Calibration and prediction improvement of imperfect subsurface flow models

by

Muzammil Hussain Rammay

Submitted for the degree of

Doctor of Philosophy

Institute of GeoEnergy Engineering

School of Energy, Geoscience, Infrastructure and Society

Heriot Watt University

January 20, 2020

Abstract

In this thesis, problems related to calibration of imperfect reservoir models, biased parameter estimation and prediction reliability have been addressed. The main objective of this thesis is to avoid overconfident, inaccurate and unreliable predictions while accounting for model-error during the calibration process. Accounting for reservoir model-error in calibration (history matching) can correct/reduce the bias in parameter estimation and improves the prediction of the subsurface flow model. In this thesis, several approaches and algorithms have been developed and investigated which could be applied at different conditions depending on the modelling assumptions. In the first approach, the parameter estimation problem is formulated as a joint estimation of the imperfect model parameters and the error-model parameters. The prior distributions of the error-model parameters are evaluated before calibration through analysis of leading sources of the modelling errors using pairs of high-fidelity and low-fidelity simulation models. A Bayesian framework is adopted for solving the inverse problem, where the ensemble smoother with multiple data assimilation (ES-MDA) is utilized as a calibration algorithm. In the second approach, two new algorithms to account for model-error during calibration are developed which are the variants of the first approach and existing algorithms. The main aim is to develop flexible algorithms that can handle strong serially correlated outputs of the physical model, variable boundary conditions (i.e. variable well open/shut schedules and rate/pressure controls) and structured model-errors (i.e. strong correlation in time). In the third approach, the model-error during calibration is accounted for without knowing any prior statistics of model-discrepancy. For this purpose, a flexible ensemble-based algorithm is developed which can reduce bias in parameter estimation after calibration of imperfect models in

order to improve the prediction capacity/reliability of the calibrated physical model. The flexible ensemble-based algorithm is quite general and has the capability to capture unknown model-error uncertainty by relaxing many of the assumptions commonly introduced in the literature.

I dedicate this thesis to my beloved mother (Riffat Riaz) and father (Muhammad Riaz Rammay) for their limitless love, prayers and moral support.

Acknowledgements

First of all, I would like to thank Almighty Allah (God) for providing me limitless mercy, blessings, curiosity to explore and motivation for research. I deeply acknowledge and thank to my supervisors Dr. Ahmed H. Elsheikh and Dr. Florian Doster. Dr. Ahmed H. Elsheikh provided me continuous guidance and support throughout my research work. I learned a lot from his guidance and complex problem solving abilities. I would like to thank Total E&P UK for the financial support to undertake my PhD work. I would also like to thank Dr. Yan Chen for her guidance and support during my placement in Geoscience Research Center, Total E&P UK related to my PhD work. I learned a lot from her practical experience and patience to address complex problems. I would also like to acknowledge Total S.A. for authorizing the publication of the papers related to this research work.

Contents

List of Figures

Chapter 1

Background and Literature Review

1.1 Introduction

Box saying, "All models are wrong, but some are useful", is considered as the aphorism for the physical and statistical models. This aphorism can also be stated as "All models are biased, some are more and some are less". In this thesis *physical model* refers to physics based mathematical model. Generally physical models are treated as if they are perfect during the calibration i.e. the parameter error is assumed as the cause of data mismatch. However, it is widely known that physical models are imperfect and approximation of reality. These approximations commonly introduce some type of model-error which is often neglected during calibration of the imperfect physical model. An imperfect model is a mathematical model which has the discrepancy from the reality in terms of description, scale, assumptions and complexity. An imperfect model exhibits data mismatch due to the parameter error and model-error. The negligence of model-error can result in biased estimated parameters after calibration of the imperfect model and as a consequence might result in inaccurate predictions (Brynjarsdottir and O' Hagan, 2014). Model-error was formally introduced as a source of uncertainty in simulator predictions by Kennedy and O'Hagan (2001), who referred to it as model inadequacy (other commonly used names are model-bias, model discrepancy, etc). In this thesis, we use these terms interchangeably and introduce the algorithms for calibration of imperfect models while accounting for model-error under different set of scenarios and assumptions.

In this chapter, a brief background and literature review are discussed with respect to approaches and methods of accounting for model-error during the calibration process.

The outline of this chapter is as follows. In section 1.2, sources of reservoir model-errors and bias are discussed. Bayesian inference for inverse problems is discussed in section 1.3, with a focus on MCMC and ensemble-based algorithms for history matching and calibration. Section 1.4 presents various approaches in the literature, which account for model-error during calibration. In section 1.5, the methods and formulation of model-error in the literature are discussed. The last section is related to the objectives and outline of the thesis with a complete summary of the rest of the chapters.

1.2 Sources of reservoir model-error/bias

Oil and gas reservoir modelling exhibits various forms of uncertainties. These uncertainties are related to reservoir parameters, reservoir geology description, measurement errors and model discrepancy/error. Often the predictions from reservoir models are not as reliable as expected, because of lack of assessment of all types of uncertainties. One of the primary reasons of unreliable and inaccurate prediction is the lack of knowledge of the reservoir model discrepancy/error. It is common practice in the oil and gas industry that model discrepancy/error is ignored during calibration (history matching) process assuming the reservoir model is perfect. This assumption results in a bias in the estimated parameters and at the end of the day, the utility of the calibrated models are significantly reduced as the model predictions would be different from the reality. Some of the most significant sources of errors, which make reservoir simulation models imperfect, are listed in the following subsections.

1.2.1 Discretization and upscaling errors

Discretization and upscaling errors in reservoir simulation modelling have been widely studied. Discretization errors are introduced when the partial differential equations describing the dynamical evolution of the system states is solved numerically and in some cases, they could be significant. The discretization errors appear due to the grid coarsening of the spatial domain or time steps selection for the time domain. In practice, this problem is solved by utilizing optimal grid size and using automatic time

step selection algorithms. Still, there is an inherent discretization error with respect to the exact solution, which in many cases is not small enough to be neglected (Ertekin et al., 2001). Upscaling is the process of substituting a heterogeneous property region consisting of fine grid cells with a homogeneous region made up of a coarse-grid cell with an effective property value such that the error related to mass balance or energy balance between fine and coarse grid cell becomes negligible or acceptable. A number of efficient upscaling techniques exists in the literature (Farmer, 2002; Durlofsky, 2003), which have corresponding pros and cons. The goal of the upscaling is also to reduce the effect of errors from discretization. However, if the find scale model has some missing information in terms of boundary conditions or spatial distribution of the grid block properties, the upscaling could also be inaccurate.

Solving with optimal grid dimensions might be very computationally expensive for history matching, optimization, and uncertainty quantification problems. Various techniques have been proposed in the literature to assess this problem, which includes reduced order models or proxy models etc (Silva et al., 2007; Rammay and Abdulraheem, 2014; Cardoso et al., 2009). The other way is to upscale the fine-scale simulation model in such a way that the up-scaled (coarse scale) model becomes computationally inexpensive. If the up-scaling error is neglected during inversion then the parameter inference would be biased and predictions of oil, gas and water rates/pressure would be inaccurate.

Omre et al. (2004) estimated the up-scaling and grid coarsening error by running pairs of fine and coarse-scale models. They modelled the reservoir model-error using multiple regression techniques and these terms were included in the coarse-scale model during the history matching process and successfully obtained the results near to the fine-scale model. Lødøen et al. (2005) proposed the same procedure as Omre et al. (2004) with different case studies and showed the benefits of the more accurate up-scaling procedure. Lødøen and Tjelmeland (2010) used multiple linear regression algorithm to model-errors, where the residual part for the multiple regression was assumed to depend on the model input parameters. The residual terms were modelled using a zero-mean Gaussian Process. They assumed the fine model as a perfect model

and did not consider measurement errors in their observations. These two assumptions can be very strong in the context of uncertainty quantification for simulation models of real reservoirs.

1.2.2 Wells location variations due to grid coarsening

Well locations are usually shifted during the up-scaling and grid coarsening process. This shift also induces some form of model-error and causes imperfection in the reservoir model. Usually, in practice, this is avoided by multiplying connection factors with Peaceman's model (Peaceman et al., 1983) or Abou-Kassem and Aziz model (Abou-Kassem et al., 1985). However, these corrections and models are approximations and have an inherent assumption, therefore, the error is still present. During up-scaling and discretization error quantification, these errors also contribute to the estimated distribution of the quantities of interest. Usually, it would be difficult to separate these errors from up-scaling and discretization errors. We can say that these three sources of errors have some sort of combined effect and can be treated as a combined source in the context of simulation models of oil and gas reservoirs.

1.2.3 Imperfect PVT model/lab data of reservoir fluids

Reservoir fluid model and data are among the important parts of simulation modelling. Oil and gas comprise of various types of the complex mixtures of organic compounds. Black oil correlations and the Equation of States (EOS) are two widely used methods to model these complex mixtures (McCain, 1990). Black oil correlations are built using linear, non-linear regression models or in some cases using neural networks (Rammay and Abdulraheem, 2016). These correlations are statistically based so they are imperfect to unknown data points. EOS are semi-empirical models and better than black oil correlations, however, they are approximate due to the complexity of the mixture of oil and gas. EOS requires tuning with lab data for a particular reservoir fluid (Ahmed, 2016).

Lab experiments are another way to generate reservoir fluid properties at the specific

pressure-temperature condition of the reservoir. The lab data is converted to reservoir conditions using specific methods (McCain, 1990; Ahmed, 2016). During reservoir simulation modelling, the converted lab data is interpolated or extrapolated to the pressure-temperature condition, where data points are not known. This PVT lab data can also be used to tune the black oil correlations or calibration of EOS for interpolation and extrapolation purposes at unknown pressure and temperature condition. PVT lab data can also have measurement errors and methods which convert PVT lab data to appropriate reservoir conditions also introduce approximations. Therefore PVT modelling is not completely perfect and can introduce considerable model-error in the reservoir simulation model.

O'Sullivan and Christie (2006) addressed the issue related to the calibration of oil viscosity in the biased simulation model. They considered up-scaling and discretization as a source of reservoir model-error and showed that the calibrated oil viscosity is significantly far from the true oil viscosity. Similarly, if the reservoir model does not contain realistic oil viscosity or other PVT properties, then the estimation of other reservoir parameters is biased and prone to produce inaccurate predictions.

1.2.4 Imperfect relative permeability model/lab data

Reservoir relative permeability model and data are also among the important parts in simulation modelling and highly sensitive to oil and gas recovery. Relative permeability involves a complex interaction between reservoir rock and fluid. Usually, they are modelled using empirical correlations or obtained using lab data measurement (Ahmed, 2006). The actual number of relative permeability models, required in simulation modelling are not fully known, therefore sometimes they introduce approximation and bias in reservoir models.

1.2.5 Imperfect reservoir geology description

Accurate reservoir geology description is one of the most difficult and challenging tasks. It always has uncertainty due to the complexity of channel geometry, fault shape, facies

proportion, stratigraphic and structural features. Unrealistic geological models could be history matched, but not capable to produce reliable predictions and prone to mislead the development plans of a particular reservoir (Hoffman and Caers, 2007). Some local faults, structural features, channels geometry, and facies discontinuities are very difficult to model on the geological scale because of the low resolution of the geophysical information. The local faults can be detected from well testing (Lee et al., 2003), but not with the complete extent of fault geometry. The imperfect geological description is one of the important reason that some calibrated reservoir models do not have predictive capability (Carter et al., 2006) and can introduce significant model-error.

1.2.6 Reservoir model assumptions

Every reservoir model has inherent assumptions of subsurface flow physics. These assumptions help to solve the complex problem with existing methods. It is not necessary that reality follows these assumptions. For example, constant or linear rock compressibility is assumed most of the time during reservoir modelling but in reality compressibility of rock can behave as non-linear way e.g. chalk reservoirs. Therefore model assumptions can introduce model-error if reality deviates from the assumptions. In oil and gas literature so far reservoir model assumptions have not been addressed as a form of model-error.

1.3 Bayesian inference for inverse problems

Bayes rule can be used in the inverse modelling or history matching of subsurface models. It is useful to estimate the posterior probability of model parameters given observed data while accounting for various forms of uncertainties. The posterior belief or probability about model parameters is updated by integrating observations of production data (Oliver et al., 2008) using the following mathematical description of Bayes rule.

$$p(m|d_{obs}) = \frac{p(d_{obs}|m)p(m)}{p(d_{obs})}, \tag{1.1}$$

where $p(m)$ is the prior probability of the model parameters m, $p(d_{obs}|m)$ is the probability of production data observation given model parameters m and also known as the likelihood function. $p(d_{obs})$ is the probability of the observation and refers to the normalizing constant in Bayes rule. Mathematically $p(d_{obs})$ can be written as,

$$p(d_{obs}) = \int p(d_{obs}|m)p(m)dm. \tag{1.2}$$

The prior and likelihood functions in Eq. 1.1 can be represented by the various types of probability distributions and among that Gaussian distribution is one of the widely used (Oliver et al., 2008). The details of the multivariate Gaussian distribution and corresponding mathematical functions are discussed in the later chapters. In recent research literature, both Markov Chain Monte Carlo (MCMC) (Gamerman and Lopes, 2006) and ensemble-based methods (Oliver et al., 2008) have been used to computationally solve the Bayesian inference problem (i.e. obtain statistical solution for the parameter estimation problem). A brief review of both types of methods is discussed in the following subsections.

1.3.1 Markov chain Monte Carlo (MCMC)

MCMC is an exact method to sample the model parameters from a complex posterior distribution. Therefore MCMC can be used to solve inverse problems or calibration of physical models. There are two major steps in MCMC methods.

- Propose new samples from multidimensional random variable or prior probability distribution.
- Update the sample, after comparing the sample probability to an acceptance probability.

There are a variety of MCMC methods that exist in the literature (Gamerman and Lopes, 2006). Most of the MCMC methods have a difference in the proposal step of the new sample and the computation of acceptance probability. In this work, the following two versions of MCMC standard and modified random walk methods are discussed (Cotter et al., 2013).

1.3.2 MCMC standard random walk

The steps for MCMC standard random walk method are given below.

- Estimate posterior probability $p(u^{(k)}|d_{obs})$ using Bayes rule by sampling of model parameters $u^{(k)}$ from prior distribution and corresponding likelihood function $p(d_{obs}|u^{(k)})$
- Propose new samples v using $v = u^{(k)} + \beta\epsilon$ where β is the step size (walk step) and ϵ is probability distribution of model parameters or if we assume Gaussian distribution then it is normal distribution $\mathcal{N}(\mu_M, C_M)$ with mean μ_M and covariance C_M.
- Estimate posterior probability $p(v|d_{obs})$ using Bayes rule of proposed samples of model parameters v and corresponding likelihood function $p(d_{obs}|v)$
- Calculate acceptance probability α using $\alpha = min(1, \frac{p(v|d_{obs})}{p(u^{(k)}|d_{obs})})$
- Set $u^{(k+1)} = v$ with acceptance probability
- Otherwise $u^{(k+1)} = u^{(k)}$
- Repeat the steps upto $k = n_{walks}$, where n_{walks} is the total number of walks or iterations and k is the iteration index

1.3.3 MCMC modified random walk

The steps for MCMC preconditioned Crank Nicolson (pCN), modified random walk method are given below.

- Estimate likelihood probability $p(d_{obs}|u^{(k)})$ by sampling of model parameters $u^{(k)}$ from prior distribution
- Propose new samples v using $v = \sqrt{1-\beta^2}u^{(k)} + \beta\epsilon$ where β is the step size (walk step) and ϵ is probability distribution of model parameters or if we assume Gaussian then it is normal distribution $\mathcal{N}(\mu_M, C_M)$ with mean μ_M and covariance C_M.

- Estimate likelihood probability $p(d_{obs}|v)$ of proposed samples of model parameters v
- Calculate acceptance probability α using $\alpha = min(1, \frac{p(d_{obs}|v)}{p(d_{obs}|u^{(k)})})$
- Set $u^{(k+1)} = v$ with acceptance probability
- Otherwise $u^{(k+1)} = u^{(k)}$
- Repeat the steps upto $k = n_{walks}$, where n_{walks} is the total number of walks or iterations and k is the iteration index

1.3.4 Ensemble smoother

Ensemble-based methods have been gaining popularity in the last two decades to perform Bayesian inference for inverse problems and data assimilation (Emerick and Reynolds, 2013). The main advantage of the ensemble-based methods is the low-computational cost for higher-dimensional data assimilation and inverse problems. Ensemble smoother ES is one of the ensemble-based techniques similar to the ensemble Kalman filter (Evensen, 2009). ES updates the model parameters by simultaneous assimilation of all available data. In the ensemble Kalman filter, state variables of the physical model are required during data assimilation, however, in ES only production data is required to evaluate the samples of model parameters from the posterior distribution. The ES update equation of model parameters can be represented as follows.

$$m_j^{(a)} = m_j^{(i)} + C_{MD}(C_{DD} + C_D)^{-1}(d_{uc,j} - d_j) \tag{1.3}$$

where $d_{uc} \sim \mathcal{N}(d_{obs}, C_D)$ is the perturbed observation, d is the output/response of physical model, C_{MD} is the covariance of production data and prior model parameters, C_{DD} is the covariance of predicted data and C_D is the covariance of measurement errors in observed data. Superscript a and i represent updated (analysed) and initial model parameters respectively. Subscript j represents ensemble member index.

1.3.5 ES-MDA

Ensemble smoother does not produce satisfactory results for non-linear inverse problems. However, the simulation models of oil and gas reservoirs are highly non-linear in nature. Emerick and Reynolds (2013) showed that ensemble smoother can be used for non-linear problems if the data is assimilated multiple times. They proposed the iterative ensemble smoother which tries to capture the non-linear behaviour of the reservoir simulation model by using multiple data assimilation or multiple iterations using inflated covariance matrix of measurement errors (Emerick and Reynolds, 2013). The steps of ES-MDA are as follows.

- Select the number of iterations (number of data assimilation) N_a and set $\alpha = N_a$
- Initialize the selected ensemble members of model parameters and perturb the observation data for each ensemble member using,

$$d_{uc,j} = d_{obs} + \sqrt{\alpha} C_D^{1/2} z_d \tag{1.4}$$

  where d_{uc} is perturbed observation and z_d is the standard normal distribution $\mathcal{N}(0, I)$
- Update each ensemble member j using,

$$m_j^{(k+1)} = m_j^{(k)} + C_{MD}(C_{DD} + \alpha C_D)^{-1}(d_{uc,j} - d_j) \tag{1.5}$$

  Superscript k is the iteration index and α is the inflated parameter during multiple data assimilation.
- Repeat the above steps for all iterations, from $k = 1$ to N_a

1.4 Approaches to account for model-error during calibration

In the literature, different approaches have been used to account for model-error during calibration. These approaches can be used for modelling of reservoir model-error. From the literature, these approaches can be classified into three categories.

1. Input parameters-dependent versus output dependent
2. Deterministic versus stochastic
3. External bias description (EBD) versus internal noise description (IND)
4. Joint (coupled) calibration versus postprocessor (uncoupled) calibration

1.4.1 Input parameters-dependent versus output dependent

In input parameters-dependent approach, the model-error is defined as the function of the uncertain parameters of the model, which are estimated after calibration of the physical model. Reservoir model-error can vary with different permeability realizations or other uncertain parameters, which have to be estimated. O' Sullivan and Christie (2005) used this type of approach of accounting for model-error in their study. They estimated the oil viscosity while accounting for viscosity dependent model-error using linear interpolation. Lødøen and Tjelmeland (2010) used the multiple linear regression algorithm to formulate model-error. In their study, the residual of multiple regression is dependent on permeability realizations and they used Gaussian process regression (GPR) to model input dependent residual. Giudice et al. (2013) used input dependent model-error to improve uncertainty estimation in urban hydrological modelling. They used the variance of model-error dependent on the input (rainfall in their case). The general input dependent model-error/bias can be written as follows.

$$y(x) = y_m(x) + b_m(x, \beta), \tag{1.6}$$

where y is the truth function, x is input parameters, y_m is model response, b_m is the model-error/bias and β represents parameters of the model-error.

Output dependent model-error is the function of the model response/output. The output from the dynamic reservoir simulation model is in the form of time series of the well rates or pressure. In reservoir modelling context this model-error is only related to the outputs and its variation with respect to time. Omre et al. (2004) used this approach of model-error using multiple linear regression. They considered there

fine model as accurate (unbiased) and used linear regression to map the relationship between the coarse and fine grid. They only used outputs, which is in the form of time series, generated from fine and coarse-scale simulations. Evin et al. (2014a) used output dependent model-error/bias to improve uncertainty estimation in urban hydrological modelling and estimated the heteroscedasticity of model-error dependent on the outputs time series. Reichert and Schuwirth (2012a) linked statistical bias description to multiobjective model calibration. They used output dependent model-error/bias to improve uncertainty estimation in environmental modelling. The general output dependent bias can be written as follows.

$$y(x) = y_m(x) + b_m(y_m, \beta), \tag{1.7}$$

where y is the truth function, x is input parameters, y_m is model response, b_m is the model-error and β represent parameters of the model-error.

1.4.2 Deterministic versus stochastic

The deterministic approach depends on the algorithm, which is used for modelling bias/error. If the algorithm is deterministic for modelling of model-error then the approach can be considered as deterministic. The deterministic approach can be both input parameter-dependent and output dependent. Originally model-error/bias was introduced as a deterministic approach by Kennedy and O'Hagan (2001). The deterministic form can be written as follows.

$$y(x) = A_m(\beta)y_m(x) + B_m(\beta) + e_i, \tag{1.8}$$

where A_m and B_m are the parameters to map imperfect physical model to truth and e_i represents the measurement error.

The stochastic approach consists of a formulation which has some form of random component for the model-error. Giudice et al. (2013) used auto-regressive (AR) error-model in order to improve uncertainty estimation for urban hydrological modelling. They accounted for model-error/bias autocorrelation by stochastic approach using the

AR model. The explanation for the AR model is given in the model-error formulation section.

1.4.3 External bias description (EBD) versus internal noise description (IND)

EBD is the description of model-error/bias externally to the simulators during calibration. The bias/model-error has already been described as EBD in the previous sections, whether it is input/output dependent or deterministic/stochastic. The EBD was developed using the background of statistical inference in a regression type framework (Giudice et al., 2015). The mathematical forms of EBD is already mentioned in Eqs. 1.6, 1.7 and 1.8.

IND is the description of model-error/bias internally into the state space of the physical model instead of adding bias externally to the simulators or model outputs (Giudice et al., 2015). This approach is also known as state space modelling or stochastic gray-box modelling (Moradkhani et al., 2012; Kristensen et al., 2004). Mathematically the use of IND with simulators can be described as follows.

$$dS = y_m(x, S, t)dt + \sigma(x, S, \beta, t)dW(t), \tag{1.9}$$

where S is the state, $dW(t)$ is the standard Wiener process, σ represents diffusion term/state noise/level disturbance and accounts for modelling errors by making the uncertain or random states. The standard Wiener process is the random variable that depends on time i.e. $dW(t) \sim \sqrt{dt}\mathcal{N}(0, 1)$. This process is used to generate realizations of the errors with respect to time. Equation 1.9 shows the state evaluation using simulator with the addition of error realizations generated from standard Wiener process. Giudice et al. (2015) concluded that EBD has some advantages over IND in terms of long-term predictions.

1.4.4 Joint (coupled) calibration versus post-processor (uncoupled) calibration

Joint (coupled) calibration refers to the parameter estimation of the physical model and error-model simultaneously. The joint calibration approach had been used in most of the research work related to model-error (Giudice et al., 2015, 2013; Kennedy and O'Hagan, 2001; Lødøen and Tjelmeland, 2010). This approach could produce good results and have the ability to reduce the bias in estimated parameters when the prior of the model discrepancy is realistic (Brynjarsdottir and O' Hagan, 2014). However, sometimes joint calibration can be non-robust due to the unrealistic prior of the model discrepancy (Brynjarsdottir and O' Hagan, 2014) and the strong interaction between the physical model and the error-model (Evin et al., 2014a).

Post-processor (uncoupled) calibration refers to the parameter estimation of the physical model and error-model sequentially. Evin et al. (2014a) used the post-processor (uncoupled) calibration, where the parameters of the hydrological model were estimated in stage 1, followed by stage 2 where the parameters of a more complex residual error-model are estimated keeping the hydrological parameters fixed at the values estimated in stage 1. Evin et al. (2014a) compared the joint calibration approach to a post-processing approach of accounting for the model-error and concluded that the post-calibration approach was found to be more robust. However, this approach has limitation in the scenarios where model-error exhibits strong structural features and non-linear heteroscedasticity.

1.5 Methods for formulation of model-error/bias

Following methods and there variants have been used in literature for formulation of model-error/bias.

1. Auto-regressive integrated moving average (ARIMA)

2. Generalized auto-regressive conditional heteroscedastic (GARCH)

3. Multiple linear regression models

4. Gaussian process regression (GPR)

5. Approximate Bayesian computation (ABC)

1.5.1 ARIMA

Auto-regressive integrated moving average (ARIMA) models are widely used in econometrics and computational finance fields (Montgomery et al., 2015) for forecasting and prediction purpose. These models have been used for the formulation of the model-error/bias related to water resources research, rainfall prediction and environmental modelling (Giudice et al., 2013, 2015). These are useful methods when the model-error/bias shows random behavior and correlated outputs with respect to time. Therefore these methods are successfully used in hydrological and environmental models because most of the times the model-error of these physical systems shows random behavior and correlated outputs with respect to time.

ARIMA models are efficient only when the time-series are stationary and have constant variance with negligible heteroscedasticity. If there is heteroscedasticity, variance volatility with respect to time, the ARIMA model may not be a good choice to model time series (Montgomery et al., 2015). Non-stationary time series can be modeled using ARIMA if data is transformed in such a way that the time-series process becomes stationary. One of the most efficient ways to obtain a stationery distribution of the data is to apply a difference operator of the data of order D. Mainly ARIMA models consist of three parts; auto-regressive (AR), moving average (MA) and difference. The autocorrelation and partial autocorrelation functions are used to find out the order of MA and AR respectively (Montgomery et al., 2015). The general form of the ARIMA model is given in the following equation.

$$\Delta^D y_t = c + \phi_1 \Delta^D y_{t-1} + ... + \phi_p \Delta^D y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}, \tag{1.10}$$

where c is the constant, ϕ is the coefficient of previous lag observations, p is the order of

AR, q is the order of the MA, ϵ_t is the uncorrelated innovation or random component, θ is the coefficient of previous lag innovations and Δ^D is the difference (D order) of the time series to deal with the non-stationary condition. The model-error shows very well structured behaviour with small noisy components and often shows heteroscedasticity for subsurface systems. Therefore ARIMA methods for the formulation of subsurface model-error may not be a good choice.

1.5.2 GARCH

It is possible that the model-error/bias has heteroscedasticity and variance volatility. (Evin et al., 2014a) discussed this effect and considered error auto-correlation and heteroscedasticity in hydrological uncertainty estimation. Generalized auto-regressive conditional heteroscedastic (GARCH) can be used to formulate the type of model-errors, which show heteroscedasticity and variance volatility. The general GARCH model of order (p,q) is given in the following equation (Montgomery et al., 2015).

$$\sigma_t^2 = \kappa + \gamma_1\sigma_{t-1}^2 + ... + \gamma_p\sigma_{t-p}^2 + \alpha_1\epsilon_{t-1}^2 + ... + \alpha_q\epsilon_{t-q}^2, \tag{1.11}$$

where κ is the constant, σ_t is the variance, γ is the coefficient of previous lag variance observations, p is the order of AR, q is the order of the error terms, ϵ represents the error terms or random component, α is the coefficient of previous lag error terms. The model for conditional variance resembles an ARMA model. However, it should be noted that the GARCH model is not a proper ARIMA model, as this would have required a white noise error term with a constant variance for the MA part as shown in the following equation.

$$\epsilon_t = \sigma_t z_t, \tag{1.12}$$

where z_t is independent and identically distributed with mean 0 and variance 1.

1.5.3 Multiple linear regression models

Linear regression is one of the classical statistical methodologies to predict outputs of variables of interest given inputs. Multiple linear regression models are the same as linear regression but instead of one type of output, it is used to predict multiple sets of outputs given inputs (Johnson et al., 2002). The general form of linear regression is given below.

$$y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3 + ... + \beta_r z_r + \epsilon, \tag{1.13}$$

where y is the output, β represents the parameters of linear regression, z is input, r is the total number of input parameters, ϵ is the residual error. Multiple linear regression models consist of more than one linear regression model. The number of regression models depends on the number of output variables. Suppose there is a number n output variables require to predict. The general form of multiple regression models for multiple outputs y_1 to y_n are given below (Johnson et al., 2002).

$$y_1 = \beta_0 + \beta_1 z_{11} + \beta_2 z_{12} + \beta_3 z_{13} + ... + \beta_r z_{1r} + \epsilon_1, \tag{1.14}$$

$$y_2 = \beta_0 + \beta_1 z_{21} + \beta_2 z_{22} + \beta_3 z_{23} + ... + \beta_r z_{2r} + \epsilon_2, \tag{1.15}$$

$$y_3 = \beta_0 + \beta_1 z_{31} + \beta_2 z_{32} + \beta_3 z_{33} + ... + \beta_r z_{3r} + \epsilon_3, \tag{1.16}$$

$$y_n = \beta_0 + \beta_1 z_{n1} + \beta_2 z_{n2} + \beta_3 z_{n3} + ... + \beta_r z_{nr} + \epsilon_n. \tag{1.17}$$

In terms of matrix notation, multiple linear regression models can be written as follows.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ ... \\ ... \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & z_{12} & z_{13} & ..... & z_{1r} \\ 1 & z_{21} & z_{22} & z_{23} & ..... & z_{2r} \\ 1 & z_{31} & z_{32} & z_{33} & ..... & z_{3r} \\ . & ..... & ..... & ..... & ..... & ..... \\ . & ..... & ..... & ..... & ..... & ..... \\ 1 & z_{n1} & z_{n2} & z_{n3} & ..... & z_{nr} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ ... \\ ... \\ \beta_r \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ ... \\ ... \\ \epsilon_n \end{bmatrix}$$

Above matrix-notation can be written as the equation form in the following way.

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.18}$$

where $\mathbf{Y}$ is the column matrix of output variables of order $n \times 1$, $\boldsymbol{\beta}$ is the column matrix of linear regression parameters of order $(r+1) \times 1$, $\mathbf{Z}$ is the matrix of input variables of order $n \times (r+1)$, $\boldsymbol{\epsilon}$ is the column matrix of residual errors of order $n \times 1$.

Omre et al. (2004) used the multiple linear regression model to fit the model-error obtained from pairs of coarse and fine-scale model outputs. They tried to fit multiple regression models at multiple points of time series outputs generated by fine and coarse-scale simulation models. They considered the diagonal input matrix, which means that they did not consider the correlation between different points of time series outputs from the reservoir model. Lødøen and Tjelmeland (2010) also used multiple regression technique. Their approach was similar to Omre et al. (2004). The main differences are, they transformed their coarse-scale model with the multiple regression technique similar to Kennedy and O'Hagan (2001) and they modelled residual errors part of multiple regression using a zero-mean Gaussian Process.

1.5.4 Gaussian process regression

Gaussian process regression has the same use as linear regression or multiple regression models. In GPR instead of defining the model as a form of linear, quadratic, trigonometric, exponential or some parametric regression model, the model is defined as a non-parametric way using kernel functions (Rasmussen, 2006). One of the advantages of GPR, we don't need to define any function with a parametric regression model. This is achieved by defining the Gaussian process as function space with associated mean and covariance (Rasmussen, 2006). The general form of GPR is written in the following way.

$$f(x) \sim GP(m(x), k(x, x')), \tag{1.19}$$

$$m(x) = E[f(x)] \tag{1.20}$$

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))]. \tag{1.21}$$

Bayesian linear regression model is a Gaussian process with mean zero and covariance $k(x, x')$ and can be implemented with any appropriate kernel function (Rasmussen, 2006). This form of GPR can be written in the following way with zero mean and covariance function of squared exponential form.

$$E[f(x)] = 0, \tag{1.22}$$

$$k(x_p, x_q) = \sigma_f^2 exp(-\frac{1}{2l^2}|x_p - x_q|^2), \tag{1.23}$$

where σ_f^2 and l are the hyper-parameters of GPR. Hyper-parameters are the free parameters in GPR, which can be adjusted or trained to match the data. Equation 1.19 can be applied by computing and sampling from the covariance function of the quantity of interest, where the function $f(x)$ can be used to map the errors with respect to time. The reservoir models produce the outputs in terms of time series which can be mapped using GPR by adjusting hyperparameters related to particular model-error/bias. For the well-structured model-error/bias, GPR has the advantage over ARIMA and GARCH, but for random and stochastic model-error/bias ARIMA and GARCH have the advantage over GPR.

1.5.5 Approximate Bayesian computation (ABC)

Sargsyan et al. (2015) introduced the approximate Bayesian computation (ABC) methodology for the statistical calibration of imperfect physical models. They described synthetic likelihood function, which considered the misfit between mean prediction from physical models and observation data, and an additional constraint of misfit between the standard deviation of physical models outputs and deviation of observation data with mean prediction. The synthetic likelihood function for ABC methodology is shown below.

$$L_{ABC}(\alpha) = \frac{1}{\epsilon\sqrt{2\pi}} \prod_{i=1}^{N} exp(-\frac{(\mu_i(\alpha) - y_i)^2 + (\sigma_i(\alpha) - \gamma|\mu_i(\alpha) - y_i|)^2}{2\epsilon}), \tag{1.24}$$

where α is the model parameters, μ_i is the mean prediction, y_i is the observations from the real system, N is the total number of observations, ϵ is the tolerance parameter and γ is the weight parameter. Approximate Bayesian computation (ABC) does not require joint calibration or model-error formulation and is useful for physical models where the realistic prior of model discrepancy is unknown. However, so far approximate Bayesian computation (ABC) is implemented using MCMC methods and the implementation is not clearly understood using ensemble-based methods. Moreover, the extent of accuracy of ABC is not known for the calibration of complex oil and gas reservoir models especially in the presence of both measurement error and model-error/bias.

1.6 Thesis objectives

The primary research objective of this thesis is to develop a flexible Bayesian framework, which accounts for model-error during the calibration process in order to improve the prediction capacity of the imperfect reservoir models. In the Bayesian framework, the modelling error can be considered as a random variable, and by using an estimate of the probability distribution of the unknown, one may estimate the probability distribution of the modelling error and incorporate it into the calibration process (Calvetti et al., 2014, 2018). However, the model-error may also have a systematic bias especially when it is related to numerical dispersion or coarse-scale representation of real physical system. Sometimes this bias is obtained as a mean but it could be deterministic. In this work, the novel algorithms are introduced which represent model-error in the form of principal component analysis based error-model in the Bayesian framework. The proposed Bayesian framework leads to a joint estimation of the physical model and error-model parameters which has the capacity to reduce or eliminate bias in the estimated approximate posterior distribution of the physical model parameters that shows the potential to capture the underlying truth with increasing accuracy. In order to achieve this objective, the research program is divided into three major parts. In the first part, sources of the model-error are assumed to be known so that the prior model-error statistics can be estimated before calibration. The first part of the thesis

objective is further elaborated by following points.

- Principle component analysis is used to formulate the model-error.
- Joint calibration approach is used to estimate subsurface flow model and error-model parameters.
- This procedure is flexible for large scale models with variable boundary conditions, which fills the significant gap in the literature.

The second part of the thesis objective is related to the investigation and comparison of the different algorithms for the calibration of imperfect reservoir models while accounting for the model-error. In this part, two new algorithms are introduced which are the variants of the proposed methodology of the first part of the thesis objective and published algorithms. These algorithms are extensively evaluated and compared to recently published algorithms which have the capacity to handle:

- Large scale models (e.g. subsurface oil and gas reservoirs).
- Physical systems which have strong serial correlated (structured) outputs.
- Physical systems with variable boundary conditions.
- Model-errors, which have strong correlation in time (i.e. structured errors).

The third part of the thesis objective is related to account for unknown sources of the model-error during calibration. These situations are challenging and different formulations/approaches are required to account for unknown model-error other than joint calibration of physical model and error-model. Therefore a flexible ensemble-based algorithm is developed for the calibration of the imperfect model which can account for the unknown model-error from the residual (data mismatch) of the simulation model.

1.7 Summary of the thesis Chapters

The summary of each chapter of the thesis is given below. Each chapter was initially written as separate paper, therefore it intends to be self-contained.

1.7.1 Summary of Chapter 2

Chapter 2 presents the first part of the objectives of the thesis, where the sources of the model-errors are known and prior model-error statistics could be estimated before calibration. The main motivation of this chapter is to account for model-error of low-fidelity reservoir models because subsurface reservoirs are more heterogeneous and complex than the simulation models in terms of scale, assumptions and description. In this chapter, the issue of prediction reliability while calibrating imperfect/low-fidelity reservoir models has been addressed. The main goal is to avoid over-confident and inaccurate predictions by including a model for the bias terms (i.e. error-model of a predefined form) during the history matching/calibration process. The aim is to obtain nearly unbiased posterior distributions of the physical model parameters thus improving the prediction capacity of the calibrated low-fidelity reservoir models. In order to achieve this aim, the Bayesian framework is adopted for the inversion of the imperfect physical model. The Bayesian approach could provide a natural framework to the inversion of the imperfect physical model if model-error is mapped in the form of function approximation and defined as the prior error-model. In this chapter, the idea is to combine the imperfect physical model and error-model in the form of one combined model such that the error-model would take into account the discrepancy of the imperfect physical model. This theory leads to the approach of the joint estimation of parameters of the imperfect physical model and error-model in order to obtain the nearly unbiased posterior distribution of the parameters of the imperfect physical model and reliable predictions. The posterior distribution of the error-model parameters could potentially yield a better approximation of the posterior distribution of imperfect physical model parameters and has the potential to provide the predictions of higher accuracy.

In this chapter, the principal component analysis (PCA) based error-model formulation is proposed to improve the calibration and prediction capacity of imperfect subsurface flow models. The model-errors of the subsurface reservoirs are not completely random in nature, therefore the error-model is formulated using PCA. The coupled calibration approach is used as a joint estimation of the parameters of the

imperfect subsurface flow-model and the error-model parameters. The structure of the error-model and the prior distributions of the error-model parameters are evaluated before calibration through analysis of leading sources of the model-errors and the prior statistics of the model-error. The Bayesian framework is adopted for solving the inverse problem using ensemble smoother with multiple data assimilation (ES-MDA) algorithm.

In this chapter, two types of history matching procedures are investigated: history matching while neglecting model-discrepancy (i.e. standard history matching procedure) and joint history matching of the model parameters and the parameters of an error/bias model. The standard history matching procedure relies on an implicit assumption that the model-errors are generally small and could be neglected (i.e. the simulation model is perfect). For the joint history matching procedure, an EBD and input/output independent error-model formulation are utilized as it is more suitable for large scale models (e.g. subsurface oil and gas reservoir models) and variable boundary conditions (e.g. different well controls and open/shut schedules). Several typical sources of model-errors are present in the test cases investigated in this chapter, including a coarse grid, less detailed geological representation (i.e. upscaling of different types of geological features including variogram based and channelized geology), discretization errors and a slight change in well locations due to grid coarsening. The model-errors of subsurface reservoirs are dominated by structured components, these errors are parameterized using smooth basis functions obtained by principle component analysis (PCA) method. The prior realizations of the model-errors are computed from the difference of the simulation output from pairs of models, accurate/high-fidelity versus approximate/low-fidelity. These prior realizations of the model-errors are used to obtain the basis functions and the prior statistics of the coefficients of the PCA basis functions. This limits the applicability of the developed approach to the cases for which an accurate/high-fidelity model is available. However, the accurate/high-fidelity model is only used to estimate the prior model-error statistics and is not used during the calibration process in the proposed methodology. The error-model formulation presented here assumes that the total modelling errors consist of two components: structural compo-

nent and noise-like component. The noise-like part of the model-error is also accounted for during the history matching process to avoid over-fitting of the error-model. The noise-like part is quantified using error-model misfits ζ which are estimated by taking the difference of the prior realizations of actual model-error and corresponding fit of the error-model. In this chapter prior to model calibration, the structure of the model-error is estimated and represented by using several basis functions. The magnitude of the noise-like component is accounted for using covariance of error-model misfit $\mathbf{C}_T$ and added into the measurement-noise covariance $\mathbf{C}_d$. This step introduced additional perturbation to the ES-MDA algorithm because of the effect of the total error covariance i.e. $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_T$. During history matching, the weights of the basis functions are jointly calibrated with the physical model parameters using data observed at well locations. The presented formulation is general and can be applied to other sources of modelling errors when dealing with low-fidelity subsurface models.

In this chapter, the subsurface reservoir has the dimensions of 7500 ft $\times$ 7500 ft $\times$ 20 ft in the x, y and z directions, respectively and has the two-phases (oil and water) in the porous media. The initial reservoir pressure is 5000 psi and the reservoir has a uniform porosity of 20%. The reservoir contains one injector well (I1) and three production wells (P1, P2, P3) and is simulated using a 2D grid with Matlab Reservoir Simulation Toolbox (MRST) (Lie, 2016). In the historical period, the flow rates at the production wells and the bottom hole pressure of the injector well are used as the historical data for the calibration process. One of the production wells is only used in the forecast phase in order to assess predictions from calibrated models on wells drilled in future development plans. Two test cases have been used in this chapter to investigate the performance of the joint calibration of simulation and error-model parameters. These test cases have typical model-errors originating from grid coarsening/up-scaling and from utilizing an imperfect geological model description. The first test case is related to the estimation of the uncertain log-permeability field originating from two-point statistics. The distributed log-permeability fields are parameterized by retaining two leading principal components obtained by singular value decomposition of the covariance of the log-permeability fields. The prior statistics of the model-error are estimated by us-

ing one hundred model discrepancy realizations generated from pairs of fine-scale/high fidelity and coarse-scale/low fidelity models. The fine-grid/high-fidelity model uses a 2D grid with 75×75 cells and used to generate the observed data perturbed with the measurement noise. The coarse-grid/low-fidelity reservoir model contains only 5×5 grid blocks and is used for the calibration/history matching procedures. The second test case is related to the estimation of the uncertain log-permeability field originating from multi-point statistics. The fine-scale/high fidelity model contains channelized features of the log-permeability fields and these channels are disappeared in coarse-model/low fidelity due to the aggressive up-scaling/coarsening of the log-permeability field. The prior statistics of the model-error are estimated by using one hundred model discrepancy realizations generated from pairs of channelized permeability images of size 75×75 grid blocks and the corresponding up-scaled permeability field with 5×5 grid blocks which do not have geological channels due to aggressive up-scaling of the log-permeability field. The observed data is generated from the fine-model with the channelized permeability field by the addition of random perturbation of the measurement noise. The coarse-grid/low-fidelity reservoir model (5×5 grid blocks) is used for the calibration/history matching procedures.

In this chapter, the results are obtained using the ES-MDA algorithm with eight iterations for calibration of test cases 1 and 2 with and without accounting for the model-discrepancy. In the case of joint calibration with error-model, two PCA components were retained per each output time series to parametrize the model-discrepancy. The results for test case 1 show that the posterior distribution from the calibration while neglecting model-error are biased and the estimated basis weights of the log-permeability field do not capture the true weights. Therefore, the predictions for all wells are inaccurate from the calibrated model neglecting model-error. However, the posterior distributions obtained by the joint inversion procedure are less biased (i.e. nearly unbiased) and successfully cover the true model parameters, which result in better matches and predictions for all wells. Similar results are obtained for test case 2, which show the generality and consistency of the proposed methodology. The calibrated models are evaluated using three different forecasting metrics to assess the

quality of the estimated parameters and the capacity of the calibrated models in making future predictions. The utilized forecasting metrics are: coverage probability (CP), mean continuous ranked probability score (CRPS) and mean square error (MSE). CP indicates the fraction of the actual data that lie within the specified confidence interval of the estimation. Mean CRPS quantifies both accuracy and precision (Hersbach, 2000) and higher values of CRPS indicate less accurate results. MSE is widely used as a metric for parameter estimation problems. However, MSE measures the quality of data-fitting and is not enough to provide a probabilistic assessment of the estimation and prediction from an ensemble of models. These metrics provide a good assessment on the consistency, reliability, and accuracy of the forecasting capacity of the calibrated models. With the joint inversion procedure, the CP is improved for all three quantities investigated (log-permeabilities, historical data, future prediction). Similarly, the mean CRPS and MSE measures also show significant improvement by accounting for model-error using the joint inversion procedure.

The methodology proposed in this chapter can be extended to different fine-scale (high fidelity) reservoir simulation, which would be very expensive to run and may not be feasible for uncertainty quantification and history matching problems, and its corresponding coarse-scale (low fidelity) simulation. The methodology is also useful for history matching of realistic geological models by accounting errors with respective up-scaled imperfect geological descriptions or parametrization. In this chapter, prior statistics of model-errors are estimated from known sources of errors and incorporated in the calibration of imperfect reservoir models. There are some limitations of the proposed methodology, that one must know the corresponding prior statistics of the model-error. If prior statistics of model-error are not known, then the error-model may dominate the simulator and estimated parameters would be under-predicted. This may also result in bad predictions. It is also possible that the fine-scale model and assumed geology are also wrong or highly biased. In that case, the first methodology can only improve parameter estimation and prediction up to the limit of the fine-scale model, which is used to learn prior statistics of model-error.

The idea of the joint calibration of the physical model with PCA based error-model

was introduced by the candidate's (main author) supervisor. The candidate implemented the joint calibration approach and obtained all corresponding results related to two different test cases of the imperfect reservoir models. The idea to account for additional covariance of the truncation error of the PCA based error-model was introduced by the candidate. The truncation error of the PCA based error-model is termed as error-model misfit in the Chapter 2. There is an additional co-author from Geoscience Research center, TOTAL (sponsors of the candidate's PhD work). She suggested the test cases for the evaluation of proposed joint calibration approach of the physical model and PCA based error-model. The test cases were inspired by the fact that the subsurface reservoirs are more heterogeneous and complex than the simulation models in terms of scale and description. Therefore the suggestion was to take the test cases which is related to the coarse-scale and imperfect geological description. Furthermore she also suggested forecasting metrics to evaluate and quantify the performance of the proposed joint calibration approach. The forecasting metrics were implemented by the candidate which allows to quantify the performance of the ES-MDA while neglecting and accounting for model-error. The candidate wrote the published paper related to Chapter 2, which was critically reviewed and edited by the second and third author.

1.7.2 Summary of Chapter 3

Chapter 3 presents the investigation and evaluation of six different algorithms for calibration (history matching) of imperfect physical models with the second part of the thesis objective. The main focus is on flexible algorithms that can handle strong serially correlated outputs of the physical model, variable boundary conditions (i.e. variable well open/shut schedules and rate/pressure controls) and structured errors (i.e. strong correlation in time). Among those flexible approaches, Algorithms 3 and 4 are related to the methodology proposed in Chapter 2. Algorithms 5 and 6 are newly introduced approaches which are variants of the methodology proposed in Chapter 2 and the approaches by Köpke et al. (2018) and Oliver and Alfonzo (2018) respectively. The novelty in Algorithms 5 and 6 is introduced in terms of identifiability of model-error parameters and the use of joint calibration of both the pre-determined error-model and the physical

model. The evaluation is performed on a test case representing model discrepancy due to grid coarsening and upscaling of geological features. This test case is motivated by the fact that generally geological models are a coarse representation of reality. Ensemble smoother with multiple data assimilation (ES-MDA) (Emerick and Reynolds, 2013) is used as a Bayesian inversion algorithm.

In this chapter, the first algorithm (base case scenario) relies on Bayesian inversion while neglecting the model-error using ES-MDA. In the second algorithm, the residual obtained after calibration is used to iteratively update the total error covariance combining the effects of both modelling errors and measurement errors. This algorithm is inspired by the work of Oliver and Alfonzo (2018). In this algorithm, the covariance matrix of the total error is initially set to equal the covariance matrix of the measurement errors. In subsequent iterations, the covariance matrix of the total error is updated based on the residual from the previous history matching iteration. The algorithm is terminated after satisfying a model diagnostic criteria or reaching a prespecified maximum number of iterations. In the third algorithm, the PCA-based error-model is used to represent the model discrepancy during history matching. This leads to a join inversion problem where both the model parameters and the parameters of a PCA-based error-model are estimated. For the joint inversion within the Bayesian framework, prior distributions have to be defined for all the estimated parameters and the prior distribution for the PCA-based error-model parameters are generally hard to define. In this study, the prior statistics of the parameters of the PCA-based error-model are estimated using the difference of outputs obtained from prior realizations of high-fidelity and low-fidelity models. In this formulation, the predictions are made by summing the simulation outputs using the estimated physical model parameters and estimated model-errors represented by the PCA-based formulation.

The fourth algorithm is similar to the third algorithm, however, an additional covariance matrix of difference between the PCA-based error-model and the corresponding actual realizations of prior error is added to the covariance matrix of the measurement error as shown in Chapter 2. The difference between prior realizations of actual model-errors and corresponding fit of the PCA-based error-model due to the truncation effect

are called as second-order errors. This remaining mismatch (i.e. second-order errors) are loosely captured by inflating the observation noise covariance with a diagonal matrix, which is the variance of the remaining mismatch before history matching $\mathbf{C}_T$ and added into the measurement-noise covariance $\mathbf{C}_d$. This step introduced additional perturbation and regularization to the ES-MDA algorithm because of the effect of the total error covariance i.e. $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_T$.

In this chapter, the fifth algorithm (first newly introduced approach) is the variant of the third algorithm and the approach proposed by Köpke et al. (2018). In this algorithm, the data residual (the difference between observations and model outputs) is filtered to remove the second-order model-error effect. Köpke et al. (2018) used an orthonormal basis, which is generated from the differences between pairs of high-fidelity (accurate) and low-fidelity models, to filter the model-error components from the calibration residuals during the history matching of imperfect models. In this algorithm, the filtering approach relies on building an orthogonal basis $\mathbf{B}$ for the error-model misfit component ζ, which is obtained from the difference between the PCA-based error-model and the corresponding actual realizations of the prior error. An additional orthonormal basis is built on second-order errors using the Gram-Schmidt GS procedure i.e. $\mathbf{B} = GS(\zeta)$. These basis functions are projected onto the residual/data mismatch $\mathbf{R}$ using $\mathbf{E} = \mathbf{B}\mathbf{B}^{\intercal}\mathbf{R}$ to filter the data residual during the joint calibration process by updating the observation using $\mathbf{D}_{obs} = \mathbf{d}_{obs}\overrightarrow{1_{N_e}} - \mathbf{E}$. This additional step introduced additional perturbation and regularization to the ES-MDA algorithm in order to prevent over-fitting the model parameters to data features that are not captured by the corrected model. This procedure differs from the work presented in (Köpke et al., 2018), where the orthonormal basis is built directly from the difference between the approximate physical model and the perfect model without introducing any functional approximation to the model discrepancy. In some sense, the second-order errors in algorithm 5 are filtered before each history matching iteration while in algorithm 4 the second-order errors are estimated before history matching and remain same during iterations.

The sixth approach (second newly introduced algorithm) is the variant of the algorithm 3 and the approach introduced by Oliver and Alfonzo (2018), where the model

discrepancy is described by the addition of physically motivated bias corrections terms with the estimation of the total error covariance matrix of data residuals $\mathbf{R}$ obtained after history matching using $\mathbf{C}_D = \frac{1}{N_e}(\mathbf{R})(\mathbf{R})^\intercal$. This procedure introduced additional perturbation to the ES-MDA algorithm during recalibration steps. Oliver and Alfonzo (2018) added physically motivated parameters for bias corrections and estimated the total error covariance matrix using an ensemble of data residuals from the calibrated models. In the current implementation, the PCA-based error-model is utilized as a physically motivated bias correction term with an iterative update of the covariance matrix of the total error during calibration.

The calibration performance of these six algorithms are evaluated on a simplified reservoir model similar to test case 2 presented in Chapter 2. The true permeability field is a channelized model of dimension 75×75 generated using multiple-point statistics (MPS) (Mariethoz and Caers, 2014). This reference model is used to generate the observation data with the addition of random perturbation of measurement-noise. The simulation model used for history matching is a much coarser model with only 5×5 cells with the corresponding up-scaled (5×5) field of the log-permeability. The up-scaled model is obtained from harmonic averaging of the 75×75 high-fidelity model. For the up-scaled model, the physical model parameters are the permeability values at the 25 cells. In this test case, the model-errors are a consequence of aggressive grid coarsening, slight change in well locations and poor geological representation. The implementation of the last four algorithms requires extraction of the PCA basis functions for the model discrepancy and the corresponding prior weights for the parameters of the error-model based on output differences between pairs of high-fidelity and low-fidelity simulation models generated from the prior realizations of the permeability field. For this step, 100 realizations of the model discrepancy are created by taking the difference of the fine-scale model output/response using 100 log-permeability fields $(75 \times 75$) generated by multi-point statistics and the corresponding up-scaled model responses (harmonic averaging). The up-scaling of the permeability fields to a 5×5 grid blocks is quite aggressive which results in large model discrepancies.

An ensemble of 100 members is used with eight number of iterations for calibra-

tion using these six algorithms. For the last four algorithms, two PCA components for each time series are retained to parameterize the model-error. The first algorithm results in the biased posterior distribution of $\ln(\mathbf{K})$ due to the negligence of model-errors during calibration. Moreover, the estimated log-permeabilities show extreme values (e.g. overshooting) and the posterior distribution failed to cover the reference coarse log-permeabilities as a consequence the quality of the predictions is very bad due to negligence of the model-errors during the model calibration process. The overshooting of the physical model parameters is reduced to a certain degree by using the second algorithm for calibration, where total errors are estimated at the end of each ES-MDA calibration step. However, the quality of predictions is also quite biased with slightly broader uncertainty intervals compared to algorithm 1. Third algorithm results in additional improvements in the estimated physical model parameters due to the joint inversion of parameters of the physical model and parameters of the error-model. However, the estimated distribution of parameters of the physical model does not cover the true solution very well due to the limited capacity of the error-model as only two PCA basis for each time series are retained. The predictions are improved using algorithm 3 due to the inclusion of the error-model term in the calibration process. However, the quality of the predictions deteriorates with time. This is likely a consequence of a poor split of the total residual between the error-model and the physical model. Algorithms 4, 5 and 6 show very good results in terms of reducing the bias/overshooting of the physical model parameters and predictions, due to the inclusion of second-order errors in addition to joint estimation of parameters of the physical model and parameters of the error-model. These six algorithms are evaluated using the forecasting measures of coverage probability (CP), mean continuous ranked probability score (CRPS) and mean square error (MSE). The obtained results show the improvement in all quantities investigated (log-permeabilities, historical data, future prediction), which indicates that a good parameterization of the error-model is needed in order to obtain a good estimate of physical model parameters and to provide better predictions.

In this chapter, the last three algorithms (i.e. 4, 5, 6) outperform the other algorithms in terms of the quality of estimated model parameters and the prediction

capability of the calibrated imperfect models. The common feature among algorithms 4, 5 and 6 is the joint calibration along with corresponding variants of accounting for an additional term that captures some remaining errors (a.k.a. second-order errors) that are not captured by the error-model. When the calibrated models are used for prediction, accounting for the second-order errors seems to allow for a better split (relative contribution) between the physical model contributed and the error-model contribution. The relative contribution of the error-model in prediction can also be used as an indicator of the magnitude of the model-error. If the contribution from the error-model is high, this could be used as an indicator of the need for model refinement and/or re-evaluation of the model assumptions. Algorithms 4, 5 and 6 scale very well with the model dimension and can be used for large scale problems because the model-error is parameterized in the output space (production profiles).

The idea of the evaluation and investigation of different algorithms was introduced by the candidate's (main author) supervisor. The candidate introduced two new algorithms based on combining the ideas of the methodology proposed in Chapter 2 and published algorithms. The candidate implemented all algorithms under consideration and obtained all corresponding results related to test case 2 of the Chapter 2. There is an additional co-author from Geoscience Research center, TOTAL (sponsors of the candidate's PhD work). She suggested the test case for the evaluation and investigation of different algorithms under consideration. The test case was inspired by the fact that generally geological models are a coarse representation of reality. Therefore the suggestion was to take the test case which is related to the coarse-scale representation of the imperfect geological description. Furthermore she also suggested forecasting metrics to evaluate and quantify the performance of the different algorithms under consideration. The forecasting metrics were implemented by the candidate which allows to quantify the performance of the algorithms. The candidate wrote the published paper related to Chapter 3, which was critically reviewed and edited by the second and third author.

1.7.3 Summary of Chapter 4

The proposed Algorithms in Chapters 2 and 3 require prior model-error statistics before

the joint calibration of the physical and error-model parameters. The prior model-error statistics is usually found by taking difference between high-fidelity/accurate model and low-fidelity/imperfect model outputs. In real life applications, an accurate model may not exist, but a high-fidelity model can still be constructed to be closer to reality compared to the low-fidelity/imperfect model and approximate prior model-error statistics could be estimated. However, in some situations one may have access to only imperfect/low-fidelity model and there is a possibility that high-fidelity model contains a large magnitude of the unknown model errors. This complex situation has been addressed in the Chapter 4, which presents the third part of the thesis objectives, where the model-error during calibration is accounted for without knowing any source and prior statistics of model discrepancy. Joint calibration can be non-robust if the prior statistics of model-error is not realistic (Brynjarsdottir and O' Hagan, 2014). Therefore, the third part of the thesis objectives is challenging and requires a flexible ensemble-based algorithm which can reduce bias after calibration of imperfect models in order to improve the prediction capability/reliability of the calibrated physical model. The aim of this chapter is to present the flexible ensemble-based algorithm which can be used to quantify the unknown model-error uncertainty during calibration of the imperfect physical model and match the data in case of the perfect model. Generally, we are uncertain about the extent of the accuracy of the physical model. The main objective of the flexible ensemble-based algorithm to reduce the uncertainty up to the limit of the accuracy of the model so that the proposed algorithm could also be used as a diagnostic tool check to identify the reliability of the physical models and the need for model refinement step.

This chapter presents the formulation of flexible iterative ensemble smoother, which can be used to calibrate imperfect models where model-errors cannot be neglected. This algorithm is the modification of the ensemble smoother with multiple data assimilation (ES-MDA) as it has been widely used for calibrating simulators of various physical systems due to the relatively low computational cost and the parallel nature of the algorithm. However, ES-MDA had been designed for perfect models under the main assumption that the specified physical/simulation models have the capability to

model the reality accurately. The main idea of the proposed algorithm is to extract the structural features of the model-error from the data mismatch (residual). In the proposed algorithm, the residual (data mismatch) is split into two parts. The first part of the residual is used to update the model-parameters and the second part is used to represent the model-error. The initial split parameter is computed based on the ratio of the norm of mean residual (mean deviation from observed data) and norm of maximum residual (maximum absolute deviation from observed data). During calibration (data assimilation) this split parameter is updated based on the ratio of the norm of mean residuals obtained in current and previous iterations. The proposed split formulation shows very close correspondence to best split (computed from true model-error) during data assimilation. In this chapter, three test cases have been used to observe the performance of the proposed algorithm. These test cases are related to the calibration of polynomial functions, simple machine, and imperfect reservoir models. For comparison purpose, calibration is performed using both the standard ES-MDA (Emerick and Reynolds, 2013) and the proposed Flexible ES-MDA algorithm. An ensemble of 100 members is used with 8 iterations for the calibration of all test cases.

In the first test case, the cubic function is considered as the perfect model (the data is generated from a cubic polynomial function and perturbed with an additive measurement noise) and imperfect models are represented by quadratic and linear functions. In this test case, the objective of the calibration of three models (cubic, quadratic and linear models) is to obtain the posterior distributions of coefficients of each model and the corresponding models' outputs and to evaluate the flexibility of the proposed algorithm in quantifying the model-error uncertainty for imperfect models as well as the ability to match the data for the perfect model case. The calibration results show that both standard ES-MDA and Flexible ES-MDA match the data for the cubic (perfect) model case. However, standard ES-MDA fails to match the data and results in bias estimation of coefficients of imperfect (quadratic and linear) models due to the negligence of model-error uncertainty during calibration. The proposed Flexible ES-MDA reduces the bias in estimated coefficients of imperfect models by capturing the unknown model-error uncertainty during calibration. The prediction interval coverage

probability (PICP) of posterior distributions is used as one of the calibration metrics. PICP is estimated by counting the number of observations in the confidence intervals (10% to 99%) of the posterior distribution normalized by the total number of observations (Xu and Valocchi, 2015). PICP is a very useful metric for quantifying both under-estimation and/or over-estimation of uncertainties in the obtained posterior distributions (Xu and Valocchi, 2015). The posterior distribution of cubic (perfect) model output obtained by the standard ES-MDA underestimates the uncertainty. However, the proposed Flexible ES-MDA shows more robust uncertainty quantification. This effect is due to the efficient coverage of the noisy features of the observation data by the proposed algorithm. PICP of the posterior distribution obtained from ES-MDA for the imperfect linear and quadratic models show a severe underestimation of uncertainties as the model-error uncertainties are missing in the formulation. A more robust PICP values obtained from the proposed algorithm for calibrating the imperfect linear and quadratic models. Calibration results from the proposed algorithm show lower CRPS values for linear, quadratic and cubic models as compare to standard ES-MDA, which indicates more reliable results in terms of precision and accuracy. MSE shows a more widespread distribution of the posterior ensemble due to the coverage of unknown model-error uncertainty using Flexible ES-MDA, however lower values of MSE is also observed as compared to standard ES-MDA due to significant reduction in model-bias.

The second test case is related to the estimation of efficiency of the simple machine model which lacks physics in terms of friction component. In this test case, the simple machine model is calibrated with the observation data obtained from the unknown complex machine. The main objective is to test the flexibility of the proposed algorithm for the cases, where the model lacks some physics. In this test case, standard ES-MDA and the proposed Flexible ES-MDA match the historical/training data, however, the predictions obtained from the calibrated model using ES-MDA are overconfident, inaccurate and unreliable. This is due to the estimation of the biased posterior distribution of efficiency (parameter) of the simple machine. This bias in the efficiency (parameter) of the machine is reduced significantly and posterior distribution covers the true efficiency of the machine using the proposed Flexible ES-MDA. Due to this effect, more reliable

predictions are obtained from a calibrated simple machine model using the proposed algorithm. The PICP of training data shows more robust uncertainty quantification of training data using the proposed Flexible ES-MDA. The prediction PICP is also improved by the Flexible ES-MDA, which shows the increase in predictions reliability from calibrated imperfect models (which lacks some physics). Similar results are observed in terms of mean CRPS and MSE metrics for both historical/training and prediction data.

The third test case is related to the calibration of the imperfect reservoir model, which has a blurred channelized geological patterns due to the up-scaling and grid coarsening of very fine geological patterns. The imperfect reservoir model has two types of model-errors, i.e. simplified geological representation and up-scaling errors. The true model consists of very fine and channelized geological patterns which consist of 75×75 grid blocks and used to generate the observation data by the addition of random perturbation of measurement noise. The imperfect reservoir model requires calibration for estimation of the uncertain log-permeability field which consists of 15×15 grid blocks so that the calibrated model can be utilized for prediction purposes. The standard ES-MDA results in over-shooting in most of the grid-block values of the estimated geological pattern of mean log-permeability. In this case, ES-MDA tried to aggressively adjust the model parameters to match the observation while neglecting the model-errors. The proposed Flexible ES-MDA results in a smooth geological pattern of the log-permeability field while accounting for unknown model-error effects using a split parameter. Moreover, relatively high values of standard deviation from Flexible ES-MDA are observed as compared to ES-MDA, due to the effect of additional uncertainty of unknown model-error. The standard ES-MDA performs well in terms of matching the observed data, however, poor quality predictions are obtained from the calibrated model. These over-confident and inaccurate predictions are due to the over-shooting of log-permeability values around production wells. This is a common problem in the petroleum industry, where the historical data is usually matched and often the calibrated models suffer from severe predictability problem (Carter et al., 2006). One of the primary reasons for this problem is failing to account for model-error effects during

the history matching/calibration process. In contrast, the proposed Flexible ES-MDA algorithm performs relatively better in terms of predictions. This could be easily attributed to the smooth mean log-permeability fields of the posterior ensemble. The Flexible ES-MDA shows very robust uncertainty estimates for the model parameters (log-permeability field). In addition, the proposed algorithm shows improved uncertainty estimate of historical production data and the prediction PICP is also improved as compared to standard ES-MDA, which shows the increase in prediction reliability from calibrated imperfect reservoir models. The mean CRPS and MSE of the historical data show lower values for standard ES-MDA as compared to Flexible ES-MDA due to relatively better matching. However, Flexible ES-MDA shows significant improvement for mean CRPS and MSE of prediction data. The good matching obtained when using ES-MDA can mislead us to over-confident and inaccurate predictions in case of imperfect models and this over-confidence in inaccurate predictions can be avoided using proposed Flexible ES-MDA.

The improvement in the physical model is recommended in terms of physics, assumptions, details, and description if additional model-error uncertainty is indicated by the proposed algorithm. Moreover, the flexibility of the proposed algorithm to account for unknown model-error uncertainty can give an idea about the reliability of the calibrated model. The reduction in uncertainty is required for effective decision making to the threshold value (i.e. minimum value which affects the decision-making process). If the model-error uncertainty is greater than the threshold value, then the model refinement step is needed to reduce the model-error uncertainty. However, if there is a time-constraint for the decision-making process and model-error uncertainty is below the threshold value then the decision can be made from the calibrated imperfect models using the proposed algorithm.

The last chapter presents the conclusions of the thesis.

Chapter 2

Quantification of prediction uncertainty using imperfect subsurface models with model error estimation

Subsurface reservoirs are far more heterogeneous and complex than simulation models in terms of scale, assumptions and description. In this work, we address the issue of prediction reliability while calibrating imperfect/low-fidelity reservoir models. The main goal is to avoid over-confident and inaccurate predictions by including a model for the bias terms (i.e. error-model of a predefined form) during the history matching process. Our aim is to obtain unbiased posterior distributions of the physical model parameters and thus improving the prediction capacity of the calibrated low-fidelity reservoir models. We formulate the parameter estimation problem as a joint estimation of the imperfect model parameters and the error-model parameters. The structure of the error-model and the prior distributions of the error-model parameters are evaluated before calibration through analysis of leading sources of the modeling errors. We adopt a Bayesian framework for solving the inverse problem, where we utilize the ensemble smoother with multiple data assimilation (ES-MDA) as a practical history matching algorithm.

We provide two test cases, where the impact of typical model errors originating from grid coarsening/up-scaling and from utilizing an imperfect geological model description is investigated. For both cases results from the ES-MDA update with and

The contents of this chapter have been published in Journal of Hydrology (2019), but includes modifications following the viva. https://doi.org/10.1016/j.jhydrol.2019.02.056

without accounting for model error are compared in terms of estimated physical model parameters, quality of match to historical data and forecasting ability compared to held out data. The test results show that calibration of the imperfect physical model without accounting for model errors results in extreme values of the calibrated model parameters and a biased posterior distribution. By accounting for modeling errors the posterior distribution of the model parameters is less biased (i.e. nearly unbiased) and improved forecasting skills with higher prediction accuracy/reliability is observed. Moreover, the consistency between the different runs of the ES-MDA is improved by including the modeling error component. Although the examples in the chapter consider the oil-water system with permeabilities being parameters of the physical model, the developed methodology is general and can be applied to typical ground water hydrology models.

2.1 Introduction

In subsurface reservoir modeling, various approximations are introduced at different stages of the modeling process which in turn render most of the models to be imperfect and low-fidelity in nature. However, these imperfect models are generally still useful for understanding the key physical interactions within the subsurface regions of interest. The sources of approximations (a.k.a. modeling errors) include: properties up-scaling (grid coarsening), discretization errors, imperfect reservoir fluid properties, relative permeability, reservoir geology description/parameterization and approximate representation of the complete complex subsurface fluid flow physics (e.g. black-oil model in place of a compositional model or constant rock compressibility assumption).

In the context of error modeling, grid up-scaling has been widely studied within the reservoir simulation community. In the published literature, there exist a number of efficient up-scaling techniques (Durlofsky, 2003), aiming to obtain optimal upscaled properties. However, up-scaling errors are not completely eliminated by most of these methods. Discretization errors also cannot be eliminated (Ertekin et al., 2001), even after selecting an optimal grid size and utilizing adaptive time stepping techniques.

Additionally, numerical simulation using an optimal fine-grid could be computationally prohibitive especially for tasks that typically requires many simulation runs, for example history matching or robust optimization problems. Various techniques have been proposed to address this computational bottleneck, for example reduced order modeling and proxy models among many other techniques [c.f., Silva et al., 2007; Rammay and Abdulraheem, 2014; Cardoso et al., 2009]. An alternative approach is to utilize an upscaled model instead of the fine scale model for multi-query computationally demanding tasks (e.g. uncertainty quantification problems). In the context of history matching, if the up-scaling errors are not modeled during the parameter inference step, the posterior distributions of the model parameters is likely to be biased and this bias will subsequently affect the future predictions of the engineering quantities of interest (e.g. oil, gas and water rates/pressure). Omre et al. (2004) approximated the up-scaling and discretization errors by computing samples or realizations of the error data using pairs of fine- and coarse-scale models. The model errors due to up-scaling were then estimated using a multiple regression technique and added to the coarse scale model predictions during the history matching process. Lødøen et al. (2005) utilized a similar procedure on a different set of test cases while employing a more accurate up-scaling procedure.

Accurate reservoir geology description is another challenging task due to various uncertainties including: channel geometry, faults shape, facies proportion, stratigraphic and/or structural frameworks. It is widely known that unrealistic geological models could be calibrated to match the historical data [c.f., Carter et al., 2006; Refsgaard et al., 2012]. However, these fitted models fail to provide reliable predictions and could ultimately mislead the reservoir development plans [c.f., Carter et al., 2006; Refsgaard et al., 2012]. Although considerable effort is often put into constructing geological models that are as realistic as possible, it is very difficult to maintain this geological realism while updating them to match the observed data (Sun and Durlofsky, 2017).

Accounting for model errors during the calibration process has attracted a large body of research [c.f., Oliver and Alfonzo, 2018; Dreano et al., 2017; Josset et al., 2015], where various approaches have been developed to account for the model-error compo-

nent during model calibration [c.f., Hansen et al., 2014; Evin et al., 2014b; Reichert and Schuwirth, 2012b]. These approaches vary according to the different behavior and complexity of the modeled physical system. For example in hydro-geophysical systems, Köpke et al. (2018) accounted for the model-error component using orthonormal basis generated from an error dictionary which is continuously enriched during the calibration process. The models of the bias or error component could be generally classified as either input dependent (Giudice et al., 2013) or output dependent (Evin et al., 2014b). Input dependent model error formulation represents the error components as a function of the model parameters. For example, a reservoir model-error can vary with permeability realizations or other input model parameters. O' Sullivan and Christie (2005) utilized an input dependent formulation for the model-error where the authors computed model-error realizations using the difference between a fine-grid and coarse-grid model outputs. During the calibration process, an interpolation of the error component was performed to estimate a correction term to the coarse-grid model predictions. Lødøen and Tjelmeland (2010) used multiple linear regression algorithm to model errors, where the residual part for the multiple regression was assumed to depend on the model input parameters. The residual terms were modeled using a zero mean Gaussian Process. Giudice et al. (2013) used an input dependent model-error representation to improve uncertainty estimation in urban hydrological models. In that application, the model error variance was set to be dependent on the input of the rainfall term. Output dependent model error formulation represents the error components as a function of the output of the physical model. For example, Evin et al. (2014b) utilized an output dependent formulation for the model-error heteroscedasticity as a function of the simulated streamflow. In a realistic setting, where large models are utilized (i.e. millions of input parameters), it is hard to relate the model errors to the high dimensional input parameter space and output dependent or input/output (I/O) independent forms of the model-error were proposed as an alternative approach that might have some advantages over the input-dependent error-models (Giudice et al., 2013).

Model-error representation can also be classified as either an external bias descrip-

tion (EBD) or an internal noise description (IND) (Giudice et al., 2015). EBD was developed using the background of statistical inference in a regression type framework (Giudice et al., 2015). In EBD, the model-error term is added externally into the forward model (approximate or inadequate model) output. In IND, the model-error is formulated as an additional term of the state space (Giudice et al., 2015). This approach is also known as state space modelling or stochastic gray-box modelling [c.f., Moradkhani et al., 2012; Kristensen et al., 2004]. Giudice et al. (2015) concluded that EBD has some advantages over IND in terms of long-term predictions.

In this chapter, we utilize an EBD and I/O independent error-model formulation as it is more suitable for large scale models (e.g. subsurface oil and gas reservoir models) relying on black-box simulators. Several typical sources of model errors are present in the test cases investigated in this chapter, including a coarse grid, less detailed geological representation (i.e. upscaling of different types of geological features including variogram based and channelized geology), discretization errors and slight change in well locations due to grid coarsening. The error-model formulation presented here, assumes that the total modeling errors consists of two components: structural component and noise-like component. We note that the structural component is often neglected in the bias correction approaches developed in hydrological literature [c.f., Maier et al., 2014; Vrugt, 2016; White et al., 2014]. In this study prior to model calibration, the structure of the model error is estimated and represented using several basis functions, and the magnitude of the noise-like component is quantified. During history matching, the weights of the basis functions are jointly calibrated with the physical model parameters using data observed at well locations. The noise-like part of the model error is also accounted for during the history matching process to avoid over-fitting of the error model. We note that the presented formulation is general and can be applied to other sources of modeling errors when dealing with low-fidelity physical models. The low-fidelity models are generally used as efficient surrogate models for computationally demanding tasks [c.f., Asher et al., 2015; Laloy et al., 2013].

For Bayesian inversion, we use a particular type of iterative ensemble smoother ES-MDA (Emerick and Reynolds, 2013). The formulation of ES-MDA has some similarities

with Kalman filtering algorithms (Sun et al., 2016). However, ES-MDA assimilates data from different times simultaneously and the same set of data is assimilated multiple times with an inflated data noise covariance matrix which is equivalent to annealing approaches (Stordal and Elsheikh, 2015). The rest of the chapter is organized as follows: In Section 2.2, we present some background on Bayesian inverse modeling followed by the proposed error-model formulation. Following that, we present the case studies with results in Section 2.3 followed by the conclusions of our work in Section 2.4.

2.2 Methodology

Bayesian inverse modeling is a generic inference framework that is widely adopted for calibration of reservoir models while accounting for different types/sources of uncertainties. In the Bayesian framework, the conditional probabilities $p(\mathbf{m}|\mathbf{d}_{obs})$ of the model parameters $\mathbf{m}$ given the observational data $\mathbf{d}_{obs}$ (a.k.a. posterior distribution of the model parameters) is estimated using Bayes rule (Oliver et al., 2008):

$$p(\mathbf{m}|\mathbf{d}_{obs}) \propto p(\mathbf{d}_{obs}|\mathbf{m})\ p(\mathbf{m}), \tag{2.1}$$

where $\mathbf{m}$ is the model parameters vector of size N_m, $\mathbf{d}_{obs}$ is the observations vector of size N_d, $p(\mathbf{m})$ is the prior probability of the model parameters and $p(\mathbf{d}_{obs}|\mathbf{m})$ is the likelihood function of the data given a specific realization of the model parameters $\mathbf{m}$. It is common to assume a Gaussian prior:

$$p(\mathbf{m}) \propto \exp\big(-\frac{1}{2}(\mathbf{m}-\mathbf{m}_{pr})^{\intercal}\mathbf{C}_M^{-1}(\mathbf{m}-\mathbf{m}_{pr})\big), \tag{2.2}$$

where $\mathbf{m}_{pr}$ is an N_m dimensional vector of the mean prior model parameters and $\mathbf{C}_M$ is the covariance matrix of the prior model parameters. It is also common to assume that data noise is Gaussian, so that the likelihood function takes the form:

$$p(\mathbf{d}_{obs}|\mathbf{m}) \propto \exp\big(-\frac{1}{2}(\mathbf{d}_{obs}-\mathbf{d})^{\intercal}\mathbf{C}_D^{-1}(\mathbf{d}_{obs}-\mathbf{d})\big), \tag{2.3}$$

where $\mathbf{d}$ is the simulated or predicted data vector using the model parameters $\mathbf{m}$ and $\mathbf{C}_D$ is the error/noise covariance matrix which is defined in Sect. 2.2.1 and 2.2.2 depending

on the utilized history matching procedure. Using these definitions, Bayes' rule defined in Eq. 2.1 could be expanded as following:

$$p(\mathbf{m}|\mathbf{d}_{obs}) \propto \exp\left(-\frac{1}{2}\big((\mathbf{d}_{obs}-\mathbf{d})^{\intercal}\mathbf{C}_D^{-1}(\mathbf{d}_{obs}-\mathbf{d})+(\mathbf{m}-\mathbf{m}_{pr})^{\intercal}\mathbf{C}_M^{-1}(\mathbf{m}-\mathbf{m}_{pr})\big)\right). \quad (2.4)$$

Several algorithms could be used to generate samples from the posterior distribution of the model parameters (Oliver et al., 2008). Among those Markov Chain Monte Carlo (MCMC) is an exact method for sampling. However, MCMC can be computationally expensive due to the large number of iterations needed to reach convergence and the sequential nature of the method. Ensemble-based methods have been widely used for calibrating subsurface flow models due to the computational feasibility and parallel nature of ensemble methods. In this study, we utilize the ensemble smoother with multiple data assimilation (ES-MDA) algorithm for the calibration step (Emerick and Reynolds, 2013). ES-MDA belongs to a class of iterative ensemble smoothing techniques that could be used to solve non-linear inverse problem iteratively with an inflated noise covariance matrix. The ES-MDA algorithm steps are summarized as follows:

- Select the number of iterations (number of data assimilation) N_a and the inflation factor α. A common choice of the inflation factor is to set it as a constant value for all iterations $\alpha = N_a$. The fixed value of alpha was taken in the original ES-MDA paper (Emerick and Reynolds, 2013). However, the fixed alpha can be non-robust in some cases (Iglesias, 2016, 2015). We have taken fixed alpha in order to further explore the regularization capability of the presented error-model approach with limited number of ensemble members.

- Initialize an ensemble of model parameters and perturb the observation data for each ensemble member using:

$$\mathbf{d}_{uc,j} = \mathbf{d}_{obs} + \sqrt{\alpha}\ \mathbf{C}_D^{1/2}\ \mathbf{z}_d, \quad (2.5)$$

  where the subscript j is the ensemble member index $j = 1 \dots N_e$ and N_e is the ensemble size, $\mathbf{d}_{uc,j}$ is N_d dimensional vector of perturbed observation, $\mathbf{z}_d$ is the N_d dimensional vector with standard Gaussian random variables as its components (i.e. $\mathbf{z}_d \sim \mathcal{N}(0, \mathbf{I}_{N_d,N_d})$).

- Update each ensemble member using,

$$\mathbf{m}_j^{(i+1)} = \mathbf{m}_j^{(i)} + \mathbf{C}_{MD}\ (\mathbf{C}_{DD} + \alpha\ \mathbf{C}_D)^{-1}(\mathbf{d}_{uc,j} - \mathbf{d}_j), \tag{2.6}$$

  where the superscript i is the iteration index, $\mathbf{C}_{DD}$ is the model output covariance matrix and $\mathbf{C}_{MD}$ is the cross covariance matrix of model parameters and model predictions.

- Repeat the above steps for all iterations, from $i = 1$ to N_a

2.2.1 Procedures for history matching of reservoir models

In this study, two types of history matching procedures are investigated: history matching while neglecting model-discrepancy (i.e. standard history matching procedure) and joint history matching of the model parameters and the parameters of an error/bias model. In this chapter, we use the term model-discrepancy and model-error interchangeably. The standard history matching procedure relies on an implicit assumption that the model-errors are generally small and could be neglected (i.e. the simulation model is perfect). Mathematically, if an accurate/high-fidelity model is utilized, the observed data is formulated as (Giudice et al., 2013):

$$\mathbf{d}_{obs} = g(\mathbf{m}_{true}) + \boldsymbol{\epsilon}_d, \tag{2.7}$$

where $g(.)$ is a nonlinear function representing the accurate/high-fidelity forward simulation model, $\mathbf{m}_{true}$ is the true model parameters, $\boldsymbol{\epsilon}_d$ is the measurement errors which is usually assumed to follow a normal distribution $\mathcal{N}(0, \mathbf{C}_d)$ and $\mathbf{C}_d$ is the measurement errors covariance matrix. In this study, uncorrelated measurement errors are considered, therefore the matrix $\mathbf{C}_d$ is a diagonal matrix. In standard history matching (i.e. neglecting modeling errors), $\mathbf{C}_D$ in Eqs. 2.3, 2.4, 2.5 and 2.6 is set to the covariance of measurement errors. Therefore,

$$\mathbf{C}_D = \mathbf{C}_d. \tag{2.8}$$

However, as noted in the introduction section, several approximations are commonly introduced in the computational model to simplify the simulation process (e.g. black-oil model versus compositional flow), or to speed-up the simulations (coarsening of the

simulation grid). During history matching if the model-error caused by these approximations is not accounted for, the obtained posterior distribution could be biased. In the case of utilizing an approximate/low-fidelity model, the observation data is related to the true model parameters $\mathbf{m}_{true}$ as (Giudice et al., 2013):

$$\mathbf{d}_{obs} = \tilde{g}(\mathbf{m}_{true}) + \boldsymbol{\epsilon}_d + \boldsymbol{\epsilon}_m, \tag{2.9}$$

where $\boldsymbol{\epsilon}_m$ is the model-error and $\tilde{g}(.)$ is a nonlinear function representing the imperfect (approximate/low-fidelity) simulation model. By subtracting Eq. 2.7 from 2.9, we obtain:

$$\boldsymbol{\epsilon}_m = g(\mathbf{m}_{true}) - \tilde{g}(\mathbf{m}_{true}). \tag{2.10}$$

In the following sub-section, we present a simple yet general parameterization of the model-error term $\boldsymbol{\epsilon}_m$.

2.2.2 Error-model formulation

In this study, EBD and I/O independent error-model approach is considered. As the model errors in our test cases were dominated by structured components, these errors are parameterized using smooth basis functions obtained by principle component analysis (PCA) method, which is an effective data-driven dimension reduction technique (Shlens, 2014; Kerschen et al., 2005). We rely on simulation output from pairs of models, accurate/high-fidelity versus approximate/low-fidelity, to obtain the basis functions and the prior statistics of the coefficients of the PCA basis functions. We acknowledge that this limits the applicability of the developed approach to the cases for which an accurate/high-fidelity model is available. However, we note that the accurate/high-fidelity model is only used to estimate the prior model-error statistics, and is not used during the calibration process. This is a notable difference between the presented framework and related studies by Josset et al. (2015) and Köpke et al. (2018). Josset et al. (2015) used the proxy and error models to accelerate the performance of two-stage MCMC and high-fidelity/accurate models were used during the evaluation of posterior distribution. Köpke et al. (2018) used orthonormal basis, which

is generated from the differences between pairs of high-fidelity and low-fidelity models, to filter the model-error components from the data residuals in the hydrogeophysical inverse problems and the high-fidelity/accurate models were used to update the error-dictionary for generation of local orthonormal basis during the calibration process.

In the current setting, the prior model-error realizations are estimated using:

$$\boldsymbol{\epsilon}_{mr} = g(\mathbf{m}_r) - \tilde{g}(\mathbf{m}_r), \tag{2.11}$$

where r is the index of the prior realizations, i.e. $r = 1$ to N_r, N_r is the total number of realizations used to estimate the model-error statistics, $\boldsymbol{\epsilon}_{mr}$ is an N_d dimensional vector of model-error for realization r. All prior model-error realizations are assembled into the matrix $\boldsymbol{\epsilon} \in \mathbb{R}^{N_d \times N_r}$. The mean of the model-error prior is,

$$\bar{\boldsymbol{\epsilon}}_m = \frac{1}{N_r} \sum_{r=1}^{N_r} (\boldsymbol{\epsilon}_{mr}). \tag{2.12}$$

The covariance of the model-error prior is (Oliver et al., 2008),

$$\mathbf{C}_e = \frac{1}{N_r - 1} (\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}_m \mathbf{I}_{N_r})(\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}_m \mathbf{I}_{N_r})^\intercal, \tag{2.13}$$

where $\mathbf{I}_{N_r}$ is an N_r dimensional row vector with all ones as its components. In this study, PCA is used to parametrize the prior model-error realizations and the weights of the obtained PCA basis vectors are jointly inferred with the model parameters during the history matching process. The basis functions are obtained by singular value decomposition (SVD) of the error covariance matrix $\mathbf{C}_e$ (Oliver et al., 2008):

$$\mathbf{C}_e = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^\intercal, \tag{2.14}$$

where $\mathbf{U}$ and $\mathbf{V}$ are the orthonormal singular vectors (basis functions) and $\boldsymbol{\Sigma}$ is a diagonal matrix of the singular values. The error-model is formulated using the leading L singular vectors as following:

$$\hat{\boldsymbol{\epsilon}}_{mr} = \boldsymbol{\Phi} \boldsymbol{\beta}_r + \bar{\boldsymbol{\epsilon}}_m, \tag{2.15}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{N_d \times L}$ are the first L orthonormal singular vectors (basis functions) from $\mathbf{U}$ and $\boldsymbol{\beta}_r \in \mathbb{R}^{L \times 1}$ are the coefficients of error-model for realization r.

The objective of the calibration process is then to find the posterior distribution of reservoir model parameters and the coefficients $\boldsymbol{\beta}$ of the PCA-based error-model. Since Bayesian inverse modelling require prior statistics of model parameters, therefore the prior statistics of the coefficients $\boldsymbol{\beta}$ should be estimated. The least square form of Eq. 2.15 is used to compute prior realizations of the coefficient vector $\boldsymbol{\beta}$ as following:

$$\boldsymbol{\beta}_r = (\boldsymbol{\Phi}^\intercal \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^\intercal (\boldsymbol{\epsilon}_{mr} - \bar{\boldsymbol{\epsilon}}_m). \tag{2.16}$$

Since $\boldsymbol{\Phi}$ is an orthonormal matrix i.e. $\boldsymbol{\Phi}^\intercal \approx \boldsymbol{\Phi}^{-1}$, therefore Eq. 2.16 can also be written as:

$$\boldsymbol{\beta}_r \approx \boldsymbol{\Phi}^\intercal (\boldsymbol{\epsilon}_{mr} - \bar{\boldsymbol{\epsilon}}_m). \tag{2.17}$$

The prior statistic, such as the mean and covariance of $\boldsymbol{\beta}$ realizations are computed using:

$$\boldsymbol{\mu}_\beta = \frac{1}{N_r} \sum_{r=1}^{N_r} (\boldsymbol{\beta}_r), \tag{2.18}$$

$$\mathbf{C}_\beta = \frac{1}{N_r - 1} \sum_{r=1}^{N_r} (\boldsymbol{\beta}_r - \boldsymbol{\mu}_\beta)(\boldsymbol{\beta}_r - \boldsymbol{\mu}_\beta)^\intercal, \tag{2.19}$$

where $\boldsymbol{\mu}_\beta \in \mathbb{R}^{L \times 1}$ is the mean of $\boldsymbol{\beta}$ realizations and $\mathbf{C}_\beta \in \mathbb{R}^{L \times L}$ is the covariance of $\boldsymbol{\beta}$ realizations. In this study, we only consider the diagonal terms of the matrix $\mathbf{C}_\beta$ to generate prior samples of error-model coefficients for the history matching purpose.

In order to avoid over-fitting the error-model, the number of coefficients of the PCA-based error-model L should be limited to a small number. Therefore the PCA-based error model is truncated in order to reduce the number of parameters of the error-model. Generally, overfitting is the historical data matching when a model is too closely fit a limited set of data points such that the model fails to provide reliable future predictions. In the ensemble-based probabilistic framework, overfitting could have occurred due to the limited ensemble members, non-robust regularization or over-parameterization, which results in an underestimation of the approximate posterior distribution or fails to cover multiple local maxima (multi-modality/non-unique solution). The residual of the error-model (truncation error) needs to be included in the inversion process which is useful to avoid overfitting in the proposed joint calibration framework. The residual

of the least square fit is defined as,

$$\boldsymbol{\zeta}_{mr} = \boldsymbol{\epsilon}_{mr} - \hat{\boldsymbol{\epsilon}}_{mr}. \tag{2.20}$$

All residual realizations are assembled into matrix $\boldsymbol{\zeta} \in \mathbb{R}^{N_d \times N_r}$. The covariance of the residual from all error-model realizations is then estimated using:

$$\mathbf{C}_T = \frac{1}{N_r - 1}\boldsymbol{\zeta}\boldsymbol{\zeta}^{\intercal}, \tag{2.21}$$

where $\mathbf{C}_T$ is denoted the error-model noise covariance. For history matching of imperfect-models, the total error covariance matrix $\mathbf{C}_D$ in Eqs. 2.3, 2.4, 2.5 and 2.6 contains both the measurement and error-model noise components as following:

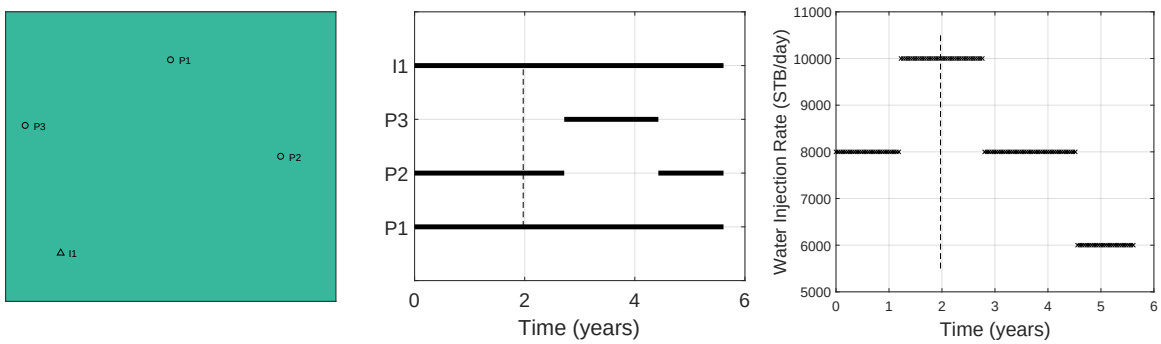
$$\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_T. \tag{2.22}$$

For simplicity, only the diagonal terms of the matrix $\mathbf{C}_T$ is considered in this study. The covariance $\mathbf{C}_T$ is computed using truncation error of the PCA-based error model. The PCA-based error-model can capture the strong correlation of the serial correlated model-errors with respect to time. The truncation part is usually weakly correlated with respect to time therefore the diagonal $\mathbf{C}_T$ is used instead of the full covariance. Conceptually, $\mathbf{C}_D$ is the total uncalibrated uncertainty, which includes both the measurement noise and the model-error noise that is not captured by the truncated PCA-based error-model. In (Hansen et al., 2014), modeling errors were considered as uncalibrated uncertainties using Gaussian distribution (i.e. accounting for mean and covariance of errors). In their approach, the model error is accounted for by replacing $\mathbf{d}_{obs}$ with $\mathbf{d}_{obs} + \overline{\boldsymbol{\epsilon}}_m$ and replacing $\mathbf{C}_d$ with $\mathbf{C}_d + \mathbf{C}_e$. However, this approach would be inconsistent/inefficient for physical systems which exhibit highly complex statistics and correlations of model errors that change significantly not only over the data space, but also as a function of the input model parameters. (Köpke et al., 2018).

2.3 Case Studies

In this section, we present the details of the case studies. The dimension of the subsurface reservoir is 7500 ft × 7500 ft × 20 ft in the x, y and z directions, respectively.

Incompressible two-phase porous media flow of oil and water is considered. The initial reservoir pressure is 5000 psi and the reservoir has uniform porosity of 20%. The reservoir contains one injector well (I1) and three production wells (P1, P2, P3) and is simulated using a 2D grid. We utilize the Matlab Reservoir Simulation Tool-box (MRST) (Lie, 2016) for the forward model simulations. Corey's power law model is used to represent relative permeabilities. Parameter values for the Corey's model and fluid properties are listed in Appendix A. The gravitational and capillary pressure effects are neglected. The wells locations and open/shut schedule are shown in Fig. 2.1(a) and (b). The production wells are operated under constant bottom hole pressure constraint of 4500 psi and the injector well is operated under constant injection rate constraint with varying control values as shown in Fig. 2.1(c). Figure 2.1 also shows end of historical period (i.e. 2 years). In the historical period, the flow rates at the production wells and the bottom hole pressure of the injector well are used as the historical data for the calibration process. We also note that one of the production wells (P3) is only used in prediction phase in order to assess predictions from calibrated models on wells drilled in future development plans.



(a) (b) (c)

Figure 2.1: Wells locations, wells open/shut schedule and Injection well control rates. Part (a) shows the locations of wells locations. Part (b) shows wells open/shut schedule. In part (b) solid back lines indicate the time periods when a well is open to flow. Part (c) shows the water injection rate of injector well. Dashed black lines show end of historical period.

The calibrated models are evaluated using three different forecasting metrics to

assess the quality of the estimated parameters and the capacity of the calibrated models in making future predictions. The utilized forecasting metrics are: coverage probability (CP), mean continuous ranked probability score (CRPS) and mean square error (MSE). CP indicates the fraction of the actual data that lie within the 95% confidence interval of the estimation. A value of 0.95 for CP indicates a consistent estimation of uncertainty and values below 0.95 indicate underestimation of uncertainty. Mean CRPS quantifies both accuracy and precision (Hersbach, 2000) and higher values of CRPS indicate a less accurate results. MSE is widely used as a metric for parameter estimation problems. However, MSE measures the quality of data-fitting and is not enough to provide a probabilistic assessment of the estimation and prediction from an ensemble of models. In this study, we observed that a combination of MSE, CP and CRPS provides a good assessment of the quality for the probabilistic forecast (Skauvold and Eidsvik, 2018). The mathematical formulations of the three forecasting metrics are listed in Appendix B.

2.3.1 Case 1: Coarse scale model

In the first case study, fine-grid/high-fidelity model uses a 2D grid with 75×75 cells. The distributed log-permeability fields are modeled as multivariate Gaussian with exponential covariance function:

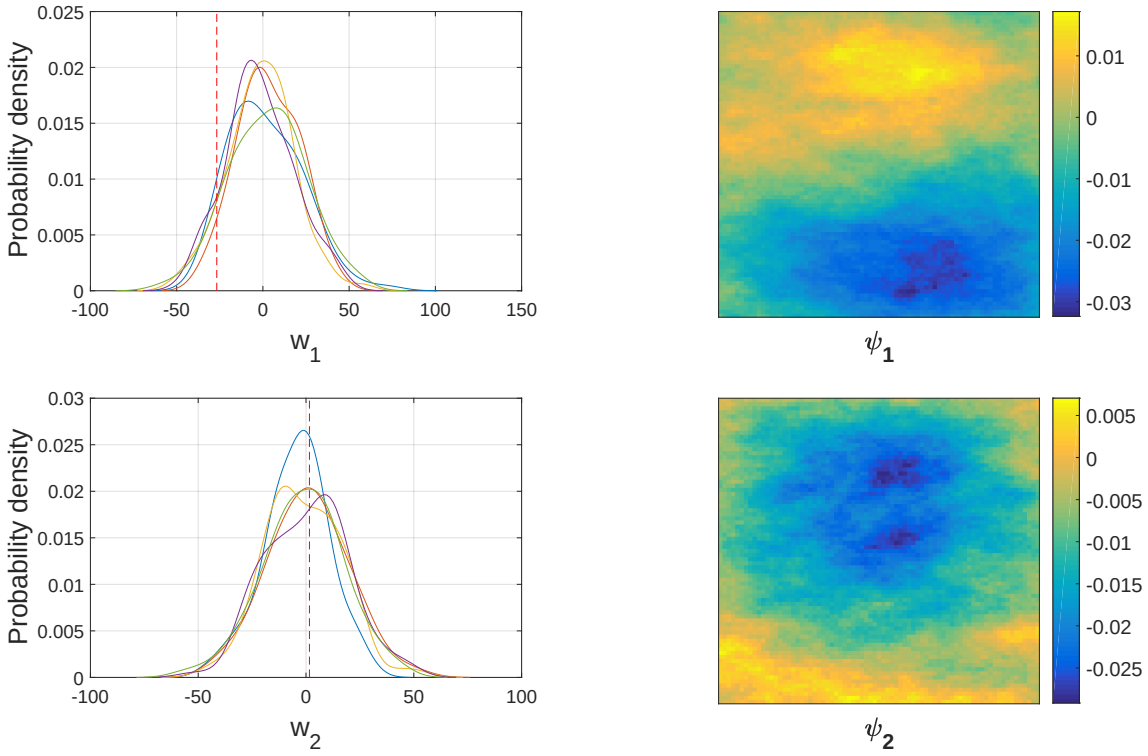
$$\mathbf{c} = \sigma^2 \exp(-3(\frac{\mathbf{s}}{r_a})^{\gamma}), \tag{2.23}$$

where $\mathbf{s}$ is the lag distance and r_a, σ^2, γ are the correlation range, variance and exponent respectively (which are 35 cells, 1 and 1 respectively in this test case). The log-permeability field $\ln(\mathbf{K})$ is parameterized using PCA and only two leading basis functions are retained:

$$\ln(\mathbf{K}) = \overline{\ln(\mathbf{K})} + \sum_{b=1}^{N_w} \mathrm{w}_b \boldsymbol{\psi}_b, \tag{2.24}$$

where $\overline{\ln(\mathbf{K})}$ is the mean log-permeability (equal to 4 in this test case), b is index of the basis weight w and basis function $\boldsymbol{\psi}$ and $N_w = 2$. Figure 2.2(b) shows the leading two principal basis functions obtained by singular value decomposition of covariance of

log-permeability fields (Eq. 2.23). Figure 2.2(a) shows the prior distribution of weights obtained by projecting the log-permeability fields into the PCA-basis functions.



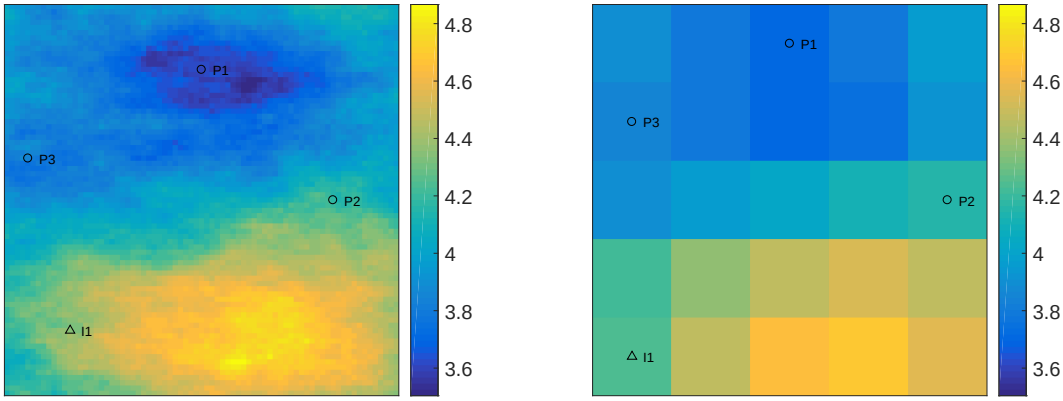
(a) Prior distribution of basis weights

(b) Two principal basis functions

Figure 2.2: Prior distribution of basis weights from five ensembles and two principle basis functions for log-permeability. Red dashed lines show reference solution and five prior ensembles distribution are shown by five different colors in part (a).

Figure 2.3(a) shows the reference fine scale log-permeability field. The fine scale reference log-permeability field is generated by the leading two-PCA basis functions and reference basis weights are shown as the red vertical lines in Fig. 2.2. The coarse-grid/low-fidelity reservoir models contain only 5×5 grid blocks. The coarsened version of the reference fine model is shown in Fig. 2.3(b), in which harmonic averaging is used to up-scale the permeability field. The harmonic average of permeability is a non-robust approach for up-scalling and it is deliberately chosen to introduce the model-error into the simulation model. The observed data are generated by the fine scale model using reference log-permeability field (Fig. 2.3(a)) with the addition of measurement noise of

2% of the reference solution. We note that except permeabilities, the rest of the static and dynamic properties (i.e. porosity, relative permeability, viscosity, and density, well controls and schedule) of the coarse scale model are the same as the fine scale model.



(a) (b)

Figure 2.3: The fine scale (75×75) reference log-permeability (a) and the corresponding up-scaled log-permeability (5×5) using harmonic average (b).

Two different procedures of history matching the coarse scale model were considered. In the case of neglecting model-discrepancy, PCA basis weights $\mathbf{w}$ of the log-permeabilities are calibrated (i.e. $\mathbf{m}_j = \mathbf{w}$ in Eq. 2.6). In the case of joint inversion with error-model, the estimated parameters consist of the combined vector of the log-permeability PCA weights and the error-model coefficients (i.e. $\mathbf{m}_j = [\mathbf{w}; \boldsymbol{\beta}]$ in Eq. 2.6).

Prior statistics of the model-discrepancy were estimated using Eq. 2.11. One hundred fine scale permeability realizations were generated through Eq. 2.24 by sampling the prior distribution of the PCA-basis weights and a corresponding number of coarse scale permeability realizations were obtained using harmonic up-scaling. Forward runs were then performed for both the coarse and fine scale models to obtain the error realizations using Eq. 2.11. A smaller number of realizations could be used to evaluate the model-discrepancy statistics. In that case, special care should be taken to select a representative set of prior realizations to cover the respective statistic. Figure 2.4 shows the prior statistics of the model-discrepancy in the simulated well production

data (bottom hole pressure of the injector well and flow rates of the producers).

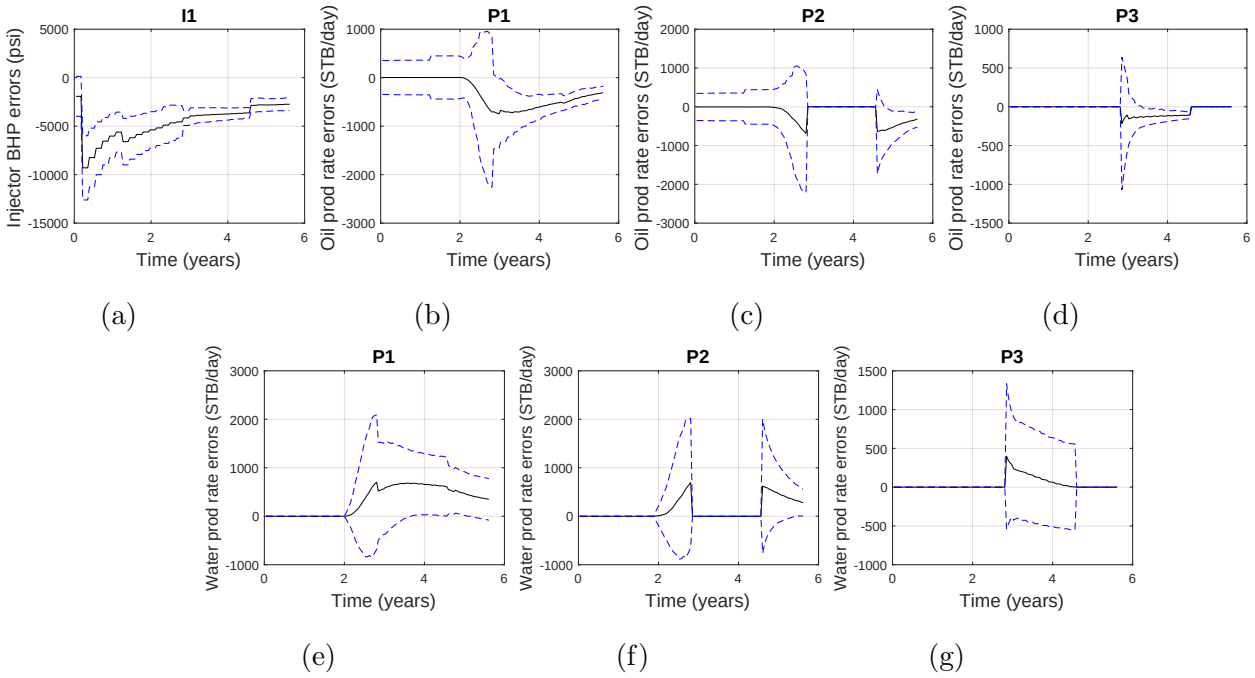


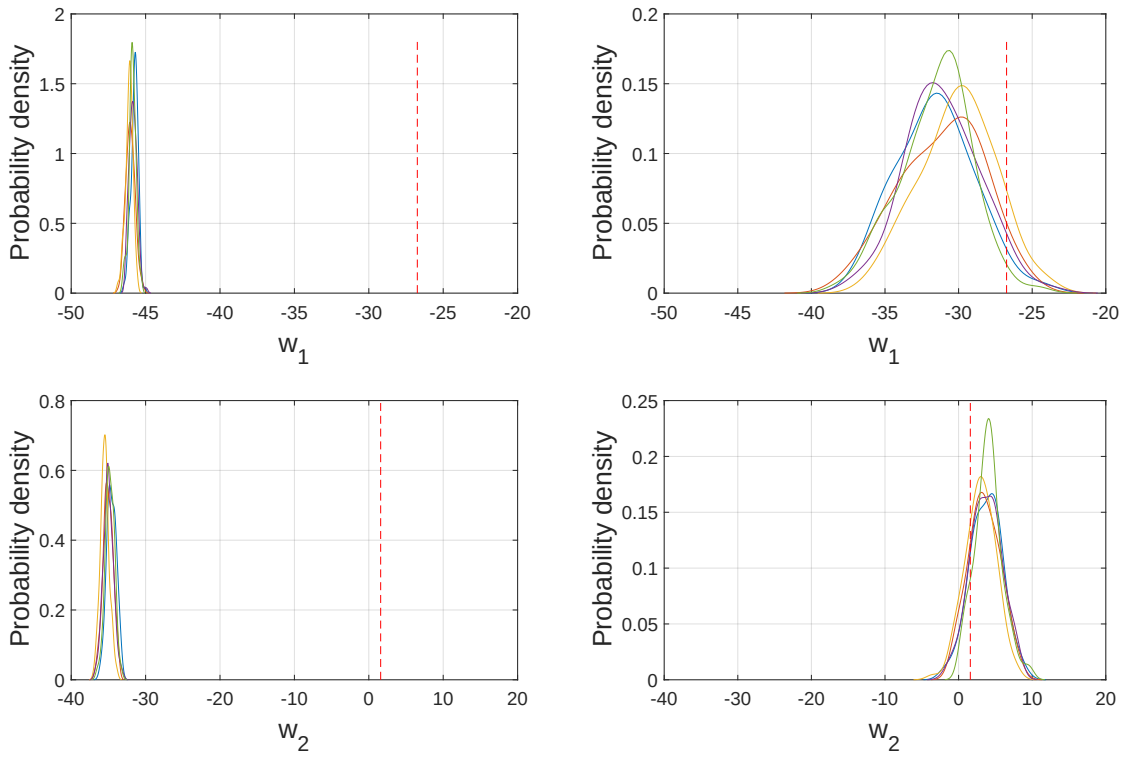
Figure 2.4: Prior model-error statistics of all wells for Case 1. Black lines show mean model errors, dashed blue lines show the 95% confidence interval (mean plus and minus two standard deviations) of model errors.

2.3.2 Case 1 Results

In this subsection, we present history matching results for test case 1 with and without accounting for the model-discrepancy. All the ensemble-based history matching results are presented for multiple runs (five independent ensembles) in order to investigate the consistency and reliability of the parameter estimation process. Each ensemble run consists of 100 ensemble members and the measurement errors are assumed to be 2% of the observation data. We utilized the ES-MDA algorithm with eight iterations ($N_a = 8$ and $\alpha = N_a$) for calibration. Generally 100 ensemble members are used as a thumb rule in oil and gas industry for ensemble based calibration. The same ensemble size was also taken in the original ES-MDA paper (Emerick and Reynolds, 2013). In the case of joint inversion with error-model, two PCA components were retained per each output time series to parametrize the model-discrepancy. Since we have seven output

time series, the total number of error-model parameters is 14.

In this test case, the log-permeability is calibrated in terms of the PCA-basis weights, i.e. w_1, w_2 as detailed in the problem description. Figure 2.5 shows the posterior distributions of the basis weights for the five runs, for both cases of neglecting and accounting for model error. The results presented in Fig. 2.5(a) show that the posterior distribution from the inversion (neglecting model-error) procedure are biased and the estimated basis weights do not capture the reference weights. In contrast, the posterior distributions obtained by the joint inversion procedure are less biased (i.e. nearly unbiased) and successfully cover the true model parameters as shown in Fig. 2.5(b).



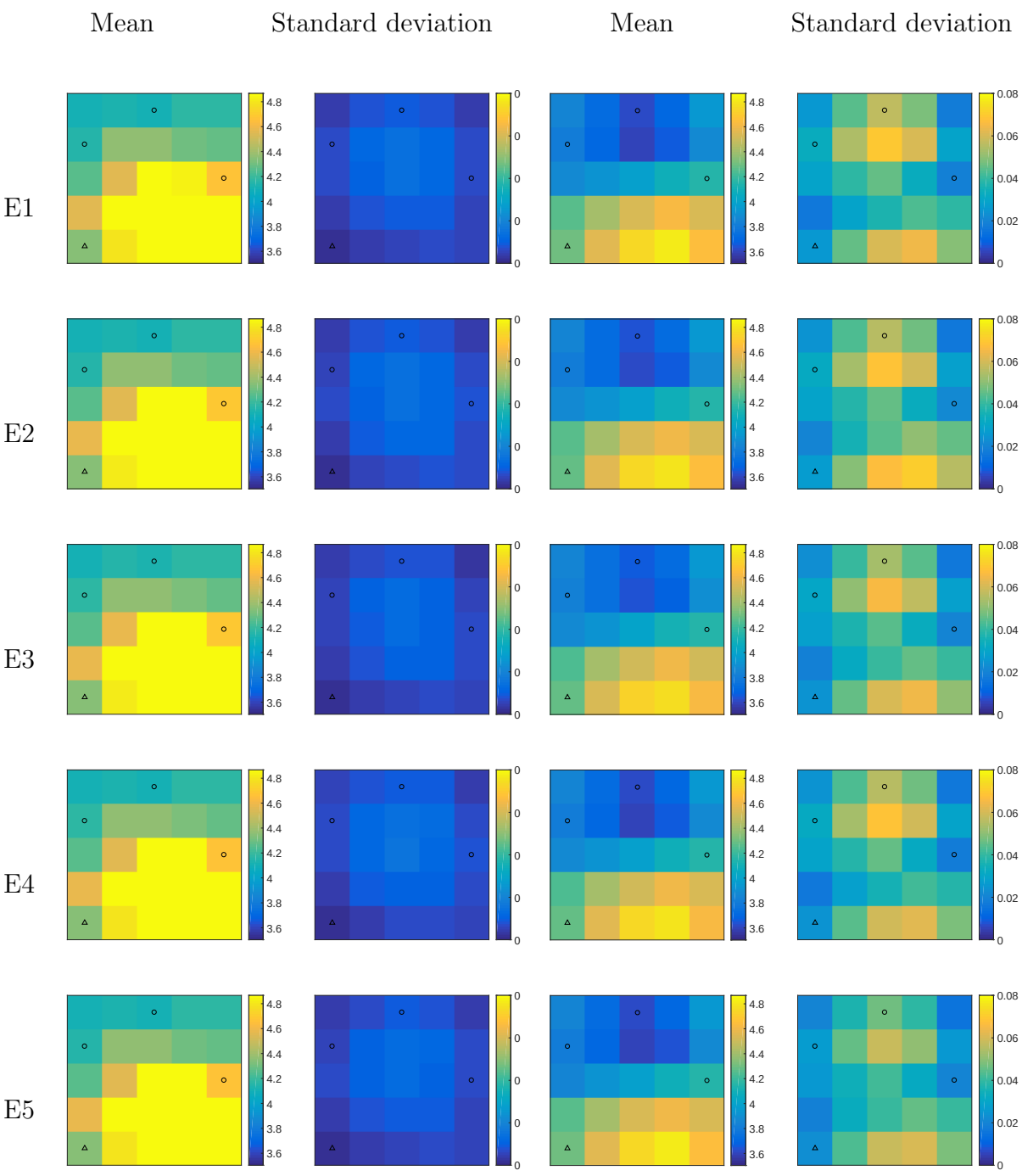
(a) inversion (neglecting model-error) (b) joint inversion with error-model

Figure 2.5: Posterior distribution of two PCA basis weights (of ln(**K**)) obtained after history matching for coarse scale model case. Dashed red lines show the reference solution and the posterior distribution of the five ensembles are shown by five different colors.

Figure 2.6 shows the mean and standard deviation of the posterior ln(**K**) for test case 1. In Fig. 2.6(a) the mean of posterior log-permeability field obtained from five

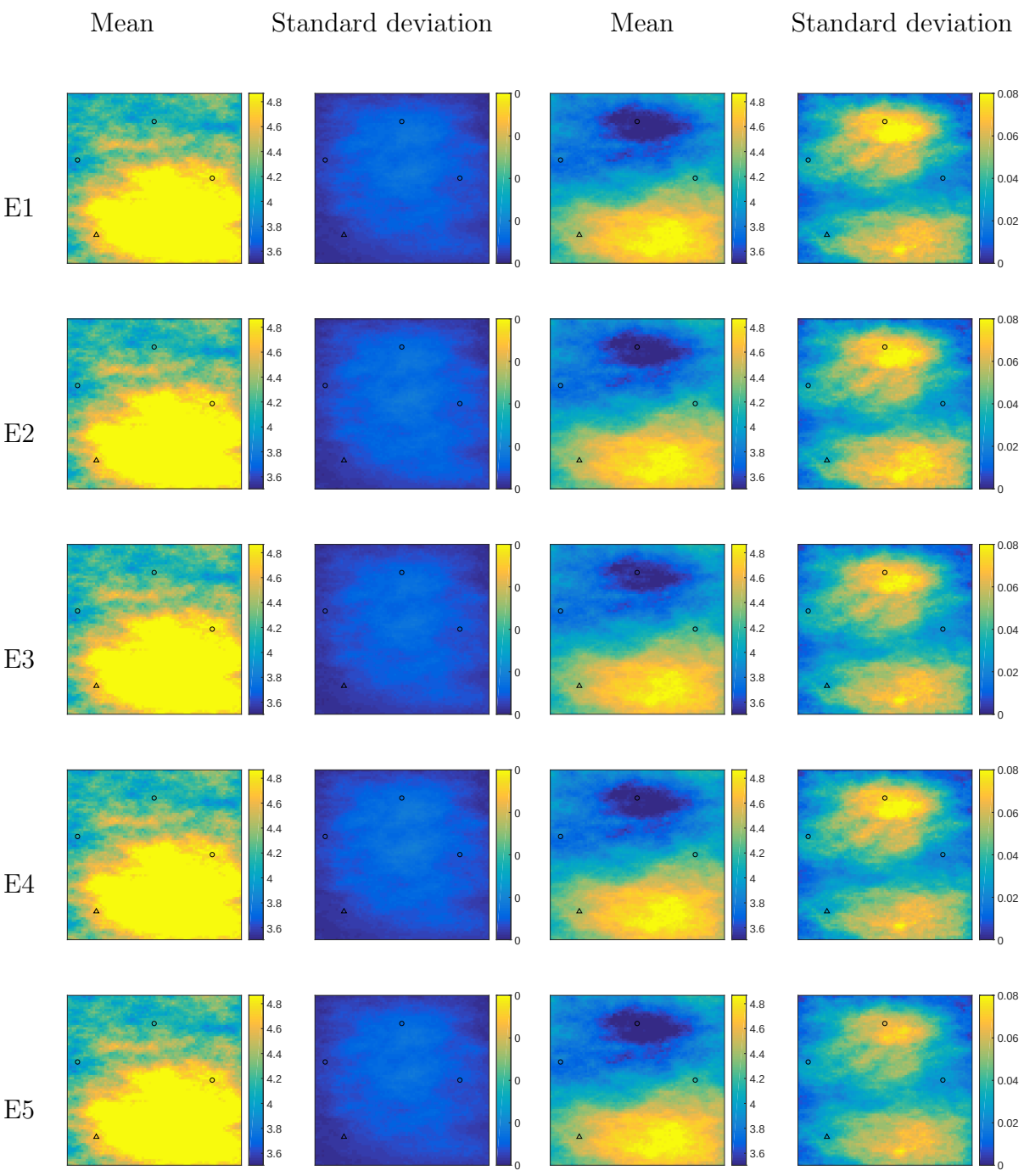
different runs are shown for the inversion procedure. This posterior mean is clearly different from the coarse scale reference log-permeability field shown in Fig. 2.3(b) due to the bias in the inferred posterior distributions. Figure 2.6(b) shows the mean posterior log-permeability fields obtained by the joint inversion procedure. These fields are quite similar to the coarse scale reference log-permeability field. We also observe that the posterior standard deviations are quite low for the inversion (neglecting model-error) procedure, which could be a sign of over-fitting the data. In contrast for the joint inversion procedure, the standard deviations of the posterior fields are much larger due to accounting for the model error and including the error-model noise covariance matrix $\mathbf{C}_T$ in Eq. 2.22.

In case 1, the log-permeability fields are parameterized using PCA with basis functions on fine-grid. The fine-grid solutions can be generated using the approximate posterior distributions of the PCA-weights of the log-permeability field, which are shown in Fig. 2.5. Figure 2.7 shows the mean and standard deviation of the posterior ensemble $\ln(\mathbf{K})$ on the fine grid. The posterior means of different ensembles are clearly different from the reference log-permeability field on fine scale (Fig. 2.3(a)) due to the negligence of model-error. However, the posterior means obtained after joint calibration are similar to the fine scale reference log-permeability field (Fig. 2.3(a)). The standard deviations of the posterior ensemble on the fine scale show the similar behaviour as the standard deviations of the posterior ensemble on the coarse scale.



(a) inversion (neglecting model-error) (b) joint inversion with error-model

Figure 2.6: Mean and standard deviation of ln($\mathbf{K}$) posterior ensembles obtained after history matching of two PCA basis weights for coarse scale model case.



(a) inversion (neglecting model-error) (b) joint inversion with error-model

Figure 2.7: Mean and standard deviation of ln(**K**) posterior ensembles on the fine grid obtained after history matching of coarse scale model.

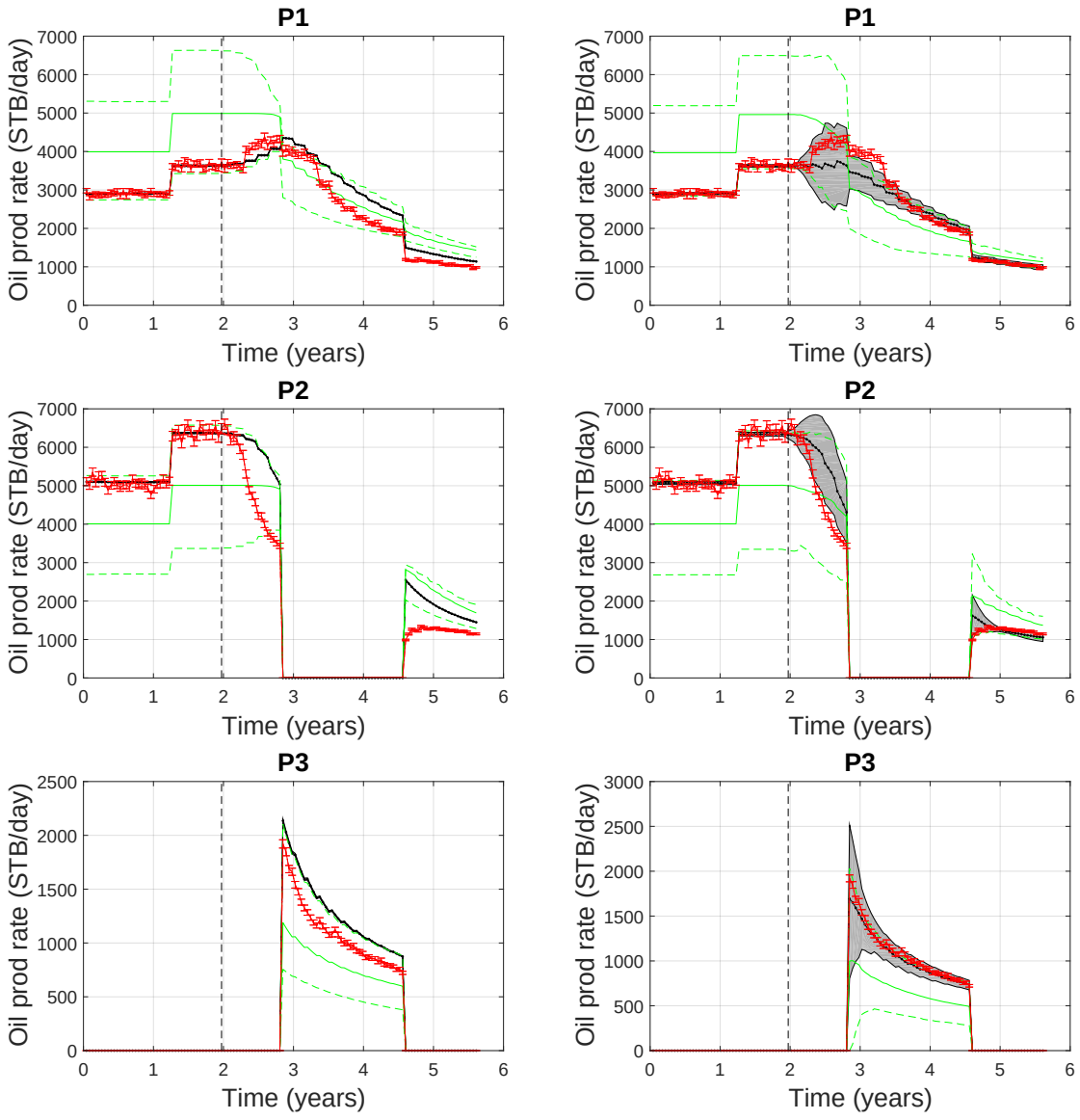
Figures 2.8 and 2.9 show the oil and water production rates of the different pro-

duction wells and the bottom hole pressure of the injector well. The 50th percentile $p50$ and 95% confidence interval (the shaded region) are obtained by combining results from all five runs. In part (a) of these figures, the results for the inversion (neglecting model-discrepancy) are presented and the results of the joint-inversion with error-model are shown in part (b). For the inversion (neglecting model-error) procedure, the results are mixed where the data are matched for some cases and not matched for others. For example in Fig. 2.8(a), the data match is quite good for the wells P1 and P2. However in Fig. 2.9(a) the data match for well I1 is not as good. Moreover, the predictions for all wells are inaccurate for the inversion (neglecting model-error) procedure. Furthermore, the prediction envelop is really narrow, resulting in invisible confidence interval in the plots, which shows over-confidence in the inaccurate predictions. In comparison, better matches and predictions are obtained by the joint inversion procedure as shown in Figs. 2.8(b) and 2.9(b).

Figure 2.10 shows the forecasting metrics (CP for the estimated log-permeabilities and mean CRPS, MSE, CP for the well data in history matching and prediction periods) for individual ensemble (E1 to E5) and for results from all five ensembles assembled together (denoted as "All" in the figure). These metrics provide a good assessment on the consistency, reliability and accuracy of forecasting capacity of the calibrated models. Figures 2.10(a) and 2.10(b) show the coverage probability of reservoir model parameters ($\ln(\mathbf{K})$), well data for both the history matching period and the forecasting period. Both the inversion (neglecting model-error) and the joint inversion results are shown. In Fig. 2.10(a) CP of $\ln(\mathbf{K})$ is zero for each individual ensemble (E1 to E5) as well as for all five ensembles combined meaning that none of the ensemble captures the true log-permeability using the inversion (neglecting model-error) procedure. In Fig. 2.10(b) CP of $\ln(\mathbf{K})$ is one for E1 to E4 as well as for the combined ensemble that means that four out of five runs managed to enclose the reference log-permeabilities completely when using the joint inversion procedure. For the inversion (neglecting model-error) procedure, the CP lies between 0.06–0.11 for the historical data and lies between 0.01–0.013 for the validation data (prediction), as shown in Fig. 2.10(a). For the joint inversion procedure, these values of CP increased to be between 0.29–0.31 and

0.77–0.82, respectively as shown in Fig. 2.10(b). Although the value of CP equal to one (higher than the correct value of 0.95) for the estimated permeability field when the joint inversion procedure is used clearly indicates that the uncertainty of the permeability field is overestimated, the overall results still show reasonable improvement from the joint inversion procedure compared to the standard inversion (neglecting model-error) procedure.

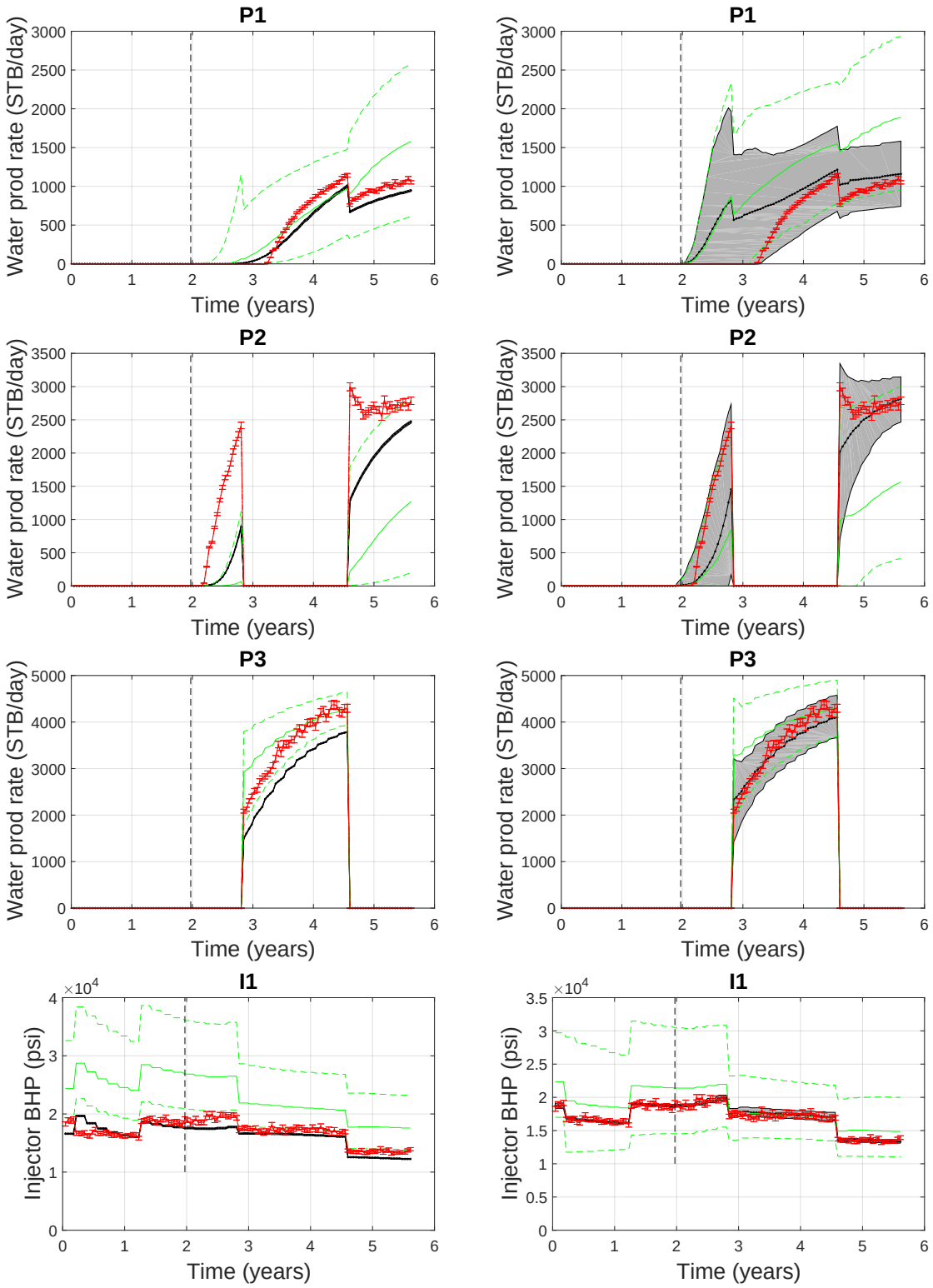
Figures 2.10(c) and 2.10(d) show the mean CRPS of history matching and prediction periods of the well data for the inversion (neglecting model-error) and the joint inversion procedures, respectively. Figure 2.10(c) shows that the mean CRPS lies between 117–118 and 466–469 for history matching and prediction periods, respectively. The results for the inversion (neglecting model-error) is unreliable and inaccurate due to the biased posterior distributions for all the different runs. Figure 2.10(d) shows that using joint inversion with error-model, the mean CRPS lies between 72–73 and 170–187 for history matching and prediction periods of the well data, respectively. A significant improvement in terms of reliability and accuracy is observed, by incorporating the error-model in the inversion process. Figures 2.10(e) and 2.10(f) show the MSE of the individual runs and the combined ensemble of all runs. With the joint inversion procedure, lower MSE values are obtained for both the history matching and the prediction periods (indicated by subscript "h" and "p" respectively in the plot).



(a) inversion (neglecting model-error)

(b) joint inversion with error-model

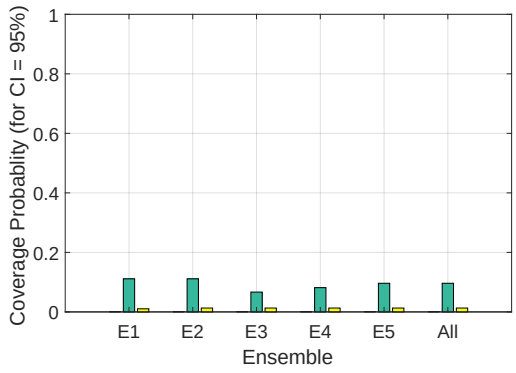
Figure 2.8: Prior and posterior of oil production data obtained from all ensembles for coarse scale model case. Red lines show observation data and bar on red lines shows measurement error. Dashed black lines show end of historical period. Solid green lines show 50th percentile $p50$ of prior distribution, dashed green lines show 95% confidence interval of prior distribution. Solid black lines show $p50$ of posterior distribution, gray shaded area shows 95% confidence interval of posterior distribution.



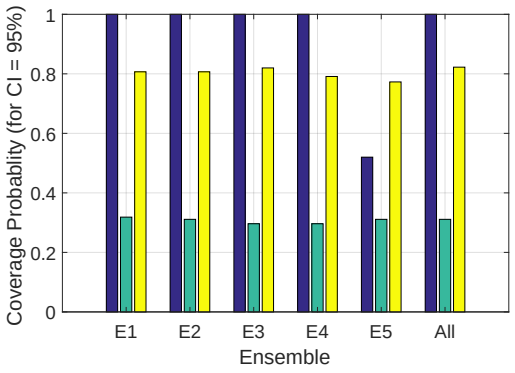
(a) inversion (neglecting model-error)

(b) joint inversion with error-model

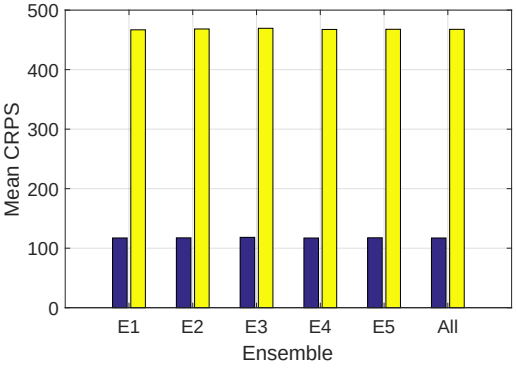
Figure 2.9: Prior and posterior of water production and injection pressure data for coarse model case. The explanation of colors and lines are the same as in Fig. 2.8.



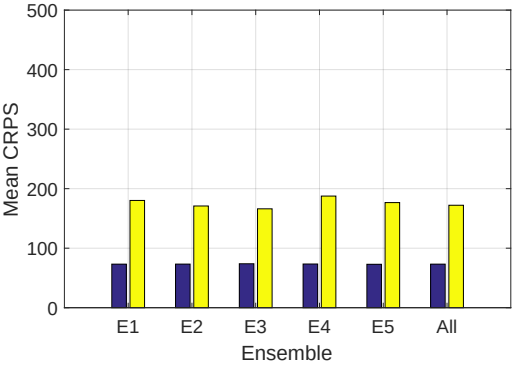
(a) inversion (neglecting model-error)



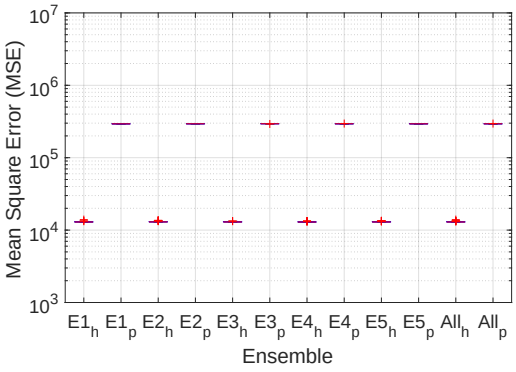
(b) joint inversion with error-model



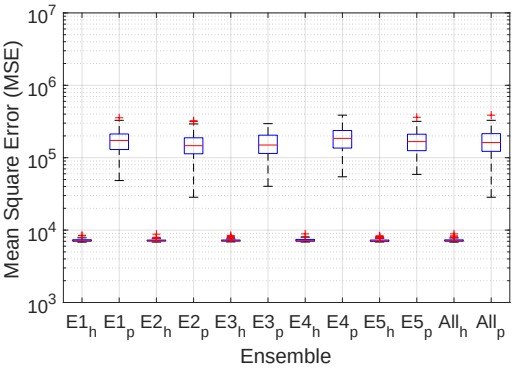
(c) inversion (neglecting model-error)



(d) joint inversion with error-model



(e) inversion (neglecting model-error)



(f) joint inversion with error-model

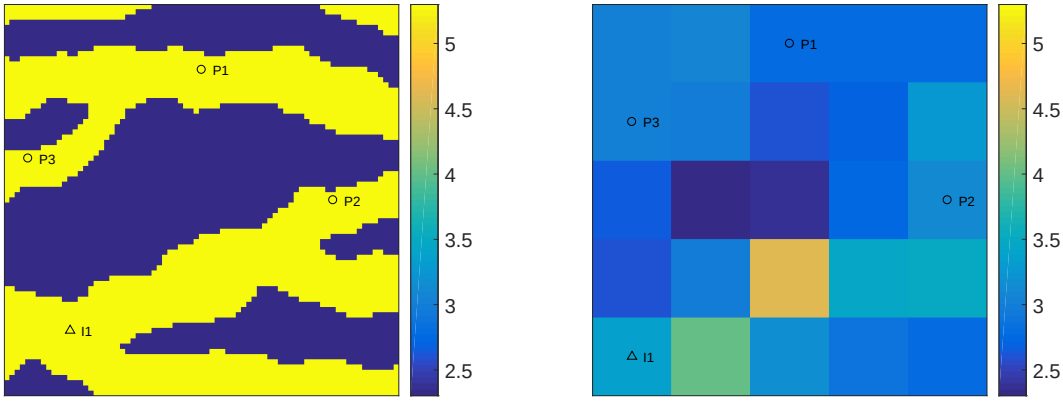
Figure 2.10: Forecasting metrics of coarse scale model case. In part (a) and (b) blue bars show the CP of true log-permeabilities, green bars show the CP of the historical data and yellow bars show the CP of prediction. In part (c) and (d) blue bars show the mean CRPS of the historical data and yellow bars show the mean CRPS of prediction. In part (e) and (f) box plots of MSE of the simulated well data from each ensemble are shown, subscript h and p are used for history and prediction respectively. On each box, the central red line indicates the median, and the bottom and top blue edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers represent extreme data points without outliers, and '+' symbol represents outliers (more than 1.5 times of interquartile range).

2.3.3 Case 2: Up-scaled imperfect geology model

Geologists commonly try to build geologically realistic prior models. However, maintaining the geological realism during the history-matching process is quite challenging (Sun and Durlofsky, 2017). For example, multipoint statistics (MPS) is widely used to represent channelized geological patterns. Geologically consistent history matching using the MPS prior is still a subject of active research (Chen et al., 2016). Sometimes the predictability of the history matched MPS models may not be satisfactory, often due to limitation of the available history matching methods in handling this type of non-Gaussian models (Chen et al., 2016). In this study we do not aim to obtain calibrated models that are consistent with the channelized geological feature, instead we focus on improving predictability of the calibrated coarse models by including the error-model.

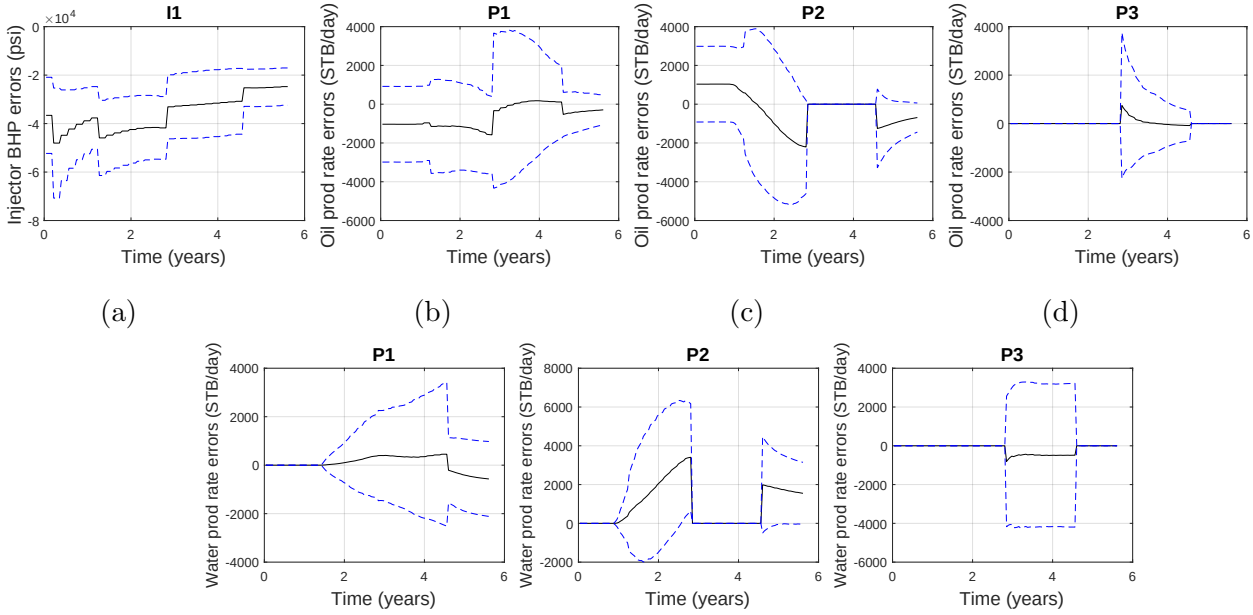
For this case, the permeability fields are based on a similar test case presented in (Chen et al., 2016). Figure 2.11(a) shows the reference fine scale log-permeability with channelized features and Fig. 2.11(b) shows the corresponding up-scaled log-permeability field in which the channelized features have been lost due to harmonic averaging. The reference and prior fine scale channelized log-permeability fields are generated using a two-facies training image with the direct sampling version of MPS (Mariethoz and Caers, 2014). The observed data are generated by the fine scale model using reference log-permeability field (Fig. 2.11(a)) with the addition of measurement noise of 2%.

Similar to the first test case, one hundred realizations of the model-discrepancy were obtained using Eq. 2.11 by running the fine scale simulation using the MPS permeability images of size 75×75 grid blocks and the corresponding up-scaled permeability field with 5×5 grid blocks. Figure 2.12 shows the prior statistics of the model-discrepancy in the simulated well production data.



(a) (b)

Figure 2.11: The fine scale (75×75) reference log-permeability with channelized features (a) and the corresponding up-scaled log-permeability (5×5) using harmonic average (b).



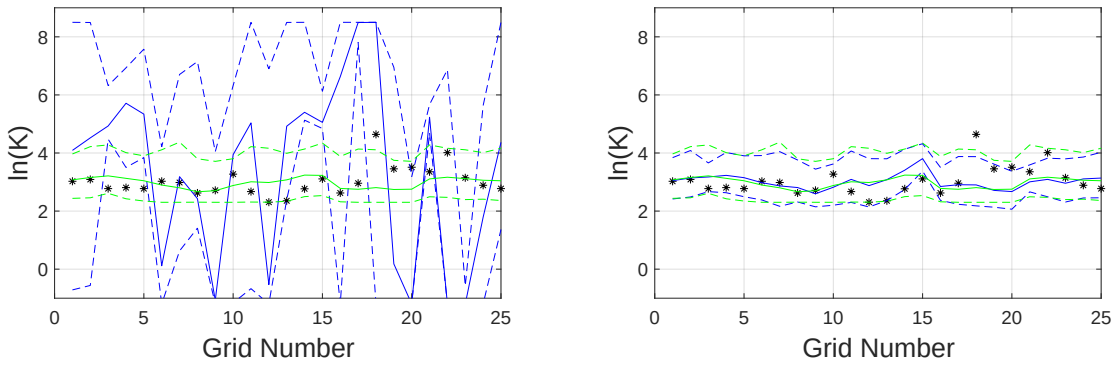
(a) (b) (c) (d)

(e) (f) (g)

Figure 2.12: Prior model-error statistics of all wells for Case 2. Black lines show mean model errors, dashed blue lines show the 95% confidence interval (mean plus and minus two standard deviations) of model errors.

2.3.4 Case 2 Results

In this test case, log-permeability of every grid cell is calibrated using both inversion (neglecting model-error) and joint inversion procedures. The ES-MDA settings for both calibration procedures are same as case 1. Figure 2.13 shows prior and posterior distributions obtained by combining realizations from all five runs for both history matching procedures. Figure 2.13(a) shows that the posterior distribution of ln(**K**) is biased where the estimated log-permeabilities show extreme values and do not capture the reference log-permeabilities. Figure 2.13(b) shows that by accounting for model error, relatively small changes have been made to the physical parameters (log-permeability in this case), and the mean of the posterior distribution of ln(**K**) remains smooth after data assimilation.



(a) inversion (neglecting model-error) (b) joint inversion with error-model

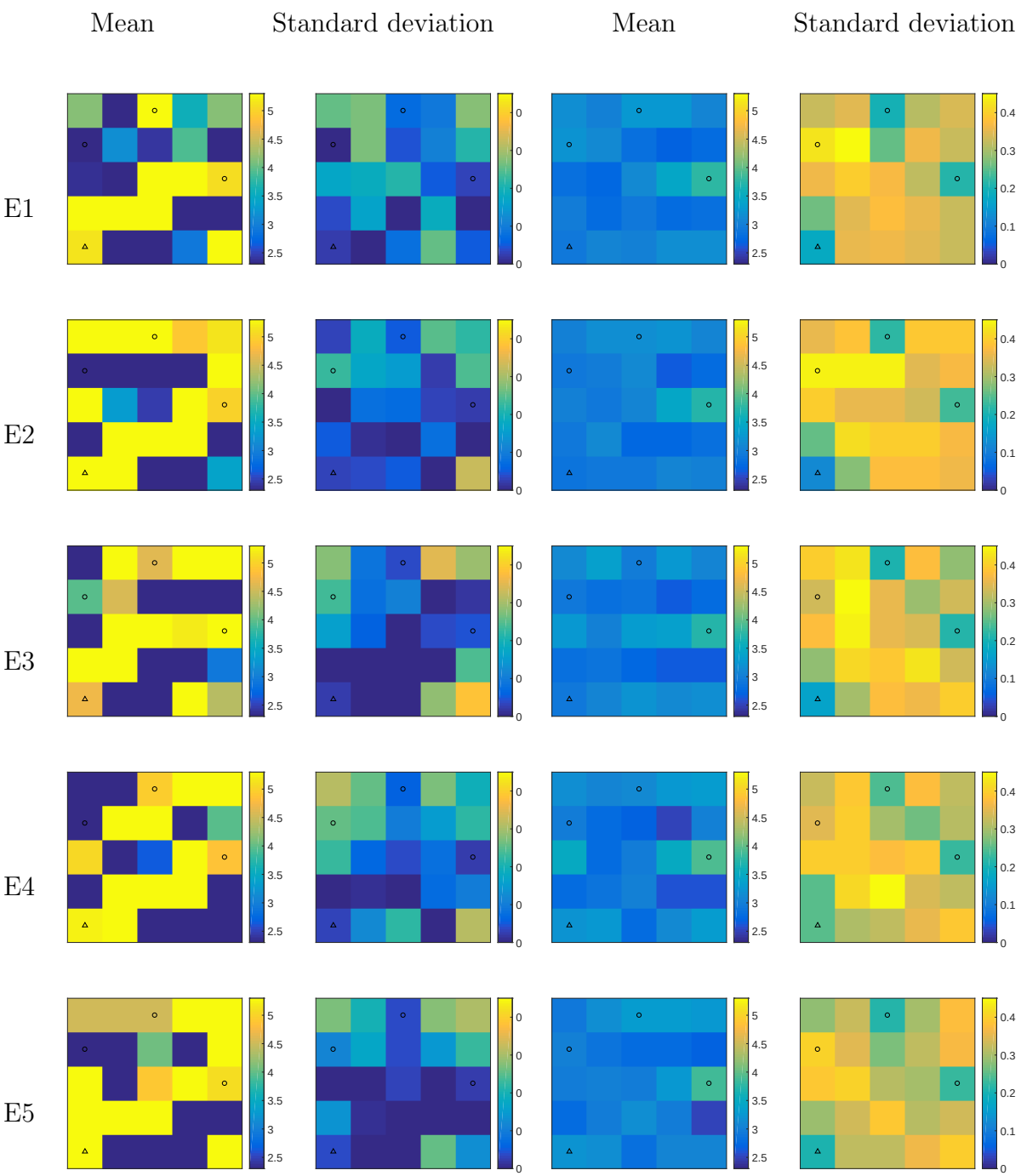
Figure 2.13: Prior and posterior distribution of ln(**K**) obtained after history matching for up-scaled imperfect geology model case using five ensembles. In both part (a) and (b), green and blue lines show the prior and posterior distribution respectively. Solid green and blue line show the $p50$ prior and posterior respectively. Dashed green and blue lines show the 95% confidence interval of prior and posterior respectively. Black asterisks show the reference solution.

Figure 2.14 shows the mean and standard deviation of ln(**K**) posterior ensembles as maps. In Fig. 2.14(a) the posterior mean log-permeability field obtained from five different runs are shown for the inversion (neglecting model-error) procedure. This

posterior mean is clearly different from the reference coarse log-permeability field as shown in Fig. 2.11(b). Moreover, in Fig. 2.14(a) the mean log-permeability field of every ensemble is different from each other, which is an indication of convergence of every ensemble to different non-unique local peak (mode) of the biased posterior. The mean of log-permeability ensemble from the joint inversion procedure is shown in Fig. 2.14(b) where no extreme features are observed. Similar to the observation from Case 1, the standard deviation is higher for the joint inversion compared to the inversion (neglecting model-error).

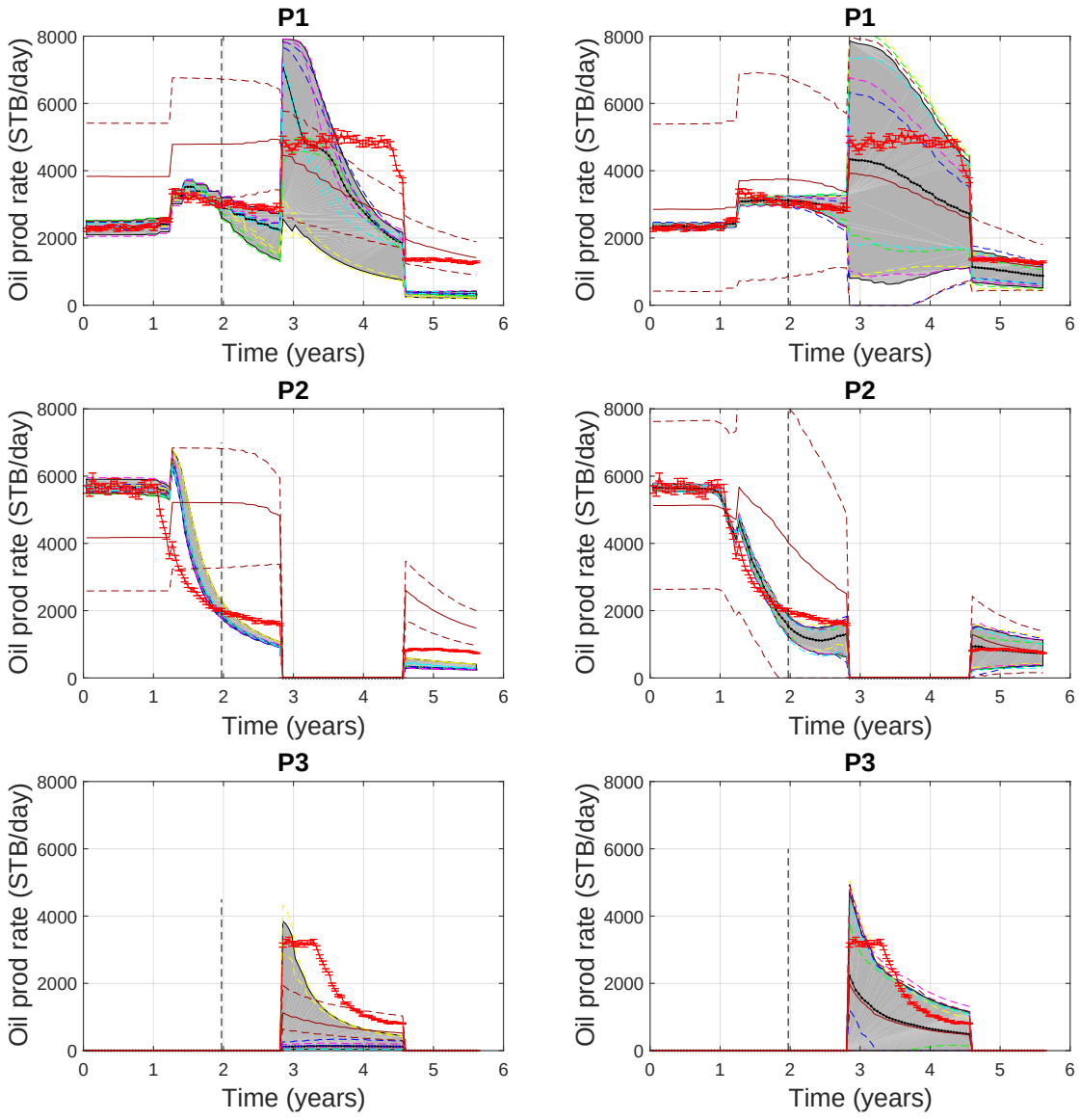
Figures 2.15 and 2.16 show the oil and water production rates of the different production wells and the bottom hole pressure of the injector well. The 50th percentile $p50$ and 95% confidence interval (the shaded region) are obtained by combining results from all five ensembles. In part (a) of these figures, the results for the inversion (neglecting model-discrepancy) are presented and the results of the joint-inversion with error-model are shown in part (b). For the inversion (neglecting model-error) procedure, only historical data at some wells are matched. For example Figs. 2.15(a) and 2.16(a) show that the data match is good for wells I1 and P1, however the data match of well P2 is not good. Moreover the future prediction from the estimated model parameters using the inversion (neglecting model-error) procedure is quite poor. A good example is the prediction of water production rate of well P1: the models predicted early water breakthrough between year 2 to 3, while the actual water breakthrough is after year 5 because P1 is separated from the injector by a low permeability region (see Fig. 2.11(a)). In comparison, better matches and predictions are obtained by the joint inversion procedure as shown in Figs. 2.15(b) and 2.16(b) for individual ensembles as well as all ensembles combined together. The prediction from the combined multiple ensembles may seem good for some well data for the case with the inversion (neglecting model-error) procedure, for example BHP pressure of well I1 in Fig. 2.16(a), even though the prediction from each individual ensemble is not good. This is often due to the fact that different ensemble converges to different local peaks (modes) of the biased posterior and the combined prediction from these multiple biased final ensembles happen to enclose the validation data.

Figure 2.17 shows the forecasting metrics (CP for the estimated log-permeabilities and mean CRPS, MSE, CP for the well data in history matching and prediction periods) for the individual ensembles and for results from all five ensembles assembled together. Figures 2.17(a) and 2.17(b) show the coverage probability of reservoir model parameters ($\ln(\mathbf{K})$), well data for both the history matching period and the forecast period. In Fig. 2.17(a) CP of $\ln(\mathbf{K})$ is between 0–0.12 for ensembles (E1 to E5), however CP of $\ln(\mathbf{K})$ is 0.6 for the combined ensembles. This relatively high coverage from the combined ensemble is due to the overshooting of $\ln(\mathbf{K})$ values and the different final estimation from each individual ensemble as shown in Fig. 2.13(a) and Fig. 2.14(a) respectively. With the joint inversion procedure, the coverage probability is improved for all three quantities investigated (log-permeabilities, historical data, future prediction). Similarly, the mean CRPS and MSE measures also show significant improvement by accounting for model error using the joint inversion procedure (2nd and 3rd row of Fig. 2.17). In addition, based on all three forecasting measures, the results from multiple ensemble runs using the joint inversion procedure are very consistent, which indicates the statistical consistency of the proposed procedure.



(a) inversion (neglecting model-error) (b) joint inversion with error-model

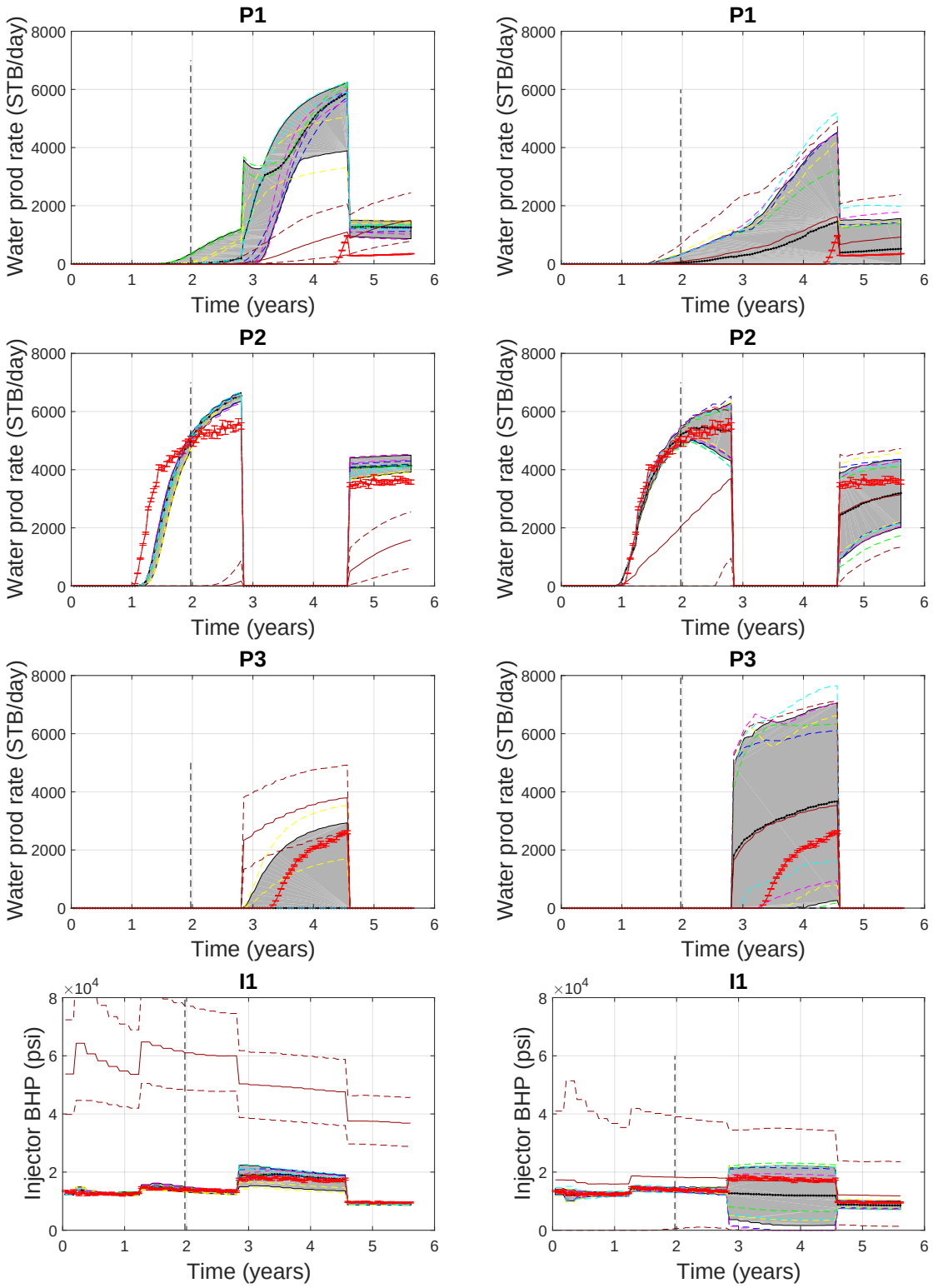
Figure 2.14: Mean and standard deviation of $\ln(\mathbf{K})$ posterior ensembles obtained after history matching of all grids log-permeabilities for up-scaled imperfect geology model case.



(a) inversion (neglecting model-error)

(b) joint inversion with error-model

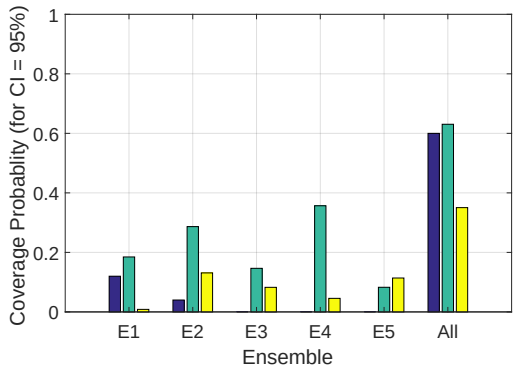
Figure 2.15: Prior and posterior of oil production data for up-scaled imperfect geological model case. Red lines show observation data and bar on red lines shows measurement error. Dashed black lines show end of historical period. Solid brown lines show 50th percentile $p50$ of the prior distribution, dashed brown lines show 95% confidence interval of prior distribution. Solid black lines show $p50$ posterior distribution obtained from all ensembles. Shaded gray area show 95% confidence interval of posterior distribution obtained from all ensembles. Dashed blue, green, yellow, magenta and cyan lines show 95% confidence interval of posterior distribution of individual ensembles.



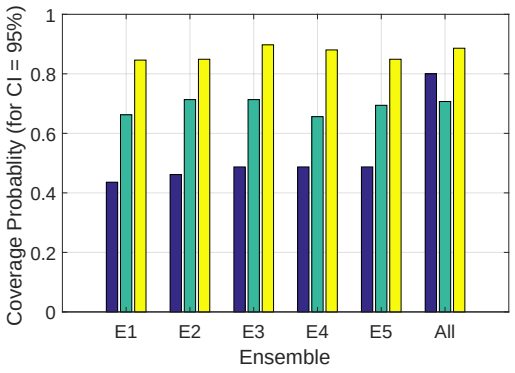
(a) inversion (neglecting model-error)

(b) joint inversion with error-model

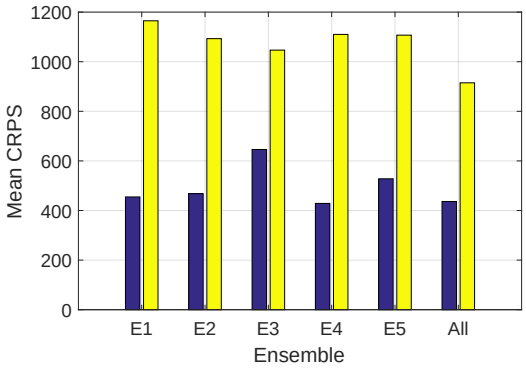
Figure 2.16: Prior and posterior of water production and injection pressure data for up-scaled imperfect geological model case. The explanation of colors and lines are the same as in Fig. 2.15.



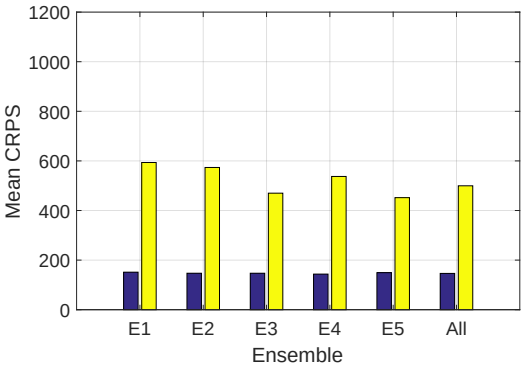
(a) inversion (neglecting model-error)



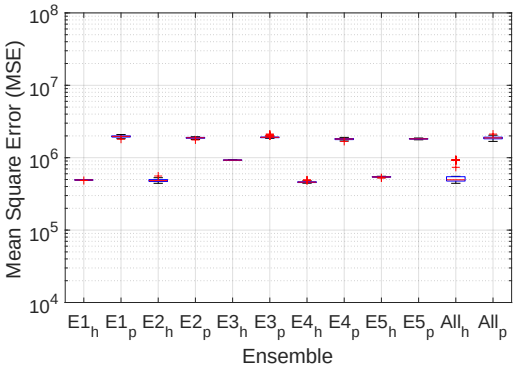
(b) joint inversion with error-model



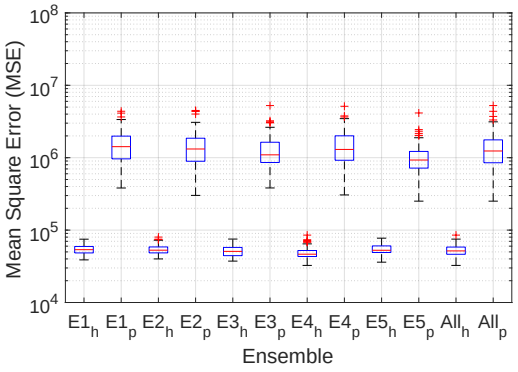
(c) inversion (neglecting model-error)



(d) joint inversion with error-model



(e) inversion (neglecting model-error)



(f) joint inversion with error-model

Figure 2.17: Forecasting metrics of up-scaled imperfect geology model case. In part (a) and (b) blue bars show the CP of true log-permeabilities, green bars show the CP of the historical data and yellow bars show the CP of prediction. In part (c) and (d) blue bars show the mean CRPS of the historical data and yellow bars show the mean CRPS of prediction. In part (e) and (f) box plots of MSE of the simulated well data from each ensemble are shown, subscript h and p are used for history and prediction respectively. On each box, the central red line indicates the median, and the bottom and top blue edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers represent extreme data points without outliers, and '+' symbol represents outliers (more than 1.5 times of interquartile range).

2.4 Conclusions

In this chapter, a generic procedure for history matching of imperfect/low-fidelity reservoir models has been developed where, we formulate the history matching problem as a joint inversion of reservoir model parameters and an error model parameters. We used principal component analysis to parameterize the error model, where the PCA basis function and prior statistics of the PCA basis weights were obtained using pairs of accurate and inaccurate models. We note that the accurate/high-fidelity model is only used for defining the prior model-error statistics and during history matching only the imperfect/low-fidelity model is used.

We evaluated the proposed history matching procedure on low-fidelity models with modeling errors due to aggressive grid coarsening/averaging of the permeability field obtained from two-point statistics and low-fidelity models where the main source of error is the grid coarsening/averaging of a channelized geology. Detailed comparison were performed against standard history matching (inversion while neglecting model error). The obtained results show that the estimated model parameters are biased using standard history matching procedure in the presence of large modeling errors. Consequently the calibrated low-fidelity model predictions are unreliable and generally inaccurate. Utilizing the developed joint inversion procedure results in significant improvements in terms of the quality of the estimated parameters, the matching capacity to historical data and prediction accuracy/reliability of the calibrated low-fidelity models. This is attributed to a reduction (and in some cases elimination) of the bias in the estimated posterior distribution of the model parameters when we included a flexible error-model terms of the inversion process. The numerical test cases were assessed using three forecasting metrics and it was observed that the consistency of ensemble-based history matching technique was also improved by including the error-model terms in the inversion procedure. We argue that this observed consistency might be due to the elimination of multiple biased peaks (modes) in the posterior distribution of the model parameters once the error modeling terms are included in the formulations.

The proposed framework is generally flexible and could be applied to large scale

models as the error-model formulation is I/O independent and the prior error-model parameters could be estimated before the history matching step. However, for general error-modeling an accurate model may be missing or the sources of the modeling errors could be unknown. It is also possible that the fine/high-fidelity model (which is assumed to be perfect) is also biased. In these cases, the proposed methodology can only improve the parameter estimation and the prediction up to the fine/high-fidelity model accuracy. Addressing the effects of unknown modeling errors without relying on an accurate (high-fidelity)/approximate (low-fidelity) model pairs is the subject of our future work.

Chapter 3

Robust algorithms for history matching of imperfect subsurface models while accounting for model error

In this work, we evaluate different algorithms to account for modeling errors while estimating the model parameters especially when the model discrepancy (a.k.a. modeling error) is large. Besides this, we introduce two new algorithms which are closely related to some of the published approaches under consideration. Considering all these algorithms, the first calibration approach (base case scenario) relies on Bayesian inversion using iterative ensemble smoothing with annealing schedules without any special treatment for the model-error. In the second approach, the residual obtained after calibration is used to iteratively update the total error covariance combining the effects of both modeling errors and measurements error. In the third approach, PCA-based error model is used to represent the model discrepancy during history matching. This leads to a joint inversion problem where both the model parameters and the parameters of a PCA-based error model are estimated. For the joint inversion within the Bayesian framework, prior distributions have to be defined for all the estimated parameters and the prior distribution for the PCA-based error model parameters are generally hard to define. In this study, the prior statistics of the model discrepancy parameters are estimated using the outputs from pairs of high-fidelity and low-fidelity models gener-

The contents of this chapter have been published in SPE Reservoir Simulation Conference (2019) and accepted in Journal, but includes modifications following the viva. https://doi.org/10.2118/193838-MS

ated from the prior realizations. The fourth approach is similar to the third approach, however an additional covariance matrix of difference between PCA-based error model and the corresponding actual realizations of prior error is added to the covariance matrix of the measurement error. This approach is same as Chapter 2 and presented as Algorithm 4 in this chapter.

The first newly introduced algorithm (fifth approach), relies on building an orthonormal basis for the misfit component of the error-model, which is obtained from difference between PCA-based error model and the corresponding actual realizations of the prior error. The misfit component of the error-model is subtracted from the data residual (difference between observations and model outputs) in order to eliminate the incorrect relative contribution to the prediction from the physical model and the error model. In the second newly introduced algorithm (sixth approach), we utilize PCA-based error model as a physically motivated bias correction term and an iterative update of the covariance matrix of the total error during history matching. All the algorithms are evaluated using three forecasting measures and the obtained results show that a good parameterization of the error-model is needed in order to obtain a good estimate of physical model parameters and to provide better predictions. In this study, the last three approaches (i.e. 4, 5, 6) outperform the other methods in terms of the quality of estimated model parameter and the prediction capability of the calibrated imperfect models.

3.1 Introduction

Subsurface reservoir modeling contains various approximations including: grid coarsening, petrophysical properties up-scaling, simplification of the complex physics of subsurface fluid flow, simplification of relative permeability and PVT model and incomplete description of reservoir structure and geology, etc. These approximations introduce considerable errors into the modeling process of subsurface reservoirs. The issue of model discrepancy is further exacerbated when the models are calibrated to match historical data as the model parameters could be adjusted to match the historical data

while compensating for the model errors. Different approaches have been developed to address this issue by accounting for the model-error component during the model calibration process (Josset et al., 2015; Dreano et al., 2017; Hansen et al., 2014; Reichert and Schuwirth, 2012b; Evin et al., 2014b).

Model errors could be formulated using different approaches and the different formulations could be generally categorized as output-dependent (Evin et al., 2014b), input-dependent (Giudice et al., 2013), Input/Output (Luo, 2019) dependent and Input/Output (I/O) independent (Rammay et al., 2019). In input-dependent formulations, the model discrepancy is defined as a function of the uncertain model parameters. During history matching of subsurface reservoir models, coarsening/upscaling errors has formulated an input-dependent function. O' Sullivan and Christie (2005) used this type of approach of accounting for model-error where the authors computed model-error realizations using the difference between a fine-grid and coarse-grid model outputs. During the calibration process, an interpolation of the error component was performed to estimate a correction term to the coarse-grid model predictions. Lødøen and Tjelmeland (2010) used the multiple linear regression algorithm to formulate the model-error. In their study, the residual of multiple regression is dependent on permeability realizations and they used Gaussian process regression (GPR) to model input dependent residual. Giudice et al. (2013) used input dependent model-error to improve uncertainty estimation in urban hydrological modelling. They used the variance of model-error dependent on the input (rainfall in their case). Input-dependent model discrepancy formulation has severe limitations in case of calibration of large scale models, because of the highly non-linear relationship between the model errors and large number of input parameters (thousands or millions). In Input/Output dependent formulations, model-errors are treated as a functional approximation problem, which can be solved using a generic machine learning method, such as kernel-based learning (Luo, 2019). However, these type of formulations are prone to be more data-driven due to the strong tendency of the machine learning algorithms to absorb a large amount of data residuals as compared to the simulation models. For large scale models, model-error formulation using output dependent or I/O independent methodologies would be be

more feasible. In output-dependent formulations, the model discrepancy is formulated as a function of the model response/output. Evin et al. (2014b) represented model discrepancy using an output-dependent formulation for hydrological systems and utilized model-error as a function of the streamflow obtained from the outputs of the hydrological model. Input/Output (I/O) independent formulations depend on the prior statistics of the model-error parameters, which is computed using prior realizations of pairs of high-fidelity and low-fidelity models response (Rammay et al., 2019).

Another aspect for classifying the various methods for the model discrepancy is related to whether the model discrepancy is formulated by post-processing the physical model outputs as in external bias description (EBD) or the physical model equations or states are modified to account for the model discrepancy as in the internal noise description (IND) (Giudice et al., 2015). IND is also known as stochastic gray-box or state space modeling [c.f., Moradkhani et al., 2012; Kristensen et al., 2004]. Some could argue about the superior performance for IND methods, however the main challenge of any IND method is the need to modify the simulator code or to use restart files, which is either difficult to perform or introduces an additional computational bottleneck. In terms of recent advances, Köpke et al. (2018) used orthonormal basis, which is generated from the differences between pairs of high-fidelity and low-fidelity models, to filter the model-error components from the calibration residuals during the history matching of imperfect models and Oliver and Alfonzo (2018) added physically motivated parameters for bias corrections and estimated the total error covariance matrix using an ensemble of data residuals from the calibrated models.

In this chapter, we investigate different algorithms for model calibration of imperfect models. These algorithms are inspired by the work of Rammay et al. (2019), Köpke et al. (2018) and Oliver and Alfonzo (2018). Besides this, we develop two new algorithms (Algorithms 5 and 6) by introducing novel modification to the published algorithms (Köpke et al., 2018; Oliver and Alfonzo, 2018) in terms of identifiability of model-error parameters and the use of joint calibration of both the pre-determined error-model and the physical model. Our focus is on flexible methods that can handle strong serially correlated outputs, variable boundary conditions (i.e. variable well

controls) and structured errors (which show specific patterns w.r.t time i.e. strong correlation in time). The evaluation is performed on a test case representing model discrepancy due to grid coarsening and upscaling of geological features. This test case is motivated by the fact that generally geological models are a coarse representation of reality. We limit the test setting to cases where a high-fidelity model without any model-error is available. This allows for proper evaluation of the different algorithms. Ensemble smoother with multiple data assimilation (ES-MDA) (Emerick and Reynolds, 2013) is used as a robust and practical Bayesian inversion algorithm. The outline of the chapter is given as follows: In Section 3.2, we present different algorithms/approaches of history matching while accounting for model-error. In Section 3.3, we present the case study. Following that, the case study results are discussed in Section 3.4 with the conclusions in Section 3.5.

3.2 Methodology

Bayesian inverse modeling is a flexible framework that has the capacity to account for the different sources of uncertainties in the model by following Bayes rule:

$$p(\mathbf{m}|\mathbf{d}_{obs}) \propto p(\mathbf{d}_{obs}|\mathbf{m})\ p(\mathbf{m}), \tag{3.1}$$

where $\mathbf{m} \in \mathbb{R}^{N_m \times 1}$ is the model parameters vector of size N_m, $p(\mathbf{m})$ is the prior probability of the model parameters, $\mathbf{d}_{obs} \in \mathbb{R}^{N_d \times 1}$ is the observations vector of size N_d and $p(\mathbf{d}_{obs}|\mathbf{m})$ is the likelihood of the data given the model parameters vector $\mathbf{m}$. Several algorithms could be used to generate samples from the posterior distribution of the model parameters (Oliver et al., 2008). In this work, the ensemble smoother with multiple data assimilation (ES-MDA) (Emerick and Reynolds, 2013), is utilized for Bayesian inversion due to the parallel nature of the algorithm and the robustness of the method in history matching of large reservoir models. Algorithm 1 presents the steps of ES-MDA in a functional form and represents the standard history matching procedure (base case scenario), in which an imperfect model is used for simulation during history matching but the model-error is neglected. In Algorithm 1, the covariance matrix of

measurement noise $\mathbf{C}_d$ is used in the update equation, i.e. $\mathbf{C}_D = \mathbf{C}_d$, as the forward model is assumed to be perfect.

Algorithm 1 History matching without accounting for model-error

$\mathbf{C}_D = \mathbf{C}_d$ $\quad \triangleright$ $\mathbf{C}_d$ is the covariance (diagonal) matrix of measurement errors

$\mathbf{M}_{prior} = [\mathbf{m}_1\ \mathbf{m}_2\ \mathbf{m}_3\ .....\ \mathbf{m}_{N_e}]$ $\quad \triangleright$ Generate prior ensemble of model parameters $\mathbf{M}_{prior} \in \mathbb{R}^{N_m \times N_e}$ consists of N_e realizations

$g = \tilde{g}$ $\quad \triangleright$ Set imperfect model $\tilde{g}$ as the forward model

$\mathbf{D}_{obs} = \mathbf{d}_{obs}\overrightarrow{1_{N_e}}$ $\quad \triangleright$ $\overrightarrow{1_{N_e}} \in \mathbb{R}^{1 \times N_e}$ is a row vector of ones

$\mathbf{M}_{post}, \mathbf{D}_{post} \leftarrow$ ESMDA($\mathbf{M}_{prior}$, $\mathbf{D}_{obs}$, $\mathbf{C}_D$, g)

function ESMDA($\mathbf{M}$, $\mathbf{D}_{obs}$, $\mathbf{C}_D$, g)

Inputs: $\mathbf{M} \in \mathbb{R}^{N_m \times N_e}$ is the ensemble of model parameters, g is the forward model, $\mathbf{C}_D$ is the error/noise covariance matrix.

Choose N_a $\quad \triangleright$ Number of data assimilations/iterations

$i \leftarrow 1$

$\alpha = N_a$

while $i <= N_a$ **do**

$\mathbf{D} = g(\mathbf{M})$ $\quad \triangleright$ Generate ensemble of model predictions $\mathbf{D} \in \mathbb{R}^{N_d \times N_e}$

$\mathbf{C}_{MD} = \frac{1}{N_e - 1}(\mathbf{M} - \bar{\mathbf{M}}\overrightarrow{1_{N_e}})(\mathbf{D} - \bar{\mathbf{D}}\overrightarrow{1_{N_e}})^\intercal$ $\quad \triangleright$ $\bar{\mathbf{M}} \in \mathbb{R}^{N_m \times 1}$ is ensemble $\mathbf{M}$ mean

$\mathbf{C}_{DD} = \frac{1}{N_e - 1}(\mathbf{D} - \bar{\mathbf{D}}\overrightarrow{1_{N_e}})(\mathbf{D} - \bar{\mathbf{D}}\overrightarrow{1_{N_e}})^\intercal$ $\quad \triangleright$ $\bar{\mathbf{D}} \in \mathbb{R}^{N_d \times 1}$ is ensemble $\mathbf{D}$ mean

$\mathbf{D}_{uc} = \mathbf{D}_{obs} + \sqrt{\alpha}\ \mathbf{C}_D^{1/2}\ \mathbf{Z}_d$, $\quad \triangleright$ Observation perturbations, $\mathbf{Z}_d = [\mathbf{z}_{d1}\ \mathbf{z}_{d2}\ \mathbf{z}_{d3}\ .....\ \mathbf{z}_{dN_e}] \in \mathbb{R}^{N_d \times N_e}$, $\quad \mathbf{z}_d \sim \mathcal{N}(0, \mathbf{I}_{N_d}) \in \mathbb{R}^{N_d \times 1}$

$\mathbf{M} \leftarrow \mathbf{M} + \mathbf{C}_{MD}\ (\mathbf{C}_{DD} + \alpha\ \mathbf{C}_D)^{-1}(\mathbf{D}_{uc} - \mathbf{D})$ $\quad \triangleright$ Update ensemble

$i \leftarrow i + 1$

end while

$\mathbf{M}_{post} = \mathbf{M}$ $\quad \triangleright$ $\mathbf{M}_{post} \in \mathbb{R}^{N_m \times N_e}$ is the posterior ensemble of model parameters

$\mathbf{D}_{post} = g(\mathbf{M}_{post})$ $\quad \triangleright$ Generate posterior ensemble of model predictions $\mathbf{D}_{post}$

return $\mathbf{M}_{post}$, $\mathbf{D}_{post}$

end function

Algorithm 2 presents the history matching procedure while accounting for the model-error by iteratively updating the total error covariance matrix (Oliver and Alfonzo, 2018). In Algorithm 2, the covariance matrix of the total error $\mathbf{C}_D$ is initially set to equal the covariance matrix of the measurement errors $\mathbf{C}_d$. In subsequent iterations, $\mathbf{C}_D$ is updated based on the residual from the previous history matching iteration. The algorithm is terminated after satisfying a model diagnostic criterion or reaching a prespecified maximum number of iterations. In Algorithm 2, the value of diagnostic criterion is estimated from the error covariance matrix and residual obtained after history matching.

Algorithm 3 presents a joint history matching of the physical model parameters and the parameters of a PCA-based error model (Rammay et al., 2019). Before history matching, PCA basis functions are extracted from the differences between simulation outputs of a accurate/high-fidelity model $g_p(\cdot)$ and an approximate model $\tilde{g}(\cdot)$. The prior statistics of the PCA basis weights (i.e. parameters of the model-discrepancy) are then estimated. A small number of the basis functions is retained for the PCA-based error-model to avoid adding too many parameters to the inversion problem and eliminate the issue of over-fitting. In this formulation, the predictions are made by summing the simulation outputs using the estimated physical model parameters and estimated model-errors represented by the PCA-based formulation. The summation of the approximate physical model and the PCA-based error model in Algorithm 3 is referred to as the corrected model ($g = \tilde{g} + \hat{\boldsymbol{\epsilon}}_m$).

Algorithm 4 accounts for the remaining mismatch between the corrected functions model and the accurate/high-fidelity model due to the truncation of the PCA basis (Rammay et al., 2019). The prior mismatch between predictions from the corrected model and predictions from the accurate/high-fidelity model g_p due to the truncation of PCA-basis of error-model (Algorithm 3) can also be called as second-order errors. This remaining mismatch (i.e. second-order errors), are loosely captured by inflating the observation noise covariance $\mathbf{C}_d$ with a diagonal matrix $\mathbf{C}_T$, that is the variance of the remaining mismatch before history matching.

The newly introduced Algorithm 5 builds on ideas from the recent work of Köpke

Algorithm 2 History matching with iterative update of total error covariance matrix

$k \leftarrow 1$

$\mathbf{C}_D{}^{(k)} = \mathbf{C}_d$

$\mathbf{M}_{prior} = [\mathbf{m}_1\ \mathbf{m}_2\ \mathbf{m}_3\ .....\ \mathbf{m}_{N_e}]$ ▷ Generate prior ensemble of model parameters

$\mathbf{M}_{prior} \in \mathbb{R}^{N_m \times N_e}$ consists of N_e realizations

$g = \tilde{g}$

$\mathbf{D}_{obs} = \mathbf{d}_{obs}\overrightarrow{1_{N_e}}$ ▷ $\overrightarrow{1_{N_e}} \in \mathbb{R}^{1 \times N_e}$ is a row vector of ones

while (True) **do**

  $\mathbf{M}_{post}, \mathbf{D}_{post} \leftarrow$ ESMDA($\mathbf{M}_{prior}$, $\mathbf{D}_{obs}$, $\mathbf{C}_D{}^{(k)}$, g) ▷ From Algorithm 1

  $\mathbf{R}^{(k)} = (\mathbf{D}_{obs} - \mathbf{D}_{post})$ ▷ Compute ensemble of residuals obtained after history matching

  $\mathbf{S}^{(k)} = (\mathbf{R}^{(k)})^{\intercal}(\mathbf{C}_D{}^{(k)})^{-1}(\mathbf{R}^{(k)})$

  $s_d{}^{(k)} = \frac{1}{2N_e}\sum(\mathrm{diag}(\mathbf{S}^{(k)}))$ ▷ Compute value of diagnostic criterion

  if $|s_d{}^{(k)} - s_d{}^{(k-1)}| < 0.1$ and $s_d{}^{(k)} < 2N_d$ **then**

    Output $\mathbf{M}_{post}$, $\mathbf{D}_{post}$

    Exit loop

  else

    $\mathbf{C}_D{}^{(k+1)} = \frac{1}{N_e}(\mathbf{R}^{(k)})(\mathbf{R}^{(k)})^{\intercal}$

    $k \leftarrow k + 1$

  end if

end while

et al. (2018) and Algorithm 3. In this algorithm the imperfect simulation model and PCA-based error-model are used in history matching while accounting for second-order errors using the algorithm proposed by Köpke et al. (2018). In Algorithm 5, an additional orthonormal basis is built on second-order errors using the Gram-Schmidt procedure (Cheney and Kincaid, 2009). In the Gram-Schmidt method, orthonormal vectors are computed based on the projection of the set of vectors under consideration (Cheney and Kincaid, 2009). In this work, the Gram-Schmidt procedure is implemented by the MATLAB orth function (MATLAB, 2019). In Algorithm 5, the orthonormal

Algorithm 3 History matching with PCA-based error model

for $r = 1, N_r$ **do**

  $\boldsymbol{\epsilon}_{mr} = g_p(\mathbf{m}_r) - \tilde{g}(\mathbf{m}_r)$ ▷ Compute N_r number of prior error realizations

end for

$\bar{\boldsymbol{\epsilon}}_m = \frac{1}{N_r}\sum_{r=1}^{N_r}(\boldsymbol{\epsilon}_{mr})$ ▷ Compute mean of prior error realizations

$\mathbf{C}_e = \frac{1}{N_r-1}\sum_{r=1}^{N_r}(\boldsymbol{\epsilon}_{mr} - \bar{\boldsymbol{\epsilon}}_m)(\boldsymbol{\epsilon}_{mr} - \bar{\boldsymbol{\epsilon}}_m)^\intercal$ ▷ Compute covariance matrix of prior error realizations

$[\mathbf{U}, \boldsymbol{\Sigma}, \mathbf{V}^\intercal] = \text{svd}(\mathbf{C}_e)$ ▷ Singular value decomposition of $\mathbf{C}_e$

$\boldsymbol{\Phi} = \mathbf{U}(:, 1:L)$ ▷ Taking L singular vectors

for $r = 1, N_r$ **do**

  $\boldsymbol{\beta}_r = \boldsymbol{\Phi}^\intercal(\boldsymbol{\epsilon}_{mr} - \bar{\boldsymbol{\epsilon}}_m)$

end for

$\boldsymbol{\mu}_\beta = \frac{1}{N_r}\sum_{r=1}^{N_r}(\boldsymbol{\beta}_r)$ ▷ Compute prior mean of PCA-error model parameters

$\mathbf{C}_\beta = \frac{1}{N_r-1}\sum_{r=1}^{N_r}(\boldsymbol{\beta}_r - \boldsymbol{\mu}_\beta)(\boldsymbol{\beta}_r - \boldsymbol{\mu}_\beta)^\intercal$ ▷ Compute prior covariance matrix of $\boldsymbol{\beta}$

$\boldsymbol{\beta}_{prior} \sim \mathcal{N}(\boldsymbol{\mu}_\beta, \mathbf{C}_\beta)$ ▷ Sample ensemble of PCA-error model parameters

$g = \tilde{g} + \hat{\boldsymbol{\epsilon}}_m$ ▷ Forward model is the combination of physical and error model $\hat{\boldsymbol{\epsilon}}_m = \boldsymbol{\Phi}\boldsymbol{\beta} + \bar{\boldsymbol{\epsilon}}_m$

$\mathbf{M}_{prior} = [\mathbf{M}_{prior}; \boldsymbol{\beta}_{prior}]$ ▷ Ensemble of model parameters is the combination of physical and error model parameters

$\mathbf{D}_{obs} = \mathbf{d}_{obs}\overrightarrow{1_{N_e}}$ ▷ $\overrightarrow{1_{N_e}} \in \mathbb{R}^{1\times N_e}$ is a row vector of ones

$\mathbf{C}_D = \mathbf{C}_d$

$\mathbf{M}_{post}, \mathbf{D}_{post} \leftarrow \text{ESMDA}(\mathbf{M}_{prior}, \mathbf{D}_{obs}, \mathbf{C}_D, g)$ ▷ From Algorithm 1

basis is projected on the data residuals which are used to filter the data residual during the joint calibration process described in Algorithm 3. This additional step provides a regularization effect and is used to prevent over-fitting the model parameters to data features that are not captured by the corrected model.

Algorithm 5 differs from the work presented in Köpke et al. (2018), where the orthonormal basis is built directly from the difference between the approximate physical model and the perfect/high-fidelity model without introducing any functional approxi-

Algorithm 4 History matching with PCA-based error-model and its noise covariance matrix

Algorithm 3 except last two lines
for $r = 1, N_r$ **do**
    $\hat{\boldsymbol{\epsilon}}_{mr} = \boldsymbol{\Phi}\boldsymbol{\beta}_r + \bar{\boldsymbol{\epsilon}}_m$
    $\boldsymbol{\zeta}_{mr} = \boldsymbol{\epsilon}_{mr} - \hat{\boldsymbol{\epsilon}}_{mr}$
end for
$\mathbf{C}_T = \frac{1}{N_r - 1} \sum_{r=1}^{N_r} (\boldsymbol{\zeta}_{mr})(\boldsymbol{\zeta}_{mr})^\intercal$
$\mathbf{C}_T = \mathrm{diag}(\mathbf{C}_T)$
$\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_T$
$\mathbf{M}_{post}, \mathbf{D}_{post} \leftarrow \mathrm{ESMDA}(\mathbf{M}_{prior}, \mathbf{D}_{obs}, \mathbf{C}_D, g)$ ▷ From Algorithm 1

Algorithm 5 History matching with PCA-based error-model while accounting for second-order errors

Algorithm 3 except last line
for $r = 1, N_r$ **do**
    $\hat{\boldsymbol{\epsilon}}_{mr} = \boldsymbol{\Phi}\boldsymbol{\beta}_r + \bar{\boldsymbol{\epsilon}}_m$
    $\boldsymbol{\zeta}_{mr} = \boldsymbol{\epsilon}_{mr} - \hat{\boldsymbol{\epsilon}}_{mr}$
end for
$\mathbf{E}_p = [\boldsymbol{\zeta}_{m1}\ \boldsymbol{\zeta}_{m2}\ \boldsymbol{\zeta}_{m3}\ .....\ \boldsymbol{\zeta}_{mN_r}]$ ▷ $\mathbf{E}_p \in \mathbb{R}^{N_d \times N_r}$ consists of N_r realizations of $\boldsymbol{\zeta}_m$
$\mathbf{B} = \mathrm{GS}(\mathbf{E}_p)$ ▷ Build orthonormal basis B using Gram-Schmidt (GS) procedure
if $\mathbf{B}\mathbf{B}^\intercal \approx \mathbf{I}$ **then**
    $\lambda = \mathrm{diag}(\boldsymbol{\Sigma})$ ▷ Taking Eigen values from Algorithm 3
    Select minimum value of L_1 such that $\frac{\sum_{l=1}^{L_1}(\lambda_l)}{\sum_{l=1}^{N_d}(\lambda_l)} = 1$
    $\mathbf{B} = \mathbf{U}(:, L+1 : L+L_1)$ ▷ $\mathbf{U}$ and L from Algorithm 3
end if
$\mathbf{E}(\mathbf{M}) = \mathbf{B}\mathbf{B}^\intercal \mathbf{R}(\mathbf{M})$ ▷ error misfit depends on $\mathbf{B}$ and residual $\mathbf{R}(\mathbf{M}) = \mathbf{d}_{obs}\overrightarrow{1_{N_e}} - g(\mathbf{M})$
$\mathbf{D}_{obs} = \mathbf{d}_{obs}\overrightarrow{1_{N_e}} - \mathbf{E}$ ▷ $\mathbf{E}$ is updated with the change of $\mathbf{M}$ during history matching
$\mathbf{M}_{post}, \mathbf{D}_{post} \leftarrow \mathrm{ESMDA}(\mathbf{M}_{prior}, \mathbf{D}_{obs}, \mathbf{C}_D, g)$ ▷ From Algorithm 1

mation to the model discrepancy. We note that the change in data residuals after each ES-MDA iteration updates the ensemble of second-order errors and as a consequence $\mathbf{D}_{obs}$ is also updated for the following iteration as shown in the lines 8 and 9 of Algorithm 5. In some sense, the second-order errors in algorithm 5 are filtered before each history matching iteration while in 4 the second-order errors are estimated before history matching and remain the same during iterations. The newly introduced Algorithm 6 builds on Algorithm 2 (Oliver and Alfonzo, 2018) and Algorithm 3, where the corrected model (approximate physical model corrected by PCA-based error model) is used in history matching with an estimate of the total error covariance at the end of each ES-MDA run until a stopping criterion is satisfied.

Algorithm 6 History matching with PCA-based error-model and iterative update of total covariance matrix

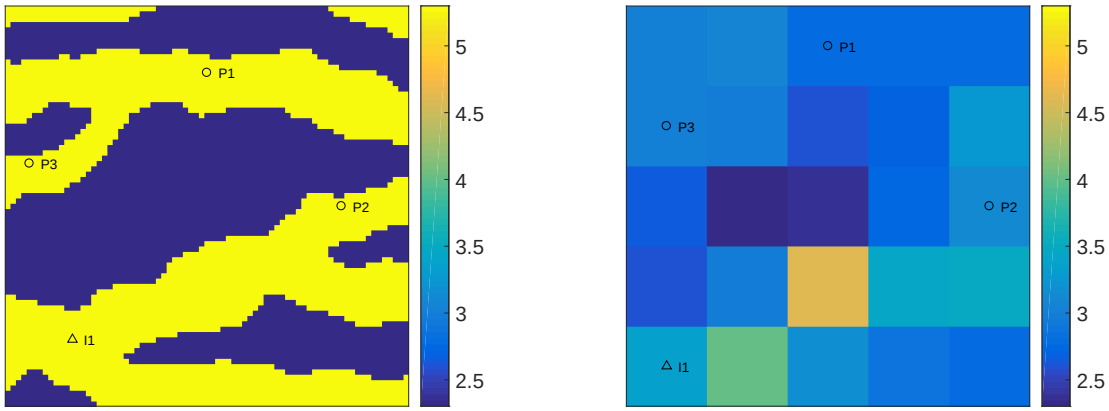
1: $k \leftarrow 1$
2: Algorithm 3 except last two lines
3: ${\mathbf{C}_D}^{(k)} = \mathbf{C}_d$
4: **while** (True) **do**
5: $\quad \mathbf{M}_{post}, \mathbf{D}_{post} \leftarrow \text{ESMDA}(\mathbf{M}_{prior}, \mathbf{D}_{obs}, {\mathbf{C}_D}^{(k)}, g)$ $\quad \triangleright$ From Algorithm 1
6: $\quad \mathbf{R}^{(k)} = (\mathbf{D}_{obs} - \mathbf{D}_{post})$ $\quad \triangleright$ Compute ensemble of residuals obtained after history matching
7: $\quad \mathbf{S}^{(k)} = (\mathbf{R}^{(k)})^{\intercal}({\mathbf{C}_D}^{(k)})^{-1}(\mathbf{R}^{(k)})$
8: $\quad {s_d}^{(k)} = \frac{1}{2N_e}\sum(\text{diag}(\mathbf{S}^{(k)}))$
9: $\quad$ **if** $|{s_d}^{(k)} - {s_d}^{(k-1)}| < 0.1$ and ${s_d}^{(k)} < 2N_d$ **then**
10: $\quad\quad$ Output $\mathbf{M}_{post}$, $\mathbf{D}_{post}$
11: $\quad\quad$ Exit loop
12: $\quad$ **else**
13: $\quad\quad {\mathbf{C}_D}^{(k+1)} = \frac{1}{N_e}(\mathbf{R}^{(k)})(\mathbf{R}^{(k)})^{\intercal}$
14: $\quad\quad k \leftarrow k + 1$
15: $\quad$ **end if**
16: **end while**

3.3 Case Study

We evaluated the various history matching algorithms on a simplified reservoir model. The reservoir model/test case is same as Case 2 of the Chapter 2. The dimension of the subsurface reservoir is 7500 ft $\times$ 7500 ft $\times$ 20 ft in the x, y and z directions, respectively. Incompressible two-phase porous media flow of oil and water is considered and simulated using a 2-D grid with Matlab Reservoir Simulation Tool-box (MRST) (Lie, 2016). The initial reservoir pressure is 5000 psi and the reservoir has uniform porosity of 20%. Corey's power law model is used to represent relative permeabilities and the parameter values for the Corey's model and fluid properties are listed in Appendix A. The capillary pressure and gravitational effects are neglected. The reference permeability model is a channelized model of dimension 75×75 generated using Multiple Point Statistics (MPS) (Mariethoz and Caers, 2014). This reference model is used to generate the observation data. The log-transformed permeability of the true model is shown in Fig. 3.1(a). The simulation model used for history matching is a much coarser model with only 5×5 cells. The corresponding up-scaled (5×5) field of the log-permeability is shown in Fig. 3.1(b). The upscaled model is obtained from harmonic averaging of the permeability field of 75×75 high-fidelity model. For the upscaled model, the physical model parameters are the permeability values at the 25 cells. In this test case, the model errors are a consequence of aggressive grid coarsening, slight change in well locations and poor geological representation.

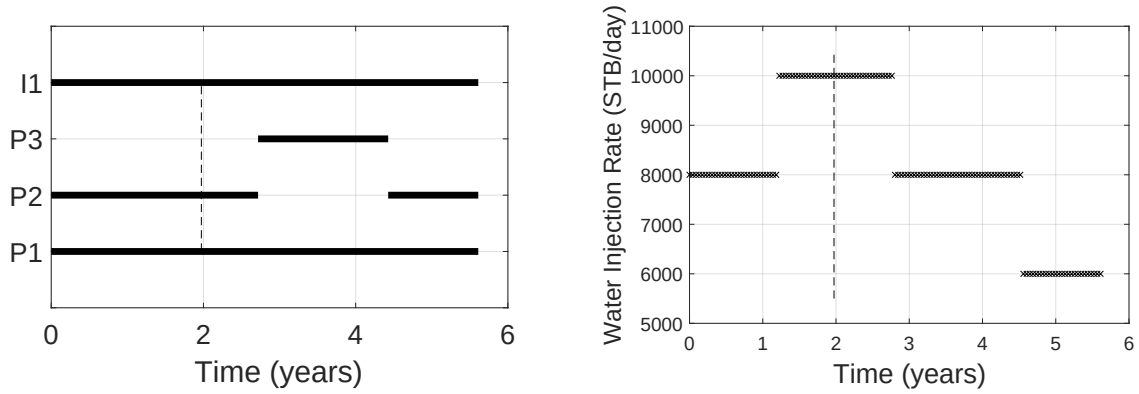
The reservoir contains four wells, three producers (P1, P2, P3) and one injector (I1). The well controls are shown in Fig. 3.2 with the end of historical period marked as a vertical dotted line. Well P3 is used only in the prediction phase and is used to assess the prediction capacity of the calibrated models. The production wells are operated under constant bottom hole pressure constraint of 4500 psi and the injector well is operated under constant injection rate constraint with varying control values as shown in Fig. 3.2(b).

The objective of this study is to investigate the performance of different algorithms presented in Section 3.2, in terms of the quality of the estimated physical model param-



(a) reference log-permeability (b) corresponding coarse model

Figure 3.1: The reference (true) log-permeability (75×75) with channelized features (a) and the corresponding (reference) up-scaled log-permeability (5×5) using harmonic average (b).



(a) (b)

Figure 3.2: Wells schedule and controls. Part (a) shows wells open (shown by solid back lines) and shut schedule. Water injection rate of the injector well I1 is shown in part (b). End of historical period is shown by dashed black lines.

eters and the quality of predictions obtained by the calibrated models. The implementation of Algorithms 3, 4, 5 and 6 requires extraction of the PCA basis functions for the model discrepancy and the corresponding prior weights for the parameters of the error-model based on output differences between pairs of high-fidelity and low-fidelity simulation models generated from the prior realizations of the permeability field. For

this step, 100 realizations of the model discrepancy are created by taking the difference of the fine scale model output/response using 100 log-permeability fields generated by MPS algorithm of size 75×75 and the corresponding upscaled model responses (harmonic averaging). The upscaling of the permeability fields to a 5×5 grid blocks is quite aggressive and is expected to result in large model discrepancies. The reference 100 fine scale MPS permeability realizations are generated in the same way as the true permeability model shown in Figure 3.1(a). The statistics of the 100 model discrepancy realizations are shown in Figure 3.3 for each time series of well production data. In this chapter the combination of multi dimensional scaling and clustering (Scheidt and Caers, 2009) are utilized to sample the 100 realizations of the model parameters to effectively estimate the prior model-error statistics.

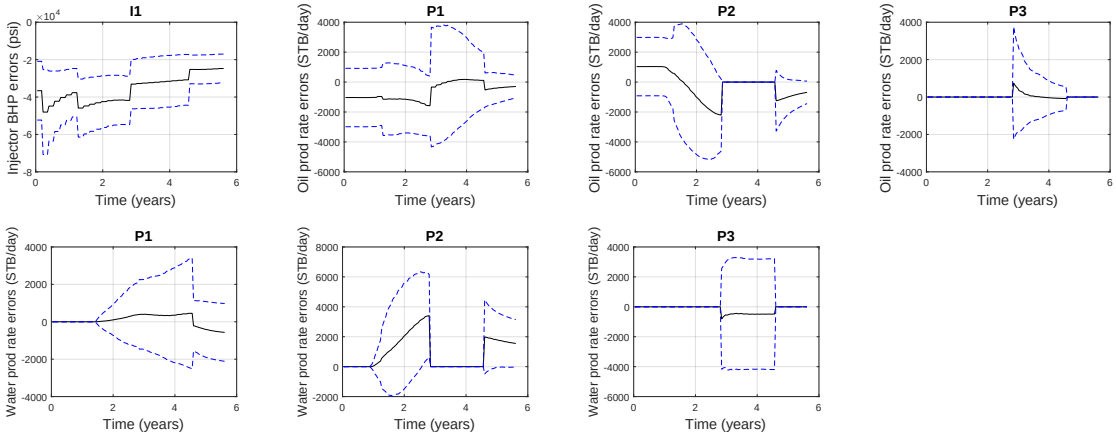


Figure 3.3: The statistic of the model discrepancy between 100 pairs of high-fidelity (75×75) and low-fidelity (5×5) models. Solid black and dashed blue lines show the mean model errors and 95% confidence interval of model errors, respectively. The 95% confidence interval obtained after adding and subtracting two standard deviations from mean.

3.4 Results and Discussion

For all algorithms, an ensemble of 100 members is used and measurements errors are assumed to be 2% of the observed reference data. For all runs, we set the number of ES-MDA iterations to 8. For algorithms 3, 4, 5 and 6, we noted that the number

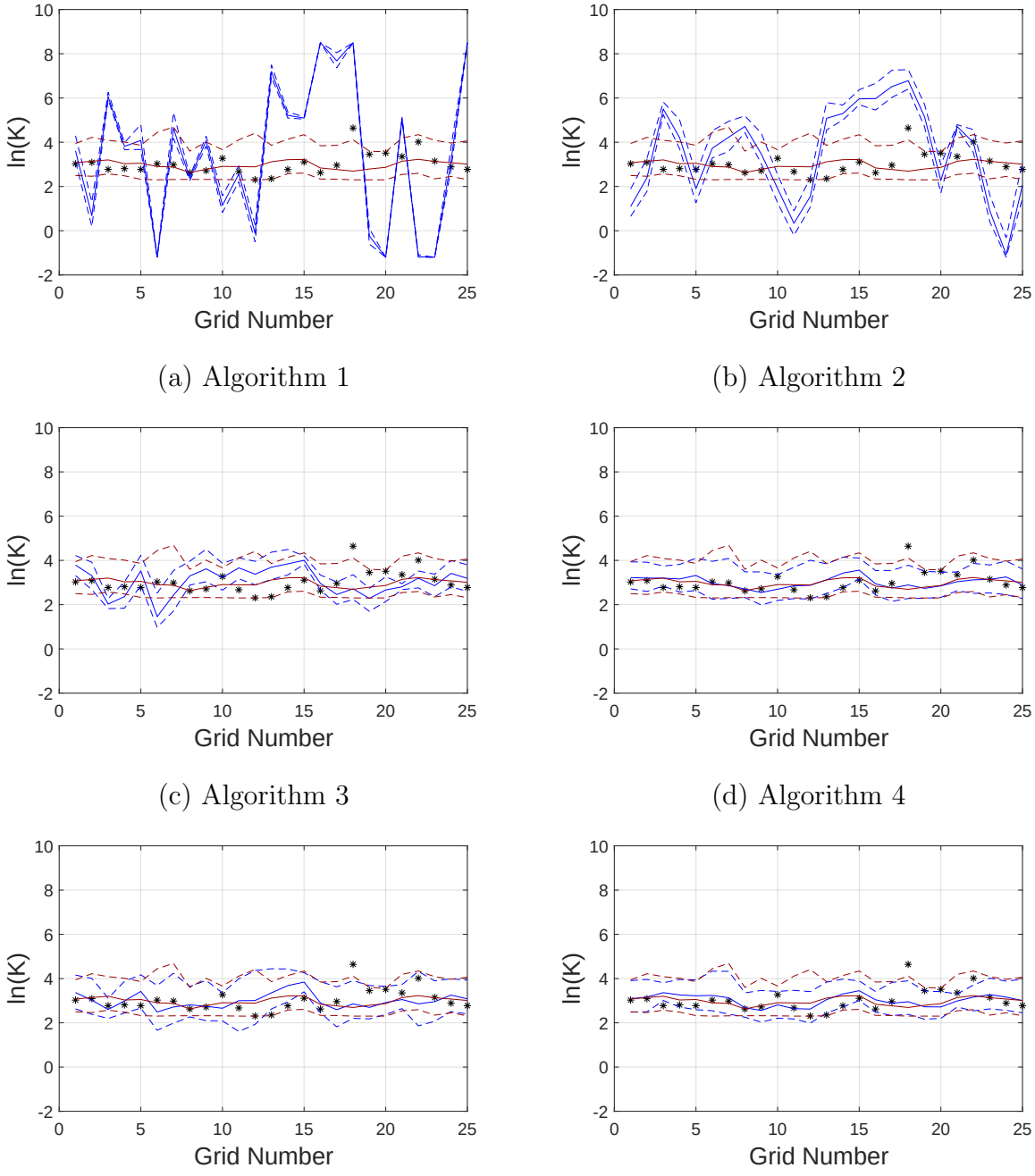
of coefficients of the PCA-based error-model should be limited to a small number in order to avoid over-fitting the error-model. We observed good fit of the error-model to the prior model-error realizations by retaining two PCA components for each time series. The remaining misfit of the PCA based error-model and prior error realizations is treated as second-order errors in algorithms 4, 5 and 6. Since the model outputs correspond to seven time series at the different wells (Fig. 3.1), 14 error-model parameters are jointly estimated along with the log-permeability $\ln(\mathbf{K})$ at every coarse grid cell in algorithms 3, 4, 5 and 6.

Figure 3.4 shows the prior (brown) and posterior (blue) distributions of $\ln(\mathbf{K})$ obtained after history matching using Algorithms 1 to 6. The coarsened version of the true permeability field (Fig. 3.1(b)) is shown in black for comparison (referred to as reference solution). We observe biased posterior distribution of $\ln(\mathbf{K})$ when the model errors are neglected (Algorithm 1), where the estimated log-permeabilities show extreme values (e.g. overshooting) and the posterior distribution failed to cover the reference coarse log-permeabilities. The parameter estimation results of the Algorithm 1 are slightly different from the case 2 of the Chapter 2, because multiple ensembles of the model parameters are used for history matching in Chapter 2.

The overshooting of the physical model parameters is reduced to a certain degree by using Algorithm 2 for calibration where total errors are estimated at the end of each ES-MDA run. However, we note that the total error is estimated from the data residual after one round of history matching, and if the physical model parameters are flexible enough, some of the model-error effect will be compensated for by adjusting the physical model parameters further from the true values (as is the case in Fig. 3.4(b)). Algorithm 3 results in additional improvements in the estimated physical model parameters due to the joint inversion of parameters of the physical model and parameters of the error model. However, we observe that estimated distribution of parameters of the physical model does not cover the reference solution very well due to the limited capacity of the error model as we only retained two PCA basis for each time series. For Algorithms 4, 5 and 6, a residual term representing second-order errors is included in the formulation in addition to joint updating of parameters of the physical model

and parameters of the error model. This results in good coverage of reference solution except for one permeability value at grid block.

Figure 3.5 shows the mean of the posterior ensemble of $\ln(\mathbf{K})$ as a map. Similar to observations from Fig.3.4, relatively large overshooting in the estimated model parameters using Algorithms 1 and 2, and this overshooting is reduced in Algorithm 3, and further reduced when using Algorithms 4, 5 and 6 for calibration.



(a) Algorithm 1

(b) Algorithm 2

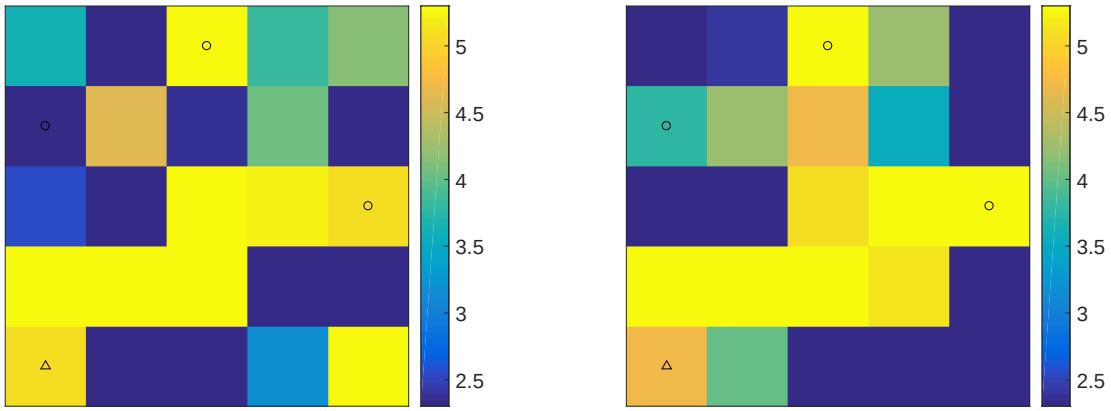
(c) Algorithm 3

(d) Algorithm 4

(e) Algorithm 5

(f) Algorithm 6

Figure 3.4: Prior and posterior distribution of ln($\mathbf{K}$) along the grid block number as one dimensional plot for results obtained using Algorithms 1 to 6. Prior distribution is shown by brown lines and posterior distribution is shown by blue lines. Solid line is the $p50$ (50th percentile) and dashed lines show the 95% confidence interval. Black asterisks show the coarsened version of the true permeability field (as in Fig. 3.1(b)) as a reference solution.



(a) Algorithm 1

(b) Algorithm 2

(c) Algorithm 3

(d) Algorithm 4

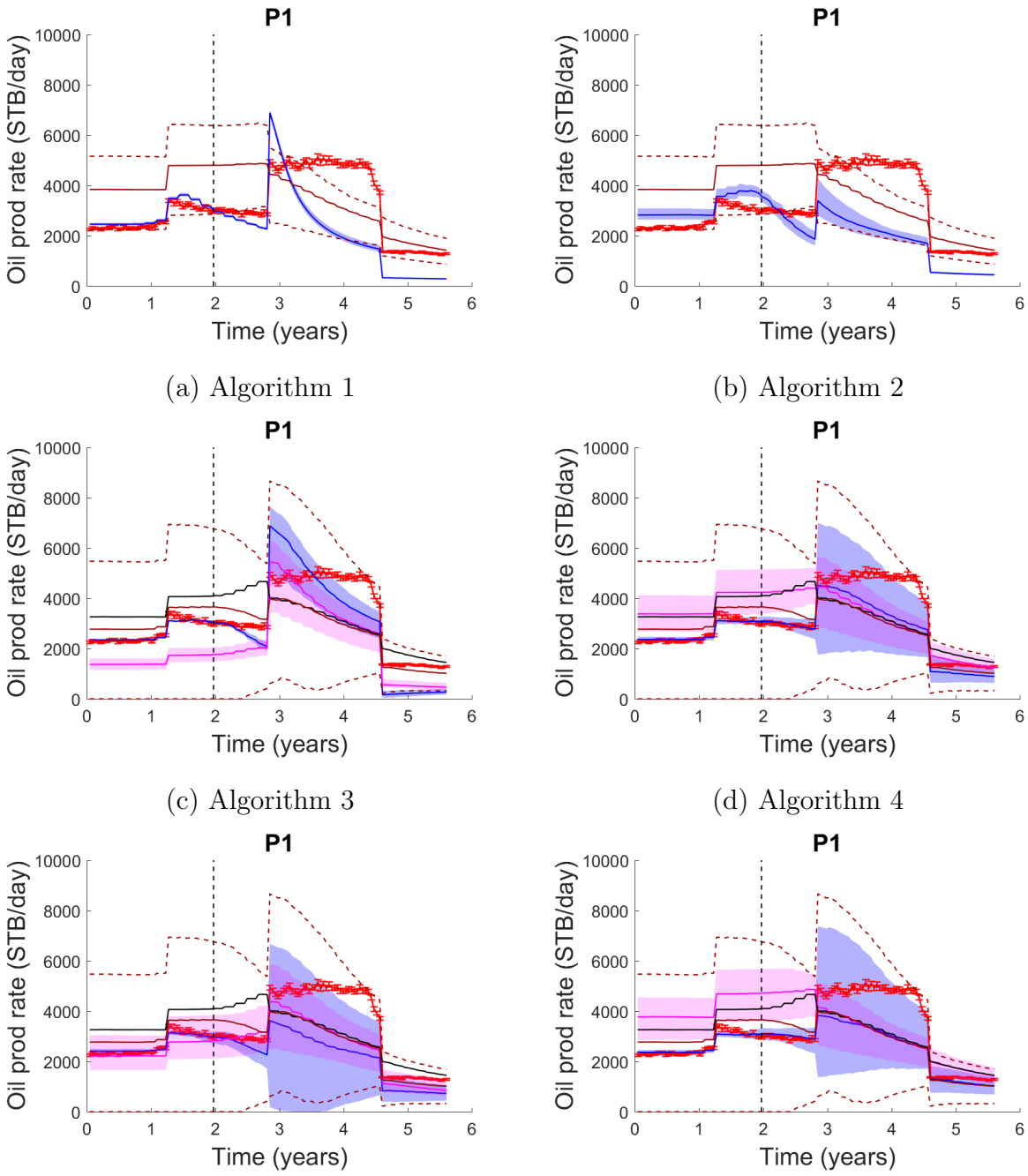
(e) Algorithm 5

(f) Algorithm 6

Figure 3.5: Mean of the posterior ensemble of $\ln(\mathbf{K})$ after history matching using Algorithms 1 to 6.

Figure 3.6 shows the prior and posterior oil production rates at well P1. For Algorithm 1, We observe that the quality of data match (left to the vertical dashed line) is relatively good. However, the future predictions (right to the vertical dashed line) significantly deviates from the reference data due to neglecting the model errors during the model calibration process. For Algorithm 2, the data match for historical data is not as tight as those obtained by Algorithm 1 due to the total error covariance effect which reduces the weight of observed data. The quality of future prediction from Algorithm 2 is also quite biased with slightly boarder uncertainty interval compared to Algorithm 1. We note that for the predictions phase, perturbed prediction should be used by adding realizations of the total error, sampled from the estimated total error covariance, to the simulation outputs from the physical model parameters. This extrapolation of the model error into the prediction period is needed as explained in Lu and Chen (2019). The error model in Lu and Chen (2019), however, has zero mean, so that effect of the error model only increases the uncertainty interval beyond the prediction from the updated physical models (the extrapolation of the error model is not implemented for results shown in this chapter).

For Algorithm 3, we obtained good data match and an improved prediction due to the inclusion of the error model term in the calibration process. The quality of the predictions, however, deteriorates with time. This is likely a consequence of a poor split of the total residual between the error model and the physical model. In order to investigate the relative contribution to the prediction from the physical model and from the error model, prediction using physical model alone with the posterior permeability ensemble (magenta curves) and predictions from the coarse reference permeability (solid black curves) are shown for Algorithms 3 to 6. Once a correct split (relative contribution) is obtained, the magenta curves should closely follow the black curve, both representing the behavior of the permeability (physical model parameter). Figure 3.6 shows that a better split is obtained by Algorithms 4, 5 and 6 with an improvement in the predictions quality.



(a) Algorithm 1

(b) Algorithm 2

(c) Algorithm 3

(d) Algorithm 4

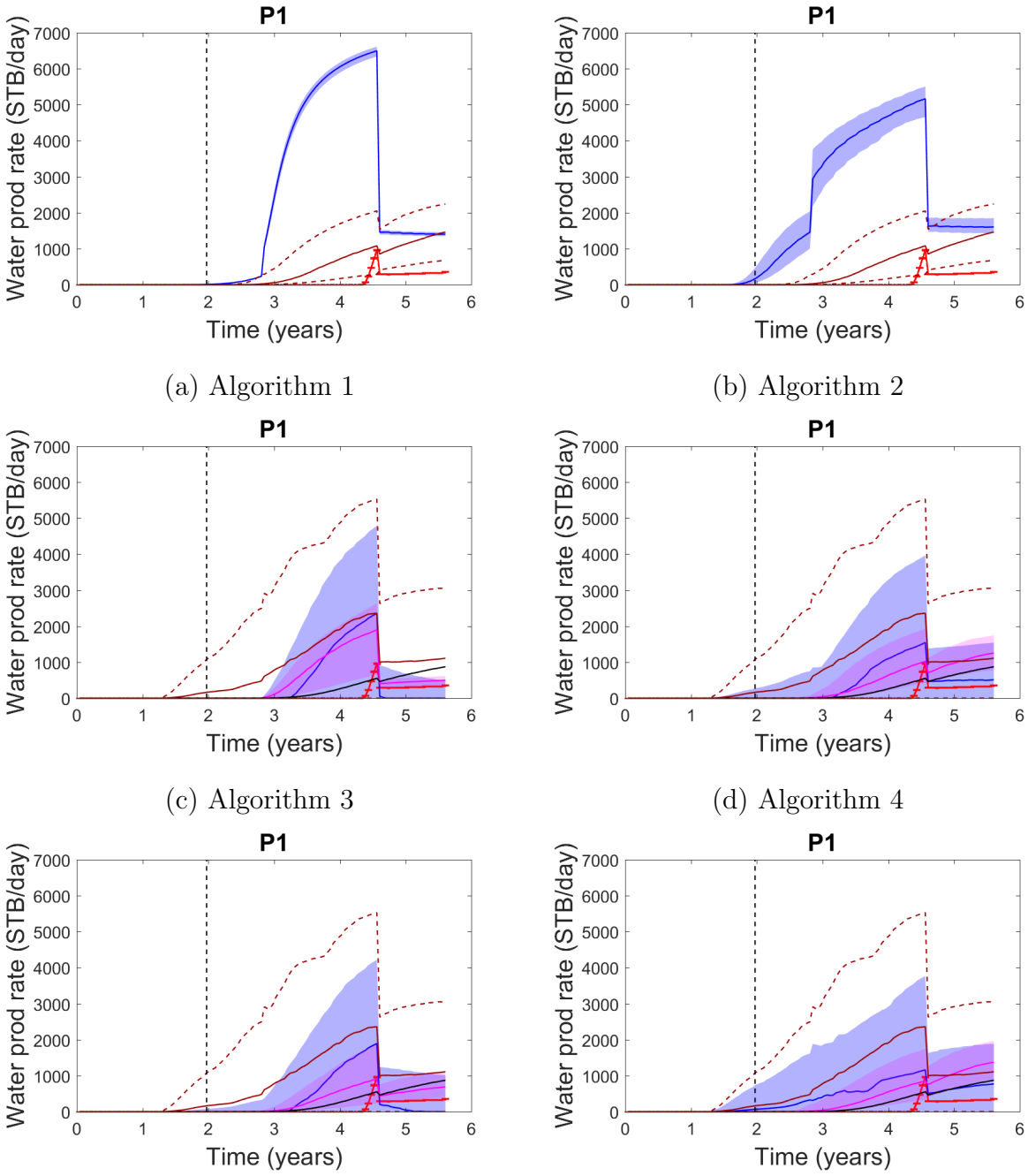
(e) Algorithm 5

(f) Algorithm 6

Figure 3.6: Prior and posterior prediction of oil production rate for well P1. Red lines show observed data and bars on red lines show measurement error. Vertical dashed black lines show the end of the historical period. Solid and dashed brown lines show 50th percentile $p50$ and 95% confidence interval of prior distribution respectively. Solid blue lines show $p50$ and shaded blue areas show 95% confidence interval of posterior distribution, obtained from Algorithms 1 to 6. In parts (c), (d), (e) and (f), solid black lines show reference simulator output (simulation output from the coarsened true permeability, as in Fig. 3.1(b)), solid magenta lines show $p50$ and shaded magenta areas show 95% confidence interval of posterior distribution of simulator output only.

Very similar observations can be made from Figs. 3.7 to 3.10 showing the prior and posterior predictions of the water production rates at well P1, oil and water production rates of P2, and injection pressure for injection well I1. We note that the prediction quality for rates at well P2 is generally better compared to rates at well P1. This is likely due to the small influence from the new well P3 that is introduced in the prediction period (see well location in Fig. 3.1 and operational well schedules shown in Fig. 3.2(a)). This implies that if the reservoir status does not change much from the history matching period to the prediction period, the predictions from the combination of the imperfect physical model and error model is more robust (Algorithms 3 to 6). Figure 3.10(c) shows that predictability between 3 and 4.5 years using Algorithm 3 is much poorer than that from Algorithms 4 to 6. This effect is attributed to the inappropriate split between the physical model and error model contributions to the total output (the distance between solid black curve and the solid magenta curve).

Figures 3.11 and 3.12 show the prior and posterior prediction of the oil and water rates at well P3, which is introduced after the history matching period. When model error is completely neglected as in Algorithm 1, the prediction quality is very poor. Algorithms 2 and 3 do not capture the true model response (red) in the prediction envelop (blue curves). The prediction envelopes from Algorithms 4 to 6 mostly enclose the observed data, but the uncertainty in the prediction is very high, only slightly reduced from the uncertainty of the initial prediction. Algorithms 5 and 6 perform relatively better to capture the observed data (Fig. 3.11) as compared to Algorithm 4. Although the predictions from Algorithms 4 to 6 are not precise, they might lack resolution for useful decision making. In this test case, the model-error is quite significant so it is not straightforward to assess the correct level of resolution that can be achieved.



(a) Algorithm 1

(b) Algorithm 2

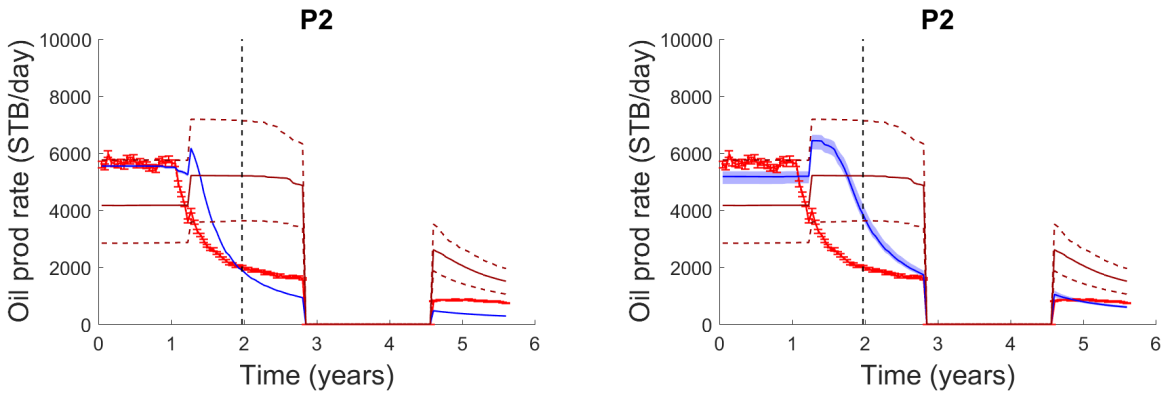
(c) Algorithm 3

(d) Algorithm 4

(e) Algorithm 5

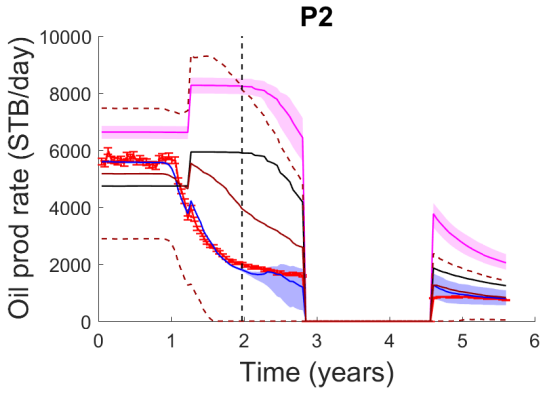
(f) Algorithm 6

Figure 3.7: Prior and posterior prediction of water production rate for well P1. The explanation of lines and colors are the same as in Fig. 3.6.

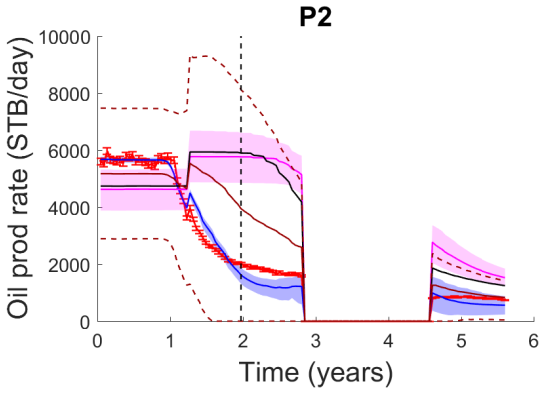


(a) Algorithm 1

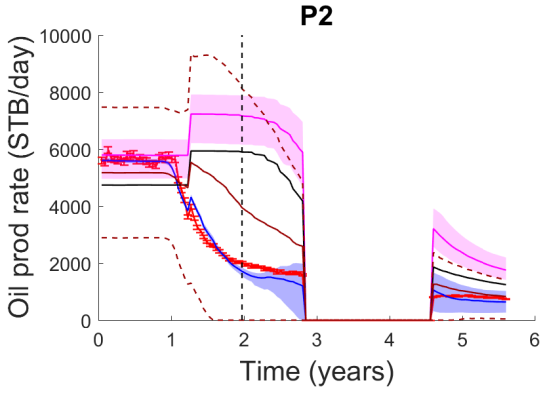
(b) Algorithm 2



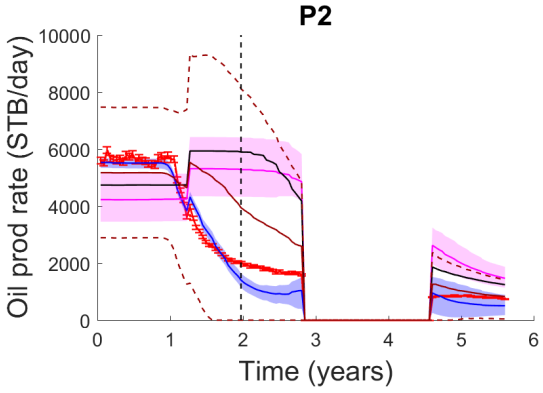
(c) Algorithm 3



(d) Algorithm 4

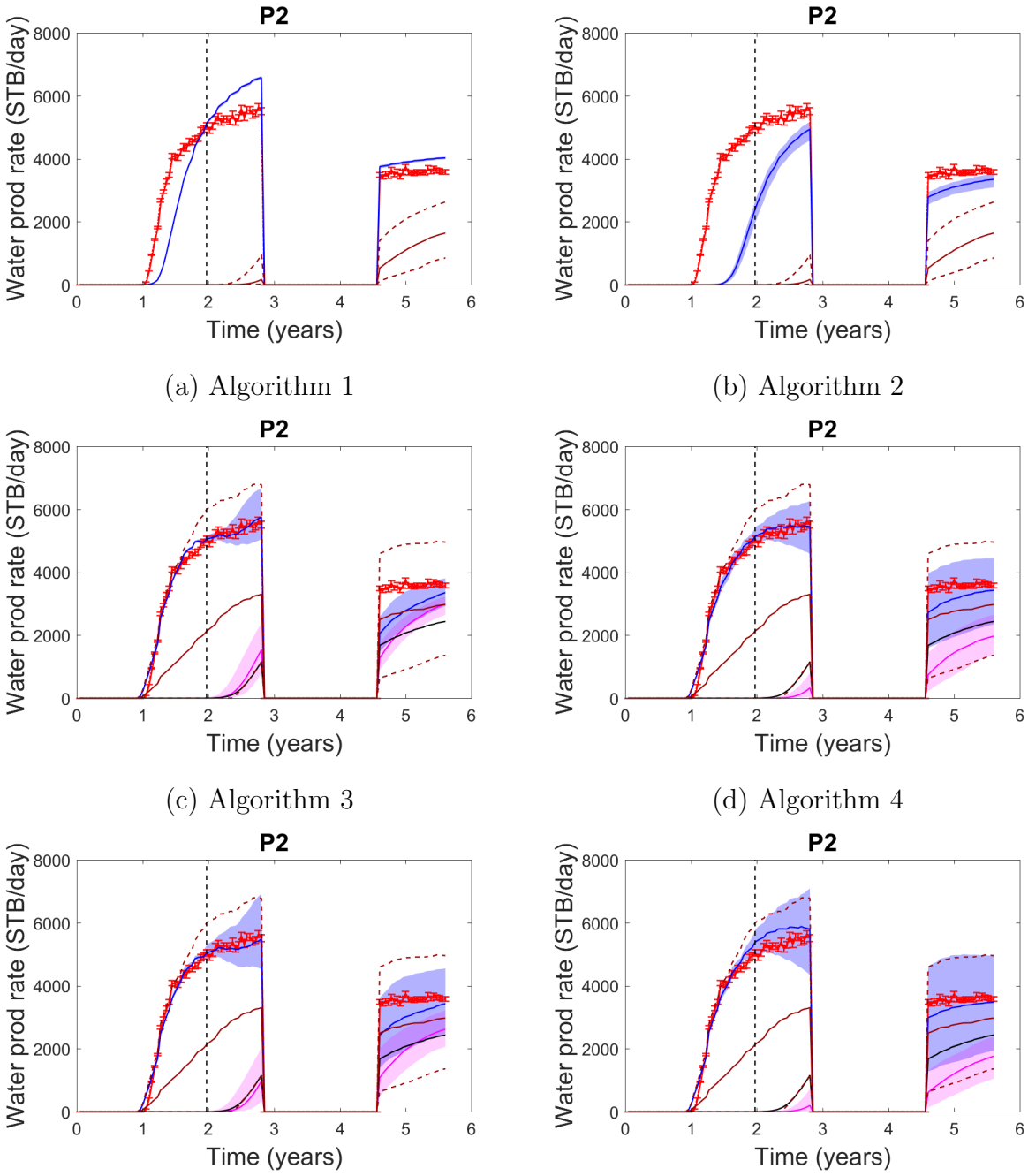


(e) Algorithm 5



(f) Algorithm 6

Figure 3.8: Prior and posterior prediction of oil production rate for well P2. The explanation of lines and colors are the same as in Fig. 3.6.



(a) Algorithm 1

(b) Algorithm 2

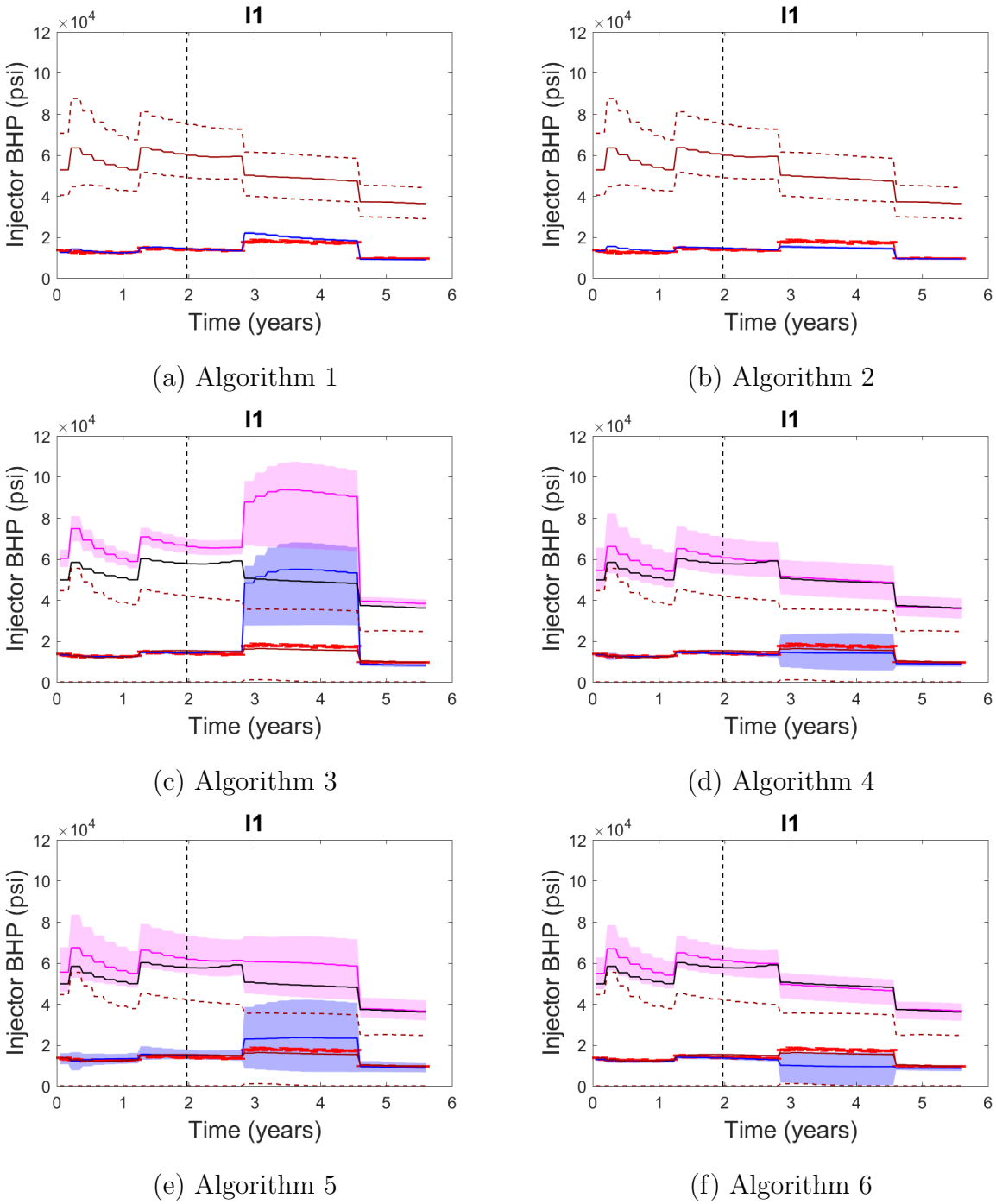
(c) Algorithm 3

(d) Algorithm 4

(e) Algorithm 5

(f) Algorithm 6

Figure 3.9: Prior and posterior prediction of water production rate for well P2. The explanation of lines and colors are the same as in Fig. 3.6.



(a) Algorithm 1

(b) Algorithm 2

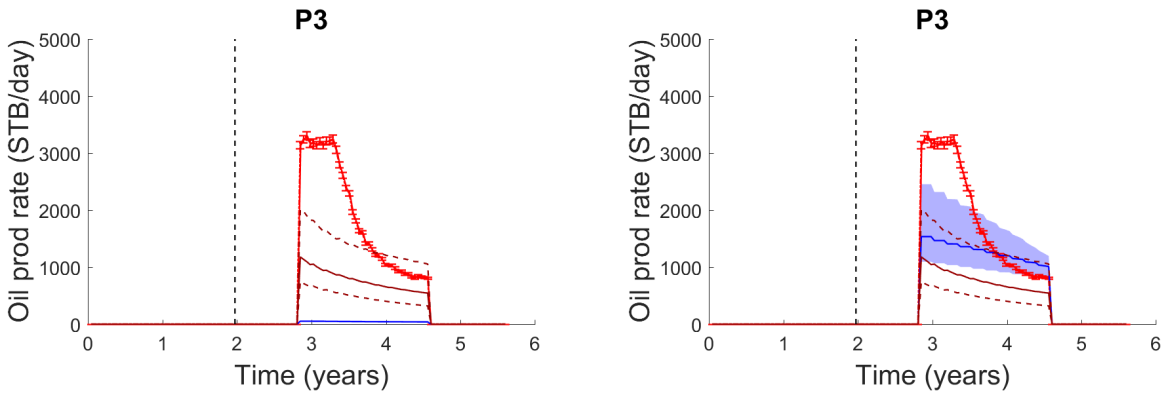
(c) Algorithm 3

(d) Algorithm 4

(e) Algorithm 5

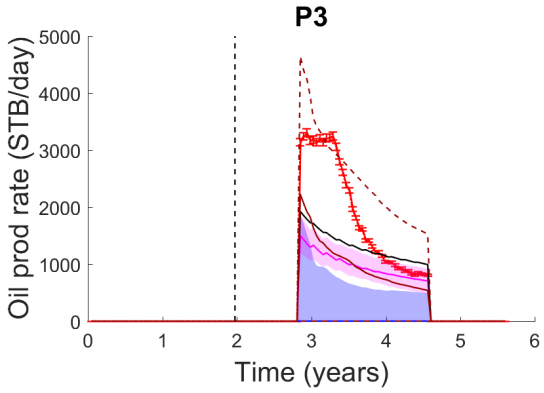
(f) Algorithm 6

Figure 3.10: Prior and posterior prediction of injection pressure for well I1. The explanation of lines and colors are the same as in Fig. 3.6.

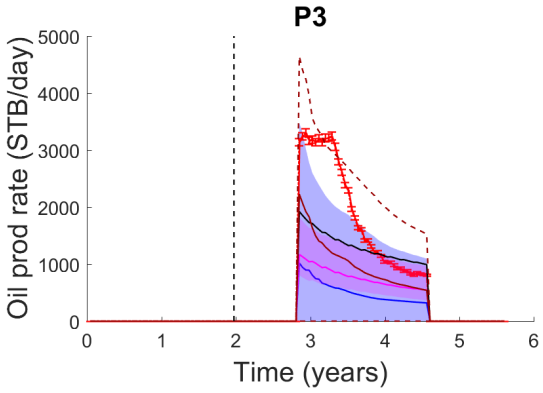


(a) Algorithm 1

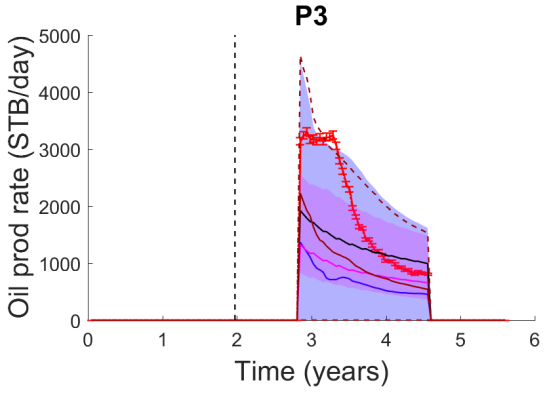
(b) Algorithm 2



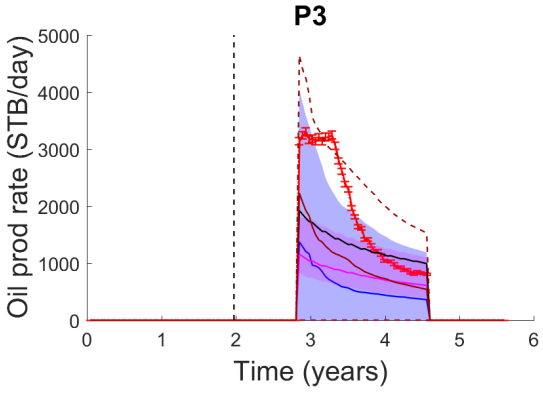
(c) Algorithm 3



(d) Algorithm 4

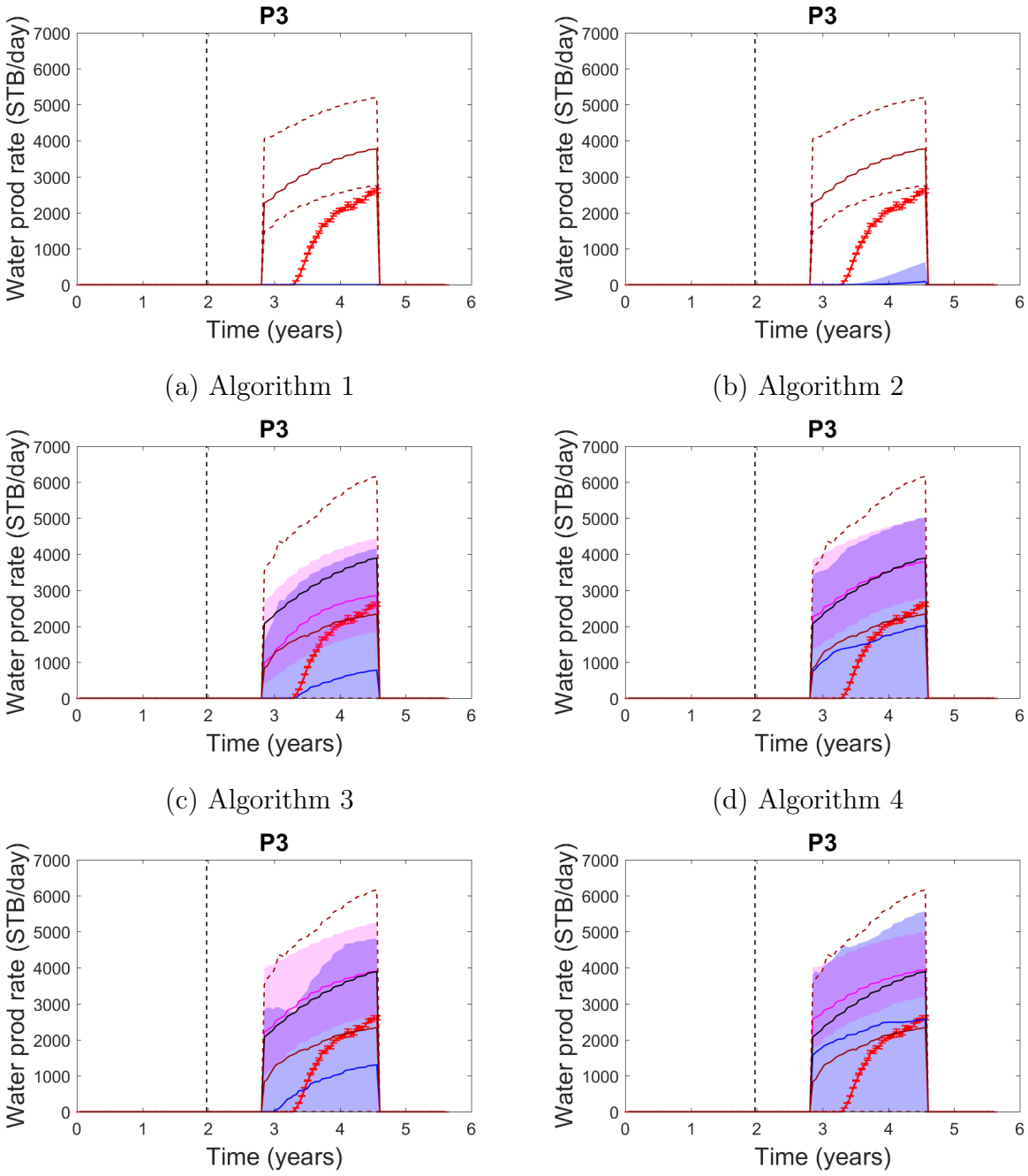


(e) Algorithm 5



(f) Algorithm 6

Figure 3.11: Prior and posterior prediction of oil production rate for well P3. The explanation of lines and colors are the same as in Fig. 3.6.



(a) Algorithm 1

(b) Algorithm 2

(c) Algorithm 3

(d) Algorithm 4

(e) Algorithm 5

(f) Algorithm 6

Figure 3.12: Prior and posterior prediction of water production rate for well P3. The explanation of lines and colors are the same as in Fig. 3.6.

We further evaluate the calibrated models using forecasting measures. We utilize three measures: coverage probability (CP), mean square error (MSE) and mean contin-

uous ranked probability score (CRPS). CP indicates the fraction of the actual data that lie within the confidence interval of the estimation. In this chapter 95% confidence was used to estimate the value of CP, therefore a value of 0.95 for CP indicates a consistent estimation of uncertainty and values below 0.95 indicate underestimation of uncertainty. Mean CRPS quantifies both accuracy and precision (Hersbach, 2000) and higher values of CRPS indicate less accurate results. MSE is widely used as a metric for parameter estimation problems. However, MSE measures the quality of data-fitting and is not enough to provide a probabilistic assessment of the estimation and prediction from an ensemble of models. Similar to Skauvold and Eidsvik (2018), we use a combination of these three forecasting measures to evaluate the quality of parameter estimation and the quality of probabilistic forecast of the calibrated model. These forecasting measures are described mathematically in the Appendix B.

Figure 3.13 shows the coverage probability of the estimated log-permeability realizations and of the simulated well data over the calibration period and over the future prediction period for all algorithms. The results obtained by using Algorithm 1 show a coverage probability of $\ln(\mathbf{K})$ less than 0.1, due to the bias (over-shooting) in the estimated physical parameters as shown in Fig. 3.4(a). Algorithm 1 also provides a low coverage probability for the historical data and the prediction data. Algorithm 2 results in a slightly better coverage probability as compared to algorithm 1, except for CP of the historical data. Accounting for modeling errors during history matching as done in Algorithms 3 to 6 results in much better coverage probabilities for all three quantities ($\ln(\mathbf{K})$, historical data and future predictions), with Algorithms 4, 5 and 6 producing the best overall results. Figures 3.14 and 3.15 show the mean CRPS and MSE measures of the simulated well data from the history matched ensembles over the historical and prediction periods. MSE and Mean CRPS measures show improvements when accounting for model error using Algorithms 2 to 6, similar to the previous results the best results are obtained by Algorithms 4, 5 and 6.

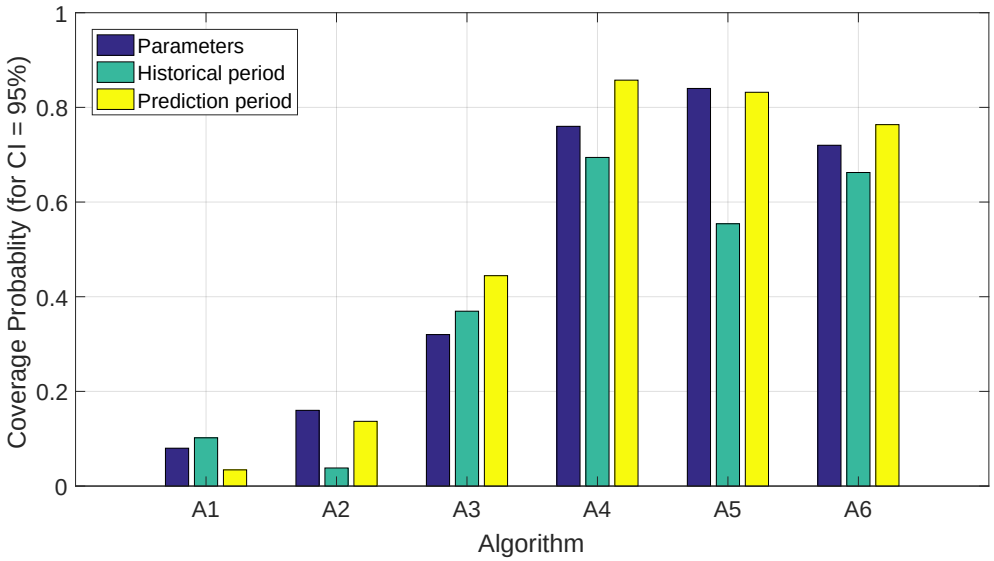


Figure 3.13: Coverage probabilities (CP) obtained after history matching using Algorithms 1 to 6. Blue bars show coverage probability of reference coarse $\ln(\mathbf{K})$; green and yellow bars show coverage probability of the historical data and prediction respectively.

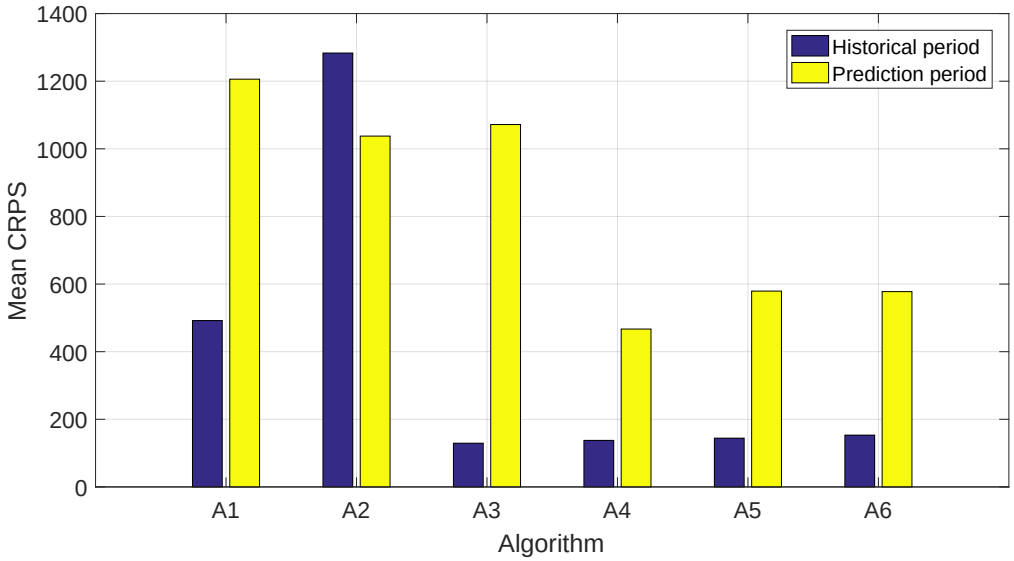


Figure 3.14: Mean continuous ranked probability score (CRPS) obtained after history matching using Algorithms 1 to 6. Blue and yellow bars show the mean CRPS of the historical and prediction data respectively.

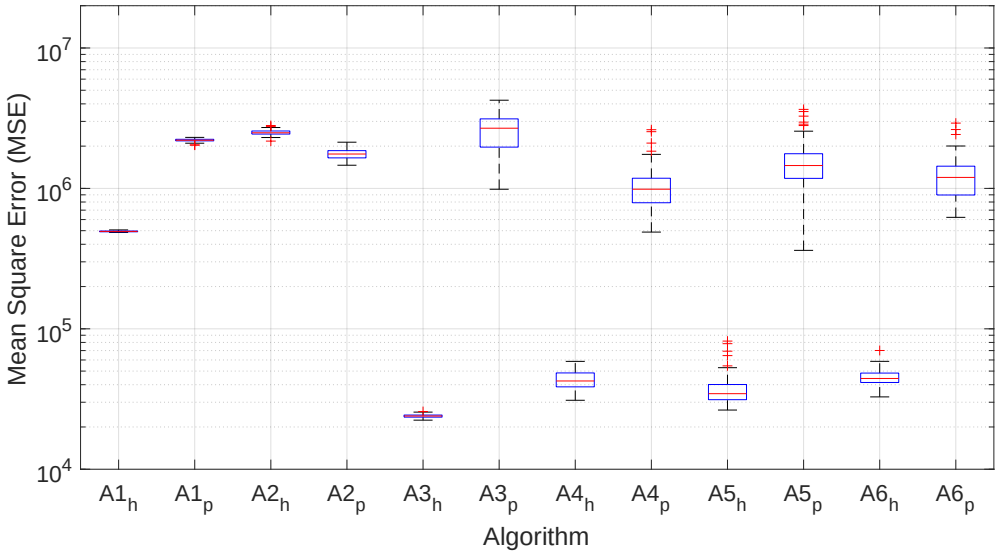


Figure 3.15: The mean square error (MSE) of the simulated model output obtained after history matching using Algorithms 1 to 6 (A1 to A6). The subscript h and p are used for indicating history matching and prediction period, respectively. In the box plot, median and 25th-75th percentiles are indicated by central red line and blue edges respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol.

3.5 Conclusions

In this chapter, we evaluated different algorithms for the calibration of imperfect models. In addition, we developed two new algorithms (5 and 6) by introducing key modifications to the published algorithms under consideration. The novelty in algorithms 5 and 6 is introduced in terms of the joint calibration of both the pre-determined error-model and the physical model. Algorithm 1 neglects the presence of model errors and is used as a base case. Algorithm 2 estimates a covariance for the total error from the data residual after history matching, and this estimated total error term is used in subsequent history matching runs in order to compensate for the model error. In Algorithms 3 and 4, the history matching problem is formulated as a joint parameter estimation of the physical model parameters (reservoir model parameters) and the parameters of an error model. Principal components (PCA) is used to build a functional

form for the error model, where the PCA basis functions and basis weights were obtained using pair of high-fidelity and low-fidelity models response/output. We note that this setting (i.e. the use of high-fidelity model) has limited applicability for practical cases when we only have access to the low-fidelity/imperfect model. In Algorithms 5 and 6, the imperfect simulation model and PCA-based error-model are used in history matching while accounting for second-order errors by building an orthogonal basis and an iterative update of the covariance of the total error respectively. We note that Algorithm 2 is quite general and does not rely on the availability of the high-fidelity/accurate model. However, Algorithm 2 has limitations for physical systems which contain high magnitudes of model-errors.

We evaluated the performance of all history matching algorithms using a 2D synthetic test case with varying well constraints and introduced a new well in the prediction phase for assessing the forecast quality. The two main sources of modeling errors in this test case are aggressive grid coarsening and a rough representation of the geological features (neglecting MPS features/connectively) compared to the synthetic true model. Although the test case is simple, the model errors involved in the test case are quite representative of a typical history matching study, as reservoir simulation models are always a coarse and incomplete representation of reality. Detailed evaluations were performed using three different forecasting metrics, i.e. coverage probability, mean square error and mean continuous ranked probability score.

The obtained numerical results show that the parameter estimation of the imperfect physical model are generally biased, when modeling errors are neglected during history matching. Consequently unreliable and inaccurate predictions are observed from the calibrated model. Methods that account for model-error during history matching provide less biased estimation and prediction. Among all the algorithms investigated, Algorithms 4, 5 and 6 showed good performance in terms of the quality of the parameter estimation, the quality of history matching and the quality of future predictions after calibration. The common feature among Algorithms 4, 5 and 6 is the joint calibration of PCA-based error-model and reservoir simulation model along with an additional term that captures some remaining errors (a.k.a. second-order errors) that are not cap-

tured by the error-model. The second-order errors are captured through the use of a diagonal error matrix (Algorithm 4), orthogonal components (Algorithm 5), and an iterative update of the covariance of the total error (Algorithm 6). When the calibrated models are used for prediction, accounting for the second-order errors seems to allow for a better split (relative contribution) between the physical model and the error-model contribution.

The relative contribution of the error-model in prediction can also be used as an indicator for the impact of the model-error. A large contribution from the error model suggests the need for model refinement and/or re-evaluation of model assumptions. Since the model-error is parameterized in the output space (production profiles), Algorithms 4, 5 and 6 scale very well with the model dimension and can be used for large scale 3D problems. In real life applications, the perfect model does not exist, but a high-fidelity model can still be constructed to be closer to reality compared to the low-fidelity model. Using algorithms (i.e. 4, 5, 6) to calibrate the low-fidelity models can improve the prediction quality up to the level to be expected from the high-fidelity model, which might not be suitable for history matching due to computational limitation. In our future work, we will test the methodology on larger scale cases where there are many sources of model error and further develop the algorithm to tackle the cases where we only have access to the low-fidelity/imperfect model.

Chapter 4

Flexible iterative ensemble smoother for calibration of perfect and imperfect models

Iterative ensemble smoothers have been widely used for calibrating simulators of various physical systems due to the relatively low computational cost and the parallel nature of the algorithm. However, iterative ensemble smoothers have been designed for perfect models under the main assumption that the specified physical models and subsequent discretized mathematical model have the capability to model the reality accurately. While significant efforts are usually made to ensure the accuracy of the mathematical model, it is widely known that the physical models are only an approximation of reality. These approximations commonly introduce some type of model-error which is generally unknown and when the models are calibrated, the effects of the model errors could be smeared by adjusting the model parameters to match historical observations. This results in a bias estimated parameters and as a consequence might result in predictions with questionable quality.

In this chapter, we formulate a flexible iterative ensemble smoother, which can be used to calibrate imperfect models where model errors cannot be neglected. We base our method on the ensemble smoother with multiple data assimilation (ES-MDA) as it is one of the most widely used iterative ensemble smoothing techniques. In the proposed algorithm, the residual (data mismatch) is split into two parts. One part

The contents of this chapter have been submitted in Journal, but includes modifications following the viva.

is used to derive the parameter update and the second part is used to represent the model-error. The proposed method is quite general and relaxes many of the assumptions commonly introduced in the literature. We observe that the proposed algorithm has the capability to reduce the effect of model-bias by capturing the unknown model-errors, thus improving the quality of the estimated parameters and prediction capacity of imperfect physical models.

4.1 Introduction

Ensemble based methods have been gaining popularity in last two decades for data assimilation and calibration of simulation models of various physical systems (Emerick and Reynolds, 2013). The main advantage of the ensemble based methods is the low-computational cost for higher dimensional data assimilation and inverse problems. A number of iterative ensemble based methods have been proposed for parameter estimation problems. The ensemble smoother with multiple data assimilation (ES-MDA) (Emerick and Reynolds, 2013), Levenberg-Marquardt ensemble randomized maximum likelihood method (LM-EnRML) (Chen and Oliver, 2013) and iterative ensemble smoother (IES) (Luo et al., 2015) are some of these techniques. In ensemble based methods, the prior ensemble members of the model-parameters are computed from the initial statistical distribution of the unknown model parameters and the objective is to find an approximate posterior distribution (i.e. posterior ensemble) of the model parameters conditioned to observation data. After calibration, the posterior ensemble of the model parameters is used for making predictions. Ensemble based methods are designed with the assumption that utilized mathematical model provides a complete representation of real physical systems and that the model errors are small enough that it could be neglected during the calibration process. This assumption might introduce bias in the estimated parameter distribution (Brynjarsdottir and O' Hagan, 2014; Schoups and Vrugt, 2010) and as a consequence results in a bad quality predictions using the calibrated models.

A large number of studies have been conducted to explore the possible ways to

account for the model-error during the calibration process. In a Bayesian inversion context, three broad lines of research have emerged in the published literature. In the first line of research, the prior model-error statistics were computed using pairs of high-fidelity and low-fidelity models. These error statistics were utilized during the calibration by using different types of algorithmic frameworks (O' Sullivan and Christie, 2005; Omre et al., 2004; Lødøen et al., 2005; Lødøen and Tjelmeland, 2010; Hansen et al., 2014). These frameworks vary according to the behavior of different physical systems and could perform poorly in high dimensional problems with variable boundary conditions and when the distributions of the model-error statistics are complex or multi-modal. The problem of complex statistics of model-error was addressed by Köpke et al. (2018), where the authors accounted for the model-error component using orthonormal basis generated from the difference between pairs of high-fidelity and low-fidelity models. These basis were evaluated locally at each update iteration as well as the model pairs (high and low fidelity) were re-run during the calibration process. More recently, an Input/Output independent formulation of model-error was introduced to handle high dimensional parameter estimation problems as well as handling problems with time varying boundary conditions (Rammay et al., 2019). However, this line of research relies on the availability of a high-fidelity models such that we could learn the statistical properties of the model errors by evaluating both the high and low fidelity model at the same set of model parameters. This assumption might not be valid for a wide range of applied problems where we only have access to one model.

The second line of research for addressing model-error/bias during calibration is related to joint calibration of physical models with a second model that is assumed to represent the model error. Different parameterization for the model-bias/error have been proposed, for example Gaussian process regression (Kennedy and O'Hagan, 2001) or autoregressive error models (Giudice et al., 2013, 2015). However, without any prior knowledge of the error-model (unknown model-error), the joint calibration may be prone to break the physical constraints of the systems and as a consequence might fail to improve the predictive capacity of the calibrated physical model (Sargsyan et al., 2015). Evin et al. (2014a) compared the joint calibration approach to a post-processing

approach of accounting for the model-error and concluded that the joint-calibration approach was found to be less robust. This leads us to the third line of research, where model-error is generally estimated from the residual (data-mismatch obtained from difference between reality and simulation models). For example, Evin et al. (2014a) estimated the model-error by using normalized residuals and autoregressive model with linear heteroscedasticity. However, this formulation has limitation in the scenarios where model-error exhibits strong structural features and non-linear heteroscedasticity. Along the same line, Sargsyan et al. (2015) utilized approximate Bayesian computation (ABC) to capture the model-error uncertainty using residuals. However, the use of ABC in ensemble based methods is not clearly understood. Recently, Oliver and Alfonzo (2018) estimated the correlated structure of an approximate model-error by computing a covariance matrix of the total residual obtained after one-round of calibration. This covariance matrix was utilized to estimate the model-error statistics and then a re-calibration step is introduced in order to compensate the model-error effects until a termination criteria is satisfied. This approach while novel requires multiple re-calibration iterations and has the limitation to handle non-Gaussian behavior of residuals. Further more, if the imperfect model is flexible enough, the total residual might vanish after one calibration step and then the model error/bias will be underestimated.

In this chapter, a flexible iterative ensemble smoother is introduced to calibrate both perfect and imperfect models. The algorithm is simple to implement and has negligible computational overhead over the standard ES-MDA algorithm. ES-MDA is reformulated by splitting of data-residual (mismatch between model output and observation) into two parts by estimating a split parameter (scalar value) during the calibration process. The first part of the residual is used to update the model-parameters and the second part is assumed to represent the model-error. The objective of the proposed algorithm is to reduce the model-bias by capturing the unknown model-error uncertainty during the calibration of imperfect models in order to improve the prediction capability of the calibrated physical model. Furthermore the proposed algorithm could be used as a diagnostic tool to check the reliability of the physical models and the need for a

model refinement step. In this work, three test cases have been used to observe the performance of the proposed algorithm. These test cases are related to the calibration of polynomial functions, simple machine and imperfect reservoir model. In the first test case, a cubic function is considered as perfect model and imperfect models are represented by quadratic and linear functions. The second test case is related to estimation of efficiency of the simple machine model which lacks physics in terms of a friction component. The third test case is related to calibration of an imperfect reservoir model, which has blurred channelized geological patterns. The imperfect reservoir model has two sources of modeling errors, i.e. simplified geological representation and up-scaling errors.

The outline of this chapter is as follows. In Section 4.2 the formulation of flexible iterative ensemble smoother is described. Following that, three test cases related to calibration of polynomial functions, simple machine model and imperfect reservoir model are described with results in Section 4.3. Section 4.4 is related to the conclusions of the chapter.

4.2 Formulation of Flexible iterative ensemble smoother

In this section, we first present the standard ES-MDA, followed by the proposed flexible formulation that could handle the cases where the model-errors cannot be neglected. In standard settings, ES-MDA update equation can be written as (Emerick and Reynolds, 2013),

$$\mathbf{\Delta M} = \mathbf{C}_{MD}\ (\mathbf{C}_{DD} + \alpha\ \mathbf{C}_D)^{-1}(\mathbf{D}_{uc} - \mathbf{D}), \tag{4.1}$$

where $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \ldots, \mathbf{m}_{N_e}] \in \mathbb{R}^{N_m \times N_e}$ is an ensemble of model parameters of size N_e, $\mathbf{m}_r$ is a realization r of the model parameters of size N_m, $\mathbf{D} \in \mathbb{R}^{N_d \times N_e}$ is an ensemble of model outputs of size N_d generated by a perfect model denoted by an operator g (i.e. $\mathbf{D} = g(\mathbf{M})$), $\mathbf{D}_{uc} \in \mathbb{R}^{N_d \times N_e}$ is the ensemble of perturbed observations ($\mathbf{d}_{obs} \in \mathbb{R}^{N_d \times 1}$), α is the noise inflation parameter and $\mathbf{C}_D$ is the measurement error/noise covariance

matrix. The covariances $\mathbf{C}_{MD}$ and $\mathbf{C}_{DD}$, representing approximate sensitivity of the model response to changes in the model parameters, are defined using the following equations:

$$\mathbf{C}_{MD} = \frac{1}{N_e - 1}(\mathbf{M} - \mathbf{M}_{mean}\overrightarrow{1_{N_e}})(\mathbf{D} - \mathbf{D}_{mean}\overrightarrow{1_{N_e}})^{\intercal}, \tag{4.2}$$

$$\mathbf{C}_{DD} = \frac{1}{N_e - 1}(\mathbf{D} - \mathbf{D}_{mean}\overrightarrow{1_{N_e}})(\mathbf{D} - \mathbf{D}_{mean}\overrightarrow{1_{N_e}})^{\intercal}. \tag{4.3}$$

where $\mathbf{M}_{mean} \in \mathbb{R}^{N_m \times 1}$ is the ensemble mean of model parameters, $\mathbf{D}_{mean}$ is the ensemble mean of model outputs and $\overrightarrow{1_{N_e}} \in \mathbb{R}^{1 \times N_e}$ is a row vector of ones. In ES-MDA, nonlinear inverse problem is solved iteratively with an inflated noise covariance matrix and the inflation factor α is normally set to the total number of data assimilations/iterations N_a.

For the proposed Flexible ES-MDA, the output of the imperfect model is related to the perfect model output using the following equation:

$$\mathbf{D} = \widetilde{\mathbf{D}} + \mathbf{E}, \tag{4.4}$$

where $\widetilde{\mathbf{D}}$ is an ensemble of model output generated from the imperfect model denoted by the operator $\widetilde{g}$ using $\widetilde{\mathbf{D}} = \widetilde{g}(\mathbf{M})$ and $\mathbf{E}$ is an ensemble of the model-error. Similarly, the mean of the previous equation over the ensemble outputs resulted in the following equation:

$$\mathbf{D}_{mean} = \widetilde{\mathbf{D}}_{mean} + \mathbf{E}_{mean}. \tag{4.5}$$

For the approximate sensitives, substituting Eqs. 4.4 and 4.5 into Eq. 4.2 and Eq. 4.3 resulted in the following modified covariances:

$$\mathbf{C}_{MD} = \mathbf{C}_{M\widetilde{D}} + \mathbf{C}_{ME}, \tag{4.6}$$

$$\mathbf{C}_{DD} = \mathbf{C}_{\widetilde{D}\widetilde{D}} + \mathbf{C}_{\widetilde{D}E} + \mathbf{C}_{E\widetilde{D}} + \mathbf{C}_{EE}. \tag{4.7}$$

where,

$$\mathbf{C}_{M\widetilde{D}} = \frac{1}{N_e-1}(\mathbf{M}-\mathbf{M}_{mean}\overrightarrow{1_{N_e}})(\widetilde{\mathbf{D}}-\widetilde{\mathbf{D}}_{mean}\overrightarrow{1_{N_e}})^{\intercal} \tag{4.8}$$

$$\mathbf{C}_{ME} = \frac{1}{N_e-1}(\mathbf{M}-\mathbf{M}_{mean}\overrightarrow{1_{N_e}})(\mathbf{E}-\mathbf{E}_{mean}\overrightarrow{1_{N_e}})^{\intercal} \tag{4.9}$$

$$\mathbf{C}_{\widetilde{D}\widetilde{D}} = \frac{1}{N_e-1}(\widetilde{\mathbf{D}}-\widetilde{\mathbf{D}}_{mean}\overrightarrow{1_{N_e}})(\widetilde{\mathbf{D}}-\widetilde{\mathbf{D}}_{mean}\overrightarrow{1_{N_e}})^{\intercal} \tag{4.10}$$

$$\mathbf{C}_{\widetilde{D}E} = \frac{1}{N_e-1}(\widetilde{\mathbf{D}}-\widetilde{\mathbf{D}}_{mean}\overrightarrow{1_{N_e}})(\mathbf{E}-\mathbf{E}_{mean}\overrightarrow{1_{N_e}})^{\intercal} \tag{4.11}$$

$$\mathbf{C}_{E\widetilde{D}} = \frac{1}{N_e-1}(\mathbf{E}-\mathbf{E}_{mean}\overrightarrow{1_{N_e}})(\widetilde{\mathbf{D}}-\widetilde{\mathbf{D}}_{mean}\overrightarrow{1_{N_e}})^{\intercal} \tag{4.12}$$

$$\mathbf{C}_{EE} = \frac{1}{N_e-1}(\mathbf{E}-\mathbf{E}_{mean}\overrightarrow{1_{N_e}})(\mathbf{E}-\mathbf{E}_{mean}\overrightarrow{1_{N_e}})^{\intercal} \tag{4.13}$$

Using these sensitivities and substitution of Eqs. 4.4, 4.6 and 4.7 into Eq. 4.1, the Flexible ES-MDA update equation while accounting for the model-error can be written as:

$$\mathbf{\Delta M} = (\mathbf{C}_{M\widetilde{D}} + \mathbf{C}_{ME})\ (\mathbf{C}_{\widetilde{D}\widetilde{D}} + \mathbf{C}_{\widetilde{D}E} + \mathbf{C}_{E\widetilde{D}} + \mathbf{C}_{EE} + \alpha\ \mathbf{C}_D)^{-1}(\mathbf{D}_{uc} - \widetilde{\mathbf{D}} - \mathbf{E}). \tag{4.14}$$

In general, the ensemble of modeling errors $\mathbf{E}$ is unknown. However, the residual of data mismatch includes the model-error effects and we postulate that this data-residual could be split between a parameter update component and a modeling error component. In this study, we approximate the ensemble of model-errors $\widetilde{\mathbf{E}}$ using a fraction of the data-residual ensemble as shown in the following equation:

$$\widetilde{\mathbf{E}} = s_p\mathbf{R}, \tag{4.15}$$

where $\mathbf{R} = \mathbf{d}_{obs}\overrightarrow{1_{N_e}} - \widetilde{\mathbf{D}} \in \mathbb{R}^{N_d\times N_e}$ is the ensemble of residuals obtained from the differences between the observation and the ensemble of imperfect model outputs and s_p is a scaler split parameter. The prior (initial) split parameter is computed based on the ratio of norm of mean residual (mean deviation from observed data) and norm of maximum residual (maximum absolute deviation from observed data).

$$s_p{}^{(1)} = \frac{\|\boldsymbol{\sigma}_m\|}{\|\boldsymbol{\sigma}_{max}\|}, \tag{4.16}$$

where $\boldsymbol{\sigma}_m = \text{mean}(\mathbf{R}) \in \mathbb{R}^{N_d\times 1}$ is the mean residual (mean deviation from observed data) and $\boldsymbol{\sigma}_{max} = \max(\text{abs}(\mathbf{R})) \in \mathbb{R}^{N_d\times 1}$ is the maximum residual (maximum absolute deviation from observed data). During calibration (data assimilation) this split

parameter is updated based on the ratio of norm of mean residuals obtained in current iteration i and previous iteration $i-1$.

$$s_p{}^{(i)} = \frac{\|\boldsymbol{\sigma}_m{}^{(i)}\|}{\|\boldsymbol{\sigma}_m{}^{(i-1)}\|}. \tag{4.17}$$

The statistical interpretation of the split parameter (Eqs. 4.16, 4.17) can be attributed to the correspondence of the mean residual to the model-error/bias. Equation 4.16 is used to normalize the initial mean residual in the range of $[0, 1]$ and Eq. 4.17 is used to represent the change in the mean residual during calibration/data-assimilation process. We note that if model response/output consists of multiple time series, the split parameter should be calculated individually for each time series. We replace the ensemble of true model-error $\mathbf{E}$ with the approximate ensemble of model-error $\widetilde{\mathbf{E}}$ in Eq. 4.14, such that:

$$\boldsymbol{\Delta}\mathbf{M} = (\mathbf{C}_{M\widetilde{D}} + \mathbf{C}_{M\widetilde{E}})\ (\mathbf{C}_{\widetilde{D}\widetilde{D}} + \mathbf{C}_{\widetilde{D}\widetilde{E}} + \mathbf{C}_{\widetilde{E}\widetilde{D}} + \mathbf{C}_{\widetilde{E}\widetilde{E}} + \alpha\ \mathbf{C}_D)^{-1}(\mathbf{D}_{uc} - \widetilde{\mathbf{D}} - \widetilde{\mathbf{E}}). \tag{4.18}$$

We argue that we want to develop a Flexible ES-MDA that could work for both good models with negligible model errors (a.k.a. perfect models) as well as for the cases where model errors cannot be neglected (a.k.a. imperfect model). Following this line of thought, we simplified the update equation (Eq. 4.18) with the objective of avoiding perturbation of the approximate sensitivities in the update equations and while keeping the symmetric regularization-like components of the error covariance.

Therefore, Equation 4.18 can be simplified to the following form:

$$\boldsymbol{\Delta}\mathbf{M} = \mathbf{C}_{M\widetilde{D}}\ (\mathbf{C}_{\widetilde{D}\widetilde{D}} + \mathbf{C}_{\widetilde{E}\widetilde{E}} + \alpha\ \mathbf{C}_D)^{-1}(\mathbf{D}_{uc} - \widetilde{\mathbf{D}} - \widetilde{\mathbf{E}}). \tag{4.19}$$

Equation 4.19 represents the final update form for flexible iterative ensemble smoother. We also note that after data assimilation, the magnitude of $\widetilde{\mathbf{E}}$ and $\mathbf{C}_{\widetilde{E}\widetilde{E}}$ could be used an indicator for the model accuracy. Higher magnitude refers to high model-inadequacy in terms of limited physics, reservoir geology, grid coarsening/upscaling, parameterization or prior realizations descriptions. Ideally, after data assimilation for the perfect model cases, $\widetilde{\mathbf{E}}$ and $\mathbf{C}_{\widetilde{E}\widetilde{E}}$ would approach zero (i.e. $\mathbf{D} = \widetilde{\mathbf{D}}$). The details of the Flexible ES-MDA are shown in Algorithm 7. We note that, if the model response/output

consists of multiple time series with different range of values, the rescaling and SVD thresholding of the inverse matrix $(\mathbf{C}_{\widetilde{D}\widetilde{D}} + \mathbf{C}_{\widetilde{E}\widetilde{E}} + \alpha\ \mathbf{C}_D)^{-1}$ should be done in the same way as shown in the paper (Emerick and Reynolds, 2012).

Algorithm 7 Flexible ensemble based algorithm for perfect and imperfect models

Choose N_a ▷ number of data assimilation/iterations
$i \leftarrow 1$
$\alpha = N_a$
while $i <= N_a$ **do**
    $\mathbf{D}_{uc} = \mathbf{d}_{obs}\overrightarrow{1_{N_e}} + \sqrt{\alpha}\ \mathbf{C}_D^{1/2}\ \mathbf{Z}_d$, ▷ Observation perturbations, $\mathbf{Z}_d = [\mathbf{z}_{d1}\ \mathbf{z}_{d2}\ \mathbf{z}_{d3}\ .....\ \mathbf{z}_{dN_e}] \in \mathbb{R}^{N_d \times N_e}$, $\mathbf{z}_d \sim \mathcal{N}(0, \mathbf{I}_{N_d}) \in \mathbb{R}^{N_d \times 1}$,
    $\widetilde{\mathbf{D}} = \widetilde{g}(\mathbf{M})$ ▷ Generate ensemble of model outputs $\widetilde{\mathbf{D}} \in \mathbb{R}^{N_d \times N_e}$
    $\mathbf{R} = \mathbf{d}_{obs}\overrightarrow{1_{N_e}} - \widetilde{\mathbf{D}}$
    $\widetilde{\mathbf{E}} = {s_p}^{(i)}\mathbf{R}$ ▷ Compute ensemble of approximate model-error, $\widetilde{\mathbf{E}} \in \mathbb{R}^{N_d \times N_e}$
    $\mathbf{M} \leftarrow \mathbf{M} + \mathbf{C}_{M\widetilde{D}}\ (\mathbf{C}_{\widetilde{D}\widetilde{D}} + \mathbf{C}_{\widetilde{E}\widetilde{E}} + \alpha\ \mathbf{C}_D)^{-1}(\mathbf{D}_{uc} - \widetilde{\mathbf{D}} - \widetilde{\mathbf{E}})$ ▷ Update ensemble
    $i \leftarrow i + 1$
end while
$\mathbf{M}_{post} = \mathbf{M}$ ▷ $\mathbf{M}_{post} \in \mathbb{R}^{N_m \times N_e}$ is the posterior ensemble of model parameters
$\widetilde{\mathbf{D}}_{post} = \widetilde{g}(\mathbf{M}_{post})$ ▷ Generate posterior ensemble of model predictions, $\widetilde{\mathbf{D}}_{post}$

4.3 Test cases

In this work, three test cases have been used to observe the performance of the proposed algorithm. These test cases are related to the calibration of polynomial functions, simple machine and imperfect reservoir model. For comparison purpose, calibration is performed using both the standard ES-MDA (Emerick and Reynolds, 2013) and the proposed Flexible ES-MDA algorithm. An ensemble of 100 members is used with 8 iterations for the calibration of all test cases. In the first test case, a cubic polynomial function is considered as a perfect model and imperfect models are represented by quadratic and linear functions. The objective is to test the performance of the proposed

algorithm for calibration of perfect and imperfect models. The second test case is related to the estimation of efficiency of the simple machine model which lacks physics in terms of friction component and used (Brynjarsdottir and O' Hagan, 2014) as a test case for a joint calibration framework with the model error parameterized by Gaussian process regression. In this work, the objective is to test the calibration and prediction improvement of simple machine model using the proposed algorithm without joint calibration. The third test case is related to calibration of imperfect reservoir model, which has blurred channelized geological patterns. The imperfect reservoir model has two sources of modeling errors, simplified geological representation and up-scaling errors.

4.3.1 Test case 1: The polynomial functions

In this test case, the data is generated from a cubic polynomial functions and perturbed with an additive measurement noise $\epsilon_d \sim \mathcal{N}(0, \mathbf{C}_D)$ using Eq. 4.20. The complexity of the calibrated models vary from a first order polynomial to a third order polynomials as shown in Eqs. 4.21, 4.22 and 4.23. The objective is to test the performance of the proposed algorithm on both scenarios where the calibrated model complexity matches the data generating process (i.e. perfect model) and when the calibrated model is less parameterized than the data generating process (i.e. imperfect models). The domain of x lies within the interval $[-1, 1]$. We calibrate three models (cubic, quadratic and linear models) to obtain the posterior distributions of each model parameters vector $\boldsymbol{\lambda}$ and the corresponding models outputs. The objective is to evaluate the flexibility of the proposed algorithm in quantifying the model-error uncertainty for imperfect models as well as the ability to match the data for the perfect model case. Prior parameters are sampled from a standard normal distribution with zero mean ($\mu = 0$) and standard deviation $\sigma = 10$.

$$d_{obs} = 2 + 2x + 3x^2 - 5x^3 + \epsilon_d, \tag{4.20}$$

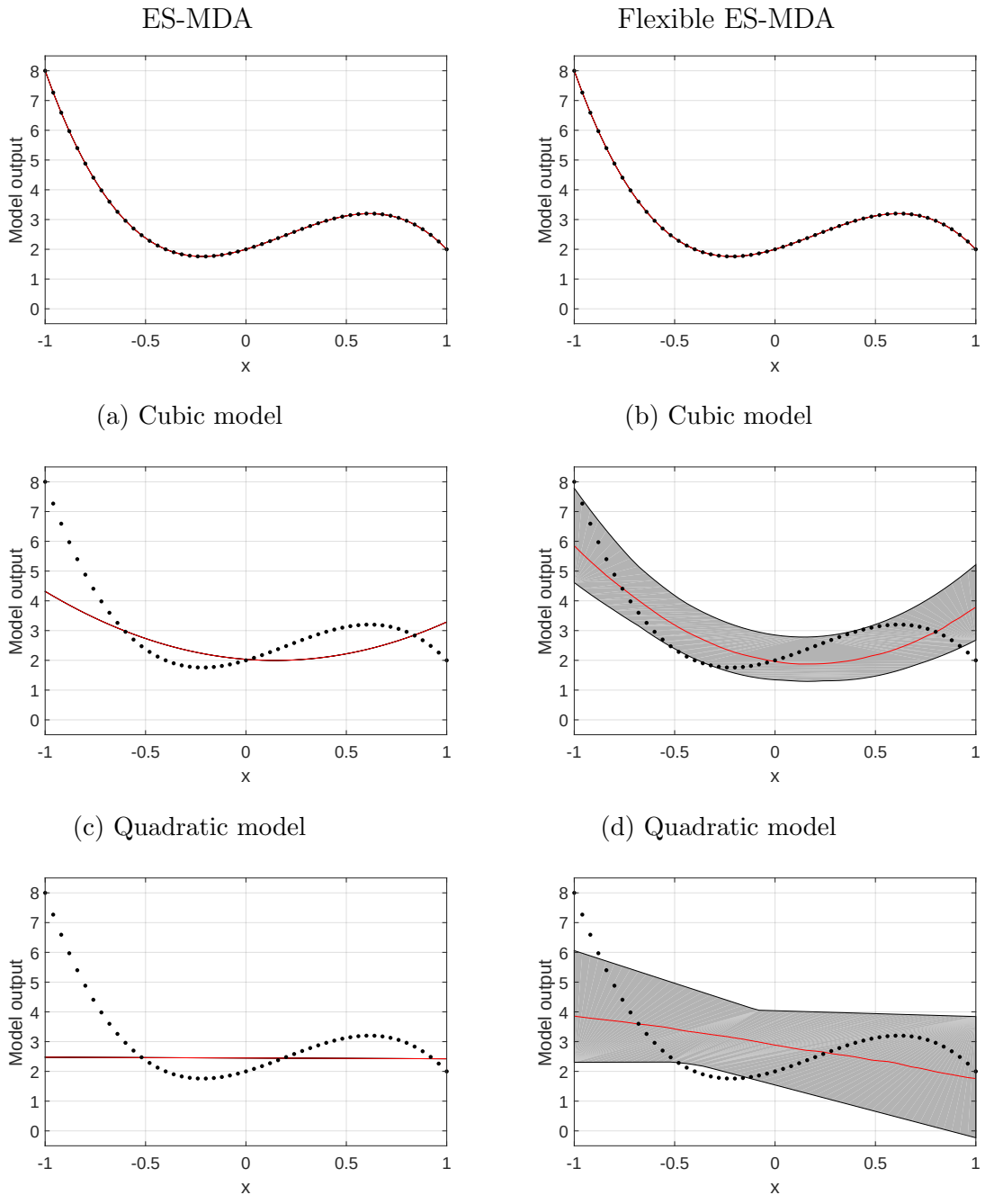
$$f_1(x) = \lambda_0 + \lambda_1 x, \tag{4.21}$$

$$f_2(x) = \lambda_0 + \lambda_1 x + \lambda_2 x^2, \tag{4.22}$$

$$f_3(x) = \lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 x^3. \tag{4.23}$$

The calibration of polynomial functions are performed using two levels of measurement noise. In the first level, negligible measurement error (i.e. order of magnitude 10^{-12}) is considered. Figure 4.1 shows the posterior distribution of cubic, quadratic and linear models outputs obtained by ES-MDA and Flexible ES-MDA algorithms. We observe that both algorithms manage to match the observations for cubic model (perfect model case), without any bias as shown in Figs. 4.1(a) and (b). This shows that the proposed algorithm did not introduce any bias and has the flexibility to match the data for the perfect model case with negligible measurement errors. The posterior distributions of cubic model output (Figs. 4.1(a) and (b)) are appeared as the point estimate (i.e. exact data match) due to the effect of negligible measurement error (i.e. order of magnitude 10^{-12}).

Figures 4.1(c) and (d) show the posterior distribution of the quadratic model output. It is clear that the standard ES-MDA fails to match the data, while the calibrated model using the Flexible ES-MDA results in model outputs that account for model-error uncertainty and provides a good coverage for the observations as shown in Fig. 4.1(d). Similar results are observed for the linear model case in Figs. 4.1(e) and (f), where a better coverage is obtained by the proposed Flexible ES-MDA. However, we note that the coverage is not as good for the quadratic model case. This shows that the capacity as well as the limitation of the proposed algorithm in accounting for the modeling errors during the calibration.



(a) Cubic model

(b) Cubic model

(c) Quadratic model

(d) Quadratic model

(e) Linear model

(f) Linear model

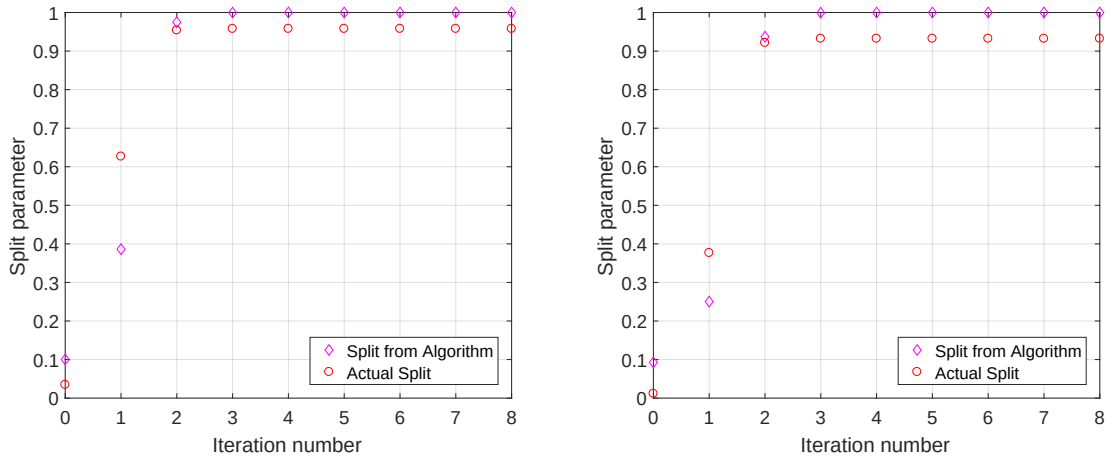
Figure 4.1: Posterior distribution of models (cubic, quadratic and linear) outputs for test case 1 (negligible measurement error). Red lines show $p50$ percentiles, grey shaded areas show 99% confidence intervals and solid black dots show observation data.

For further validation of the proposed algorithm, we compare the best split factor

(computed using the true model-error ensemble $\mathbf{E}_{actual}$) with the approximated split factor s_p estimated using the formulation proposed in this manuscript. The best split is computed by minimizing the Frobenius norm of the difference between true model-error ensemble and the residual ensemble i.e. $\min \|\mathbf{E}_{actual} - s_p\mathbf{R}\|$. This simple minimization problem is solved using differential evolution algorithm (Storn and Price, 1997). The Frobenius norm of $m \times n$ matrix $\mathbf{X}$ is defined by,

$$\|\mathbf{X}\|_F = \sqrt{\text{trace}(\mathbf{X}^\intercal \mathbf{X})}. \tag{4.24}$$

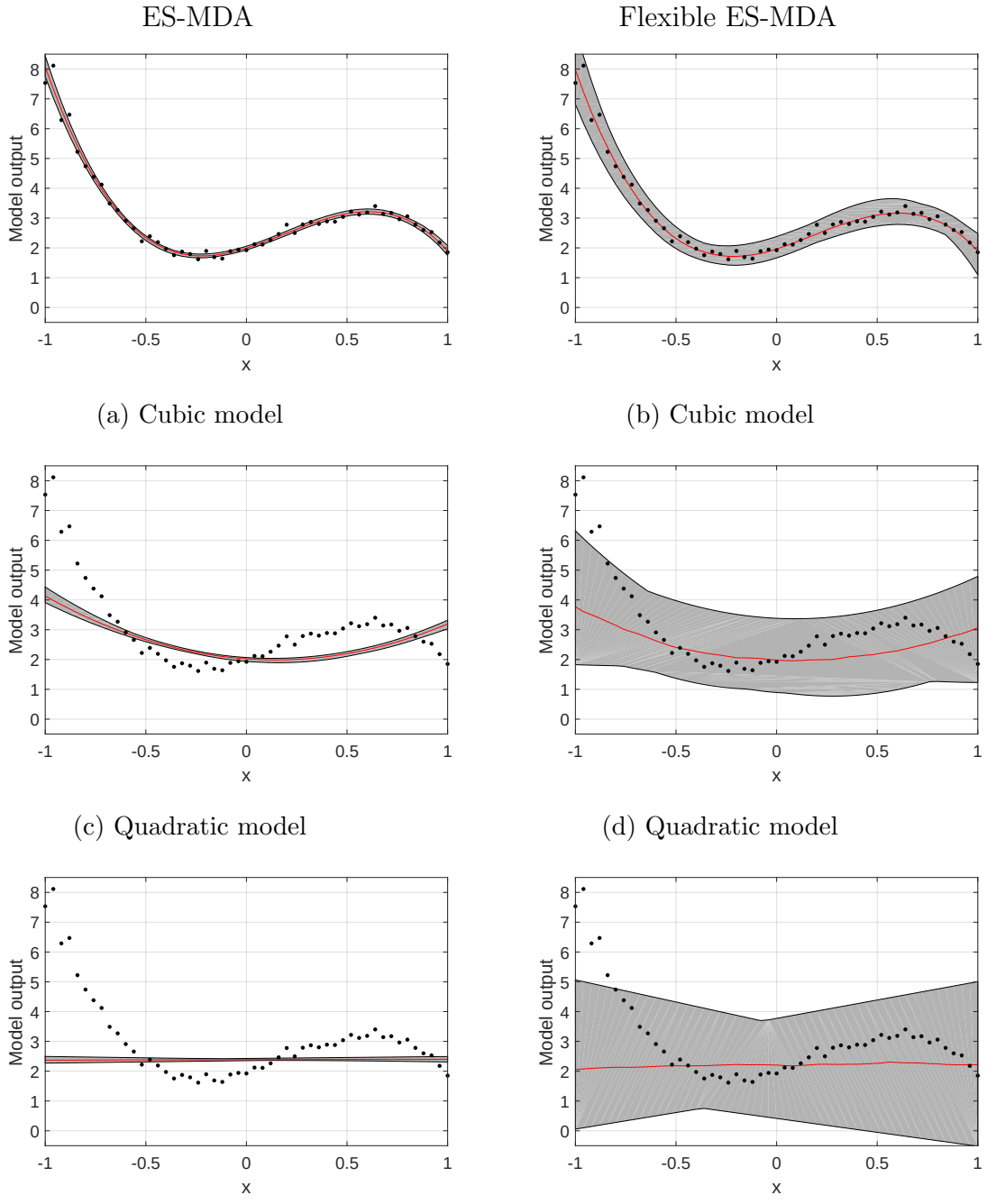
Figure 4.2 shows both the best split factor s_p when the true model error is known and the approximated split versus the number of ES-MDA iterations. At iteration 0 (i.e. prior), both the best and proposed split factors show very low values, because the prior residual is relatively larger than the model-error. For subsequent iterations, the split parameter is increasing with respect to the iteration number, which is attributed to the convergence of the proposed flexible algorithm to the level of model-error uncertainty. At iteration 2, the proposed split is very close to best split value for both linear and quadratic imperfect models, which show the robustness of the proposed equation in approximating the split factor. After iteration 2, the proposed split parameter approached 1, which show that all the data mismatch (residual ensemble) is treated as a model-error ensemble and further reduction of uncertainty is not possible due to the limited capacity of both the linear and quadratic models in matching the data. These results demonstrate the ability of the proposed algorithm in capturing the unknown model-error uncertainties during the calibration of imperfect models.



(a) Calibration of linear model (b) Calibration of quadratic model

Figure 4.2: Comparison between splits obtained from actual model-error and proposed algorithm during calibration.

Using the same problem setup, we add a measurement error of 5% of true function value (Eq. 4.20). The objective of this experiment is to evaluate the proposed algorithm when measurement errors are present while calibrating perfect and imperfect models. Figure 4.3 shows the posterior outputs of the calibrated cubic, quadratic and linear models. For the cubic models case, both the standard ES-MDA and the Flexible ES-MDA algorithms managed to match the structural feature of the data as shown in Figs. 4.3(a) and (b). However, the proposed Flexible ES-MDA algorithm performs significantly better in terms of capturing the noisy features of the data (Fig. 4.3(b)). This effect shows that the split formulation of proposed algorithm act as a adaptive regularizer for the perfect model scenario. Similar results (w.r.t negligible measurement noise case) are observed for both the quadratic and the linear models in the presence of both measurement and model errors (Figs. 4.3(c), (d), (e) and (f)), where the results from the Flexible ES-MDA are clearly better than the standard ES-MDA in quantifying the unknown model-error uncertainty along in the presence of measurement errors.



(a) Cubic model

(b) Cubic model

(c) Quadratic model

(d) Quadratic model

(e) Linear model

(f) Linear model

Figure 4.3: Posterior distribution of models (cubic, quadratic and linear) outputs for test case 1 (with measurement error of 5%). The descriptions of lines and colors are same as in Fig. 4.1.

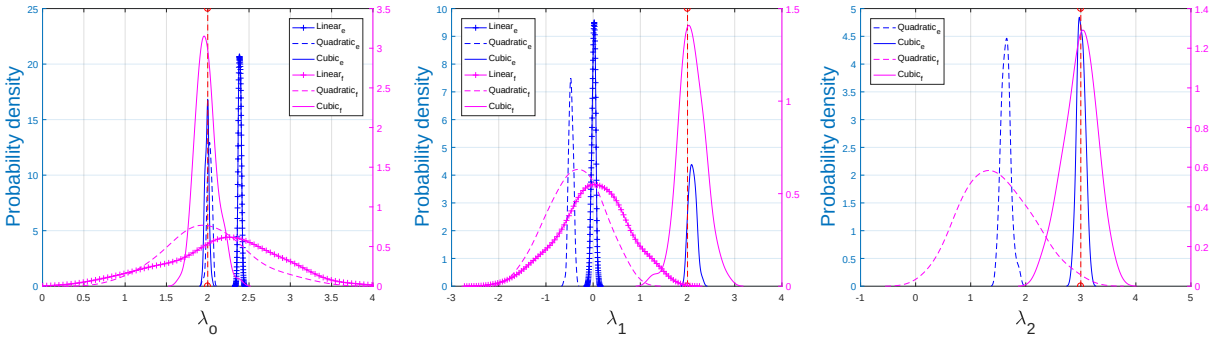
Figure 4.4 shows the posterior distribution of parameters for the case of 5% measurements errors. In this figure, blue and magenta colors show the posterior distribution

obtained from ES-MDA and Flexible ES-MDA, respectively. ES-MDA results in a biased posterior distribution of the model parameters as shown in Figs. 4.4(a), (b) and (c) except for the perfect model case and constant coefficient of the quadratic model. Flexible ES-MDA reduces model-bias in parameter estimation of linear and quadratic models by capturing the model-error effects and covers the reference solution as shown in Figs. 4.4(a), (b) and (c) except for coefficient of x (i.e. λ_1) of the quadratic model.

We use the Prediction Interval Coverage Probability (PICP) of posterior distributions as one of the calibration metric. PICP is estimated by counting the number of observations in the confidence intervals (10% to 99%) of the posterior distribution normalized by the total number of observations (Xu and Valocchi, 2015). PICP is a very useful metric for quantifying both under-estimation and/or over-estimation of uncertainties in the obtained posterior distributions (Xu and Valocchi, 2015). The mathematical description of Coverage Probability (CP) is shown in the Appendix B. PICP values close to 45° line (dashed red line in Figure 4.5) indicates a perfect posterior distribution. Figure 4.5 shows the PICP of the prior and posterior distribution of the linear, quadratic and cubic models outputs. We observe that the posterior distribution of cubic (perfect) model output obtained by the standard ES-MDA underestimates the uncertainty. However, the proposed Flexible ES-MDA shows more robust uncertainty quantification as shown in Fig. 4.5(a). This effect is due to the efficient coverage of the noisy features of the observation data by the proposed algorithm (Fig. 4.3(b)). PICP of posterior distribution obtained from ES-MDA for the imperfect linear and quadratic models show severe underestimation of uncertainties (Figs. 4.5(b) and (c)) as the model-error uncertainties are missing in the formulation. Further, we observe a more robust PICP values obtained from the proposed algorithm for calibrating the imperfect linear and quadratic models as shown in Figs. 4.5(b) and (c).

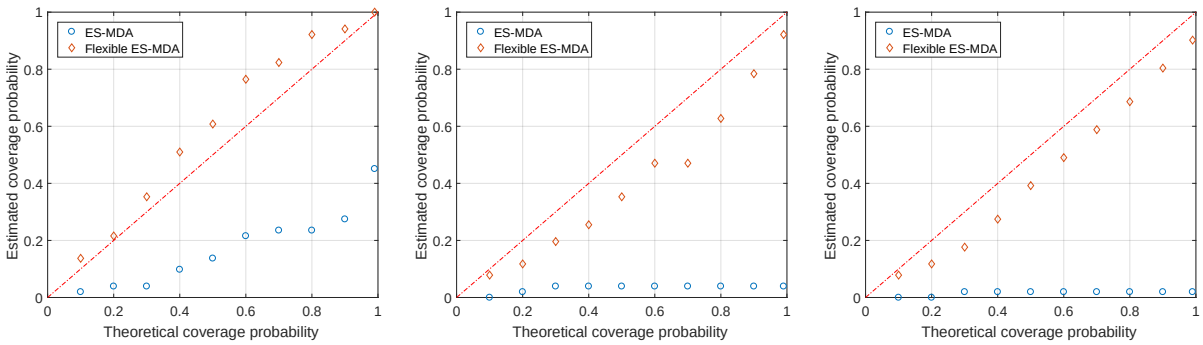
Mean continuous ranked probability score (CRPS) is one of the most useful calibration and prediction metric to evaluate the precision and accuracy of the posterior ensemble. Less accurate or poor quality results are indicated by the higher values of CRPS. The complete detail of CRPS for ensemble prediction system can be found in Hersbach (2000). The mathematical descriptions of mean CRPS along with Mean

square error (MSE) are presented in the Appendix B. Figure 4.6 shows the comparison of mean CRPS and MSE of the posterior ensemble of models outputs obtained by the ES-MDA and the proposed algorithm. Calibration results from the proposed algorithm show lower CRPS values for linear, quadratic and cubic models (Fig. 4.6(a)), which indicates more reliable results in terms of precision and accuracy. In Fig. 4.6(b), MSE shows more wide distribution of posterior ensemble due to the coverage of unknown model-error uncertainty using Flexible ES-MDA, however lower values of MSE is also observed using the proposed algorithm due to significant reduction in model-bias.



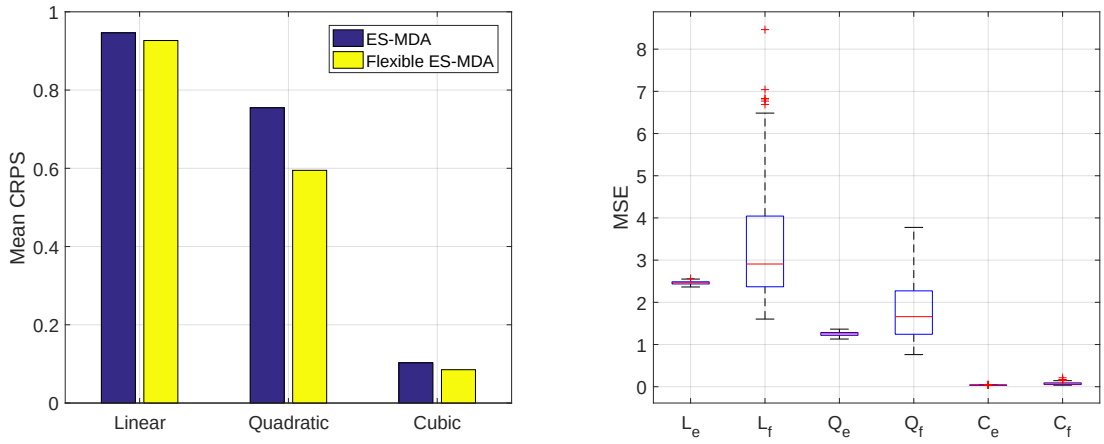
(a) Constant coefficient (b) Coefficient of x (c) Coefficient of x^2

Figure 4.4: Posterior distribution of models (cubic, quadratic and linear) parameters for test case 1 (with measurement error of 5%). Red dashed lines with small circles on ends show the reference model parameters, which are used to generate observations. In figure legends, subscripts 'e' and 'f' show the calibration using ES-MDA and Flexible ES-MDA respectively. Blue and magenta show the y-axes correspond to PDF values of posterior distribution obtained using ES-MDA and Flexible ES-MDA respectively.



(a) PICP of Cubic model (b) PICP of Quadratic model (c) PICP of Linear model

Figure 4.5: Prior and posterior PICP of linear, quadratic and cubic models (with measurement error of 5%).



(a) Mean CRPS Comparison (b) MSE Comparison

Figure 4.6: Mean CRPS and MSE of posterior ensemble for test case 1 (with measurement error of 5%). In part (b), 'L', 'Q' and 'C' show MSE from linear, quadratic and cubic models respectively. Subscripts 'e' and 'f' show the calibration using ES-MDA and Flexible ES-MDA respectively.

Comparison with Algorithms 4, 5, 6 and MCMC reference solution

In this test case, Algorithms 4, 5, 6 are also compared with respect to the MCMC reference solution. This comparison allows us to understand the difference between the

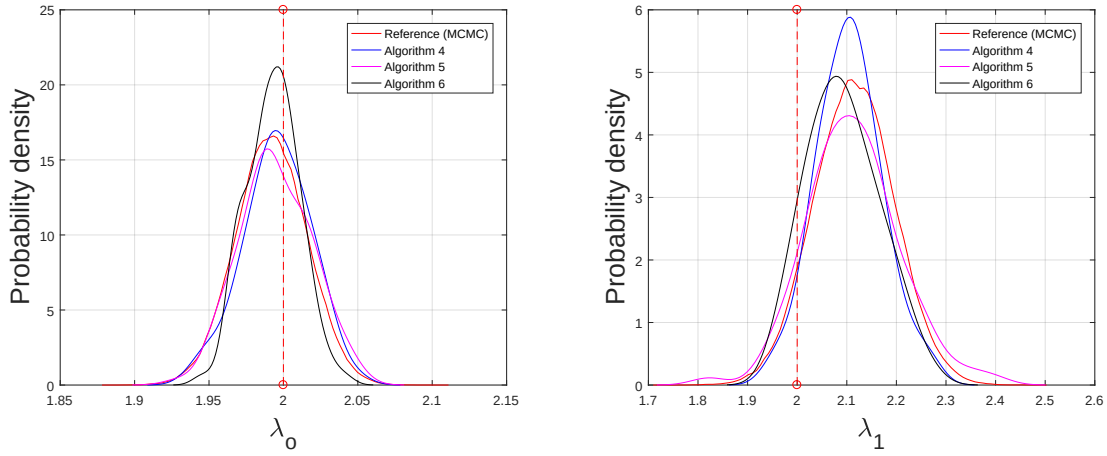
performance and requirements of the proposed algorithms in this thesis. Algorithms 4, 5 and 6 are proposed for the joint calibration of imperfect models and error-model, therefore the linear model (Eq. 4.21) is used as the imperfect model and error-model is formulated using the prior statistics of the model-error, which is evaluated based on the difference between the cubic (Eq. 4.23) and linear models (Eq. 4.21) using 100 samples. The PCA based error-model is formulated for Algorithms 4, 5 and 6 similar to Chapter 3 by retaining two basis function and basis weights from prior model-error statistics. The prior parameters are sampled from a normal distribution with zero mean ($\mu = 0$) and standard deviation $\sigma = 10$. The same prior mean and standard deviation are taken for ES-MDA and Flexible ES-MDA in the previous section of the test case 1, which allows the consistent comparison between the algorithms. The prior samples of the common parameters between linear and cubic models are same so that the actual prior model-error statistics can be computed. The initial settings for the Algorithms 4, 5 and 6 are same as Chapter 3, where 100 ensemble members are used with 8 ES-MDA iterations.

MCMC modified random walk method (discussed in Chapter 1) is used to evaluate the reference posterior distribution of the parameters by inverting the linear model (Eq. 4.21) and the error-model. The joint inversion of the linear model and error-model using MCMC allows us to evaluate the reference posterior distribution and to verify the solutions from Algorithms 4, 5 and 6. The reference posterior distribution is obtained using one million samples of the linear model and error-model parameters with the step size of 0.005 and 10% burn-in period.

Figure 4.7 shows the posterior distribution of the linear model parameters. Algorithms 4, 5 and 6 perform relatively well to reduce the bias in parameter estimation and nearly obtain an unbiased solution as compared to the reference solution from MCMC. The solutions are not exactly same as reference solution from MCMC because of the limitations of the ensemble based techniques. However Fig. 4.8 shows the posterior solution are a robust approximation in case of a calibration of imperfect model. Figure 4.8 shows the combined result of linear model and error-model using Algorithms 4, 5, 6 and MCMC, which show similar posterior behaviour as cubic model because the

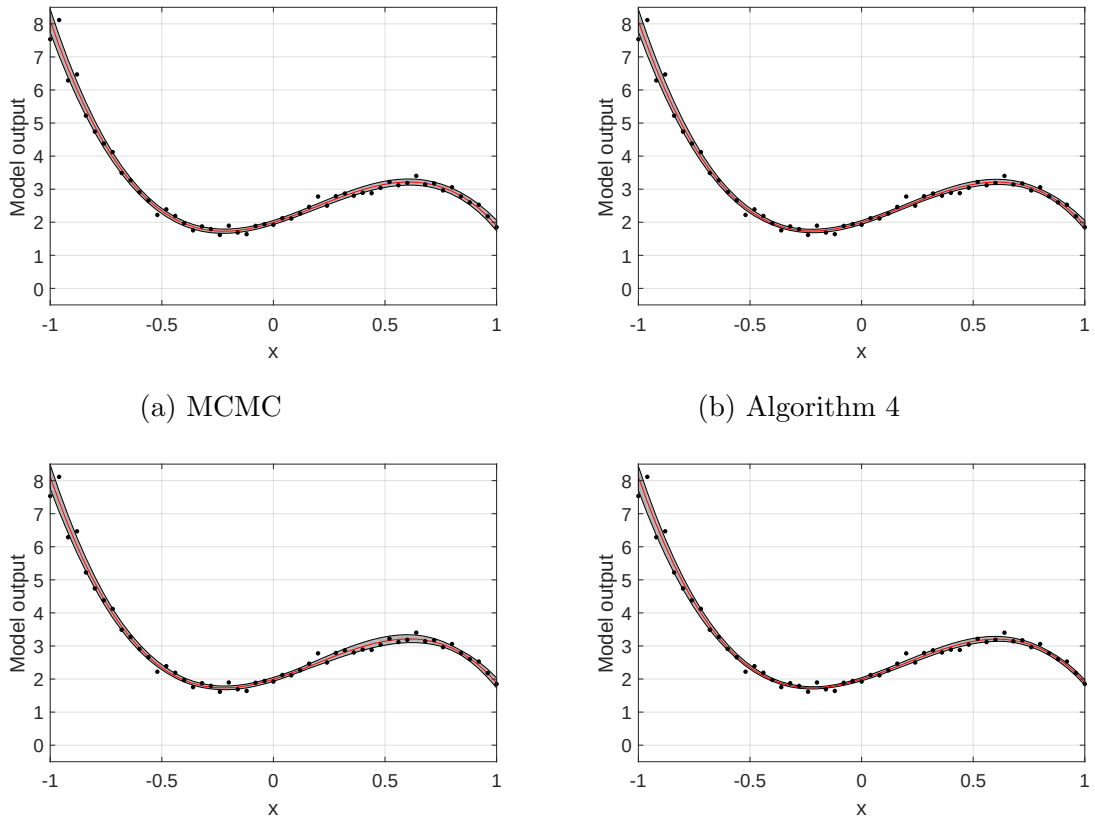
outputs (errors) from the error-model are added into the linear model solution.

The approximate posterior distributions from Algorithms 4, 5 and 6 are clearly far better than the solution of ES-MDA and Flexible ES-MDA as shown in Fig. 4.4 in terms of accuracy and precision in uncertainty quantification. This is due to the fact that Algorithms 4, 5 and 6 require prior model-error statistics for performing joint calibration of the linear and error-model. However Flexible ES-MDA relaxes the assumption of requirement of prior model-error statistics and joint calibration. The Flexible ES-MDA can only reduce the bias in the parameter estimation in order to cover reference solution (Figs. 4.3(f), 4.4(a, b)) by capturing unknown model-error uncertainty, which is a kind of trade-off between reliability and precision.



(a) Constant coefficient (b) Coefficient of x

Figure 4.7: Comparison of the posterior distribution of parameters of the linear model. Red vertical dashed lines with small circles on ends show the reference solution (true model parameters).



(a) MCMC (b) Algorithm 4

(c) Algorithm 5 (d) Algorithm 6

Figure 4.8: Posterior distribution of linear model and error-model outputs for test case 1 (with measurement error of 5%). The descriptions of lines and colors are same as in Fig. 4.1.

4.3.2 Test case 2: The simple machine

In this test case, the simple machine model is calibrated with the observation data obtained from the unknown complex machine. The main objective is to test the flexibility of proposed algorithm for the cases, where the model lacks some physics. The imperfect model for simple machine is shown in eq. 4.25.

$$y = \theta x, \tag{4.25}$$

where y is the work obtained from machine, x is the effort on machine and θ is the efficiency of machine. True complex machine $T(x)$ includes the friction effect, which

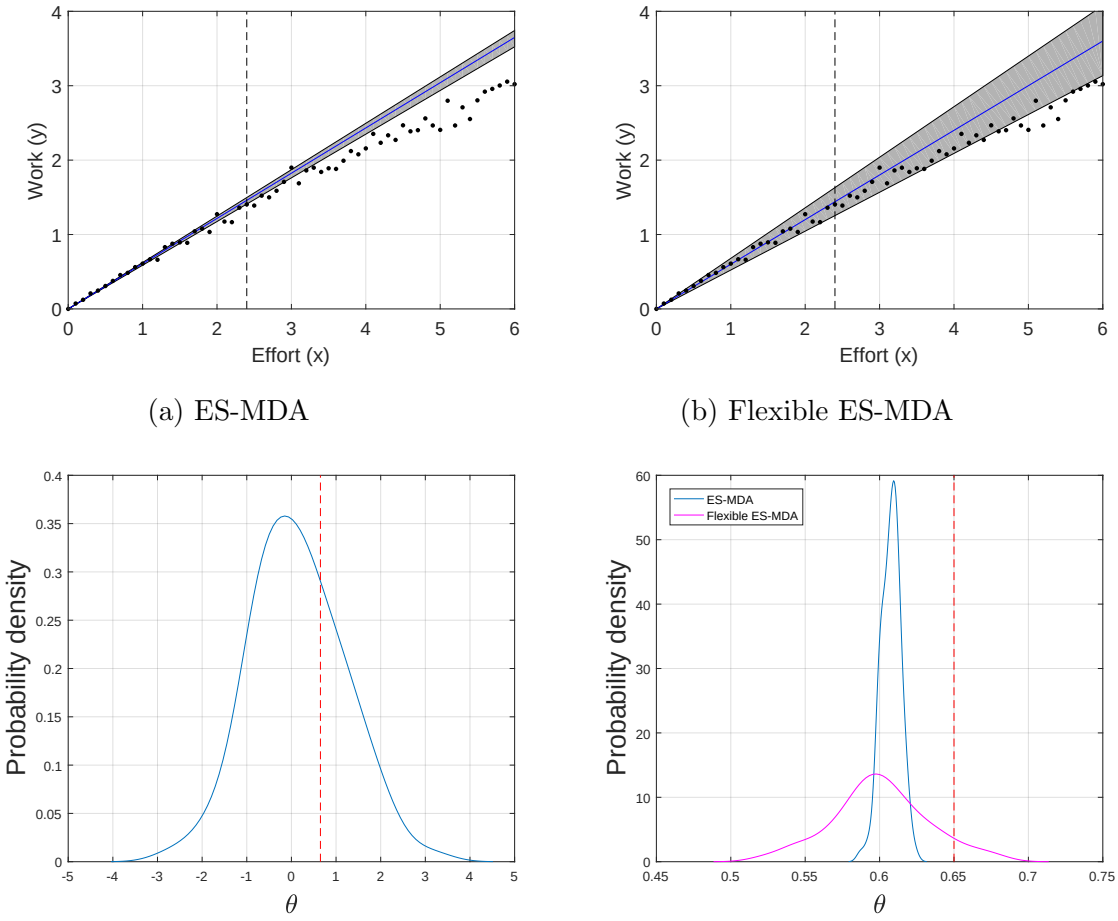
is unknown. The observed data d_{obs} is generated using the function of true complex machine using:

$$T(x) = \frac{\theta x}{1 + x/a}, \tag{4.26}$$

$$d_{obs} = T(x) + \epsilon_d, \tag{4.27}$$

where $a = 20$ and ϵ_d is the measurement error, which is taken as 5% of true function. In this test case, the task is to estimate the efficiency of the machine θ and the corresponding output of the calibrated model. The prior distribution for the model parameters is assumed to follow a standard normal distribution i.e. mean $\mu = 0$ and standard deviation $\sigma = 1$. The domain of effort variable x lies within the interval $[0, 6]$. Total number of points are 61 i.e. x contains values from 0 to 6 with the difference of 0.1 between two consecutive points, where 40% are used for calibration/parameter estimation purpose and 60% are used for predictions.

In this test case, an ensemble of 100 members is used with 8 iterations for the calibration of the simple machine model. Figure 4.9 shows the posterior distribution of efficiency (parameter) of the machine and model output obtained by ES-MDA and Flexible ES-MDA algorithms. Both algorithms match the historical/training data, however predictions obtained from the calibrated model using ES-MDA are overconfident, inaccurate and generally unreliable (Fig. 4.9(a)). This is due to the biased posterior distribution of efficiency (parameter) of machine as shown in Fig. 4.9(d). This bias in the estimated efficiency parameter of the machine is reduced significantly and posterior distribution covers the true efficiency of the machine using proposed Flexible ES-MDA (Fig. 4.9(d)). Due to this effect more reliable predictions are obtained from the calibrated simple machine model using the proposed Flexible ES-MDA algorithm (Fig. 4.9(b)).



(a) ES-MDA

(b) Flexible ES-MDA

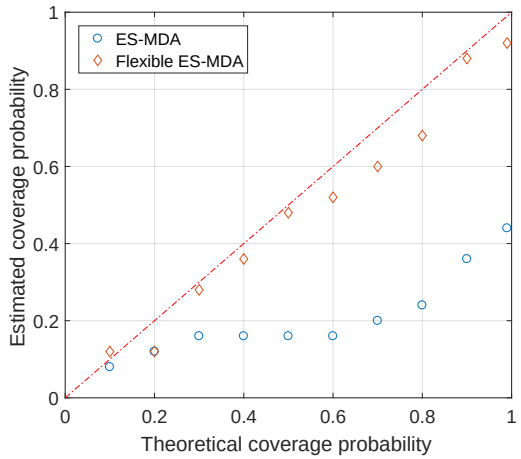
(c) Prior distribution

(d) Posterior distribution

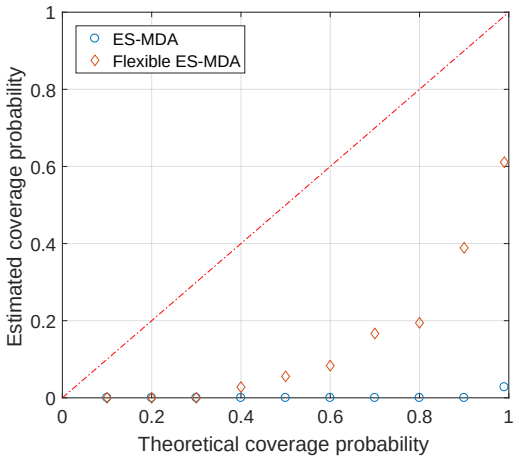
Figure 4.9: Parameter estimation, calibration and prediction results for test case 2. Parts (a) and (b) show the posterior distribution of model output. In parts (a) and (b) blue lines show $p50$ percentiles, grey shaded areas show 99% confidence intervals, dashed black line show the end of historical/training (i.e. use for parameter estimation) data and solid black dots show observation data. Parts (c) and (d) show prior and posterior distribution of model parameter. In parts (c) and (d) red dashed line show the true efficiency of the machine.

Figure 4.10 shows the PICP of training data and predictions. The proposed algorithm shows more robust uncertainty quantification of training data as shown in Fig. 4.10(a). The prediction PICP is also improved by the Flexible ES-MDA (Fig. 4.10(b)), which shows the increase in predictions reliability from calibrated imperfect models (which lacks some physics). Similar results are observed in terms of mean CRPS and

MSE metrics for both historical/training and prediction data as shown in Fig. 4.11.

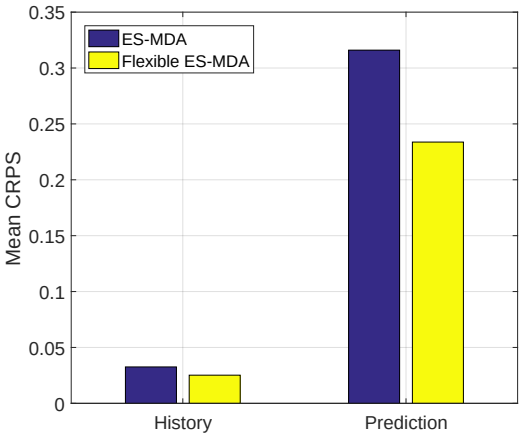


(a) PICP of training/historical data

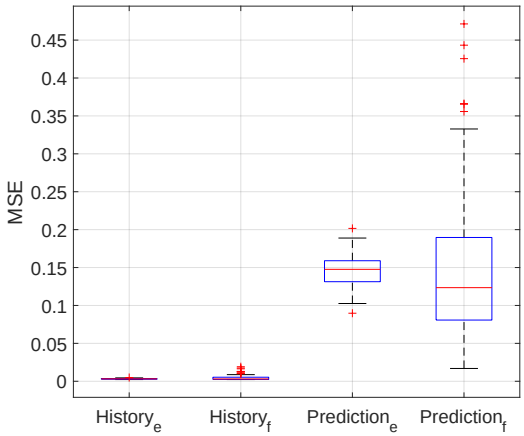


(b) PICP of prediction data

Figure 4.10: Posterior PICP of training/historical data and prediction from simple machine.



(a) Mean CRPS comparison



(b) MSE comparison

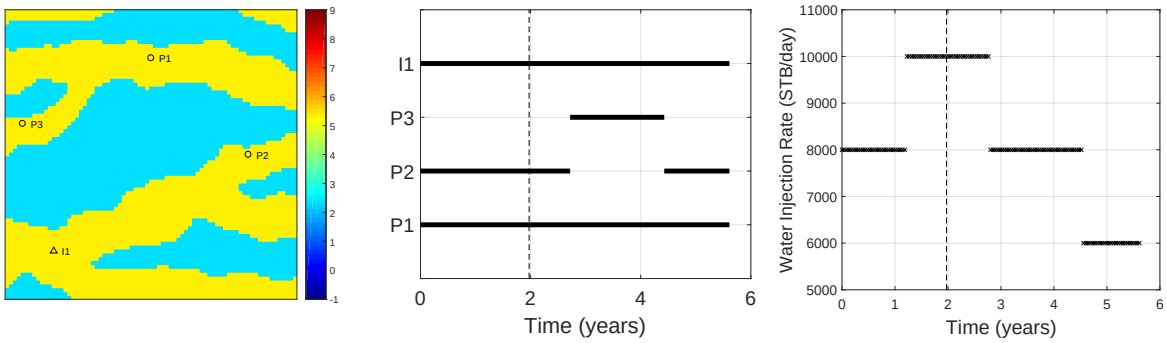
Figure 4.11: Mean CRPS and MSE of posterior ensemble for historical and prediction data of test case 2. In part (b) subscripts 'e' and 'f' show the calibration from ES-MDA and Flexible ES-MDA respectively.

4.3.3 Test case 3: The imperfect reservoir model

This test case is related to the calibration of imperfect model of subsurface oil reservoir with channelized geological patterns of log-permeability field. The subsurface reservoir has dimensions of 7500 ft $\times$ 7500 ft $\times$ 20 ft in the x, y and z directions, respectively. The reservoir contains oil and water phases with in-compressible flow in porous media. The reservoir has uniform porosity of 20% and the initial reservoir pressure is 5000 psi. Figure 4.12 shows the true model, which consists of 75×75 grid blocks, along with different wells open/shut schedule (Fig. 4.12(b)) and controls (Fig. 4.12(c)). The true model is used to generate the observed data and the observations are perturbed by an additive measurement noise, which is taken as 5% of true model response. The reservoir contains one injector well (I1) and three production wells (P1, P2, P3). The production wells are operated under constant bottom hole pressure constraint of 4500 psi and the injector well is operated with time varying constraint of constant injection rate as shown in Fig. 4.12(c). The relative permeability is represented by Corey's power law model, which is described along with parameter values and fluid properties in Appendix A. The capillary pressure and gravitational effects are neglected.

The reservoir is simulated using a 2-D grid with the Matlab Reservoir Simulation Toolbox (MRST) (Lie, 2016). Well P3 is used in the prediction period in order to assess the prediction capabilities of the calibrated reservoir model for future/new wells. The reference and prior fine scale channelized log-permeability fields are generated using a two facies training image as an input to a direct sampling version of MPS algorithm (Mariethoz and Caers, 2014). The imperfect reservoir model is an up-scaled version of true model with a size of 15×15 grid block with no parameterization of the geological features (parameters corresponds to grid-block values). The prior realizations of the imperfect reservoir model are obtained by the harmonic averaging of the permeability field of 75×75 high-fidelity. We note that harmonic average is non-robust upscalling procedure and it is deliberately chosen to introduce the model-error into the simulation model. The upscaled model (a.k.a. imperfect reservoir model) contains two major sources of modeling errors, simplified geological representation and up-scaling

errors. Figure 4.12(a) shows the log-permeability field of true model with the channelized features and Fig. 4.13(a) shows the corresponding up-scaled log-permeability field where the channelized features are blurred due to harmonic up-scaling of the grid properties. For the upscaled model, the physical model parameters are the permeability values at the 225 cells. Except for permeability, all other inputs to the reservoir simulator take the same value for both the reference fine model and the up-scaled models used in history matching, for example relative permeability and porosity.

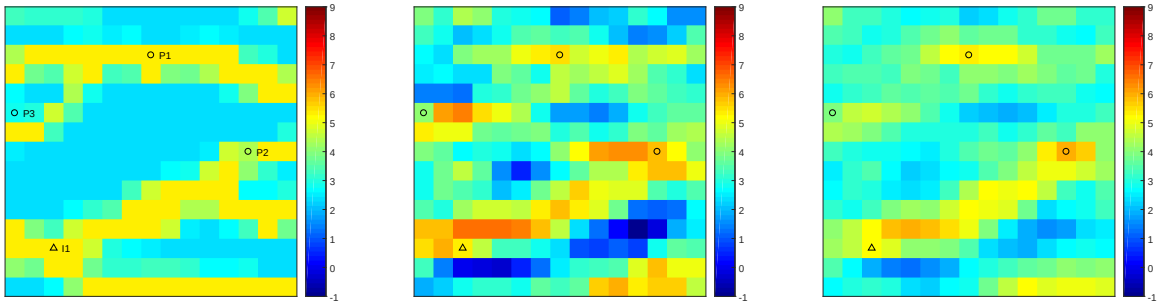


(a) Reference fine (b) Well schedule (c) Injection well control

Figure 4.12: The true reservoir model (use to generate observations). Part (a) shows the fine scale (75×75) reference log-permeability with channelized features. Part (b) shows wells open/shut schedule. In part (b) solid back lines indicate the time periods when a well is open to flow. Part (c) shows water injection rate of injector well I1. Vertical dashed black lines show the end of the historical period in part (b) and (c).

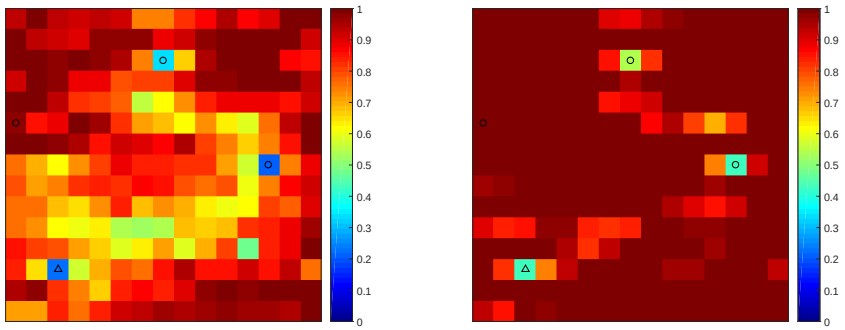
In this test case, an ensemble of 100 members (i.e. one hundred realizations of log-permeability field) is used with 8 iterations for the calibration of imperfect reservoir model. Figure 4.13 shows the mean of the posterior distribution of the log-permeability field obtained by standard ES-MDA and the proposed Flexible ES-MDA algorithms. For the standard ES-MDA, we observe over-shooting (red and blue values) in the mean log-permeability field at a large number of grid blocks. In this case, ES-MDA aggressively tried to adjust the model parameters to match the observation while neglecting the model errors. For the Flexible ES-MDA, we obtained relatively smooth log-permeability fields as shown in Fig. 4.13(c), as the algorithm only updates the model

parameters using a percentage of the data mismatch defined by the split parameter s_p to account for unknown model-error effects. Figure 4.14 shows the standard deviation of log-permeability field obtained using both the standard ES-MDA and the proposed Flexible ES-MDA algorithms. We observe relatively high values of standard deviation from Flexible ES-MDA (Fig. 4.14(b)) as compared to ES-MDA (Fig. 4.14(a)), due to the effect of additional uncertainty of unknown model-error.



(a) Reference (15x15) (b) ESMDA (c) Flexible ESMDA

Figure 4.13: Reference log-permeability field of the imperfect reservoir model with parameter estimation results. Part(a) shows the reference log-permeability field. Part(b) shows the posterior mean obtained from ES-MDA calibration. Part(c) shows the posterior mean obtained from the proposed Flexible ES-MDA.



(a) ESMDA (b) Flexible ESMDA

Figure 4.14: Standard deviation of log-permeability field obtained from posterior ensemble.

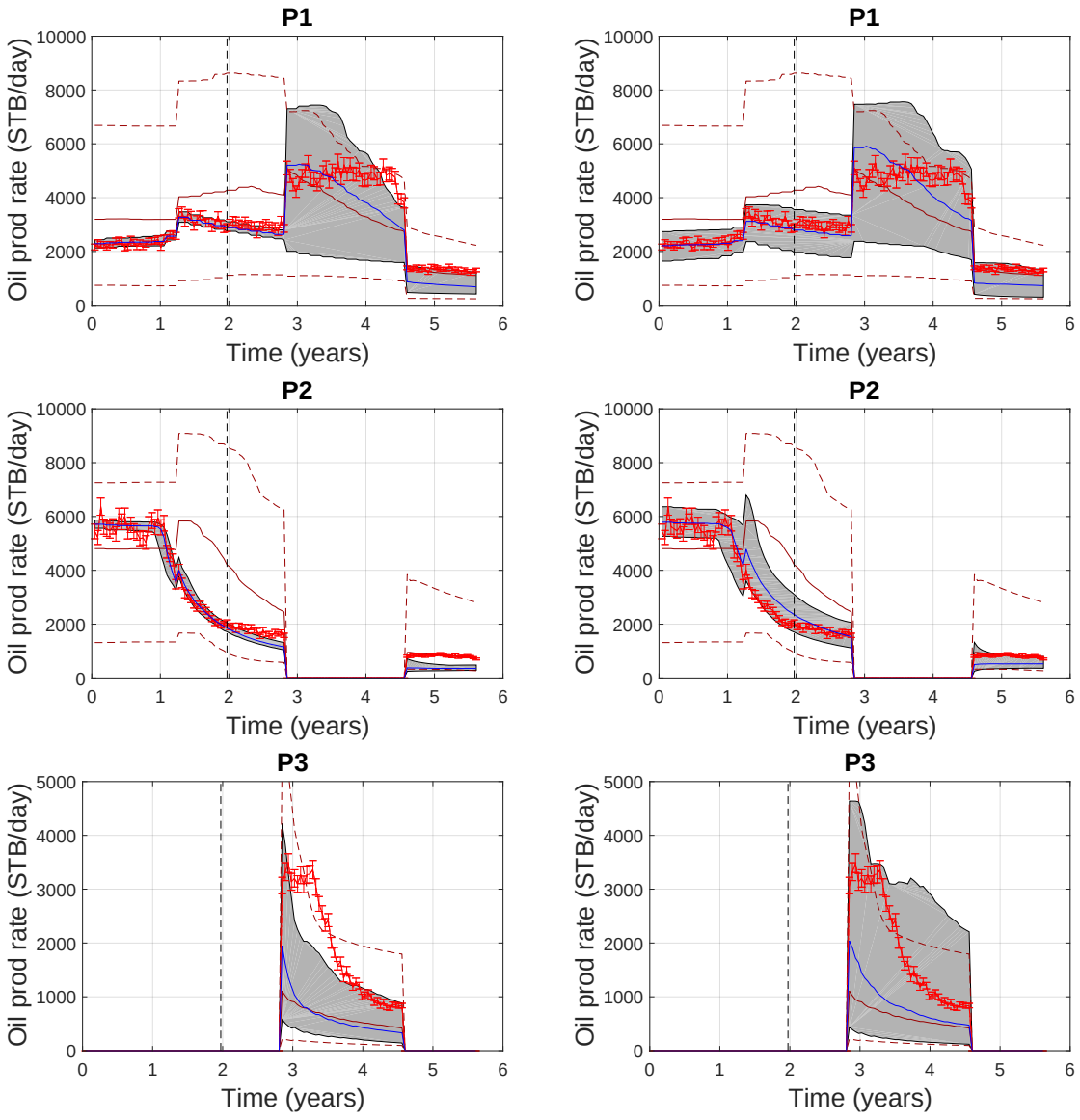
Figure 4.15 shows the prior and posterior distribution of oil rates from the produc-

tion wells. ES-MDA performs well in terms of matching the observations as shown in Fig. 4.15(a), however poor quality predictions are observed for wells P2 and P3. For well P1, the predictions start to deviate after 4.2 years. These over-confident and inaccurate predictions are due to the over-shooting of log-permeability values around production wells (Fig. 4.13(b)). This is a common problem in the petroleum industry, where the historical data is usually matched and often the calibrated models suffers from severe predictability problem (Carter et al., 2006). One of the primary reason of this problem is failing to account for model-error effects during the history matching/calibration process as evident in these results. In contrast, the proposed Flexible ES-MDA algorithm performs relatively better in terms of predictions quality for production wells P1, P2 and P3 as shown in Fig. 4.15(b). This could be easily attributed to the smooth mean log-permeability fields of posterior ensemble as shown in Fig. 4.13(c).

Similar results are observed in Fig. 4.16 for the water rates of production wells and injector well I1. However, we observe that the historical data for well I1 is not fully covered by the posterior distribution obtained by the Flexible ES-MDA as shown in the bottom row of Fig. 4.16(b). This is due to the bad prior distribution (shown by brown lines) of the injection pressure for well I1. We note that this low quality prior is a consequence of modeling errors effect (upscaling) and this can be avoided by using model/grid refinement around wells that are difficult to match. Figure 4.17 shows the PICP of parameter estimation, historical data and predictions. The Flexible ES-MDA shows very robust uncertainty estimates for the model parameters (log-permeability field) as the PICP values lie close to the reference line as shown in Fig. 4.17(a). Additionally, the proposed algorithm shows improved uncertainty estimate of historical production data as shown in Fig. 4.17(b), despite failing to match the historical data for well I1. The prediction PICP is also improved by the Flexible ES-MDA (Fig. 4.17(b)), which shows an increase in predictions reliability from calibrated imperfect reservoir models.

Figure 4.18 shows the mean CRPS and MSE metrics for both historical and prediction data. We observe that mean CRPS and MSE of the historical data show lower values for ES-MDA due to relatively better matching and higher values for Flexible ES-MDA due to the uncovered historical data of well I1. However, the Flexible ES-MDA

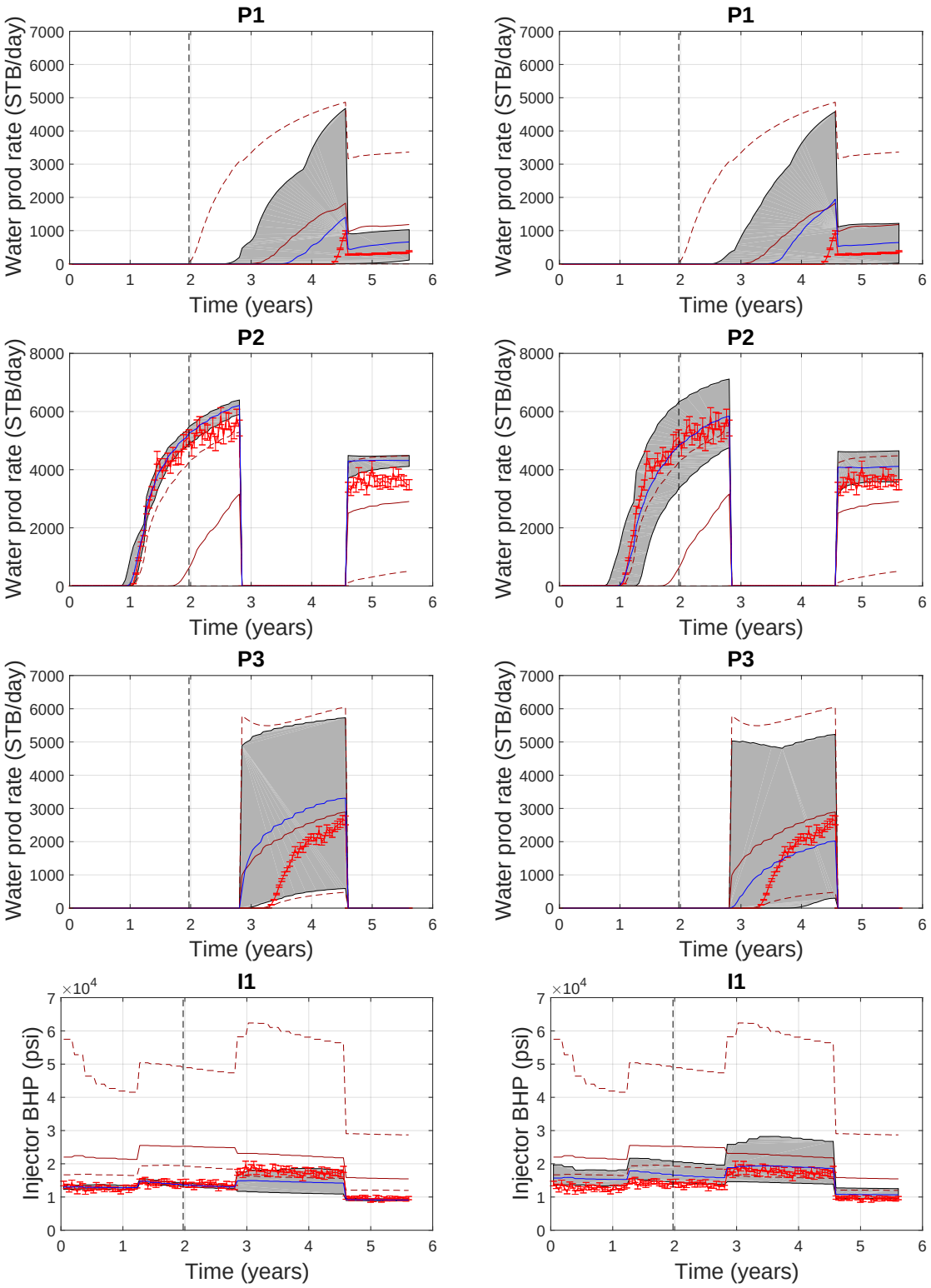
shows significant improvement for mean CRPS and MSE of prediction data. The good matching using ES-MDA can mislead us to over-confident and inaccurate predictions in case of imperfect models and this over-confidence in inaccurate predictions could be avoided using the proposed Flexible ES-MDA.



(a) ESMDA

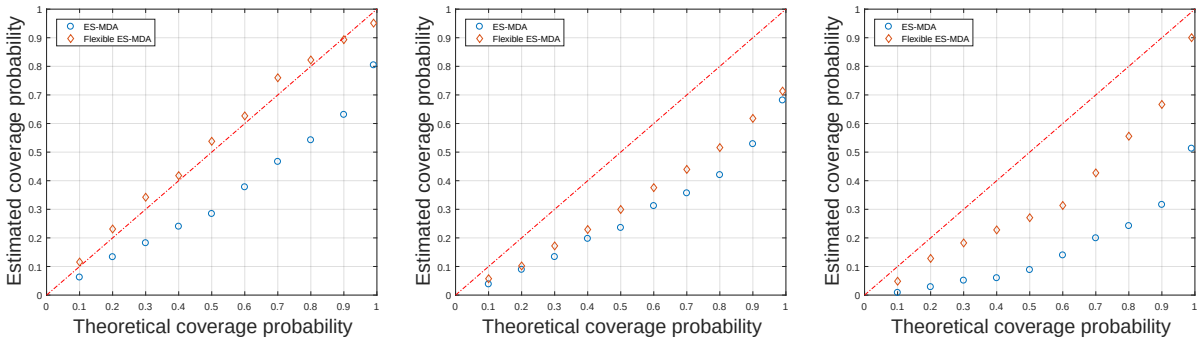
(b) Flexible ESMDA

Figure 4.15: Prior and posterior of oil production data for production wells P1, P2 and P3. Red lines show observed data and bars on red lines show measurement error. Vertical dashed black lines show the end of the historical period. Solid and dashed brown lines show 50th percentile $p50$ and 99% confidence interval of prior distribution respectively. Solid blue lines and gray shaded area show $p50$ and 99% confidence interval of posterior distribution.



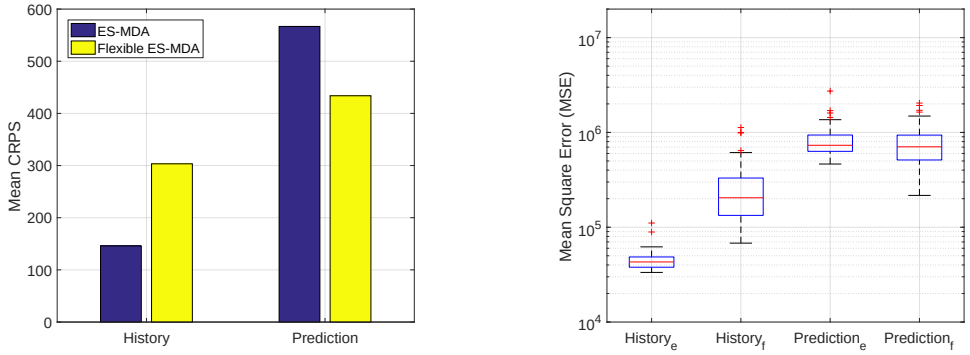
(a) ES-MDA (b) Flexible ES-MDA

Figure 4.16: Prior and posterior of water production and injection pressure data. The descriptions of lines and colors are same as in Fig. 4.15.



(a) PICP of model parameters (b) PICP of historical data (c) PICP of prediction data

Figure 4.17: Posterior PICP of model parameters, historical and prediction data of test case 3.



(a) (b)

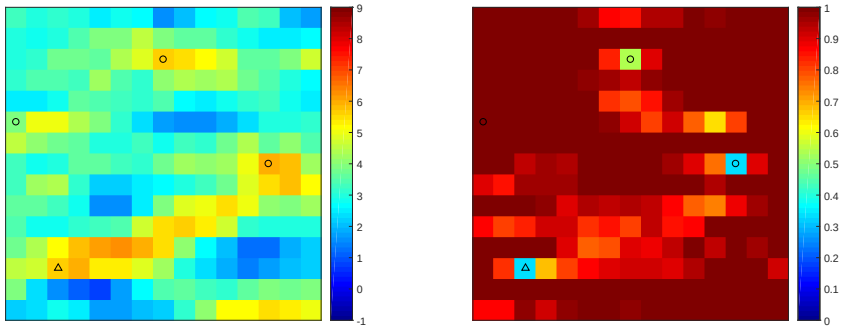
Figure 4.18: Mean CRPS and MSE of posterior ensemble for historical and prediction data of test case 3. In part (b) subscripts 'e' and 'f' shows the MSE obtained from ES-MDA and Flexible ES-MDA respectively.

In cases where model-refinement/correction is not feasible, we propose a variant of the Flexible algorithm where the maximum split factor s_p is bounded by the coverage probability of the 99.99% confidence interval. This results in an adaptive algorithm where the split parameter is adaptively adjusted for each time series using the following simple equation,

$$\text{If: } \; s_p > CP: \quad \text{then } \; s_p = CP, \tag{4.28}$$

where CP is the coverage probability of the 99.99% confidence interval. The only additional modification is the computation of the CP parameter for each time series at every iteration of Flexible ES-MDA algorithm. The mathematical description of CP is presented in the Appendix B. The idea behind the comparison of split parameter with coverage probability that both have same range from $[0, 1]$, and in the case of a time series showing bad prior (failed to cover the data, i.e. $CP = 0$) then the Flexible ES-MDA performs like standard ES-MDA for this output time series during the initial few iterations of the ES-MDA. During later iterations it is expected that the CP becomes non-zero and the split parameter would be adjusted according to the value of CP.

Figure 4.19 shows the posterior results of log-permeability field using the proposed Flexible ES-MDA with adaptive adjustment of the split parameter with respect to coverage probability. We note that the quality of the mean log-permeability field is slightly decreased (Fig. 4.19(a)) as compared to earlier results shown in Fig. 4.13(c) without any adaptive adjustment of the split parameter. Figure 4.20 shows the calibration and prediction results of well outputs obtained from Flexible ES-MDA with adaptive adjustment of the split parameter. In this figure, we only present the results that are affected by the adaptive adjustment of the split parameter. We observe a better data match of the historical data for well I1 (Fig. 4.20) due to the adaptive adjustment of the split parameter in Flexible ES-MDA as compared to without any adjustment of split parameter (Fig. 4.16(b)). In addition, we note that the confidence intervals of production wells P2 and P3 are reduced (Fig. 4.20) due to the relatively lower values of the standard deviation of the posterior ensemble (Fig. 4.19(b)) obtained from Flexible ES-MDA with adaptive split parameter as compared to Fig. 4.14(b).



(a) Mean (b) Standard deviation

Figure 4.19: Mean and standard deviation of log-permeability field of posterior ensemble obtained from Flexible ESMDA with adaptive adjustment of split parameter.

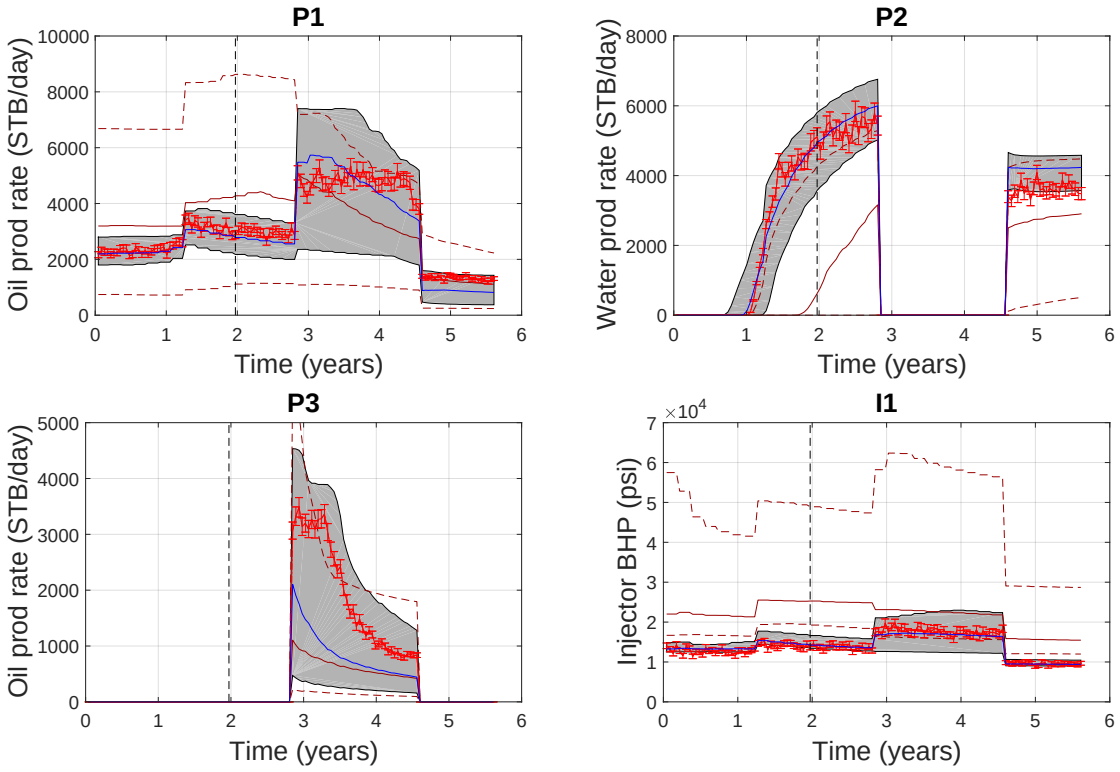
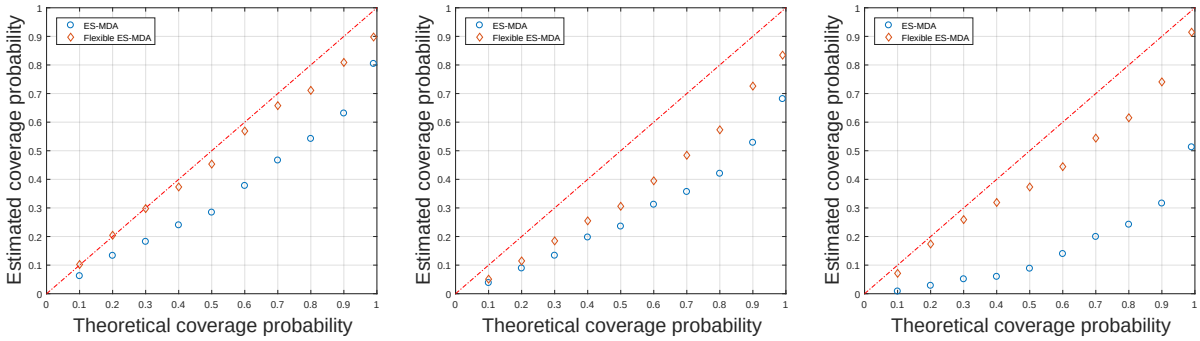


Figure 4.20: Calibration and prediction results of production and injection data using Flexible ESMDA with adaptive adjustment of split parameter. The descriptions of lines and colors are same as in Fig. 4.15.

Figures 4.21 shows the PICP of model parameters, historical data and predictions.

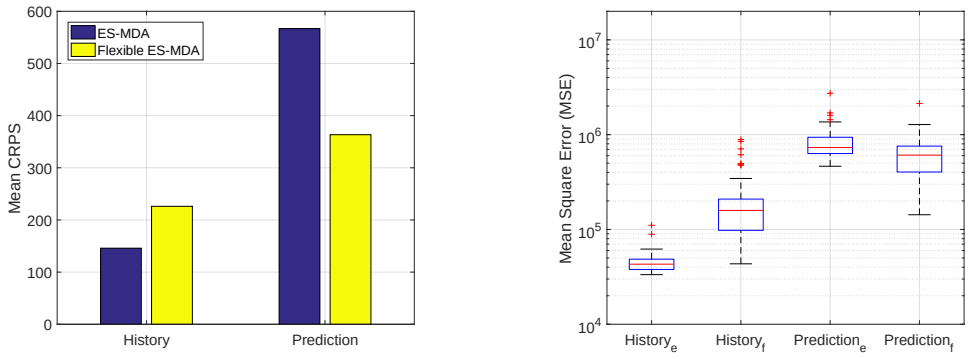
We observe that PICP of model parameters is slightly decreased as compared to the results in Fig. 4.17(a). This is due to the adjustment of split parameter in order to match bad prior time series of well I1. However, the PICP of historical data and predictions are improved as compared to Figs. 4.17(b) and (c) using Flexible ES-MDA. This effect shows some trade-off between quality of estimated parameters and corresponding outputs (specially bad prior) of imperfect reservoir model using adaptive adjustment of the split parameter in the proposed algorithm.



(a) PICP of model parameters (b) PICP of historical data (c) PICP of prediction data

Figure 4.21: Posterior PICP of model parameters, historical and prediction data of test case 3 with adaptive adjustment of split parameter in Flexible ES-MDA.

Figure 4.22 shows the mean CRPS and MSE metrics for both historical and prediction data with adaptive adjustment of split parameter. We observe that mean CRPS and MSE of the historical and prediction data are improved using Flexible ES-MDA as compared to Fig. 4.18 due to matching of historical data of well I1 and reduction in confidence intervals of water production of well P2 and oil production of well P3 by adaptive adjustment of split parameter. These results show that the adaptive adjustment of split parameter in Flexible ES-MDA can be used as an alternative option if model-refinement is not feasible and output time series cannot be matched due to bad prior.



(a) (b)

Figure 4.22: Mean CRPS and MSE of posterior ensemble for historical and prediction data of test case 3 with adaptive adjustment of split parameter in Flexible ES-MDA. In part (b) subscripts 'e' and 'f' shows the MSE obtained from ES-MDA and Flexible ES-MDA respectively.

4.4 Conclusions

In this chapter, a flexible algorithm is proposed for calibration of perfect and imperfect models. This flexible algorithm builds on the ensemble smoother with multiple data assimilation, which has the assumption that models are perfect i.e. accurate representation of the real systems. However, it is widely known that mathematical models are approximations of the real systems and some inherent errors always exist in the computational models. Neglecting the model-error during calibration causes bias in the estimated physical parameter and often results in unreliable predictions. In the proposed algorithm, the residual (data mismatch) is split into two parts. One part is used for parameter estimation/update and the second part is used to represent the model-error. The initial split parameter is computed based on the ratio of norm of mean residual (mean deviation from observed data) and norm of maximum residual (maximum absolute deviation from observed data). During calibration (data assimilation) this split parameter is updated based on the ratio of norm of mean residuals obtained in current and previous iterations. The proposed split formulation shows very

close correspondence to the best split value computed from true model-error (see Test case 1).

In this work, three test cases have been used to observe the performance of the proposed algorithm. These test cases are related to the calibration of polynomial functions, a simple machine and an imperfect reservoir model. In the first test case, cubic function is considered as perfect model and imperfect models are represented by quadratic and linear functions. We observe that if the model is perfect, the proposed algorithm exactly match the data and for imperfect models the algorithm has the flexibility to capture the unknown model-error, which is very useful to avoid over-confidence and generally inaccurate predictions. The second test case is related to estimation the efficiency parameter of a simple machine model which lacks physics in terms of friction component. The calibration result from proposed algorithm shows more reliable estimation of the efficiency of the machine which provides significant improvement in prediction. The third test case is related to calibration of an imperfect reservoir model, which has blurred channelized geological patterns. The imperfect reservoir model has two sources of modelling errors, a simplified geological representation and up-scaling errors. We note that the proposed algorithm reduces the model-bias in parameter estimation and as a consequence predictions are improved significantly from the calibrated imperfect models. We observe improvements in PICP, mean CRPS and MSE metrics after evaluating the parameter estimation, calibration and prediction performance of the proposed algorithm for three test cases of different physical systems. However, the well output which shows a bad prior due to the model-error effect can be difficult to match using the proposed algorithm. This can be avoided by adaptive adjustment of the split parameter in the proposed algorithm with some trade-off between quality of estimated parameters and corresponding model outputs.

The proposed algorithm has the assumption that the data misfit/residual can be split into a parameter error and a model error according to the ratio of mean deviations between iterations, and strongly recommend model improvement in terms of physics, assumptions, details and description, if a large magnitude of model-error is indicated by the proposed algorithm. However, we argue that the proposed algorithm provides

good indicators about the reliability of the calibrated models especially for the cases with unknown model-error. Sometimes for effective decision making we need to reduce the uncertainty to a threshold value (i.e. a minimum value which effects the decision making process). If the model-error uncertainty is greater than the threshold value, we strongly recommend to improve the model in order to reduce model-error uncertainty. In cases where model-error uncertainty is less than acceptable threshold value and there is a time-constraint, the proposed algorithm provides a robust method for calibrating imperfect models that can be used for decision support. In future, we would try to address different types of uncertain parameters and simulation models with different forms of model-errors in terms of complex physics, structural geology, fluid dynamics and phase behaviour.

Chapter 5

Thesis conclusions

In this thesis, different approaches and algorithms are developed to account for model-error during calibration of imperfect reservoir models in order to reduce bias in parameter estimation and to obtain reliable predictions. In Chapter 2, a Bayesian framework for history matching of imperfect reservoir models has been developed where, the calibration is performed as joint estimation of reservoir model parameters and an error-model parameters. Principal component analysis is used to formulate the model-errors, where the PCA basis function and prior statistics of the PCA basis weights were obtained using pairs of accurate/high-fidelity and inaccurate/low-fidelity models. The accurate/high fidelity model is only used for defining the prior model-error statistics and during calibration only the imperfect model is used. The obtained results show that the estimated model parameters are biased when modelling errors are neglected. Consequently the calibrated imperfect model predictions are unreliable and inaccurate. Utilizing the developed joint inversion procedure results in significant improvements in terms of the quality of the estimated parameters, the matching capacity to historical data and prediction accuracy/reliability of the calibrated imperfect models.

In Chapter 3, six different algorithms for calibration of imperfect models are investigated and evaluated. In this chapter, two new algorithms (5 and 6) are introduced which are the variants of the proposed methodology of the first part of the thesis objective and published algorithms. In Algorithms 5 and 6, the novelty is introduced in terms of the identifiability of model-error parameters and the use of joint calibration of both the predetermined error-model and the physical model. Algorithm 3 and 4 are related to the methodology proposed in Chapter 2. Algorithm 1 neglects the presence of model errors and is used as a base case for comparison. Algorithm 2 estimates a

covariance for the total error from the data residual after history matching, and this estimated total error term is used in subsequent history matching runs in order to compensate for the model error. In Algorithms 3, 4, 5 and 6, the history matching problem is formulated as a joint parameter estimation of the physical model parameters (reservoir model parameters) and the parameters of an error-model. In addition, Algorithms 4, 5 and 6 account for the remaining second-order errors through the use of a diagonal error matrix, orthogonal components, and an iterative update of the covariance of the total error respectively. The obtained numerical results show that the Algorithms 4, 5 and 6 perform better in terms of the quality of the parameter estimation, the quality of history matching and the quality of future predictions after calibration. Algorithms 4, 5 and 6 could easily be scaled to large scale models as the error-model formulation is I/O independent and the prior distribution of the error-model parameters could be estimated before the calibration step. Usually the prior realizations (ensemble members) of model parameters are estimated from geo-statistical techniques. These techniques can be broadly classified into two-point (Kriging, SGSIM) and multi-point geostatistics (SNESIM or object based modelling) etc. In two-point statistics the correlation length is estimated based on the variogram modelling which describe the spatial relationship of the model parameters. If the uncertainty of the flow region is reduced then the uncertainty of non-flow region uncertainty could be reduced for a certain range based on the prior correlation length.

The proposed Algorithms in Chapters 2 and 3 require prior model-error statistics before the joint calibration of the physical and error-model parameters. Joint calibration can be non-robust if the prior statistics of model-error is not realistic. In real life applications, an accurate model may not exist, but a high-fidelity model can still be constructed to be closer to reality compared to the low-fidelity/imperfect model and approximate prior model-error statistics could be estimated. However, in some situations one may have access to only imperfect/low-fidelity model and there is a possibility that high-fidelity model contains a large magnitude of the unknown model errors. This complex situation has been addressed in the Chapter 4 where the model-error during calibration is accounted for without knowing any source and prior statistics of model

discrepancy. In Chapter 4, a flexible ensemble-based algorithm is proposed for calibration of perfect and imperfect models. In the proposed algorithm, the residual (data mismatch) is split in two parts. One part is used for parameter estimation/update and second part is used to represent the model-error. The initial split parameter is computed based on the ratio of norm of mean residual (mean deviation from observed data) and norm of maximum residual (maximum absolute deviation from observed data). During calibration (data assimilation) this split parameter is updated based on the ratio of norm of mean residuals obtained in current and previous iterations of the proposed algorithm in Chapter 4. The numerical results show that if the model is perfect, the proposed algorithm exactly match the data and for imperfect models the algorithm has the flexibility to capture the unknown model-error uncertainty, which is very useful to avoid over-confidence in inaccurate predictions. The proposed algorithm reduces the model-bias in parameter estimation and as a consequence predictions are improved significantly from the calibrated imperfect models. The obtained results show significant improvements in PICP, mean CRPS and MSE metrics after evaluating the parameter estimation, calibration and prediction performance of the proposed algorithm for different physical systems. Algorithms 4, 5 and 6 perform relatively well with respect to reference MCMC solution for test case 1 in the Chapter 4. The approximate posterior distributions are nearly unbiased as compared to reference MCMC solution. These results show that the Algorithms 4, 5 and 6 outperform as compared to the methods in the literature (Köpke et al., 2018; Calvetti et al., 2018) by obtaining nearly unbiased results from the imperfect model without evaluating the accurate/high-fidelity model during the calibration process. The methods in the literature (Köpke et al., 2018; Calvetti et al., 2018) can produce nearly unbiased results for a particular type of real systems, however these methods require the evaluation of accurate/high-fidelity model during the calibration process. The Flexible ES-MDA can only reduce the bias in parameter estimation. This leads us to the conclusion that Algorithms 4, 5 and 6 are better than the flexible iterative ensemble smoother. However, it is important to note that the proposed Flexible ES-MDA does not require prior model-error statistic. If the prior model-error statistic is not realistic then Flexible ES-MDA could perform better than

Algorithms 4, 5 and 6.

There are few limitations of this work which could be avoided and addressed in the future work. In this thesis the reference posterior distribution of perfect reservoir model has not been obtained. In future, the calibration of perfect reservoir model could be performed using MCMC in order to obtain reference posterior distribution of the reservoir model and could allow more robust comparison between the proposed methodologies. In Algorithms 4, 5 and 6 only two PCA basis functions per time series were taken for joint calibration of imperfect model and error-model. The number of PCA basis functions could vary for different types of reservoir models. The general criteria for the number of PCA basis functions could be explored further. One could argue that 99% of the singular values criteria can be chosen, which might be prone to over-fitting especially in case of Algorithm 4 due to the reduction of second-order error. However, Algorithms 5 and 6 could overcome this over-fitting problem. These issues would deserve another compressive study on different types of reservoir models. In Algorithm 4 the full covariance of the second-order error could also be taken instead of diagonal matrix, especially in situations when the second-order errors would have systematic bias or strong correlation w.r.t time. Algorithm 6 could be computationally expensive as compared to other proposed algorithms due to the re-calibration step in addition to ES-MDA iterations. This could be avoided by defining different types of stopping criteria. The joint calibration could be evaluated using the prior model-error statistics obtained from imperfect high-fidelity model. In this scenario, the proposed Flexible ES-MDA could be used instead of ES-MDA in joint-calibration framework (Algorithm 3) in order to avoid over-fitting and to determine the reliability of the joint-calibration of the physical and error-model.

5.1 Remarks and future directions

In practical situations we don't know about the realistic reliability of the physical model i.e. how much data mismatch from a physical model would be contributed by the parameter error and model-error. In other words, we don't know to what degree a

model is perfect or imperfect. In these situations it is recommended to use the proposed flexible iterative ensemble smoother which can serve as a tool to identify the reliability of the physical model. If we have an idea about the reliability of the final model, then we could do forecast far into the future such that the level of uncertainty (imprecision) is acceptable for the decision making process. Fortunately the proposed flexible iterative ensemble smoother can be used to identify the reliability of the model in terms of quantification of model-error uncertainty which results in the imprecise (uncertain) behaviour of forecast. Therefore model improvement is strongly recommended in terms of physics, assumptions, details and description, if a large magnitude of model-error is indicated by the proposed algorithms in order to reduce model error uncertainty. The proposed flexible iterative ensemble smoother for calibration of imperfect models can be used in situations, when an accurate or high fidelity model may be missing or the sources of the modelling errors are unknown. Sometimes for effective decision making the uncertainty has to be less than threshold value (i.e. minimum value which effects decision making process). If the model-error uncertainty is greater than the threshold value, the model-improvement is recommended in order to reduce model-error uncertainty. The time limitation in decision making process may restrict the model-improvement step in practical situations. In these scenarios, the imperfect models calibrated using flexible iterative ensemble smoother could be used for decision making process.

In future, the proposed flexible iterative ensemble smoother can be evaluated and extended to the calibration of imperfect models of different physical systems with various types of uncertain parameters and model inadequacy. In addition, the split based formulation of the model-error could be applied to other iterative ensemble smoothers (Chen and Oliver, 2013; Luo et al., 2015) in the literature and could be compared with the proposed flexible iterative ensemble smoother.

If this work was started again then Flexible ES-MDA would be used first to identify the level of reliability of the physical model. If the high model-error is obtained using Flexible ES-MDA then the model refinement step in terms of the scale, description, physics and assumptions should be taken. The refined and unrefined models can be

used as joint calibration of parameter estimation in the scenario when the refined model is computationally very expensive for calibration (history matching) and uncertainty quantification task. The proposed framework of joint parameter estimation of the imperfect reservoir model and error-model could be used because the approximate prior model-error statistics can be estimated before calibration and uncertainty quantification task. The proposed joint calibration framework could easily be applied to large scale models as the error-model formulation is I/O independent and the prior distribution of the error-model parameters could be estimated before the calibration step. In this work, a high-fidelity model without model-error is considered in order to estimate the prior model-error statistics for joint calibration framework. It is also possible that the high-fidelity model contains some model-error. In these cases, the joint calibration can only improve the parameter estimation and the prediction up to the level expected from high-fidelity model. In future, the application of the proposed framework of the joint calibration can be extended to the cases with different types of known sources of errors. Moreover, the low fidelity models can be used for computationally expensive uncertainty quantification task using proposed algorithmic framework instead of high fidelity models, which could be only used to quantify the prior model-error statistics.

In future, different forms of model-errors, uncertain parameters, high-fidelity and low-fidelity models in terms of complex physics, structural geology, fluid dynamics and phase behaviour could be addressed using the proposed framework. The prior model-error statistics for the large scale models could be challenging to obtain especially when a few runs are feasible due to the computation limitation of high fidelity models. This challenging situation could be addressed using combination of multi-dimensional scaling and clustering (Scheidt and Caers, 2009) or machine learning algorithms (Murphy, 2012) to select the best possible realizations for the estimation of prior model-error statistics for large scale models. Moreover, an Input/Output dependent approach could also be used to formulate model discrepancy where the model-errors is treated as a functional approximation problem, which can be solved using a generic machine learning method, such as kernel-based learning (Schölkopf and Smola, 2002) or deep neural networks (Murphy, 2012). These approaches could be used to the development

of hybrid algorithms which would be combination of physics based models and data driven (machine learning) algorithms. However, these type of formulations are prone to be more data-driven due to the strong tendency of the machine learning algorithms for absorbing large amount of data residuals as compared to the simulation models. This problem could be avoided by setting the realistic prior of the model-error using pairs of high-fidelity and low-fidelity models and the use of multi-modal calibration approach (Elsheikh et al., 2013).

Appendix A

Reservoir properties

A.1 Reservoir fluid and relative permeability properties

Corey model in form of power law is used to generate relative permeability data for the reservoir model. Mathematically Corey model in form of power law is written as follows.

$$k_{rw} = (\hat{S_w})^{n_w} k_w^0. \tag{A.1}$$

$$k_{ro} = (1 - \hat{S_w})^{n_o} k_o^0. \tag{A.2}$$

$$\hat{S_w} = \frac{S_w - S_{wc}}{1 - S_{or} - S_{wc}}. \tag{A.3}$$

The notations of above equations are described in MRST manual Lie (2016). The fluid data and corey relative permeability model parameters used in the reservoir model are shown in the Table A.1.1.

Table A.1.1: Reservoir fluid data and Corey relative permeability model parameters

Fluid properties		Corey relative permeability model parameters			
water viscosity	0.5 cp	S_{or}	0.2	k_o^0	1
oil viscosity	1 cp	S_{wc}	0.2	k_w^0	1
water density	1000 kg/m^3	n_w	2		
oil density	700 kg/m^3	n_o	2		

Appendix B

Forecasting metrics

B.1 Mean Square Error (MSE)

The Mean Square Error (MSE) is obtained using,

$$MSE = \frac{1}{N_d}\sum_{n=1}^{N_d}(d_n - d_{obs,n})^2, \tag{B.1}$$

where n is index of observation or model prediction at corresponding time.

B.2 Coverage Probability (CP)

$$CP = \frac{N_{CI}}{N_t}. \tag{B.2}$$

N_{CI} = Number of samples, parameters or observations in Confidence Interval

N_t = Total number of samples, parameters or observations

B.3 Continuous Ranked Probability Score (CRPS)

The details of CRPS for ensemble prediction system were described by Hersbach (2000). In this section summary of CRPS is explained.

Mathematically CRPS can be defined as,

$$CRPS = \int_{-\infty}^{\infty}[p(x) - H(x - x_{obs})]^2 dx, \tag{B.3}$$

where $p(x) = \int_{-\infty}^{x} \rho(y)dy$ Cumulative distribution of quantity of interest, $H(x - x_{obs})$ = Heaviside function (Step function) i.e.

$$H(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

For an ensemble system with N_e realizations, the CRPS can be written as follows,

$$CRPS = \sum_{i=0}^{N_e} c_i. \tag{B.4}$$

$$c_i = \alpha_i p_i^2 + \beta_i (1 - p_i)^2. \tag{B.5}$$

where $p_i = P(x) = i/N_e$, for $x_i < x < x_{i+1}$ (Cumulative distribution is a piece wise constant function).

$$\alpha_i = \begin{cases} 0 & \text{if } x_{obs} < x_i \\ x_{obs} - x_i & \text{if } x_i < x_{obs} < x_{i+1} \\ x_{i+1} - x_i & \text{if } x_{obs} > x_{i+1} \\ x_{obs} - x_{N_e} & \text{if } x_{obs} > x_{N_e} \\ 0 & \text{if } x_{obs} < x_1 \end{cases}$$

$$\beta_i = \begin{cases} x_{i+1} - x_i & \text{if } x_{obs} < x_i \\ x_{i+1} - x_{obs} & \text{if } x_i < x_{obs} < x_{i+1} \\ 0 & \text{if } x_{obs} > x_{i+1} \\ 0 & \text{if } x_{obs} > x_{N_e} \\ x_1 - x_{obs} & \text{if } x_{obs} < x_1 \end{cases}$$

Bibliography

Abou-Kassem, J. H., Aziz, K., et al. (1985). Analytical well models for reservoir simulation. *Society of Petroleum Engineers Journal*, 25(04):573–579.

Ahmed, T. (2006). *Reservoir Engineering Handbook*. Elsevier Science.

Ahmed, T. (2016). *Equations of State and PVT Analysis: Applications for Improved Reservoir Modeling*. Elsevier Science.

Asher, M. J., Croke, B. F., Jakeman, A. J., and Peeters, L. J. (2015). A review of surrogate models and their application to groundwater modeling. *Water Resources Research*, 51(8):5957–5973.

Brynjarsdottir, J. and O' Hagan, A. (2014). Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, 30(11):114007.

Calvetti, D., Dunlop, M., Somersalo, E., and Stuart, A. (2018). Iterative updating of model error for bayesian inversion. *Inverse Problems*, 34(2):025008.

Calvetti, D., Ernst, O., and Somersalo, E. (2014). Dynamic updating of numerical model discrepancy using sequential sampling. *Inverse Problems*, 30(11):114019.

Cardoso, M., Durlofsky, L., and Sarma, P. (2009). Development and application of reduced-order modeling procedures for subsurface flow simulation. *International journal for numerical methods in engineering*, 77(9):1322–1350.

Carter, J. N., Ballester, P. J., Tavassoli, Z., and King, P. R. (2006). Our calibrated model has poor predictive value: An example from the petroleum industry. *Reliability Engineering & System Safety*, 91(10):1373–1381.

Chen, Y., Lallier, F., and Moncorge, A. (2016). On uncertainty quantification of history matched facies models. In *ECMOR XV-15th European Conference on the Mathematics of Oil Recovery*.

Chen, Y. and Oliver, D. S. (2013). Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Computational Geosciences*, 17(4):689–703.

Cheney, E. and Kincaid, D. (2009). *Linear Algebra: Theory and Applications*. Jones and Bartlett Publishers.

Cotter, S. L., Roberts, G. O., Stuart, A. M., White, D., et al. (2013). MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science*, 28(3):424–446.

Dreano, D., Tandeo, P., Pulido, M., Ait-El-Fquih, B., Chonavel, T., and Hoteit, I. (2017). Estimating model-error covariances in nonlinear state-space models using Kalman smoothing and the expectation–maximization algorithm. *Quarterly Journal of the Royal Meteorological Society*, 143(705):1877–1885.

Durlofsky, L. J. (2003). Upscaling of geocellular models for reservoir flow simulation: a review of recent progress. In *7th International Forum on Reservoir Simulation Bühl/Baden-Baden, Germany*, pages 23–27. Citeseer.

Elsheikh, A. H., Wheeler, M. F., and Hoteit, I. (2013). Clustered iterative stochastic ensemble method for multi-modal calibration of subsurface flow models. *Journal of Hydrology*, 491:40 – 55.

Emerick, A. A. and Reynolds, A. C. (2012). History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Computational Geosciences*, 16(3):639–659.

Emerick, A. A. and Reynolds, A. C. (2013). Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, 55:3–15.

Ertekin, T., Abou-Kassem, J. H., and King, G. R. (2001). *Basic applied reservoir simulation*. Number Sirsi i9781555630898.

Evensen, G. (2009). *Data assimilation: the ensemble Kalman filter*. Springer Science & Business Media.

Evin, G., Thyer, M., Kavetski, D., McInerney, D., and Kuczera, G. (2014a). Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resources Research*, 50(3):2350–2375.

Evin, G., Thyer, M., Kavetski, D., McInerney, D., and Kuczera, G. (2014b). Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resources Research*, 50(3):2350–2375.

Farmer, C. (2002). Upscaling: a review. *International journal for numerical methods in fluids*, 40(1-2):63–78.

Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall/CRC.

Giudice, D. D., Honti, M., Scheidegger, A., Albert, C., Reichert, P., and Rieckermann, J. (2013). Improving uncertainty estimation in urban hydrological modeling by statistically describing bias. *Hydrology and Earth System Sciences*, 17(10):4209–4225.

Giudice, D. D., Löwe, R., Madsen, H., Mikkelsen, P. S., and Rieckermann, J. (2015). Comparison of two stochastic techniques for reliable urban runoff prediction by modeling systematic errors. *Water Resources Research*, 51(7):5004–5022.

Hansen, T. M., Cordua, K. S., Jacobsen, B. H., and Mosegaard, K. (2014). Accounting for imperfect forward modeling in geophysical inverse problems – exemplified for crosshole tomography. *GEOPHYSICS*, 79(3):H1–H21.

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570.

Hoffman, B. T. and Caers, J. (2007). History matching by jointly perturbing local facies proportions and their spatial distribution: Application to a north sea reservoir. *Journal of Petroleum Science and Engineering*, 57(3):257–272.

Iglesias, M. A. (2015). Iterative regularization for ensemble data assimilation in reservoir models. *Computational Geosciences*, 19(1):177–212.

Iglesias, M. A. (2016). A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems. *Inverse Problems*, 32(2):025002.

Johnson, R. A., Wichern, D. W., et al. (2002). *Applied multivariate statistical analysis*, volume 5. Prentice hall Upper Saddle River, NJ.

Josset, L., Demyanov, V., Elsheikh, A. H., and Lunati, I. (2015). Accelerating Monte Carlo Markov chains with proxy and error models. *Computers and Geosciences*, 85:38 – 48.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464.

Kerschen, G., Golinval, J.-c., VAKAKIS, A. F., and BERGMAN, L. A. (2005). The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: An overview. *Nonlinear Dynamics*, 41(1):147–169.

Köpke, C., Irving, J., and Elsheikh, A. H. (2018). Accounting for model error in Bayesian solutions to hydrogeophysical inverse problems using a local basis approach. *Advances in Water Resources*, 116:195–207.

Kristensen, N. R., Madsen, H., and Jørgensen, S. B. (2004). Parameter estimation in stochastic grey-box models. *Automatica*, 40(2):225–237.

Laloy, E., Rogiers, B., Vrugt, J. A., Mallants, D., and Jacques, D. (2013). Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov

chain Monte Carlo simulation and polynomial chaos expansion. *Water Resources Research*, 49(5):2664–2682.

Lee, J., Rollins, J. B., and Spivey, J. P. (2003). *Pressure transient testing*. Richardson, Tex.: Henry L. Doherty Memorial Fund of AIME, Society of petroleum engineers.

Lie, K.-A. (2016). *An Introduction to Reservoir Simulation Using MATLAB - User Guide for the Matlab Reservoir Simulation Toolbox (MRST)*. Sintef ICT, Department of Applied Mathematics.

Lødøen, O. P., Omre, H., Durlofsky, L. J., and Chen, Y. (2005). Assessment of uncertainty in reservoir production forecasts using upscaled flow models. In *Geostatistics Banff 2004*, pages 713–722. Springer.

Lødøen, O. P. and Tjelmeland, H. (2010). Bayesian calibration of hydrocarbon reservoir models using an approximate reservoir simulator in the prior specification. *Statistical Modelling*, 10(1):89–111.

Lu, M. and Chen, Y. (2019). Improved estimation and forecast through model error estimation – Norne field example. In *International Petroleum Technology Conference*.

Luo, X. (2019). Ensemble-based kernel learning for a class of data assimilation problems with imperfect forward simulators. *PLOS ONE*, 14(7):1–40.

Luo, X., Stordal, A. S., Lorentzen, R. J., Naevdal, G., et al. (2015). Iterative ensemble smoother as an approximate solution to a regularized minimum-average-cost problem: Theory and applications. *SPE Journal*, 20(05):962–982.

Maier, H. R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L. S., Cunha, M. C., Dandy, G. C., Gibbs, M. S., Keedwell, E., Marchi, A., et al. (2014). Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions. *Environmental Modelling & Software*, 62:271–299.

Mariethoz, G. and Caers, J. (2014). *Multiple-point geostatistics: stochastic modeling with training images*. John Wiley & Sons.

MATLAB (2019). Matlab function reference (R2019b). *The MathWorks Inc.*, pages 10761–10763.

McCain, W. (1990). *The Properties of Petroleum Fluids*. PennWell Books.

Montgomery, D., Jennings, C., and Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting*. Wiley Series in Probability and Statistics. Wiley.

Moradkhani, H., DeChant, C. M., and Sorooshian, S. (2012). Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method. *Water Resources Research*, 48(12).

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

O' Sullivan, A. and Christie, M. (2005). Error models for reducing history match bias. *Computational Geoscience*, 9(2-3):125–153.

Oliver, D. S. and Alfonzo, M. (2018). Calibration of imperfect models to biased observations. *Computational Geosciences*, 22(1):145–161.

Oliver, D. S., Reynolds, A. C., and Liu, N. (2008). *Inverse theory for petroleum reservoir characterization and history matching*. Cambridge University Press.

Omre, H., Lødøen, O. P., et al. (2004). Improved production forecasts and history matching using approximate fluid-flow simulators. *SPE Journal*, 9(03):339–351.

O'Sullivan, A. and Christie, M. (2006). Simulation error models for improved reservoir prediction. *Reliability Engineering & System Safety*, 91(10):1382–1389.

Peaceman, D. W. et al. (1983). Interpretation of well-block pressures in numerical reservoir simulation with nonsquare grid blocks and anisotropic permeability. *Society of Petroleum Engineers Journal*, 23(03):531–543.

Rammay, M. H. and Abdulraheem, A. (2014). Automated history matching using combination of adaptive neuro fuzzy system (anfis) and differential evolution algorithm. In *SPE Large Scale Computing and Big Data Challenges in Reservoir Simulation Conference and Exhibition*. Society of Petroleum Engineers.

Rammay, M. H. and Abdulraheem, A. (2016). PVT correlations for Pakistani crude oils using artificial neural network. *Journal of Petroleum Exploration and Production Technology*, pages 1–17.

Rammay, M. H., Elsheikh, A. H., and Chen, Y. (2019). Quantification of prediction uncertainty using imperfect subsurface models with model error estimation. *Journal of Hydrology*, 576:764 – 783.

Rasmussen, C. E. (2006). *Gaussian processes for machine learning*. MIT Press.

Refsgaard, J. C., Christensen, S., Sonnenborg, T. O., Seifert, D., Højberg, A. L., and Troldborg, L. (2012). Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. *Advances in Water Resources*, 36:36–50.

Reichert, P. and Schuwirth, N. (2012a). Linking statistical bias description to multiobjective model calibration. *Water Resources Research*, 48(9).

Reichert, P. and Schuwirth, N. (2012b). Linking statistical bias description to multiobjective model calibration. *Water Resources Research*, 48(9). W09543.

Sargsyan, K., Najm, H., and Ghanem, R. (2015). On the statistical calibration of physical models. *International Journal of Chemical Kinetics*, 47(4):246–276.

Scheidt, C. and Caers, J. (2009). Uncertainty quantification in reservoir performance using distances and kernel methods–application to a west africa deepwater turbidite reservoir. *SPE Journal*, 14(04):680–692.

Schölkopf, B. and Smola, A. J. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

Schoups, G. and Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, 46(10).

Shlens, J. (2014). A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*.

Silva, P. C., Maschio, C., and Schiozer, D. J. (2007). Use of neuro-simulation techniques as proxies to reservoir simulator: application in production history matching. *Journal of Petroleum Science and Engineering*, 57(3):273–280.

Skauvold, J. and Eidsvik, J. (2018). Data assimilation for a geological process model using the ensemble Kalman filter. *Basin Research*, 30(4):730–745.

Stordal, A. S. and Elsheikh, A. H. (2015). Iterative ensemble smoothers in the annealed importance sampling framework. *Advances in Water Resources*, 86:231 – 239.

Storn, R. and Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359.

Sun, L., Seidou, O., Nistor, I., and Liu, K. (2016). Review of the Kalman-type hydrological data assimilation. *Hydrological Sciences Journal*, 61(13):2348–2366.

Sun, W. and Durlofsky, L. J. (2017). A new data-space inversion procedure for efficient uncertainty quantification in subsurface flow problems. *Mathematical Geosciences*, 49(6):679–715.

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the dream software package: Theory, concepts, and matlab implementation. *Environmental Modelling & Software*, 75:273–316.

White, J. T., Doherty, J. E., and Hughes, J. D. (2014). Quantifying the predictive consequences of model error with linear subspace analysis. *Water Resources Research*, 50(2):1152–1173.

Xu, T. and Valocchi, A. J. (2015). Data-driven methods to improve baseflow prediction of a regional groundwater model. *Computers & Geosciences*, 85:124–136.