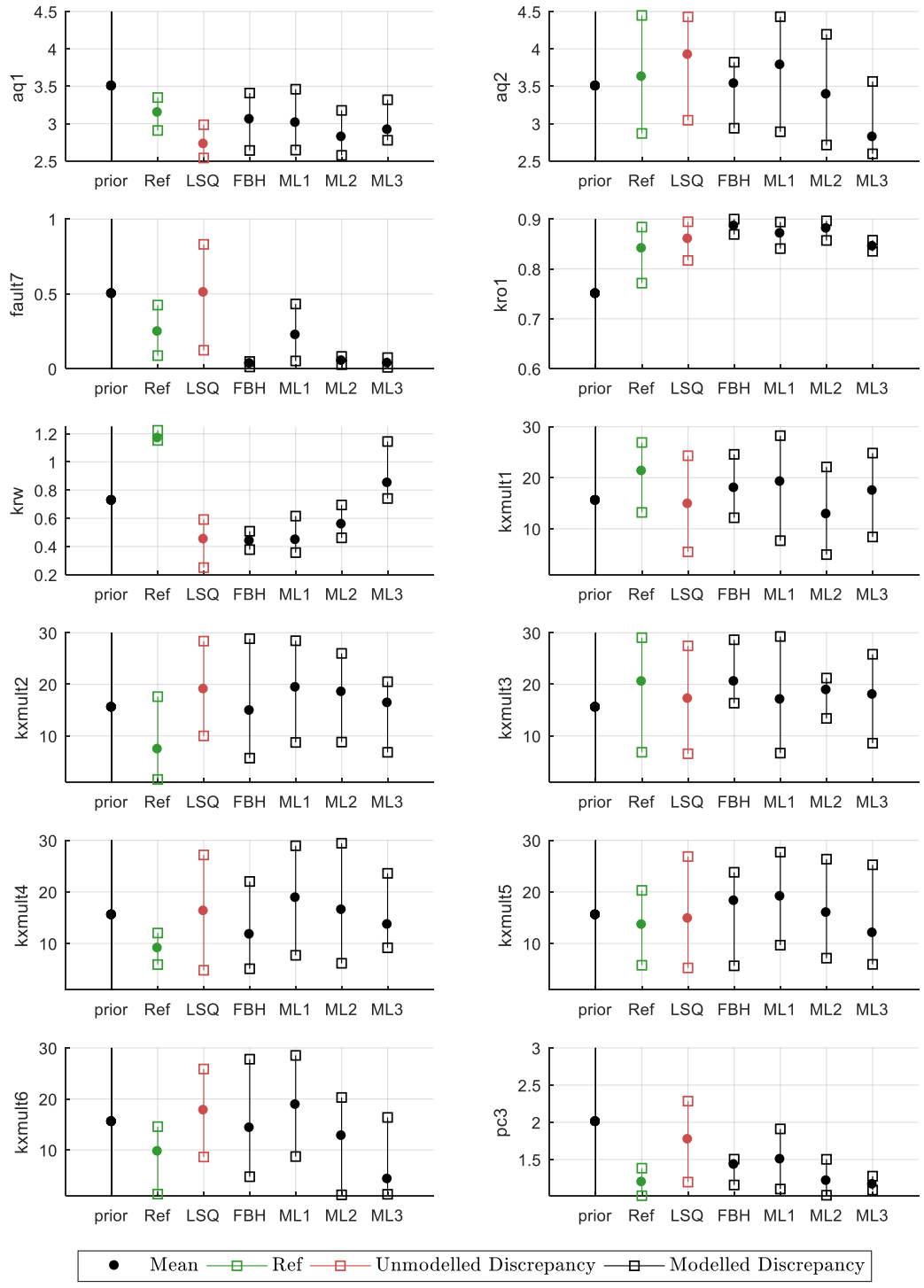

Figure 6.12 Posterior distribution of model parameters: 12 model parameters are estimated by LSQ, FBH and ML solutions, each having the same prior. The circles and squares refer to the posterior mean and P10-P90 credible interval, respectively.

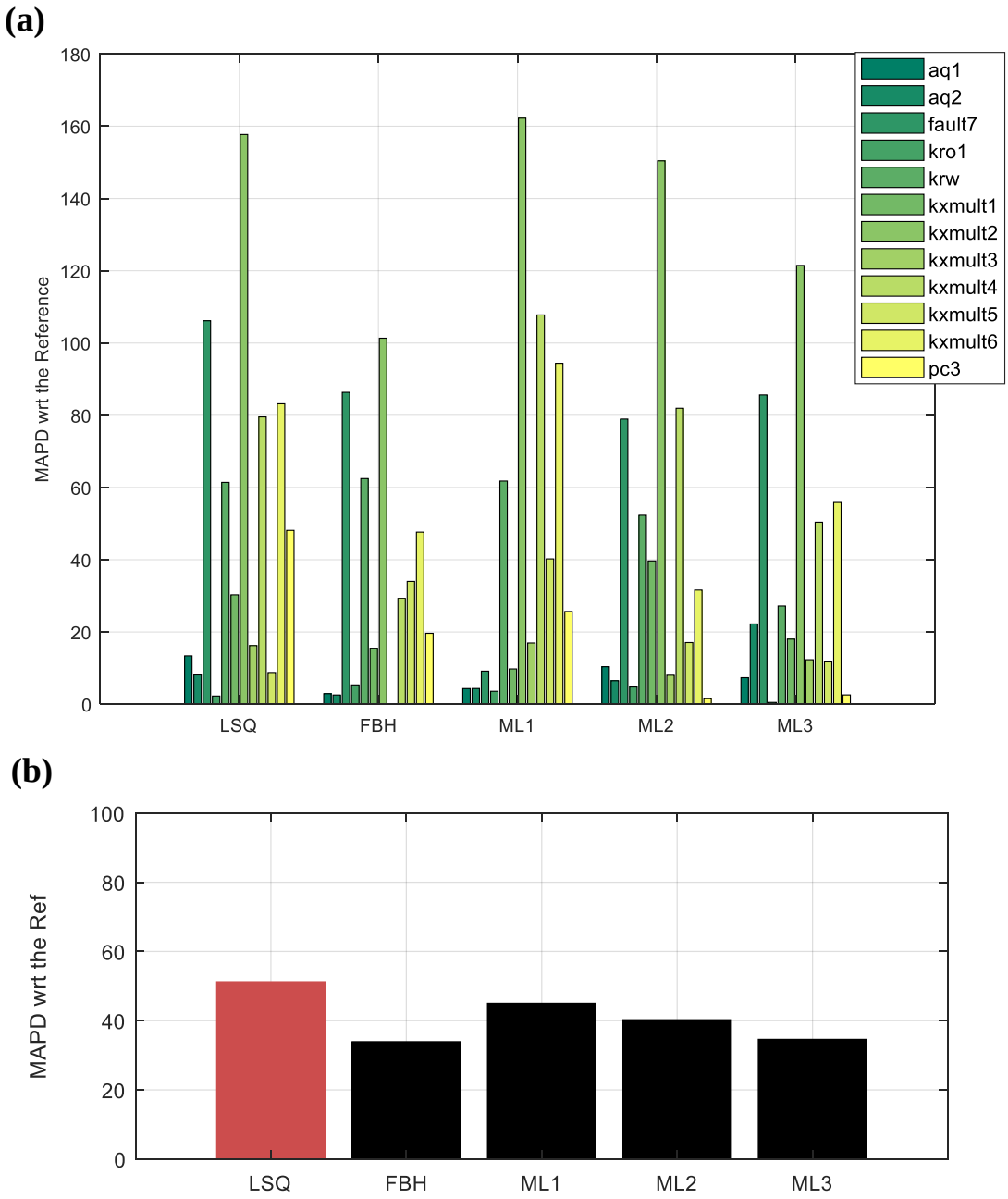

Figure 6.13 The deviation of all model predictions from the reference case calculated by Mean Absolute Percentage Deviation (MAPD) for **a)** each model parameters, and **b)** all model parameters.

Figure 6.13-a illustrates pivot between heterogeneity (macroscopic sweep) and relative permeability/wettability (microscopic sweep) wherein ML3 has optimised towards a lower permeability multipliers than ML1 or ML2.

Finally, looking at each parameter estimates individually, hierarchical models do not always reduce the bias when averaging over multiple restarts of history match in our case study. This implies that making a more accurate prediction about the future field behaviour does not always give a better estimation of reservoir model parameters. It might be that there are other essential factors influencing the parameters estimates which are not involved in the objective function. Not including pressure in the objective function may lead to optimisation of the oil and water rates in a way that may be inconsistent with the pressure and material balance.

6.6 Conclusion

This chapter aimed to explore two significant aspects of error modelling throughout the history matching of oil reservoirs: the non-parametric hierarchical modelling of model inadequacy, and the model selection problem. Because the data from our case study followed non-stationary behaviour, we implemented a segmentation technique which allows using stationary kernels. The stationary kernels should infer the correlation structures hidden in the datasets without overfitting the data. Therefore, we exerted the BIC model selection to pick the best kernels that fit the data appropriately. Note that in our study, less attention has been paid to apply different non-stationary GP kernels. This is because our central focus is on an appropriate way of error modelling rather than different solutions to non-stationarity.

As for the error modelling, we parametrised uncertainty within Kennedy O'Hagan (KOH) framework in that a new source of uncertainty, modelling discrepancy, adds up to the total uncertainty and compared it to linear least-squares LSQ model which ignores model discrepancy. Then we performed model selection to evaluate the best hierarchical models (kernels) representing modelling discrepancy. Furthermore, Maximum Likelihood (ML) and Full Bayesian (FBH) solution to the hierarchical models gave the posterior estimate of model parameters and future response of the field variables. ML and FBH solution predicted the future behaviour of the field variables with insignificant variation, and both improved the LSQ prediction.

Application of hierarchical models along with BIC model selection to the Zagadka oilfield demonstrated the importance of modelling strategies in modelling error and prediction of field variables when there exists a correlation between their residual errors.

In 3 history match trials, the model prediction with hierarchical modelling of discrepancy reduced bias caused by ignoring model discrepancy. Our results demonstrate how incorporating a proper correlation structure of errors improves the uncertainty quantification of the model for the deterministic aspect of reservoir modelling.

As far as model parameters are concerned, hierarchical models rendered less global bias concerning the reference case with the FBH being the best solution. However, looking at each parameter estimates individually, hierarchical models do not always reduce the bias when averaging over multiple restarts of history match. This implies that making a more accurate prediction about the future field behaviour does not always give a better estimation of reservoir model parameters.

Consequently, our case study confirms previous findings of chapter 5 and contributes additional evidence for enhancement of field variables prediction when the discrepancy is modelled. However, the evidence for better prediction of each of the model parameters by error modelling is inconclusive. The limitation originates from the ill-posed nature of our history match problem where improving the estimate of a parameter may result in poor estimates of another parameter.

Chapter 7 – Concluding Remarks

7.1 Summary of key findings

This thesis presents a comprehensive framework for modelling discrepancy in the Bayesian calibration and probabilistic forecasting of reservoir models. The framework efficiently implemented data-driven approaches to handle uncertainty caused by ignoring the modelling discrepancy in reservoir predictions using two major hierarchical strategies, parametric and non-parametric hierarchical models.

In parametric hierarchical modelling of discrepancy, we examined the impact of unknown standard deviations of field production variables on posterior probabilities. Therefore, we added 6 unknown standard deviations (tuning parameters for the likelihood function) at the higher level of the physical parameter of the simulation model and established a hierarchical model (FEM).

The parametric hierarchical model was tested on three experimental cases of the Teal South reservoir and compared to a standard linear least-square model (LSQ) that ignores the model discrepancy.

In the first case study, we built an idealised model (without uncertainty) by simulation of the Teal South reservoir model at the actual value of 2 model parameters. Then, LSQ and FEM history match tuned the 2 parameters until finding their actual values. Both of FEM and LSQ recovered the real value of input variables with the same speed of convergence and the same distribution.

In the second case study, we added Gaussian white noise to the idealised model data while the model configuration remained unchanged. Our findings demonstrated that LSQ is very sensitive to the changes in the standard deviation of the noise, whereas FEM is more consistent with changes to the assumed standard deviation.

In the thirds case study, we utilised the original observation data from the Teal South reservoir model with all 6 model parameters. In addition to model parameters, FEM made use of 6 unknown time-varying standard deviations for the likelihood function which tracked down the uncertainty in field variables predictions. Regardless of the type of the sampling method, the FEM reached a better model prediction score compared to the LSQ models.

Although our parametric hierarchical model was able to predict the trained data, it failed to generalise since the parameters are only defined for particular instances of the time domain. Therefore, if the objective was to find the error model at any instance, we should deploy non-parametric emulators to generalise towards the entire input space.

In contrast to parametric methods, nonparametric methods do not assume any parametric form of the function other than certain smoothness assumption. This enables our non-parametric Gaussian process models to generalise towards the forecast which is of high importance for reservoir engineering problems (Nezhad Karim Nobakht et al., (2018)).

In advance of building our non-parametric model to capture discrepancy, we exerted Kennedy and O'Hagan, (2001) notation (KOH) for modelling discrepancy within the Bayesian framework. According to KOH, the simulator response, albeit at the true input value, can never trace the true response of a physical process. However, in a reservoir model, the true value of the parameters can never be learned with certainty. Hence, following KOH modelling error strategy, we assumed that there exists a true $\hat{\theta}$ estimated by a best-fitting model with the highest marginal likelihood.

To retain the probabilistic treatment, we must use a non-parametric regression model that enables us to choose from a large class of functions. In this context, we used Gaussian processes (GPs) models as non-parametric emulators to build our statistical error models.

The major focus of our non-parametric error model strategy was to avoid overfitting while finding the best solution to the correlation structure of the GP models. Therefore, we made use of the BIC model selection code to obtain the best solution to the form of GP kernels which were then used to predict the error for the unseen future data. This allowed us to select those models that best fit the data without overfitting. Furthermore, Maximum Likelihood (ML) and Full Bayesian (FBH) solution to the hierarchical models gave the posterior estimate of model parameters and future response of the field variables.

Application of hierarchical models along with BIC model selection to the real case study, the Zagadka oilfield, demonstrated the importance of modelling strategies in modelling discrepancy and prediction of field variables when there exists a correlation between their residual errors. ML and FBH solution predicted the future behaviour of the field variables with insignificant variation, and both improved the LSQ prediction. Overall, the model prediction with the hierarchical modelling of discrepancy reduced bias caused by ignoring model discrepancy.

As for the model parameters, hierarchical models rendered less global bias with respect to the reference case with the FBH being the best solution. However, looking at each parameter estimates individually, hierarchical models do not always reduce the bias when averaging over multiple restarts of history match. This implies that making a more accurate prediction about the future field behaviour does not always give a better estimation of reservoir model parameters.

Consequently, all our case studies confirmed the enhancement of field variables prediction when the discrepancy is modelled. However, in the considered case studies, the evidence for better prediction of each of the model parameters by error modelling was inconclusive.

It is highly questionable that statistical methods, no matter how sophisticated, can predict events or features which are not represented in the reservoir model nor for which there is information in the observed data.

The underlying issue concerns the extent of the predictive domain discussed in the introduction of the thesis. We assumed that there is some form of continuity of the represented drainage volumes and displacement processes, e.g. forecasting under a continuation of current well production and displacement/depletion process within the predictive domain. In practice, this work applies to short-term forecasting, for which a quick turnaround time is required as conditions change (e.g. well availability), and there is no time to rematch models to the most recent data or run other realisations.

7.2 Recommendation for future work

The hypothesis of data stationarity has been the most common framework in time series analysis. However, in the real world processes, the relationship between a contemporary response variable and its past may follow a non-stationary stochastic process. On the other hand, modelling a non-stationary process with stationary techniques to capture the correlation structure of data may lead to an unjustified approximation. Consequently, prediction of a non-stationary class of data using a stationary model is high-risk (Korkas and Fryzlewicz, 2017) as it is unable to realistically quantify main sources of non-stationarity (i.e. changes in mean, trend and standard deviation).

The central focus of this thesis is on an appropriate way of error modelling and the importance of model selection in controlling overfitting rather than different solutions to non-stationarity and different noise models. Therefore, our framework has paid less

attention to the application of different non-stationary emulators and different noise models in error modelling.

Figure 7.1 depicts a stationary Gaussian process model emulating a non-stationary time series error y where the estimated total uncertainty $\sigma = \sigma_\delta + \sigma_\varepsilon$ includes both measurement uncertainty σ_ε and model discrepancy σ_δ. The noise model is assumed unknown Gaussian white noise $N(0, \sigma_\varepsilon)$ and needs to be learned from data.

The predicted uncertainty in Figure 7.1-a seems to have overestimated the total uncertainty σ at the early time steps and underestimated the total uncertainty at the last time steps. This implies that the calibration will fail to penalise the discrepancy at late time steps (x=4800,…, 5600) with higher intensity followed by an overconfident forecast.

The statistical properties of the stationary error model shown in Figure 7.1-b remain constant throughout the entire time domain wherein the Gaussian white noise model assumption may not be realistic.

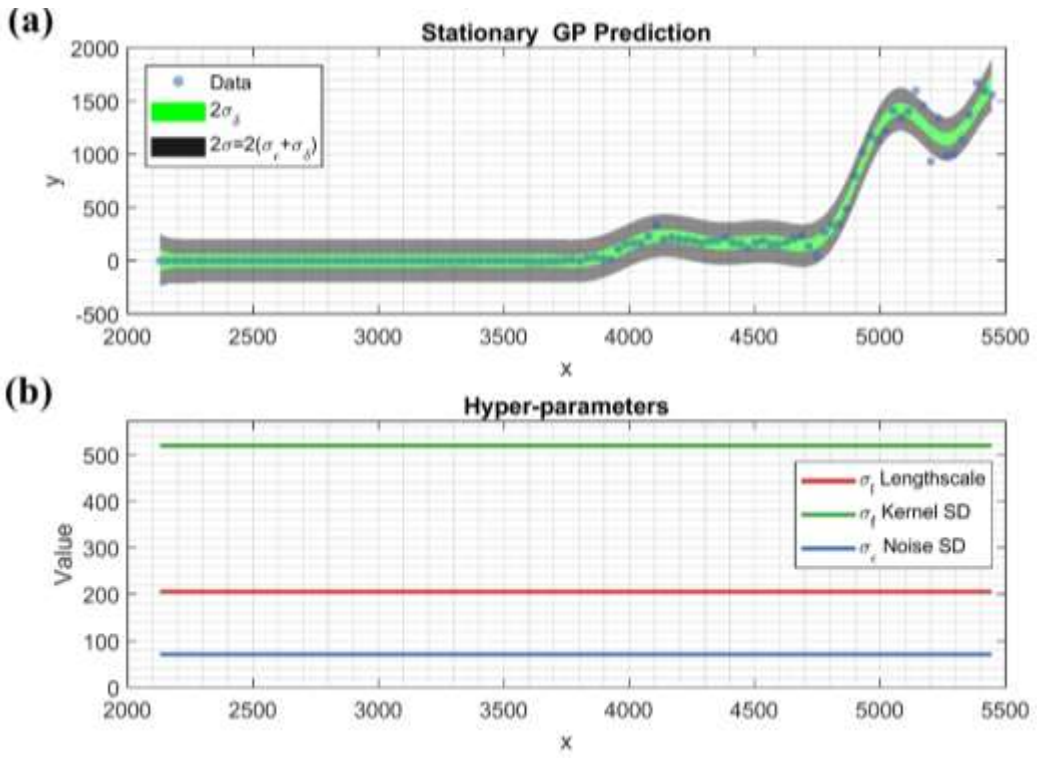

Figure 7.1 **a)** Error model constructed by stationary Gaussian Process model Gaussian white noise model. **b)** The error model hyper-parameters are constant all over the time domain

Several attempts have been made to construct heteroscedastic noise models (Quadrianto et al., 2009; Kersting et al., 2007; Goldberg et al., 1998; Heinonen et al., 2016). Heinonen

et al. (2016)'s approach is the first to model the heteroscedastic noise and non-stationary length scale while allowing the kernel standard deviation to vary over time.

We suggest a non-stationary form of squared exponential kernel models presented by Heinonen et al. (2016) whose parameters are time-varying (input-dependent). Heinonen et al. (2016) models the length scale σ_{l}, kernel standard deviation σ_{f} and the noise standard deviation σ_{ε} with latent functions by placing separate GP priors on them as well:

$$\sigma_{\mathrm{l}} \sim GP(m_l(x), k_l(x, x')) \tag{7-1}$$

$$\sigma_{\mathrm{f}} \sim GP(m_{\mathrm{f}}(x), k_{\mathrm{f}}(x, x')) \tag{7-2}$$

$$\sigma_{\varepsilon} \sim GP(m_{\varepsilon}(x), k_{\varepsilon}(x, x')) \tag{7-3}$$

Now, if we assume an unknown but constant noise model and place GP priors only on the kernel length scale and the kernel sigma (Eq. (7-1) and Eq. (7-2)), we arrive at LSGP model that is still incapable of reliably predicting the uncertainty (Figure 7.2-a,b). The predicted latent variables σ_{l} and σ_{f} shown in Figure 7.2-b vary in time with the kernel sigma σ_{f} being very high at the late time steps. However, since the noise sigma σ_{ε} is very small all over the input, the total uncertainty predicted by the LSGP model remains small even at the late time steps.

Now, if we also posit priors on noise (Eq. (7-3)) and allow all the hyper-parameters σ_{l}, σ_{f} and σ_{ε} vary in time, we can build a fully non-stationary Gaussian process model (LSOGP model in Figure 7.2-c, d) where all three hyper-parameters can be input-dependent.

As can be seen from Figure 7.2-c, LSOP model can reliably make predictions without underestimating/overestimating the total uncertainty. In Figure 7.2-d, the latent function $\sigma_{\varepsilon} \sim GP(m_{\varepsilon}(x), k_{\varepsilon}(x, x'))$ seems to have controlled the small and large fluctuations of time-series errors at early stages and late stages of calibration, respectively.

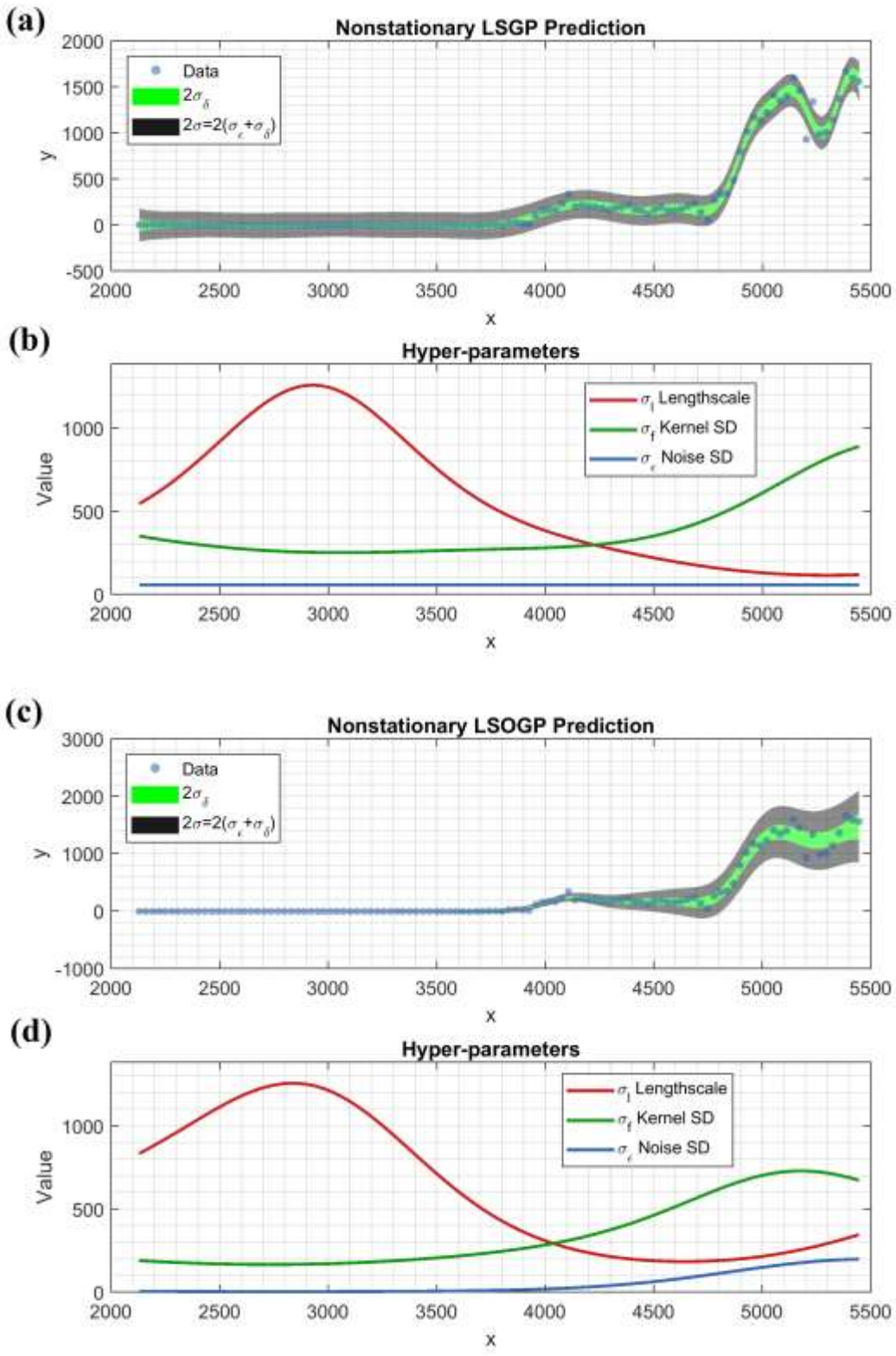

Figure 7.2 Non-stationary GP models: a) LSGP model assumes a constant Gaussian noise model σ_ε, and latent σ_l, σ_f. b) The maximum likelihood estimates of LSGP hyper-parameters. c) LSOGP model assume latent σ_ε, σ_l and σ_f. d) The maximum likelihood estimates of LSOGP hyper-parameters

Further extensions would be to go for the model selection of the best error models among varieties of non-parametric models with a high penalty on model complexity. The resulting error models integrate it into the Bayesian approach for inferring hyper-parameters. However, any assumption about the noise depends on the nature of our physical parameters and must be used with careful consideration.

To overcome difficulties associated with high dimensional non-stationary spatially-referenced data, (Risser and Calder, 2015) introduced a nonstationary covariance function by building a Bayesian model for continuously-indexed spatial data using process convolution techniques. Their resulting model provided a practical compromise between stationary and highly parameterised nonstationary.

An alternative emulator for GPs can be non-parametric generalised additive models (GAM) developed by Hastie and Tibshirani, 1987. The GAM models do not need the assumption of linear relationships between independent variables (input parameters) and the response variables (outputs). Instead, they provide a non-parametric estimation of model response and determine the relationship between the continuous predictor and the outcome (Hin et al., 1999, West, 2012). GAM is a more flexible extension of the Generalised Linear Model (GLM) where the modelling of the mean functions relaxes the assumption of linearity (Barrio et al., 2013).

Appendix A: Kernel Functions

Table A- 1 Mathematical description of several stationary kernel functions

Kernel Function Name	***Kernel Formulae***
Exponential	$\sigma_f{}^2\, exp\left(\frac{-\lvert x-x'\rvert)}{\sigma_l}\right)$
Squared Exponential	$\sigma_f{}^2\, exp\left(\frac{-(x-x')^2}{2\,\sigma_l{}^2}\right)$
Mattern 3/2	$\sigma_f{}^2\left(1+\frac{\sqrt{3}\lvert x-x'\rvert}{\sigma_l}\right)exp\left(\frac{-\sqrt{3}\lvert x-x'\rvert}{\sigma_l}\right)$
Mattern 5/2	$\sigma_f{}^2\left(1+\frac{\sqrt{5}\lvert x-x'\rvert}{\sigma_l}+\frac{5(x-x')^2}{3\,\sigma_l{}^2}\right)exp\left(\frac{-\sqrt{5}\lvert x-x'\rvert}{\sigma_l}\right)$
Rational Quadratic	$\sigma_f{}^2\left(1+\frac{(x-x')^2}{2\,\alpha\sigma_l{}^2}\right)^{-\alpha}$, $\alpha: scale\ mixture\ paramater$
ARD Exponential	$\sigma_f{}^2\, exp\left(-\sqrt{\sum_{m=1}^{d}\frac{(x_m-x'_m)^2}{\sigma_m{}^2}}\right)$, $\sigma_m: length\ scale\ for\ predictor\ m$
ARD Squared Exponential	$\sigma_f{}^2\, exp\left(-\frac{1}{2}\sum_{m=1}^{d}\frac{(x_m-x'_m)^2}{\sigma_m{}^2}\right)$
ARD Mattern 3/2	$\sigma_f{}^2\left(1+\sqrt{3}\sqrt{\sum_{m=1}^{d}\frac{(x_m-x'_m)^2}{\sigma_m{}^2}}\right)exp\left(-\sqrt{3}\sqrt{\sum_{m=1}^{d}\frac{(x_m-x'_m)^2}{\sigma_m{}^2}}\right)$
ARD Mattern 5/2	$\sigma_f{}^2\left(1+\sqrt{5}\sqrt{\sum_{m=1}^{d}\frac{(x_m-x'_m)^2}{\sigma_m{}^2}}+\frac{5}{3}\sum_{m=1}^{d}\frac{(x_m-x'_m)^2}{\sigma_m{}^2}\right)exp\left(-\sqrt{5}\sqrt{\sum_{m=1}^{d}\frac{(x_m-x'_m)^2}{\sigma_m{}^2}}\right)$
ARD Rational Quadratic	$\sigma_f{}^2\left(1+\frac{1}{2\alpha}\sum_{m=1}^{d}\frac{(x_m-x'_m)^2}{\sigma_m{}^2}\right)^{-\alpha}$

References

[1] Aanonsen, S. I., Nævdal, G., Oliver, D. S., Reynolds, A. C. & Vallès, B. 2009. The ensemble Kalman filter in reservoir engineering--a review. *Spe Journal,* 14**,** 393-412.

[2] Aarnes, J. E., Hauge, V. L. & Efendiev, Y. 2007. Coarsening of three-dimensional structured and unstructured grids for subsurface flow. *Advances in Water Resources,* 30**,** 2177-2193.

[3] Abdollahzadeh, A., Reynolds, A., Christie, M., Corne, D. W., Davies, B. J. & Williams, G. J. 2012. Bayesian optimization algorithm applied to uncertainty quantification. *SPE Journal,* 17**,** 865-873.

[4] Abdollahzadeh, A., Reynolds, A., Christie, M., Corne, D. W., Williams, G. J. & Davies, B. J. 2013. Estimation of distribution algorithms applied to history matching. *SPE Journal,* 18**,** 508-517.

[5] Adams, D. C. & Markus, D. 2013. Systematic error in instantaneous attributes. *SEG Technical Program Expanded Abstracts 2013.* Society of Exploration Geophysicists.

[6] Ahmadi, M. 2012. *Modelling and quantification of structural uncertainties in petroleum reservoirs assisted by a hybrid cartesian cut cell/enriched multipoint flux approximation approach, PhD thesis.* Heriot-Watt University.

[7] Ahmadi, M., Christie, M. & Gerritsen, M. Structural uncertainty quantification with immersed interface methods. SPE Reservoir Simulation Symposium, 2013. Society of Petroleum Engineers.

[8] Akaike, H. 1974. A new look at the statistical model identification. *IEEE transactions on automatic control,* 19**,** 716-723.

[9] Al-Busafi, B., Fisher, Q., Harris, S. & Kendall, M. The Impact of Faults Representation on History Match and Future Generated Seismic Impedance

Response in Reservoir Models–Case Study for Pierce Field, North Sea. 67th EAGE Conference & Exhibition, SPE-93429-MS, 2005.

[10] Al-Yahya, S. A. 2010. Reservoir Monitoring and Performance Using Simbest II Black Oil Simulator Middle East Reservoir Case Study. *Brazil Oil & Gas, tt_nrg and Norway Oil & Gas***,** 27.

[11] Arendt, P. D., Apley, D. W. & Chen, W. 2012. Quantification of model uncertainty: Calibration, model discrepancy, and identifiability. *Journal of Mechanical Design,* 134**,** 100908.

[12] Arnold, D., Demyanov, V., Christie, M., Bakay, A. & Gopa, K. 2016. Optimisation of decision making under uncertainty throughout field lifetime: A fractured reservoir example. *Computers & Geosciences,* 95**,** 123-139.

[13] Asadollahi, M. & Naevdal, G. Waterflooding optimization using gradient based methods. SPE/EAGE Reservoir Characterization & Simulation Conference, SPE-125331-MS, 2009.

[14] Bajkowski, B., Cyrankowski, M. & Osipiuk, J. 2013. Types and causes of errors in automatic measurement of roundwood. *Annals of Warsaw University of Life Sciences-SGGW. Forestry and Wood Technology,* 82.

[15] Baker, R., Regier, C. & Sinclair, R. PVT error analysis for material balance calculations. Canadian International Petroleum Conference, 2003. Petroleum Society of Canada.

[16] Barrio, I., Arostegui, I. & Quintana, J. M. 2013. Use of generalised additive models to categorise continuous variables in clinical prediction. *BMC medical research methodology,* 13**,** 83.

[17] Bazargan, H., Christie, M. & Tchelepi, H. Efficient Markov chain Monte Carlo sampling using polynomial chaos expansion. SPE Reservoir Simulation Symposium, 2013. Society of Petroleum Engineers.

[18] Bergosh, J., Marks, T. & Mitkus, A. New core analysis techniques for naturally fractured reservoirs. SPE California Regional Meeting, 1985. Society of Petroleum Engineers.

[19] Bertolini, A. C. & Schiozer, D. J. 2011. Influence of the objective function in the history matching process. *Journal of Petroleum Science and Engineering,* 78**,** 32-41.

[20] Betz, J. 2015. Data Integration Enables Quicker Decisions. *Journal of Petroleum Technology,* 67**,** 76-77.

[21] Bhark, E. W. & Dehghani, K. Assisted history matching benchmarking: Design of experiments-based techniques. SPE Annual Technical Conference and Exhibition, SPE-0415-0142-JPT, 2014. Society of Petroleum Engineers.

[22] Bhushan, V., Lee, R. & Mukerji, P. Mature Field Production Optimisation Through Standardisation of Operating Procedures for Reservoir Monitoring. Offshore Europe, 2009. Society of Petroleum Engineers.

[23] Bissell, R., Killough, J. & Sharma, Y. Reservoir history matching using the method of gradients on a workstation. European Petroleum Computer Conference, 1992. Society of Petroleum Engineers.

[24] Bouska, J., Cooper, M., O'donovan, A., Corbett, D., Malinverno, A., Prange, M. & Ryan, S. 1999. Validating Reservoir Models to Improve Recovery. *Oilfield Review,* 11**,** 20-35.

[25] Bouzarkouna, Z. & Nobakht, B. A Better Formulation of Objective Functions for History Matching Using Hausdorff Distances. EUROPEC 2015, 2015. Society of Petroleum Engineers.

[26] Bouzarkouna, Z., Verdiere, S., Jaulneau, P., Le Reun, J. & Corpel, V. How to Improve Efficiency in Multiple History Matching: A Gas Field Case Study. International Petroleum Technology Conference, 2014. International Petroleum Technology Conference.

[27] Bp 2017. BP energy outlook energy 2017. *BP Stat. Rev. World Energy.*

[28] Brier, G. W. 1950. VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY. *Monthly Weather Review,* 78**,** 1-3.

[29] Brynjarsdóttir, J. & O'hagan, A. 2014. Learning about physical parameters: The importance of model discrepancy. *Inverse Problems,* 30**,** 114007.

[30] Bu, T. & Damsleth, E. 1996. Errors and uncertainties in reservoir performance predictions. *SPE Formation Evaluation,* 11**,** 194-200.

[31] Bukkapatnam, S. T. & Cheng, C. 2010. Forecasting the evolution of nonlinear and nonstationary systems using recurrence-based local Gaussian process models. *Physical Review E,* 82**,** 056206.

[32] Bybee, K. 2008. Transforming Production Data Into Knowledge. *Journal of Petroleum Technology,* 60**,** 68-69.

[33] Cancelliere, M., Verga, F. & Viberti, D. Benefits and limitations of assisted history matching. Offshore Europe, 2011. Society of Petroleum Engineers.

[34] Cao, R. & Van Keilegom, I. 2006. Empirical likelihood tests for two-sample problems via nonparametric density estimation. *Canadian Journal of Statistics,* 34**,** 61-77.

[35] Cardone, V., Broccoli, A., Greenwood, C. & Greenwood, J. 1980. Error characteristics of extratropical-storm wind fields specified from historical data. *Journal of Petroleum Technology,* 32**,** 872-880.

[36] Casella, G., Robert, C. P. & Wells, M. T. 2004. Generalized accept-reject sampling schemes. *A Festschrift for Herman Rubin.* Institute of Mathematical Statistics.

[37] Castellini, A., Vahedi, A., Singh, U., Sawiris, R. S. & Roach, T. Reconciling History Matching and Assessment of Uncertainty in Production Forecasts: A Study Combining Experimental Design, Proxy Models and Genetic Algorithms. International Petroleum Technology Conference, 2008. International Petroleum Technology Conference.

[38] Chen, W. H., Gavalas, G. R., Seinfeld, J. H. & Wasserman, M. L. 1974. A new algorithm for automatic history matching. *Society of Petroleum Engineers Journal,* 14**,** 593-608.

[39] Choi, H., Ohmori, S., Yoshimoto, K. & Ohtake, H. Improvement of particle swarm optimization: Application of the mutation concept for the escape from local minima. Supply Chain Management and Information Systems (SCMIS), 2010 8th International Conference on, 2010. IEEE, 1-5.

[40] Christie, M. & Clifford, P. 1998. Fast procedure for upscaling compositional simulation. *SPE Journal,* 3**,** 272-278.

[41] Christie, M., Demyanov, V. & Erbas, D. 2006. Uncertainty quantification for porous media flows. *Journal of Computational Physics,* 217**,** 143-158.

[42] Christie, M., Eydinov, D., Demyanov, V., Talbot, J., Arnold, D. & Shelkov, V. Use of multi-objective algorithms in history matching of a real field. SPE reservoir simulation symposium, 2013. Society of Petroleum Engineers.

[43] Christie, M., Macbeth, C. & Subbey, S. 2002. Multiple history-matched models for Teal South. *The Leading Edge,* 21**,** 286-289.

[44] Christie, M. A., Glimm, J., Grove, J. W., Higdon, D. M., Sharp, D. H. & Wood-Schultz, M. M. 2005. Error analysis and simulations of complex phenomena. *Los Alamos Science,* 29.

[45] Cosentino, L. 2001. *Integrated reservoir studies*, Editions Technip.

[46] Cunha, L. Integrating static and dynamic data for oil and gas reservoir modelling. Canadian International Petroleum Conference, 2003. Petroleum Society of Canada.

[47] Curran, J. M. 2005. An introduction to Bayesian credible intervals for sampling error in DNA profiles. *Law, Probability and Risk,* 4**,** 115-126.

[48] Datta-Gupta, A. & King, M. J. 2007. *Streamline simulation: theory and practice*, Society of Petroleum Engineers Richardson.

[49] Davis, R. A., Lee, T. C. M. & Rodriguez-Yam, G. A. 2006. Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association,* 101**,** 223-239.

[50] Del Giudice, D., Reichert, P., Bareš, V., Albert, C. & Rieckermann, J. 2015. Model bias and complexity–Understanding the effects of structural deficits and input errors on runoff predictions. *Environmental modelling & software,* 64**,** 205-214.

[51] Dennis Jr, J. E. & Schnabel, R. B. 1996. *Numerical methods for unconstrained optimization and nonlinear equations*, Siam.

[52] Deutsch, C. 2000. Integration of geological, geophysical, and historical production data in geostatistical reservoir modeling: Canadian Society of Exploration Geophysicists Annual Meeting Abstracts, 5 p.

[53] Domingos, P. 1999. The role of Occam's razor in knowledge discovery. *Data mining and knowledge discovery,* 3**,** 409-425.

[54] Doublet, L., Nevans, J., Fisher, M., Heine, R. & Blasingame, T. Pressure transient data acquisition and analysis using real time electromagnetic telemetry. Permian Basin Oil and Gas Recovery Conference, 1996. Society of Petroleum Engineers.

[55] Dupin, M., Reynaud, P., Jarošík, V., Baker, R., Brunel, S., Eyre, D., Pergl, J. & Makowski, D. 2011. Effects of the training dataset characteristics on the performance of nine species distribution models: application to Diabrotica virgifera virgifera. *PLoS One,* 6**,** e20957.

[56] Eberhart, R. & Kennedy, J. A new optimizer using particle swarm theory. Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on, 1995. IEEE, 39-43.

[57] Ehigie, S. NMR-Openhole Log Integration: Making the Most of NMR Data Deliverables. Nigeria Annual International Conference and Exhibition, 2010. Society of Petroleum Engineers.

[58] Eia, U. 2016. The International Energy Outlook 2016 (IEO2016).

[59] Elsheikh, A. H., Demyanov, V., Tavakoli, R., Christie, M. A. & Wheeler, M. F. 2015. Calibration of channelized subsurface flow models using nested sampling and soft probabilities. *Advances in water resources,* 75**,** 14-30.

[60] Erbas, D. 2007. *Sampling strategies for uncertainty quantification in oil recovery prediction, , PhD thesis.* Heriot-Watt University.

[61] Erbas, D. & Christie, M. A. Effect of sampling strategies on prediction uncertainty estimation. SPE Reservoir Simulation Symposium, 2007. Society of Petroleum Engineers.

[62] Estublier, A., Bachaud, P., Michel, A., Maurand, N. & Deflandre, J.-P. 2014. Long-term fate of CO2 in a saline aquifer: modeling issues. *Energy Procedia*, 63**,** 3464-3474.

[63] Evensen, G., Hove, J., Meisingset, H., Reiso, E., Seim, K. S. & Espelid, Ø. Using the EnKF for assisted history matching of a North Sea reservoir model. SPE reservoir simulation symposium, 2007. Society of Petroleum Engineers.

[64] Exxonmobil 2004. ExxonMobil Corporation. Energy outlook to 2030, technical report from www.exxonmobil.com/energyoutlook. 2004.

[65] Gao, G., Li, G. & Reynolds, A. C. A stochastic optimization algorithm for automatic history matching. SPE Annual Technical Conference and Exhibition, SPE-90065-MS, 2004. Society of Petroleum Engineers.

[66] Gao, G., Vink, J. C., Chen, C., Alpak, F. O. & Du, K. 2016. A parallelized and hybrid data-integration algorithm for history matching of geologically complex reservoirs. *SPE Journal,* 21**,** 2,155-2,174.

[67] Garb, F. A. 1988. Assessing Risk in Estimating Hydrocarbon Reserves and in Evaluating Hydrocarbon-Producing Properties (includes associated papers 18606 and 18610). *Journal of Petroleum Technology*, 40**,** 765-778.

[68] Garbuno-Inigo, A., Diazdelao, F. & Zuev, K. 2016. Gaussian process hyper-parameter estimation using parallel asymptotically independent Markov sampling. *Computational Statistics & Data Analysis,* 103**,** 367-383.

[69] Gearhart, M., Ziemer, K. A. & Knight, O. M. 1981. Mud pulse MWD systems report. *Journal of Petroleum Technology,* 33**,** 2,301-2,306.

[70] Geman, S. & Geman, D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. Pattern Analysis and Machine Intelligence, 6, 721-741.

[71] Gershman, S. J. & Blei, D. M. 2012. A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology,* 56**,** 1-12.

[72] Gill, P. E., Murray, W. & Saunders, M. A. 2005. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM review,* 47**,** 99-131.

[73] Gilman, J. R. & Ozgen, C. 2013. *Reservoir simulation: history matching and forecasting*, Society of Petroleum Engineers Richardson, TX.

[74] Glimm, J., Hou, S., Lee, Y.-H., Sharp, D. H. & Ye, K. 2004. Sources of uncertainty and error in the simulation of flow in porous media. *Computational & Applied Mathematics,* 23**,** 109-120.

[75] Glimm, J. & Sharp, D. 1999. Prediction and the quantification of uncertainty. *Physica D: Nonlinear Phenomena,* 133**,** 152-170.

[76] Goldberg, P. W., Williams, C. K. & Bishop, C. M. Regression with input-dependent noise: A Gaussian process treatment. Advances in neural information processing systems, 1998. 493-499.

[77] Gouveia, W. P. 1996. Bayesian seismic waveform data inversion: Parameter estimation and uncertainty analysis. *Ph. D. thesis, Colo. Sch. of Mines*.

[78] Hajizadeh, Y. 2011. *Population-based algorithms for improved history matching and uncertainty quantification of petroleum reservoirs, PhD thesis.* Heriot-Watt University.

[79] Hajizadeh, Y., Christie, M. A. & Demyanov, V. Ant Colony Optimization Algorithm for History Matching. EUROPEC/EAGE Conference and Exhibition, 2009. Society of Petroleum Engineers, SPE-121193-MS.

[80] Hajizadeh, Y., Christie, M. A. & Demyanov, V. History matching with differential evolution approach; a look at new search strategies. SPE EUROPEC/EAGE annual conference and exhibition, 2010. Society of Petroleum Engineers, SPE-130253-MS.

[81] Hamdi, H. & Sousa, M. C. Calibrating Multi-Point Geostatistical Models Using Pressure Transient Data. SPE Europec featured at 78th EAGE Conference and Exhibition, 2016. Society of Petroleum Engineers, SPE-180163-MS.

[82] Hastie, T. & Tibshirani, R. 1987. Generalized additive models: some applications. *Journal of the American Statistical Association,* 82**,** 371-386.

[83] Hategan, F. & Hawkes, R. 2007. The importance of initial reservoir pressure for tight gas completions and long-term production forecasting. *Journal of Canadian Petroleum Technology,* 46.

[84] Hazelrigg, G. 1999. On the role and use of mathematical models in engineering design. *Journal of Mechanical Design,* 121**,** 336-341.

[85] Heinonen, M., Mannerström, H., Rousu, J., Kaski, S. & Lähdesmäki, H. Non-stationary gaussian process regression with hamiltonian monte carlo. Artificial Intelligence and Statistics, 2016. 732-740.

[86] Heppner, F. & Grenander, U. 1990. A stochastic nonlinear model for coordinated bird flocks. *The ubiquity of chaos***,** 233-238.

[87] Heyen, G., Maréchal, E. & Kalitventzeff, B. 1996. Sensitivity calculations and variance analysis in plant measurement reconciliation. *Computers & chemical engineering,* 20**,** S539-S544.

[88] Hin, L., Lau, T., Rogers, M. & Chang, A. 1999. Dichotomization of continuous measurements using generalized additive modelling–application in predicting intrapartum caesarean delivery. *Statistics in medicine,* 18**,** 1101-1110.

[89] Hox, J. J. 1995. *Applied multilevel analysis,* TT-publikaties.

[90] Hu, B., Sagen, J., Chupin, G., Haugset, T., Ek, A. & Sommersel, T. Integrated wellbore-reservoir dynamic simulation. Asia Pacific Oil and Gas Conference and Exhibition, 2007. Society of Petroleum Engineers.

[91] Hutahaean, J. J. J. 2017. *Multi-objective methods for history matching, uncertainty prediction and optimisation in reservoir modelling, PhD Thesis.* Heriot-Watt University.

[92] Ibrahimov, T. History of History Match in Azeri Field. SPE Annual Caspian Technical Conference & Exhibition, 2015. Society of Petroleum Engineers, SPE-177395-MS.

[93] Ilk, D., Anderson, D. M., Stotts, G. W., Mattar, L. & Blasingame, T. 2010. Production data analysis--Challenges, pitfalls, diagnostics. *SPE Reservoir Evaluation & Engineering,* 13**,** 538-552.

[94] Jackman, S. 2009. Hierarchical statistical models. *Bayesian Analysis for the Social Sciences***,** 299-378.

[95] Jackson, M., Hampson, G., Saunders, J., El-Sheikh, A., Graham, G. & Massart, B. 2013. Surface-based reservoir modelling for flow simulation. *Geological Society, London, Special Publications,* 387**,** 271-292.

[96] Jackson, M., Percival, J., Mostaghimi, P., Tollit, B., Pavlidis, D., Pain, C., Gomes, J., Elsheikh, A. H., Salinas, P. & Muggeridge, A. 2015. Reservoir modeling for flow simulation by use of surfaces, adaptive unstructured meshes, and an overlapping-control-volume finite-element method. *SPE Reservoir Evaluation & Engineering,* 18**,** 115-132.

[97] Jaeger, T. F. & Levy, R. P. Speakers optimize information density through syntactic reduction. Advances in neural information processing systems, 2007. 849-856.

[98] Jahanshahi, E., Salahshoor, K. & Kharrat, R. 2008. Online LQG stabilization of unstable gas-lifted oil wells. *Computer Aided Chemical Engineering,* 25**,** 381.

[99] Jahns, H. O. 1966. A rapid method for obtaining a two-dimensional reservoir description from well pressure response data. *Society of Petroleum Engineers Journal,* 6**,** 315-327.

[100] Jefferys, W. H. & Berger, J. O. 1992. Ockham's razor and Bayesian analysis. *American Scientist,* 80**,** 64-72.

[101] Johnson, M. J. & Willsky, A. S. 2013. Bayesian nonparametric hidden semi-Markov models. *Journal of Machine Learning Research,* 14**,** 673-701.

[102] Jones, E. & Mitchell, T. 1978. Design criteria for detecting model inadequacy. *Biometrika,* 65**,** 541-551.

[103] Jung, S., Lee, K., Park, C. & Choe, J. 2018. Ensemble-based data assimilation in reservoir characterization: a review. *Energies,* 11**,** 445.

[104] Kabir, C. & Young, N. 2004. Handling production-data uncertainty in history matching: the Meren reservoir case study. *SPE Reservoir Evaluation & Engineering,* 7**,** 123-131.

[105] Kaipio, J. & Somersalo, E. 2007. Statistical inverse problems: discretization, model reduction and inverse crimes. *Journal of computational and applied mathematics,* 198**,** 493-504.

[106] Kalyanaraman, J., Kawajiri, Y., Lively, R. P. & Realff, M. J. 2016. Uncertainty quantification via Bayesian inference using sequential Monte Carlo methods for CO2 adsorption process. *AIChE Journal,* 62**,** 3352-3368.

[107] Kang, S., Bhark, E., Datta-Gupta, A., Kim, J. & Jang, I. 2015. A hierarchical model calibration approach with multiscale spectral-domain parameterization: Application to a structurally complex fractured reservoir. *Journal of Petroleum Science and Engineering,* 135**,** 336-351.

[108] Keelan, D. K. Core analysis techniques and applications. SPE Eastern Regional Meeting, 1972. Society of Petroleum Engineers.

[109] Kennedy, J. 2011. Particle swarm optimization. *Encyclopedia of machine learning.* Springer.

[110] Kennedy, M. C. & O'hagan, A. 2001. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 63**,** 425-464.

[111] Kersting, K., Plagemann, C., Pfaff, P. & Burgard, W. Most likely heteroscedastic Gaussian process regression. Proceedings of the 24th international conference on Machine learning, 2007. ACM, 393-400.

[112] King, M. J., Burn, K. S., Wang, P., Muralidharan, V., Alvarado, F. E., Ma, X. & Datta-Gupta, A. 2006. Optimal coarsening of 3D reservoir models for flow simulation. *SPE Reservoir Evaluation & Engineering,* 9**,** 317-334.

[113] Korkas, K. K. & Fryzlewicz, P. 2017. Multiple change-point detection for non-stationary time series using wild binary segmentation. *Statistica Sinica,* 27**,** 287-311.

[114] Kuss, M. 2006. *Gaussian process models for robust regression, classification, and reinforcement learning, PhD Thesis.* Technische Universität.

[115] Kuznetsova, A. A. 2017. *Heirarchical geological realism in history matching for reliable reservoir uncertainty predictions, PhD thesis.* Heriot-Watt University.

[116] Lafferty, J., Mccallum, A. & Pereira, F. C. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

[117] Landa, J. L. Technique to Integrate Production and Static Data in a Self-Consistent Way. SPE Annual Technical Conference and Exhibition, 2001. Society of Petroleum Engineers.

[118] Lavielle, M. 2005. Using penalized contrasts for the change-point problem. *Signal processing,* 85**,** 1501-1510.

[119] Le Cam, L. & Yang, G. L. 2012. *Asymptotics in statistics: some basic concepts*, Springer Science & Business Media.

[120] Leo, T.-Y., Kravaria, C. & Seinfeld, J. H. 1986. History matching by spline approximation and regularization in single-phase areal reservoirs. *SPE Reservoir Engineering,* 1**,** 521-534.

[121] Li, C. & King, M. J. Integration of pressure transient data into reservoir models using the Fast Marching Method. SPE Europec featured at 78th EAGE Conference and Exhibition, 2016. Society of Petroleum Engineers, SPE-180148-MS.

[122] Li, R., Reynolds, A. C. & Oliver, D. S. History matching of three-phase flow production data. SPE reservoir simulation symposium, 2001. Society of Petroleum Engineers.

[123] Li, X., Qiu, M. & Zheng, S. Integration of Numerical Well Testing and Deconvolution Algorithm for Analyzing Permanent Down-hole Gauge (PDG) Data. OTC Brasil, 2011. Offshore Technology Conference.

[124] Liebling, T. M. & Pournin, L. 2012. Voronoi diagrams and Delaunay triangulations: Ubiquitous siamese twins. *Documenta Mathematics, ISMP***,** 419-431.

[125] Lima, R., Abreu, A. C. & Pacheco, M. A. Optimization of reservoir development plan using the system octopus. OTC Brasil, 2015. Offshore Technology Conference.

[126] Ling, Y., Mullins, J. & Mahadevan, S. 2014. Selection of model discrepancy priors in Bayesian calibration. *Journal of Computational Physics,* 276**,** 665-680.

[127] Liseo, B. 2005. The elimination of nuisance parameters. *Handbook of Statistics,* 25**,** 193-219.

[128] Liu, N., Betancourt, S. & Oliver, D. S. Assessment of uncertainty assessment methods. SPE Annual Technical Conference and Exhibition, 2001. Society of Petroleum Engineers, SPE-71624-MS.

[129] Lucas, S. R. 2014. An inconvenient dataset: bias and inappropriate inference with the multilevel model. *Quality & quantity*, 48**,** 1619-1649.

[130] Mattar, L. & Mcneil, R. 1998. The" flowing" gas material balance. *Journal of Canadian Petroleum Technology*, 37.

[131] Mattax, C. C. & Dalton, R. L. 1990. Reservoir Simulation (includes associated papers 21606 and 21620). *Journal of Petroleum Technology,* 42**,** 692-695.

[132] Mcintire, M., Ratner, D. & Ermon, S. Sparse Gaussian Processes for Bayesian Optimization. UAI, 2016.

[133] Miroshnikov, A., Wei, Z. & Conlon, E. M. 2015. Parallel Markov chain Monte Carlo for non-Gaussian posterior distributions. *Stat,* 4**,** 304-319.

[134] Mishev, I., Fedorova, N., Terekhov, S., Beckner, B., Usadi, A., Ray, M. & Diyankov, O. Adaptive control for solver performance optimization in reservoir simulation. ECMOR XI-11th European Conference on the Mathematics of Oil Recovery, 2008.

[135] Mohamed, L., Christie, M. A. & Demyanov, V. History matching and uncertainty quantification: multiobjective particle swarm optimisation approach. SPE EUROPEC/EAGE annual conference and exhibition, 2011. Society of Petroleum Engineers, SPE-143067-MS.

[136] Mohamed, L., Christie, M. A., Demyanov, V., Robert, E. & Kachuma, D. Application of particle swarms for history matching in the Brugge reservoir. SPE Annual Technical Conference and Exhibition, SPE-135264-MS, 2010. Society of Petroleum Engineers.

[137] Mohamed, L. M. Y. 2011. *Novel sampling techniques for reservoir history matching optimisation and uncertainty quantification in flow prediction, , PhD thesis.* Heriot-Watt University.

[138] Mohammad, J., Mohammad, J. & Siavash, A. 2014. Reservoir evaluation in Undersaturated oil reservoirs using modern production data analysis; a simulation study. *Science International,* 26**,** 1089-1094.

[139] Morrison, R. E., Oliver, T. A. & Moser, R. D. 2018. Representing model inadequacy: A stochastic operator approach. *SIAM/ASA Journal on Uncertainty Quantification,* 6**,** 457-496.

[140] Mu, L. & Kuang, L. Maritime data integration using standard ISO 15926. The Twentieth International Offshore and Polar Engineering Conference, 2010. International Society of Offshore and Polar Engineers.

[141] Muggeridge, A., Cockin, A., Webb, K., Frampton, H., Collins, I., Moulds, T. & Salino, P. 2014. Recovery rates, enhanced oil recovery and technological limits. *Phil. Trans. R. Soc. A,* 372**,** 20120320.

[142] Murray-Smith, R. & Girard, A. Gaussian Process priors with ARMA noise models. Irish Signals and Systems Conference, Maynooth, 2001. 147-152.

[143] Nandram, B. & Xu, H. 2011. Bayesian Corrections of a Selection Bias in Genetics. *J Biomet Biostat,* 2**,** 2.

[144] Nezhad Karim Nobakht, B., Christie, M. & Demyanov, V. Model Selection for Error Generalization in History Matching. SPE Europec featured at 80th EAGE Conference and Exhibition, 2018. Society of Petroleum Engineers, SPE-190778-MS.

[145] Nghiem, L., Mirzabozorg, A., Yang, C. & Chen, Z. Differential evolution for assisted history matching process: Sagd case study. SPE Heavy Oil Conference-Canada, 2013. Society of Petroleum Engineers, SPE-165491-MS.

[146] Nguyen, H. H. & Chan, C. W. 2005. Applications of data analysis techniques for oil production prediction. *Engineering Applications of Artificial Intelligence,* 18**,** 549-558.

[147] Nicotra, G., Godi, A., Cominelli, A. & Christie, M. Production data and uncertainty quantification: A real case study. SPE Reservoir Simulation Symposium, 2005. Society of Petroleum Engineers, SPE-93280-MS.

[148] Nobakht, B. N. K. & Christie, M. Model Prediction under Uncertainty Using Hierarchical Models. 79th EAGE Conference and Exhibition 2017, 2017.

[149] O'sullivan, A. & Christie, M. 2005a. Error models for reducing history match bias. *Computational Geosciences,* 9**,** 125-153.

[150] O'sullivan, A. & Christie, M. Solution error models: a new approach for coarse grid history matching. SPE Reservoir Simulation Symposium, 2005b. Society of Petroleum Engineers.

[151] O'sullivan, A. & Christie, M. 2006. Simulation error models for improved reservoir prediction. *Reliability Engineering & System Safety,* 91**,** 1382-1389.

[152] Oberkampf, W. L., Deland, S. M., Rutherford, B. M., Diegert, K. V. & Alvin, K. F. 2002. Error and uncertainty in modeling and simulation. *Reliability Engineering & System Safety,* 75**,** 333-357.

[153] Okano, H. Reservoir Model History-Matching and Uncertainty Quantification in Reservoir Performance Forecast Using Bayesian Framework. SPE Reservoir Characterization and Simulation Conference and Exhibition, 2013. Society of Petroleum Engineers, SPE-165970-MS.

[154] Oliver, D. S. & Chen, Y. 2011. Recent progress on reservoir history matching: a review. *Computational Geosciences,* 15**,** 185-221.

[155] Orbanz, P. & Teh, Y. W. 2011. Bayesian nonparametric models. *Encyclopedia of Machine Learning.* Springer.

[156] Ordaz-Hernandez, K., Fischer, X. & Bennis, F. 2007. A mathematical representation for mechanical model assessment: numerical model qualification method. *International Journal of Mathematics Sciences,* 1**,** 216--226.

[157] Ouenes, A., Meunier, G., Pelce, V. & Lhote, I. Enhancing Gas Reservoir Characterization by Simulated Annealing Method (SAM). European Petroleum Conference, 1992. Society of Petroleum Engineers.

[158] Ouenes, A. & Saad, N. A new, fast parallel simulated annealing algorithm for reservoir characterization. SPE Annual Technical Conference and Exhibition, 1993. Society of Petroleum Engineers, SPE-26419-MS.

[159] Pacheco, M. A. & Vellasco, M. M. 2009. *Intelligent systems in oil field development under uncertainty*, Springer.

[160] Paffenholz, J., Monk, D. & Fryar, D. 1994. Random and systematic navigation errors: How do they affect seismic data quality? *First Break,* 12**,** 505-513.

[161] Panda, M. & Lake, L. Parallel simulated annealing for stochastic reservoir modeling. SPE Annual Technical Conference and Exhibition, SPE-26418-MS, 1993. Society of Petroleum Engineers.

[162] Parish, E. J. & Duraisamy, K. 2016. A paradigm for data-driven predictive modeling using field inversion and machine learning. *Journal of Computational Physics,* 305**,** 758-774.

[163] Pederson, J. M., Moon, M. S. & Al-Ajeel, H. Data validation: Key to development of an integrated reservoir model for the Wara Formation, Greater Burgan field. Middle East Oil Show and Conference, 1997. Society of Petroleum Engineers.

[164] Pelikan, M., Goldberg, D. E. & Cantú-Paz, E. BOA: The Bayesian optimization algorithm. Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation-Volume 1, 1999. Morgan Kaufmann Publishers Inc., 525-532.

[165] Pernot, P. & Cailliez, F. 2017. A critical review of statistical calibration/prediction models handling data inconsistency and model inadequacy. *AIChE Journal,* 63**,** 4642-4665.

[166] Perrone, A., Pennadoro, F., Tiani, A., Della Rossa, E. & Saetrom, J. Enhancing the Geological Models Consistency in Ensemble Based History Matching an Integrated Approach. SPE Reservoir Characterisation and Simulation Conference and Exhibition, SPE-186049-MS, 2017. Society of Petroleum Engineers.

[167] Pickup, G., Valjak, M. & Christie, M. Model complexity in reservoir simulation. ECMOR XI-11th European Conference on the Mathematics of Oil Recovery, 2008.

[168] Polyak, B. T. 2007. Newton's method and its use in optimization. *European Journal of Operational Research,* 181**,** 1086-1096.

[169] Pritchett, Y. L., Menon, S., Marchenko, O., Antonijevic, Z., Miller, E., Sanchez-Kam, M., Morgan-Bouniol, C. C., Nguyen, H. & Prucka, W. R. 2015. Sample size re-estimation designs in confirmatory clinical trials—current state, statistical considerations, and practical guidance. *Statistics in Biopharmaceutical Research,* 7**,** 309-321.

[170] Qu, D., Røe, P. & Tveranger, J. 2015. A method for generating volumetric fault zone grids for pillar gridded reservoir models. *Computers & Geosciences,* 81**,** 28-37.

[171] Quadrianto, N., Kersting, K., Reid, M. D., Caetano, T. S. & Buntine, W. L. Kernel conditional quantile estimation via reduction revisited. 2009 Ninth IEEE International Conference on Data Mining, 2009. IEEE, 938-943.

[172] Rabosky, D. L. & Goldberg, E. E. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Systematic Biology,* 64**,** 340-355.

[173] Rasmussen, C. E. 2004. Gaussian processes in machine learning. *Advanced lectures on machine learning.* Springer.

[174] Rasmussen, C. E. & Williams, C. K. 2006. *Gaussian process for machine learning,* MIT press.

[175] Ratmann, O., Andrieu, C., Wiuf, C. & Richardson, S. 2009. Model criticism based on likelihood-free inference, with an application to protein network evolution. *Proceedings of the National Academy of Sciences,* 106**,** 10576-10581.

[176] Rausch, S., Odendahl, S., Kersting, P., Biermann, D. & Zabel, A. 2012. Simulation-based prediction of process forces for grinding free-formed surfaces on machining centers. *Procedia CIRP,* 4**,** 161-165.

[177] Reynolds, C. 1987. Flocks, herds and schools: A distributed behavioral model. SIGGRAPH Comput. Graph. 21 (4), 25–34 (1987).

[178] Riani, M., Atkinson, A. & Cerioli, A. 2012. Problems and challenges in the analysis of complex data: static and dynamic approaches. *Advanced Statistical Methods for the Analysis of Large Data-Sets.* Springer.

[179] Rissanen, J. 1989. Stochastic complexity in statistical inquiry. *World Scientific Series in Computer Science,* 15**,** 79-93.

[180] Risser, M. D. & Calder, C. A. 2015. Regression-based covariance functions for nonstationary spatial modeling. *Environmetrics*, 26**,** 284-297.

[181] Roberts, S., Osborne, M., Ebden, M., Reece, S., Gibson, N. & Aigrain, S. 2013. Gaussian processes for time-series modelling. *Phil. Trans. R. Soc. A,* 371**,** 20110550.

[182] Romero, C., Carter, J., Zimmerman, R. & Gringarten, A. Improved reservoir characterization through evolutionary computation. SPE Annual Technical Conference and Exhibition, 2000. Society of Petroleum Engineers, SPE-62942-MS.

[183] Rotondi, M., Nicotra, G., Godi, A., Contento, F. M., Blunt, M. J. & Christie, M. Hydrocarbon production forecast and uncertainty quantification: A field application. SPE Annual Technical Conference and Exhibition, SPE-102135-MS, 2006. Society of Petroleum Engineers.

[184] Rushing, J. & Newsham, K. An Integrated Work-Flow Model to Characterize Unconventional Gas Resources: Part II-Formation Evaluation and Reservoir Modeling. SPE Annual Technical Conference and Exhibition, SPE-71352-MS, 2001. Society of Petroleum Engineers.

[185] Sambridge, M. 1999a. Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space. *Geophysical Journal International,* 138**,** 479-494.

[186] Sambridge, M. 1999b. Geophysical inversion with a neighbourhood algorithm—II. Appraising the ensemble. *Geophysical Journal International,* 138**,** 727-746.

[187] Sanghyun, L. & Stephen, K. Optimizing Automatic History Matching for Field Application Using Genetic Algorithm and Particle Swarm Optimization. Offshore Technology Conference Asia, 2018. Offshore Technology Conference.

[188] Saracco, P., Pia, M. G. & Batic, M. 2014. Theoretical grounds for the propagation of uncertainties in Monte Carlo particle transport. *IEEE Transactions on Nuclear Science,* 61**,** 877-887.

[189] Sargsyan, K., Huan, X. & Najm, H. N. 2018. Embedded Model Error Representation for Bayesian Model Calibration. *arXiv preprint arXiv:1801.06768*.

[190] Satter, A. & Thakur, G. C. 1994. *Integrated petroleum reservoir management: a team approach*, PennWell Books.

[191] Schulze-Riegert, R., Axmann, J., Haase, O., Rian, D. & You, Y.-L. 2002. Evolutionary algorithms applied to history matching of complex reservoirs. *SPE Reservoir Evaluation & Engineering,* 5**,** 163-173.

[192] Schwarz, G. 1978. Estimating the dimension of a model. *The annals of statistics,* 6**,** 461-464.

[193] Seeger, M. 2004. Gaussian processes for machine learning. *International journal of neural systems,* 14**,** 69-106.

[194] Seiler, A., Evensen, G., Skjervheim, J., Hove, J. & Vabø, J. 2011. Using the EnKF for history matching and uncertainty quantification of complex reservoir models. *Computational Methods for Large-Scale Inverse Problems and Quantification of Uncertainty***,** 247-271.

[195] Seinfeld, J. & Kravaris, C. 1982. Distributed parameter identification in geophysics—petroleum reservoirs and aquifers. *Distributed Parameter Control Systems*. Elsevier.

[196] Sharifi, M., Kelkar, M., Bahar, A. & Slettebo, T. 2014. Dynamic ranking of multiple realizations by use of the fast-marching method. *SPE Journal,* 19**,** 1,069-1,082.

[197] Shearer, P. M. 2009. *Introduction to seismology*, Cambridge University Press.

[198] Singh, V., Yemez, I. & Sotomayor, J. 2013. Key factors affecting 3D reservoir interpretation and modelling outcomes: Industry perspectives. *British Journal of Applied Science & Technology,* 3**,** 376.

[199] Skjervheim, J.-A., Evensen, G., Aanonsen, S. I., Ruud, B. O. & Johansen, T.-A. Incorporating 4D seismic data in reservoir simulation models using ensemble Kalman filter. SPE Annual Technical Conference and Exhibition, 2005. Society of Petroleum Engineers, SPE-95789-PA.

[200] Slater, G. & Durrer, E. 1971. Adjustment of reservoir simulation models to match field performance. *Society of Petroleum Engineers Journal,* 11**,** 295-305.

[201] Spall, J. C. & Garner, J. 1990. Parameter identification for state-space models with nuisance parameters. *IEEE Transactions on Aerospace and Electronic Systems,* 26**,** 992-998.

[202] Stephen, K. D., Shams, A. & Macbeth, C. Faster seismic history matching in a UKCS reservoir. EUROPEC/EAGE Conference and Exhibition, 2007. Society of Petroleum Engineers.

[203] Stephen, K. D., Soldo, J., Macbeth, C. & Christie, M. A. 2006. Multiple model seismic and production history matching: A case study. *SPE Journal,* 11**,** 418-430.

[204] Stine, R. A. 2004. Model selection using information theory and the MDL principle. *Sociological Methods & Research,* 33**,** 230-260.

[205] Subbey, S., Christie, M. & Sambridge, M. 2004. Prediction under uncertainty in reservoir modeling. *Journal of Petroleum Science and Engineering,* 44**,** 143-153.

[206] Sue, L. 1987. Linear models, random censoring and synthetic data. *Biometrika,* 74**,** 301-309.

[207] Svensson, A., Dahlin, J. & Schön, T. B. Marginalizing Gaussian process hyperparameters using sequential Monte Carlo. Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on, 2015. IEEE, 477-480.

[208] Tarantola, A. 2005. *Inverse problem theory and methods for model parameter estimation*, siam.

[209] Tavassoli, Z., Carter, J. N. & King, P. R. 2004. Errors in history matching. *SPE Journal,* 9**,** 352-361.

[210] Thomas, L. K., Hellums, L. & Reheis, G. 1972. A nonlinear automatic history matching technique for reservoir simulation models. *Society of Petroleum Engineers Journal,* 12**,** 508-514.

[211] Tokuda, N., Takahashi, S., Watanabe, M. & Kurose, T. Application of genetic algorithm to history matching for core flooding. SPE Asia Pacific Oil and Gas Conference and Exhibition, 2004. Society of Petroleum Engineers, SPE-88621-MS.

[212] Vaart, E., Prangle, D. & Sibly, R. M. 2018. Taking error into account when fitting models using Approximate Bayesian Computation. *Ecological Applications,* 28**,** 267-274.

[213] Valjak, M. 2008. *History matching and forecasting with uncertainty: challenges and proposed solutions for real field applications, , PhD thesis.* Heriot-Watt University.

[214] Van Der Geest, R., Broman Jr, W. H., Johnson, T. L., Fleming, R. H. & Allen, J. O. Reliability through data reconciliation. Offshore Technology Conference, 2001. Offshore Technology Conference.

[215] Van Der Waart, A. 1998. Asymptotic statistics, volume 27 of Cambridge Series in Statistical and Probabilistic Mathematics 03. Cambridge Univ. Press, New York.

[216] Varhaug, M. 2016. Basic Well Log Interpretation. *Oilfield Review***,** 52-53.

[217] Vasco, D. & Datta-Gupta, A. 1999. Asymptotic solutions for solute transport: A formalism for tracer tomography. *Water Resources Research,* 35**,** 1-16.

[218] Vink, J. C., Gao, G. & Chen, C. Bayesian Style History Matching: Another Way to Under-Estimate Forecast Uncertainty? SPE Annual Technical Conference and Exhibition, SPE-175121-MS, 2015. Society of Petroleum Engineers.

[219] Wathelet, M. 2008. An improved neighborhood algorithm: parameter conditions and dynamic scaling. *Geophysical Research Letters,* 35.

[220] Watkins, A. & Modine, A. A stochastic role for engineering input to reservoir history matching. SPE Latin America Petroleum Engineering Conference, 1992. Society of Petroleum Engineers, SPE-23738-MS.

[221] Watson, A., Gavalas, G. & Seinfeld, J. 1984. Identifiability of estimates of two-phase reservoir properties in history matching. *Society of Petroleum Engineers Journal,* 24**,** 697-706.

[222] Watts, J. 1997. Reservoir simulation: Past, present, and future. *SPE Computer Applications,* 9**,** 171-176.

[223] Weiss, G. 1977. Shot noise models for the generation of synthetic streamflow data. *Water Resources Research,* 13**,** 101-108.

[224] West, R. M. 2012. Generalised additive models. *Modern methods for epidemiology.* Springer.

[225] Wising, U., Vrielynck, B., Kalitventzeff, P.-B. & Campan, J. Improving Operations Through Increased Accuracy of Production Data. Offshore Europe, 2009. Society of Petroleum Engineers.

[226] Wood, D. & Mokhatab, S. 2007. Optimize reservoir management in real time. *Hart's E & P,* 80**,** 23-24.

[227] Yusuf, N. O., Silpngarmlers, L. & Eme, V. O. Addressing The Impossible History Match Problem: Using a Systematic Approach to Eliminate Model Bias. SPE Nigeria Annual International Conference and Exhibition, 2018. Society of Petroleum Engineers, SPE-193521-MS.

[228] Zheng, S.-Y., Stewart, G. & Corbett, P. Analyzing pressure transient test in semi-Infinite and finite reservoirs using de-superposition method. International Oil and Gas Conference and Exhibition in China, 2000. Society of Petroleum Engineers.

[229] Zingg, D. W., Nemec, M. & Pulliam, T. H. 2008. A comparative evaluation of genetic and gradient-based algorithms applied to aerodynamic optimization.

European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique, 17**,** 103-126.