UNIVERSITY OF CAMPINAS
SCHOOL OF MECHANICAL ENGINEERING
AND INSTITUTE OF GEOSCIENCES

DURHAM UNIVERSITY
DEPARTMENT OF MATHEMATICAL
SCIENCES

HELENA NANDI FORMENTIN

UNCERTAINTY QUANTIFICATION OF PETROLEUM RESERVOIRS: A STRUCTURED BAYESIAN APPROACH

QUANTIFICAÇÃO DE INCERTEZAS EM RESERVATÓRIOS DE PETRÓLEO: UMA ABORDAGEM BAYESIANA ESTRUTURADA

CAMPINAS – BRAZIL
DURHAM – UNITED KINGDOM
2020

HELENA NANDI FORMENTIN

UNCERTAINTY QUANTIFICATION OF PETROLEUM RESERVOIRS: A STRUCTURED BAYESIAN APPROACH

A thesis presented to:

The School of Mechanical Engineering and the Institute of Geosciences of the University of Campinas in partial fulfilment of the requirements for the degree of Doctor in Petroleum Sciences and Engineering in the area of Reservoirs and Management,

and

The Department of Mathematical Sciences of Durham University in partial fulfilment of the requirements for the degree of Doctor in Mathematical Sciences,

Under the cotutelle agreement signed between University of Campinas and Durham University.

Supervisors:

Prof. Dr Denis José Schiozer (University of Campinas)

Prof. Dr Camila Caiado (Durham University)

Prof. Dr Ian Vernon (Durham University)

Prof. Dr Michael Goldstein (Durham University)

Co-supervisor:

Prof. Dr Guilherme Daniel Avansi (University of Campinas)

This document corresponds to the final version of the thesis defended by the student Helena Nandi Formentin and supervised by the supervisors and co-supervisor named above.

CAMPINAS – BRAZIL

DURHAM – UNITED KINGDOM

2020

HELENA NANDI FORMENTIN

QUANTIFICAÇÃO DE INCERTEZAS EM RESERVATÓRIOS DE PETRÓLEO: UMA ABORDAGEM BAYESIANA ESTRUTURADA

Tese apresentada:

à Faculdade de Engenharia Mecânica e Instituto de Geociências da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutora em Ciências e Engenharia de Petróleo na área de Reservatórios e Gestão,

e

ao Departamento de Ciências Matemáticas da Universidade de Durham como parte dos requisitos exigidos para a obtenção do título de Doutora em Ciências Matemáticas,

No âmbito do Acordo de Cotutela firmado entre a Universidade Estadual de Campinas e a Universidade de Durham.

Orientadores:

Prof. Dr. Denis José Schiozer (Universidade Estadual de Campinas)

Prof. Dr. Camila Caiado (Universidade de Durham)

Prof. Dr. Ian Vernon (Universidade de Durham)

Prof. Dr. Michael Goldstein (Universidade de Durham)

Co-orientador:

Prof. Dr. Guilherme Daniel Avansi (Universidade Estadual de Campinas)

Este exemplar corresponde à versão final da Tese defendida pela aluna Helena Nandi Formentin e orientada pelo orientadores e co-orientador nomeados acima.

CAMPINAS – BRASIL

DURHAM – REINO UNIDO

2020

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Luciana Pietrosanto Milla - CRB 8/8129

N153u

Nandi Formentin, Helena, 1987-
Uncertainty quantification of petroleum reservoirs : a estructure Bayesian approach / Helena Nandi Formentin. – Campinas, SP : [s.n.], 2020.

Orientadores: Denis José Schiozer; Camila Caiado; Ian Vernon e Michael Goldstein.
Coorientador: Guilherme Daniel Avansi.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Mecânica.
Em cotutela com: Durham University.

1. Engenharia de petróleo. 2. Reservatórios (Simulação). 3. Estatística. 4. Inferência bayesiana. 5. Análise de dados. I. Schiozer, Denis José, 1963-. II. Caiado, Camila. III. Vernon, Ian. IV. Goldstein, Michael. V. Avansi, Guilherme Daniel, 1984-. VI. Universidade Estadual de Campinas. Faculdade de Engenharia Mecânica. VII. Título.

Informações para Biblioteca Digital

Título em outro idioma: Quantificação de incertezas em reservatórios de petróleo : uma abordagem Bayesiana estruturada
Palavras-chave em inglês:
Petroleum engineering
Reservoirs (Simulation)
Statistics
Bayesian inference
Data analysis
Área de concentração: Reservatórios e Gestão
Titulação: Doutora em Ciências e Engenharia de Petróleo
Banca examinadora:
Denis José Schiozer [Orientador]
Peter Craig
Guilherme Palermo
Hailiang Du
Philip Jonathan
Data de defesa: 20-02-2020
Programa de Pós-Graduação: Ciências e Engenharia de Petróleo

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-4696-8514
- Currículo Lattes do autor: http://lattes.cnpq.br/5759035899083199

UNIVERSITY OF CAMPINAS
SCHOOL OF MECHANICAL ENGINEERING AND INSTITUTE OF GEOSCIENCES

DURHAM UNIVERSITY
DEPARTMENT OF MATHEMATICAL SCIENCES

PHD THESIS

UNCERTAINTY QUANTIFICATION OF PETROLEUM RESERVOIRS: A STRUCTURED BAYESIAN APPROACH

Student: Helena Nandi Formentin
Supervisors:
Prof. Dr Denis José Schiozer
Prof. Dr Camila Caiado
Prof. Dr Ian Vernon
Prof. Dr Michael Goldstein
Co-supervisor:
Prof. Dr Guilherme Daniel Avansi

The Examination Board composed by the members below has approved the Thesis:

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA MECÂNICA E INSTITUTO DE GEOCIÊNCIAS
UNIVERSIDADE DE DURHAM
DEPARTAMENTO DE CIÊNCIAS MATEMÁTICAS

TESE DE DOUTORADO

QUANTIFICAÇÃO DE INCERTEZAS EM RESERVATÓRIOS DE PETRÓLEO: UMA ABORDAGEM BAYESIANA ESTRUTURADA

Estudante: Helena Nandi Formentin
Orientadores:
Prof. Dr. Denis José Schiozer
Prof. Dr. Camila Caiado
Prof. Dr. Ian Vernon
Prof. Dr. Michael Goldstein
Co-orientador:
Prof. Dr. Guilherme Daniel Avansi

A Banca Examinadora composta pelos membros abaixo aprovou esta Tese:

Prof. Dr Denis José Schiozer – UNICAMP

Prof. Dr Peter Craig - Durham University

Prof. Dr Guilherme Palermo – UNICAMP

Prof. Dr Hailiang Du - Durham University

Prof. Dr Philip Jonathan - Lancaster University/Shell

The Minute of the Viva with the respective signatures of the members is in the UNICAMP student's records.

Campinas/Durham, February 20th, 2020

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no processo de vida acadêmica do aluno UNICAMP.

Campinas/Durham, 20 de Fevereiro de 2020

DECLARATION

The work in this thesis is based on research carried out at the University of Campinas (Brazil, 2016-2017) and Durham University (United Kingdom, 2018-2019). No part of this thesis has been submitted elsewhere for any other degree or qualification, and it is all my own work unless referred to the contrary in the text. Chapters 2 to 5 are based on work produced in collaboration with the supervisory team and collaborators, as stated at the beginning of each Chapter. Chapter 2 follows Formentin *et al.* (2019-a); Chapters 3 and 4 closely follow Formentin *et al.* (2019-b), to be submitted for publication as two separate papers as per suggestion of the editor (Formentin *et al.* 2020-a and -b), and Chapter 5 is in preparation for submission in 2020 (Formentin *et al.* 2020-c).

STATEMENT OF COPYRIGHT

DEDICATION

To the time, always gently and generously offering wisdom.

To my family for all the smiles and unconditional support.

To my partner, a truly curious person who gives me a good reason to laugh and care every day.

ACKNOWLEDGEMENTS

This research was carried out in association with the ongoing project registered as ANP 19708-7, “Fomento à Formação de Recursos Humanos em Gestão de Incertezas e Tomada de Decisão: Um Programa BG Fellowship" (UNICAMP/ Shell Brazil /ANP), co-funded by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brazil) through Science Without Borders Programme and Shell Brazil under the ANP R&D levy as “Compromisso de Investimentos com Pesquisa e Desenvolvimento”. The author received a CNPq scholarship process number 206985/2017-7.

Thanks to my supervisors and Shell E&P Brazil team for having designed the project that my PhD was inserted. I am grateful for having the opportunity to research simultaneously two challenging and exciting scientific topics (Statistics and Reservoir Engineering) while working with a supervisory team from a diverse background.

Thank you my supervisors: Guilherme Avansi, for uncountable hours of technical discussions with openness to disagreement and his example of good willingness to help; Camila Caiado, for empowering me to develop this project with autonomy and supporting through multiple bureaucracies involved in this programme; Michael Goldstein, for projecting me as a Bayesian statistician ☺ and persistently sharing his insights, often miles ahead, always offering enthusiasm and peace in any discussion; Prof. Denis Schiozer, for hardworking to manage the UNISIM group and incessantly challenging ideas and concepts; Ian Vernon, for the uncountable hours of collaborative discussions and being open to listen as much as to talk. Thanks to Dr Célio Maschio for promptly sharing knowledge in the beginning of this project.

I am grateful for the professionalism of the staff of the Durham University and the University of Campinas, particularly Michelle, Bruna, Rachel (dual-degree agreements), Daniel (IT support), Tonin (office), Adelle, David (career advisers); Eleanor Loughlin (Maths and Stats Lab); Bev, Rachel, Trish and Maths Office colleagues; colleagues in Unicamp and Durham who were there to help each other finding solutions and made pleasant this PhD journey.

I would like to thank the Examiners for their constructive comments and corrections which have contributed to a more precise and in-depth presentation and discussion: Peter Craig, Philip Jonathan, Hailiang Du, Denis Schiozer, Guilherme Palermo and André Fioravante, thank you.

Ærlighed i små ting er ikke nogen lille ting.

[Honesty in small things is not a small thing.]

Danish saw

ABSTRACT

This thesis proposes a systematic Bayesian approach for uncertainty quantification with an application for petroleum reservoirs. First, we demonstrated the potential of additional misfit functions based on specific events in reservoir management, to gain knowledge about reservoir behaviour and quality in probabilistic forecasting. Water breakthrough and productivity deviation were selected and provided insights of discontinuities in simulation data when compared to the use of traditional misfit functions (*e.g.* production rate, BHP) alone. Second, we designed and implemented a systematic methodology for uncertainty reduction combining reservoir simulation and emulation techniques under the Bayesian History Matching for Uncertainty Reduction (BHMUR) approach. Flexibility, repeatability and scalability are the main features of this high-level structure, incorporating innovations such as phases of evaluation and multiple emulation techniques. This workflow potentially turns the practice of BHMUR more standardised across applications. It was applied for a complex case study, with 26 uncertainties, outputs from 25 wells and 11+ years of historical data based on a hypothetical reality, resulting in the construction of 115 valid emulators and a small fraction of the original search space appropriately considered non-implausible by the end of the uncertainty reduction process. Third, we expanded methodologies for critical steps in the BHMUR practice: (1) extension of statistical formulation to two-class emulators; (2) efficient selection of a combination of outputs to emulate; (3) validation of emulators based on multiple criteria; and (4) accounting for systematic and random errors in observed data. Finally, a critical step in the BHMUR approach is the quantification of model discrepancy which accounts for imperfect models aiming to represent a real physical system. We proposed a methodology to quantify the model discrepancy originated from errors in target data that are set as boundary conditions in a numerical simulator. Its application demonstrated that model discrepancy is dependent on both time and location in the input space, which is a central finding to guide the BHMUR practice in case of studies based on real fields.

Keywords: Uncertainty Quantification, Petroleum Reservoirs, Bayesian Emulator, Numerical Simulation, Model Discrepancy, Bayesian History Matching.

RESUMO

Essa tese propõe uma abordagem Bayesiana sistemática para quantificação de incertezas de reservatórios de petróleo. No primeiro artigo, demonstramos o potencial de funções-objetivo adicionais que são baseadas em eventos específicos da fase de gerenciamento de reservatórios, a fim de melhorar a representação do comportamento do reservatório e a qualidade da previsão probabilística. Irrupção de água e desvio de produtividade foram selecionados, proporcionando um entendimento de descontinuidades no modelo numérico e nos dados de simulação quando comparado com o uso exclusivo de funções objetivo tradicionais (por exemplo, taxa de produção). No segundo artigo, definimos e implementados uma metodologia sistemática para redução de incertezas que combina simulação de reservatórios e técnicas de emulação em uma abordagem de Ajuste de Histórico Bayesiano para Redução de Incertezas (BHMUR, *Bayesian History Matching for Uncertainty Reduction*, acrônimo em inglês). Flexibilidade, repetitividade e escalabilidade são as características principais dessa estrutura geral que incorpora inovações tais como fases de avaliação e múltiplas técnicas de emulação. Esse procedimento potencialmente transforma a prática de BHMUR em uma mais padronizada para diversas aplicações. Aplicamos em um estudo de caso com 26 atributos incertos, dados de produção de 25 poços e 11+ anos de dados de histórico de produção baseado em uma realidade hipotética, resultando na construção de 115 emuladores validados e uma pequena fração do espaço de busca apropriadamente considerada não-implausível ao final do processo de redução de incertezas. No terceiro artigo, expandimos metodologias para estágios críticos na prática de BHMUR: (1) extensão da formulação estatística de BHMUR para acomodar emuladores do tipo classificadores; (2) seleção efetiva de uma combinação de dados de produção para emulação; (3) validação de emuladores baseados em múltiplos critérios; e (4) consideração de erros sistemáticos e aleatórios em dados observados. No último artigo, avaliamos um passo crítico para a prática de BHMUR, que é a quantificação de discrepância do modelo para contabilizar a representação de sistemas físicos a partir de modelos imperfeitos. Propusemos uma metodologia para quantificar a discrepância do modelo originada em erros de dados medidos e informados ao simulador numérico como condição de contorno (*target*). A aplicação da metodologia demonstrou que a discrepância do modelo é simultaneamente dependente de tempo e da posição no espaço de busca: uma descoberta importante para orientar o processo de quantificação de incertezas em estudos de caso baseados em reservatórios de petróleo reais.

Palavras Chave: Quantificação de Incertezas, Reservatório de Petróleo, Emulador Bayesiano, Simulação Numérica, Discrepância de Modelo, Ajuste de Histórico Bayesiano.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Problem Outline

Uncertainty Quantification (UQ) is an interdisciplinary rational science that provides a better understanding of complex systems by characterising, propagating and reducing uncertainties associated with a problem. In petroleum reservoir applications, performing an appropriate UQ process depends on the integration of scientific elements such as statistics and mathematical models, numerical models or simulators, physical system (*e.g.* the flow in the porous media and then flowing from the reservoir to surface), experimental data, expert judgement and multiple sources of model and reservoir uncertainties.

Similarly, exploring efficiently a hydrocarbon asset relies on a work of collaborative and integrative nature, to consolidate and share information within the organisations involved. Reservoir simulation models are central in exploration and production workflows: they incorporate data acquired, the technology available and an understanding and interpretation about sub-surface physics and information. The difficulty level to produce hydrocarbons depends on exploratory environment (onshore or offshore), formation and reservoir conditions, among others. Some decisions within the field life cycle are:

- Reserves evaluation, composing in the organisation competitiveness and share value;
- Field development plan, defining facilities, recovery and drilling plan and schedule;
- Reservoir management: reviewing model/grid and evaluating infill drilling options.

Most activities of the asset team aim to offer realistic and on-schedule forecasting with appropriate levels of uncertainty and ultimately provide a better basis for decision making. Schiozer *et al.* (2019) described a methodology for integrated decision analysis in the development and management of petroleum fields considering reservoir simulation, risk analysis, history matching, uncertainty reduction, representative models, and production strategy selection under uncertainties.

In practice, the complexity of uncertainty propagation in a reservoir life cycle often leads to simplifications of the processes, which may lead to an unrealistic consideration of the uncertainties in forecasting. In this context, and enabled by enhanced affordability of computational power, UQ gained a prominent position in technological innovation, tools and research in the energy industry, particularly examining topics such as:

(a) Collection of simulator evaluations consistent with historical observations and levels of uncertainty associated with the problem;

(b) Speed up the process to find this collection of scenarios; and

(c) Limitation of the use of numerical models that are imperfect representations of the real physical system.

This thesis encompasses three corresponding areas of research, namely Bayesian History Matching, emulation and model discrepancy. Combining reservoir sciences with uncertainty quantification defines an interdisciplinary field of study which aims to provide consistent production forecasting and enhance the decision-making process.

An extensive list of references is integrated in Chapter 7 of this thesis. Here, we highlight four schools of research that significantly contributed to the development of this thesis:

- Durham: In 1995, Craig, Smith, Goldstein, Seheult (Craig *et al.* 1995 and 1997) made significant progress in UQ area by formulating a technique that combines simulation models and emulators under a coherent statistical framework, referred to as Bayesian History Matching (BHM). Several fundamental and applied advances derived from this work, notably Vernon *et al.* (2018), Caiado and Goldstein. (2015) and Vernon *et al.* (2010).

- Sheffield: A sequence of remarkable works in the area of emulation, validation, and diagnostics is provided by O'Hagan (1996), Oakley (1999), O'Hagan (2004) and Bastos (2009).

- Campinas: Benchmarking cases publically available are applicable for works in reservoir simulation, history matching and uncertainty reduction (UNISIM webpage). Notably, a partnership with Durham enabled to apply BHM to a case study with uncertain parameters related to a channel and reservoir permeability (Ferreira *et al.* 2014).

- Bergen: More recently, Evensen and Eikrem (2019) started a discussion with the community about the impact of errors in observed data and models applied to reservoir simulation of petroleum fields, which supported the definition of the focus of the paper 4 of this thesis.

As mentioned above, a suitable literature is offered by the end of the thesis.

1.2 Justification

Despite these efforts and continued research from several recognised groups, the problem of using historical data to gain more understanding about uncertain attributes and to provide more reliable forecasting is still challenging in practice. The main challenges are:

- Find multiple sets of scenarios consistent with observed data and the uncertainties associated with the problem;

- Overcome bottlenecks related to people, investment, time and computational power;

- Scale techniques in high dimensional input and output spaces;

- Characterise model discrepancy and account for imperfect reservoir models.

This thesis aims to integrate solutions for these four challenges in a systematic way, proposing linked steps that turn the practice of Bayesian History Matching for Uncertainty Reduction (BHMUR) less manual and more standardised across applications. A more standardised approach is important to frame incremental innovation for specific steps, facilitate cross-discipline contributions and speed-up the steep learning curve for new practitioners of BHMUR.

1.3 Objectives

In order to progress this interdisciplinary research area, the main objective of this thesis is the formulation of a general and systematic strategy of uncertainty quantification for petroleum reservoirs, using BHMUR and emulation techniques.

In specific terms, the methodology is expected to incorporate:

- An understanding of data structure associated with the physical system, incorporating particularities of physical behaviour (*e.g.* production history in early and late stages) to boost the uncertainty quantification analysis;
- The construction of emulators; able to mimic the reservoir simulation behaviour at appropriate levels of uncertainty (*certainly* not over-confident, but also avoiding under-confident emulators);
- An application for a complex petroleum reservoir to validate the techniques proposed.

The combination of these elements in a unique strategy is expected to establish a robust BHMUR framework applicable to speed up one part of the process of uncertainty quantification of petroleum reservoirs.

1.4 Case Studies

The benchmark reservoir model UNISIM-I-H (Avansi *et al.* 2015) is the case study of the first paper, with the parameterisation proposed by Maschio and Schiozer (2016). For the second and third papers, a case study with hypothetical reality inspired in Avansi *et al.* (2015) and Maschio and Schiozer (2016) was defined through a sequence of steps and named as HR-82:

1. Select a geostatistical realisation, *e.g.* with a specific spatial distribution of porosity, horizontal and vertical permeability and facies;
2. Define the uncertain parameters, ranges and initial distribution (except the geo-realisation which was considered deterministic); this step defines the initial input or search space;
3. Select one scenario (*i.e.* a combination of the uncertain attributes) as the hypothetical reality, representing the values set in the physical system; the

selection of the hypothetical reality can be random or based on an experiment with simulations;

4. Run the hypothetical reality under operational conditions by setting, for example, minimal and maximal bottom-hole pressure of wells, platform constraints, well control conditions. This simulation results in data of the hypothetical reality without an observational error or model discrepancy;

5. Noise-up the hypothetical reality to mimic the observational error in the real process, which is the first step in the direction of a more realistic application and generate the target file in the required format;

6. Select a number of attributes to be considered uncertain, depending on the reservoir characterisation and project requirements (complexity required in the stage of development); *e.g.* the complete set of uncertainties has 82 attributes, and we deliberately considered 26 as uncertain in papers 2 and 3.

Hypothetical reality is part of the work proposal because it allows validating the complex methodology, providing maturity to applications in more complex settings of reservoir simulation. The main limitation is that a hypothetical reality allows us to consider models as a perfect representation of the physical system – a clear disadvantage since all models are imperfect representations of the physical system.

The work of article four was developed based on a simple case study, with one producer and one injection well, kindly provided by Maschio et al. (2018). A hypothetical reality was also defined to validate the proposed methodology.

1.5 Outline and Structure

Four scientific articles are laid out to accomplish the main objective of this thesis. The focus of the first article was on gaining an understanding of the case study, data analysis and tools available while providing to the research community a relevant contribution in the uncertainty analysis framework.

The second and third articles require some contextualisation. After a couple of months in Durham University working with BHMUR, our abstract submitted to Europec Conference 2019 was accepted. As an opportunity to speed up the project

and bring it back to schedule (multiple reasons lead to a project delay, including approximately one and a half year of postponement to start the period in Durham University), we decided to present the complete systematic procedure in this relevant conference. Even if it was long (45 pages) and dense work, we needed to extend Formentin *et al.* (2019-b) to capture relevant aspects of the methodology proposed. Moreover, shortly after submitting it to an SPE Journal, we received the suggestion to split the conference paper into two articles to comply with the regulations of the selected journal. We decided to divide Formentin *et al.* (2019-b) into two articles that capture the advancements of the methodology proposed and are in an appropriate format to journal publication. These two documents consist of articles two and three of this thesis (Chapters 3 and 4).

The second article proposed the high-level structure of the systematic procedure combining reservoir simulation, emulation techniques through the BHM approach. In the third article, statistical techniques were implemented for an appropriate data analysis and critical steps in the high-level structure presented in the second article were expanded. Finally, in the last article, we proposed a procedure to account for one source of model discrepancy associated with reservoir studies.

1.5.1 Article 1: Gaining more understanding about reservoir behaviour through assimilation of breakthrough time and productivity deviation in the history matching process

Helena Nandi Formentin, Forlan la Rosa Almeida, Guilherme Daniel Avansi, Célio Maschio, Denis José Schiozer, Camila Caiado, Ian Vernon, Michael Goldstein. Journal of Petroleum Science and Engineering, 2019, v. 173(1), p.1080-96

In this work, we evaluated the impact of the integration of additional objective functions in the process of uncertainty reduction. The additional objective functions quantitatively evaluate the mismatch in (1) water breakthrough time and (2) well productivity. We applied the Iterative Discrete Latin Hypercube (Maschio and Schiozer 2016), which is fully simulation-based method (*i.e.* do not implement emulation techniques).

Besides demonstrating the potential of additional objective functions to improve the results of uncertainty reduction processes, this work was relevant to the thesis for three main reasons:

1. Develop the programming skills required to produce a systematic procedure: from an understanding about reservoir parameterisation with multipliers until post-processing techniques to use the reservoir simulator data;

2. Identify boundary conditions defined in the simulator during the historical period and their impact in the history matching procedures. For example, from several discussions, we understood and agreed about the difference between operating pressure and limit pressures. This understanding was fundamental to, later – in articles two and three, identify the need of two-class emulators;

3. Observe that the water breakthrough time is relevant for the uncertainty reduction process. The understanding of the impact and how to capture this type of data was a facilitator when defining a binary output associated with the water breakthrough time, important for articles 2 and 3, where we applied two-class emulators.

In this way, article 1 provided both (a) scientific contribution to the community of reservoir engineering, and (b) deep-level of skills and an understanding of reservoir data and behaviour.

1.5.2 Article 2: Systematic uncertainty reduction for petroleum reservoirs combining reservoir simulation and Bayesian emulation techniques: Part I

Helena Nandi Formentin, Ian Vernon, Guilherme Daniel Avansi, Camila Caiado, Célio Maschio, Michael Goldstein, Denis José Schiozer

To be submitted.

We proposed a systematic procedure for uncertainty reduction in petroleum reservoirs. This workflow is the core of the thesis, responding to its main objectives. The systematic procedure provides a general strategy for uncertainty reduction in petroleum reservoirs. The sequence of steps logically linked advances the applicability of the BHMUR approach by offering scalability, repeatability and flexibility.

The proposed methodology with 20 linked steps was applied in a complex model with well-known behaviour, resulting in 115 valid emulators through 15 waves. The decision to use a hypothetical reality was an important approach to guarantee the correctness of the techniques implemented. The hypothetical reality remained as non-

implausible by the end process, *i.e.* coherent with the historical data and uncertainties in the problem, which is a relevant consistency check of the applied technique. We compared the forecasting obtained from initial scenarios and non-implausible scenarios by the end of the process, corroborating the potential of the high-level structure to reduce the uncertainty of complex systems, such as reservoir simulation models.

1.5.3 Article 3: Systematic uncertainty reduction for petroleum reservoirs combining reservoir simulation and Bayesian emulation techniques: Part II

Helena Nandi Formentin, Ian Vernon, Guilherme Daniel Avansi, Camila Caiado, Michael Goldstein, Denis José Schiozer, Carla Janaína Ferreira

To be submitted.

The third paper focuses on statistical analysis applied to the systematic procedure for uncertainty reduction. We advanced the applicability of BHMUR with four contributions:

1. The extension of the Bayesian History Matching technique to structures in the simulation data that are discontinuous across the search space and require two-class emulators that appropriately capture the behaviour evaluated (*e.g.* water breakthrough, production and injection targets);

2. A procedure for selection of outputs to emulate, especially relevant in high dimensional spaces with a large number of outputs;

3. A systematic combination of quality indicators for the validation of emulators and selection of valid emulators;

4. A framework for the estimation of observational errors.

Although this work is a stand-alone paper, *i.e.* the techniques proposed can be applied independently, we illustrate the four contributions of the third article with their application in the systematic procedure proposed in article two.

1.5.4 Article 4: Characterising Imperfect Models by Combining Numerical Simulation and Bayesian Emulation Techniques – An Application to Petroleum Reservoirs

Helena Nandi Formentin, Ian Vernon, Michael Goldstein, Camila Caiado, Guilherme Daniel Avansi, Denis José Schiozer

To be submitted.

One of the key steps in the BHMUR methodology is to identify and characterise the uncertainties originated from diverse sources, including the uncertainty of (a) measured data to represent the data from the physical reservoir; (b) the reservoir model to represent the physical system, named as model discrepancy.

This work aims to advance the applicability of BHMUR for case studies incorporating elements of the model discrepancy. The main objective is to develop the background and techniques aiding the future application of the BHMUR in a case study closer to a real case compared to the hypothetical reality.

We focus on exploring a source of model discrepancy, identified in articles 2 and 3. In these works, we considered as historical data the hypothetical reality noised with errors containing random and systematic portions. The choice to noise the hypothetical reality was a step in the direction to a more realistic case study. We expose that measurements in real reservoirs are susceptible to a sequence of processes that can give rise to observational errors.

While simulating in the historical period, we apply simulation targets. Targets are boundary conditions to the solution of differential equations implemented in the simulator. Random and systematic errors contained in target data generates an inadequacy in the simulation model, *i.e.* observational error propagates as a model discrepancy. For example, when liquid production targets are overestimated compared to the physical reservoir, a bias in the bottom hole pressure is expected.

We proposed a procedure to quantify the model discrepancy originated by errors in target data. We combined emulation and simulation techniques to demonstrate that model discrepancy is dependent on the time and location of the scenario in the input space and that it can be a significant source of model discrepancy.

In the next sections, the full version of the papers is presented, followed by conclusions and recommendation for future work.

2 ARTICLE 1: GAINING MORE UNDERSTANDING ABOUT RESERVOIR BEHAVIOR THROUGH ASSIMILATION OF BREAKTHROUGH TIME AND PRODUCTIVITY DEVIATION IN HISTORY MATCHING PROCESS

Helena Nandi Formentin, Forlan la Rosa Almeida, Guilherme Daniel Avansi, Célio Maschio, Denis José Schiozer, Camila Caiado, Ian Vernon, Michael Goldstein

Journal of Petroleum Science and Engineering, 2019, v. 173(1), p.1080-96

Contribution by the candidate Helena Nandi Formentin: Conceptualisation, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualisation, as described in the Statement of Contribution submitted to the Journal of Petroleum Science and Engineering (fully reproduced in Appendix B – Contribution Statements). The original IDLHC software in Matlab (Maschio and Schiozer, 2016) was made available by Célio Machio and Denis Schiozer and was the foundation for this paper. CMG was selected as reservoir simulation software.

Abstract – Article 1

History matching (HM) is an inverse problem where uncertainties in attributes are reduced by comparison with observed dynamic data. Typically, normalised misfit summarises dissimilarities between observed and simulation data. Especially for long-time series, objective functions (OFs) aggregate multiple events and tendencies relevant to field performance in a single indicator (*e.g.* water rate and breakthrough time). To capture the attributes influencing the reservoir behaviour, we evaluate the assimilation of data series through additional OFs, obtained from splitting time-series data. In this study, two additional OF groups supplement the time-series misfits: Breakthrough Deviation (BD) indicating dissimilarities in water breakthrough time; Productivity Deviation (PD), representing mismatches of the well potential, mainly impacting the transition from history to forecast conditions. The Productivity Deviation (PD) is adapted from previous studies. Instead of simulating the last time of the historical period under forecast conditions, we propose keeping it under historical data. The change is the historical data used as target condition to the simulator: Bottom Hole Pressure (BHP) in place of liquid production and water injection rates; with this, we estimate a mismatch in well productivity, while avoiding the influence of other boundary conditions in the evaluation. Two applications (1 & 2), assimilating different OF quantities, highlight the influence of the additional groups. Application 1 only computes time-series misfit (64 OFs) whereas Application 2 includes the BD and PD (counting 128 OFs). The iterative HM method presents flexibility regarding OFs assimilated and incorporation of uncertain attributes. UNISIM-I-H case allows us to evaluate the HM considering history and forecast data. We examine differences between the 450 scenarios resulting of data assimilation for each application through four perspectives. Application 2 resulted in scenarios with better predictability of the field behaviour and smoother transitions between field history and forecast periods. Field cumulative oil production of Application 2 is also forecasted closer to the reference data when compared to Application 1; all forecast periods (1, 5 and 19 years) emphasise this impact. Some wells presented breakthrough time closer to the reference for Application 2. The challenging achievement of exact BD matches leads to the third advantage of the additional indicators. These OFs supply supplementary information to the diagnosis of scenarios, identifying unnoticed problems in the traditional approach. Finally, even with overall better performance, some of the well OFs presented poorer matches for Application 2. To explain this, we analysed the relationship between attributes and the OFs used to update the attributes. In conclusion, the improved forecast of the simulation scenarios indicates that superior performance of the HM

process is possible by splitting the available dynamic data in relevant additional OFs. This study presents a case application with 11 years of field history, in which additional OFs, derived from dynamic data, add value to the reservoir characterisation. They allow the influence of uncertain attributes to be captured for relevant events in reservoir performance. We also show how the increased quantity of OFs assimilated makes the HM harder for some OFs.

Keywords: History Matching; Iterative Discrete Latin Hypercube methodology; Breakthrough Time; Well Productivity; Reservoir Characterisation; Transition between Historical and Forecast periods.

2.1 Introduction

Reservoir simulation models are representations of real petroleum fields used in production forecast and decision-making process. Closed-Loop Reservoir Development and Management (CLRDM) endorses the application of simulation techniques in all stages of the field lifetime. CLRDM methodologies (Jansen *et al.* 2009; Wang *et al.* 2009; Schiozer *et al.* 2015) integrate model-based optimisation and data assimilation to support decisions about the physical problem with uncertainties. Silva *et al.* (2017) propose a closed-loop workflow constructed with an ensemble-based method. They demonstrate the effectiveness of CLRM to improve the predictability of the models, in contrast to ensemble-based separated applications.

Data assimilation is a stage in the CLRDM known as History Matching (HM) in the petroleum industry. It uses the observed dynamic data to afford a better representation and predictability of the physical model through simulation models. The HM is an inverse problem with multiple possible solutions. The complexity to solve the problem increases with dimensionality in terms of the number of inputs and outputs.

A wide undestanding of the inverse theory and history matching, including explanatory examples, is available in the book of Oliver *et al.* (2008). Oliver and Chen (2011) discuss the progress of diverse HM processes in their paper, detailing the advantages and disadvantages of manual, evolutionary, Ensemble Kalman Filter based and Adjoint methods. Rwechungura *et al.* (2011) summarises the evolution of HM techniques through the time and highlights aspects to the integration of 4D seismic. Maschio and Schiozer (2016) offer a more recent overview of HM methods, classifying optimisation, probabilistic and mixed methods.

In the HM process, parameters of the reservoir characterisation, which are inputs into the reservoir numerical model, are uncertain and represent undetermined reservoir features

(fault transmissibility, for instance). These uncertainties in the attributes influence dynamic production estimated by the simulator and the asset team used this data to understand flow and transport in the real petroleum field. The closer the simulator output is to the dynamic data measured in the field (production rate in a specific period, for example), the better we expect that the model represents the physical field. In this context, objective functions (OFs) computes the difference between observed and simulation data.

A reservoir analysis based on a deterministic approach considers one or more scenarios that represent a partial set of the possible production scenarios. Nevertheless, this approach can present biased results since it generates production forecasts without adequately exploring the range of production scenarios (Goodwin 2015). In contrast, the probabilistic approach represents the uncertainty toward the reservoir behaviour. It supports reliable forecast by addressing questions of risk and uncertainty in reservoir management. This approach incorporates the consideration of several sources of uncertainties involved in the reservoir characterisation process and measurement errors in observed data (Maschio and Schiozer 2017).

Some probabilistic methods allow the redefinition of the probability distribution based on the OFs misfit, improving the reservoir knowledge in terms of reservoir characterisation. An example of a methodology with this characteristic is the Iterative Discrete Latin Hypercube (IDLHC), a method developed by Maschio and Schiozer (2016). The IDLHC is an automated probabilistic method to reduce uncertainty and update the probability of the uncertain attributes with nonparametric density estimation. The process consists of applying a correlation matrix to automatically identify relationships between uncertain attributes and OFs. Due to its flexibility in terms of quantity of uncertain attributes and OFs assimilated, it can be adapted to several scenarios of reservoir characterisation and information available.

In order to offer a broader understanding and representation of the reservoir model, multi-objective and probabilistic HM processes have been employed. These processes simultaneously evaluate the reservoir behaviour through multiple quality indicators associated with observed data in the production and injection wells (Almeida *et al.* 2014, Kam *et al.* 2017). Hutahaean *et al.* (2015) showed that an ensemble of matched scenarios from multi-objective HM provides a more diverse set of matched-scenarios, which leads to a better comprehension of the forecast behaviour.

Nevertheless, since multi-objective-HM performance (convergence speed and match quality) can deteriorate under an increasing number of objective functions, Hutahaean *et al.* (2017) investigate the selection of objective grouping for multi-objective HM. Min *et al.* (2014) developed an evolutionary procedure to overcome inefficiencies of multiple-objective constraints by introducing preference-ordering and successive objective reduction to the conventional multi-objective optimisation module.

Several studies evaluate the influence of the OF definition in the HM process. For example, Tillier *et al.* (2013) focused in defining a formulation for incorporating seismic data in the process; Bertolini and Schiozer (2011) compared eight global OFs in the history matching process by assessing the matching quality of synthetic reservoir model.

A normalised misfit called Normalized Quadratic Deviation with Sign computes the difference between simulated and observed data (Avansi *et al.* 2016). This OF summarises time-series curves for a scenario (Figure 2.1-a) in a single indicator (Figure 2.1-b) and is useful in probabilistic and multi-objective HM approaches (more details in the NQDS section). An acceptance range [$-\gamma$, $+\gamma$] supports the classification of the scenarios taking into account the sources of errors considered (*e.g.* noise in the history data, measurement errors, level of fidelity of the reservoir simulation model).

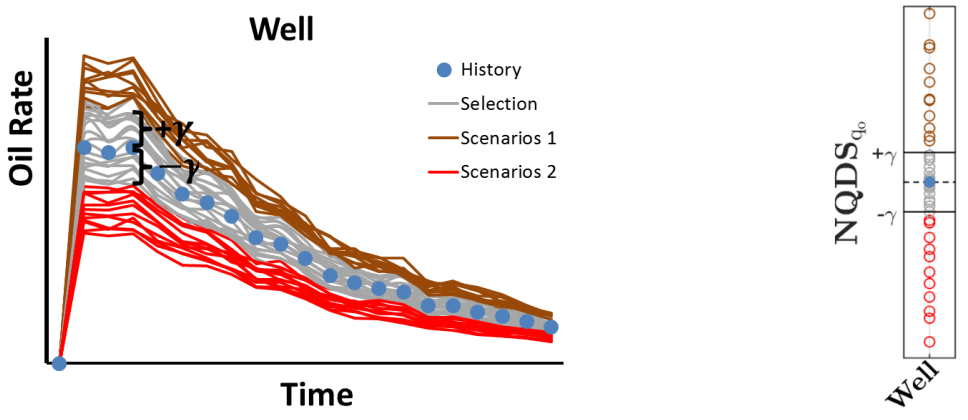


Figure 2.1. Typical NQDS graphic summarising data from several scenarios: (a) Curves of oil production rate plotted against time, adapted from Avansi *et al.* 2016: History data (blue points), selected scenarios that are within an acceptance range ±γ (in grey lines), scenarios with production rates higher and lower than the acceptance range (in brown and red lines respectively); (b) NQDS plot applying the same legend colours, where each dot corresponds to a production rate curve.

Due to the high quantity of observed data, especially for long time series, these OFs aggregate into a single indicator, events and temporary trends relevant to reservoir performance. For example, water breakthrough time and changes in the Gas-Oil Rate (GOR) are relevant for the field management; well production trends evolve over time under distinct reservoir

conditions (*e.g.* recovery mechanism from natural flow to water/gas injection to pressure maintenance). Different uncertain attributes can influence these events and temporary trends. Once aggregated in a single OF, the relationship between uncertain attributes and OFs may be difficult to capture with mathematical structures as a correlation matrix.

Splitting the conventional NQDS into more objective functions is an alternative approach to better understand the reservoir from the dynamic data available. Almeida *et al.* (2018) presented an introductory study with the application of unconventional OFs to measure the deviation of specific events (Breakthrough Deviation and Productivity Deviation). Each additional OF captures specific well behaviours (not mapped by the conventional OFs) that are influenced by distinct uncertain attributes. Then, the uncertain attributes update process considers the constraints established by both conventional and unconventional OFs. Because of this, the relationships identified between the unconventional OFs and uncertain attributes improved the reservoir calibration and uncertainty reduction process.

2.1.1 Objectives

This paper aims to evaluate the assimilation of dynamic data series in a way to capture deviations in the breakthrough time and the well productivity. With that, we aim to assess the possibility of gathering more information from available dynamic data series in the HM process, which improves the reservoir behaviour predictability.

When compared to the definitions of Almeida *et al.* 2018, we propose a distinct way to simulate the scenarios to better capture the physics that surround the well productivity. The proposed computation of Productivity Deviation avoids the influence of other sources of information, such as platform and well capacities, required in the previous work of Almeida *et al.* 2018. Moreover, this study assesses the additional OFs as a source of information to reveal reservoir behaviour, not explored in previous works.

We adapt a history matching methodology (IDLHC from Maschio and Schiozer, 2016) to consider the additional groups of Objective Functions for updating the uncertain attributes and use the same parameterisation presented in that paper. Maschio and Schiozer 2016 and 2018 tested the IDLHC methodology and compared it to other methodologies, assuring the quality of the history matching procedure.

2.2 Theoretical background

After describing the main aspects of the probabilistic HM methodology, this section details the objective functions applied to this proposed work.

2.2.1 Iterative Discrete Latin Hypercube (IDLHC)

The main advantage of the probabilistic IDLHC methodology proposed by Maschio and Schiozer (2016) is to simultaneously assimilate a large number of OFs to update probability distributions of uncertain attributes. Additionally, the process is flexible in terms of quantity of uncertain attributes and OFs assimilated, being adapted to several scenarios of reservoir characterisation and information available. This HM process generates multiple history-matched scenarios per iteration and the last set of scenarios is useful for prediction and optimisation studies. In the IDLHC general workflow (Figure 2.2), the uncertain attributes parameterised at the beginning of the process (*STEP 2*) are the same until the last pre-defined iteration ($Iter_{max}$). In each iteration, a set of scenarios representing the distribution of uncertain attributes is generated with Discrete Latin Hypercube (DLHC) sampling (*STEP 3*) conceived by Schiozer *et al.* (2017).

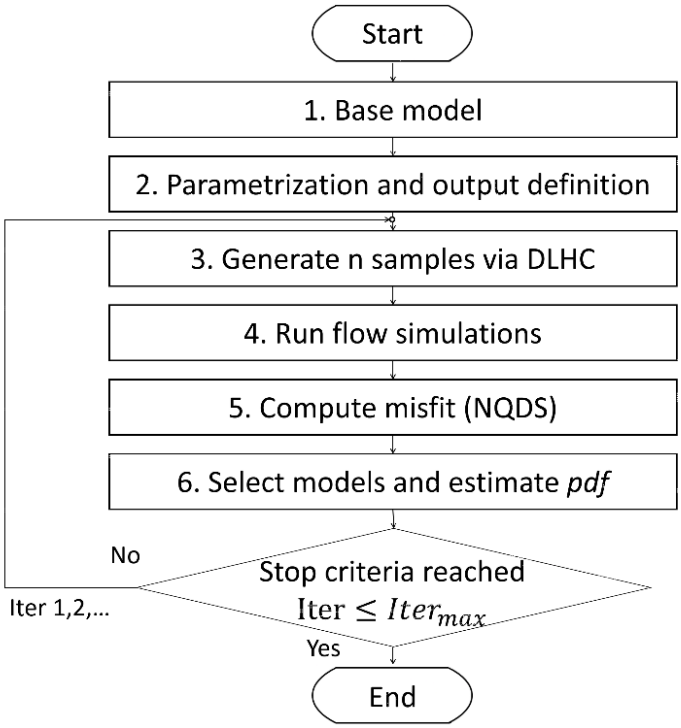


Figure 2.2. General workflow for probabilistic history matching (Maschio and Schiozer, 2016).

After running these scenarios in the flow simulator (*STEP 4*), NQDS computation quantifies the misfit between scenarios and observed data for each scenario and objective function (*STEP 5*). In *STEP 6*, selected scenarios are used for the generation of the posterior distribution for each uncertain attribute (*e.g.* posterior distribution in the sense that these are the distribution after the assimilation of the observed data in a given iteration). Maschio and Schiozer (2016) proposed three approaches to update the probability density function (*pdf)* of the uncertain attributes. Figure 2.3 details method 3, chosen for this study.

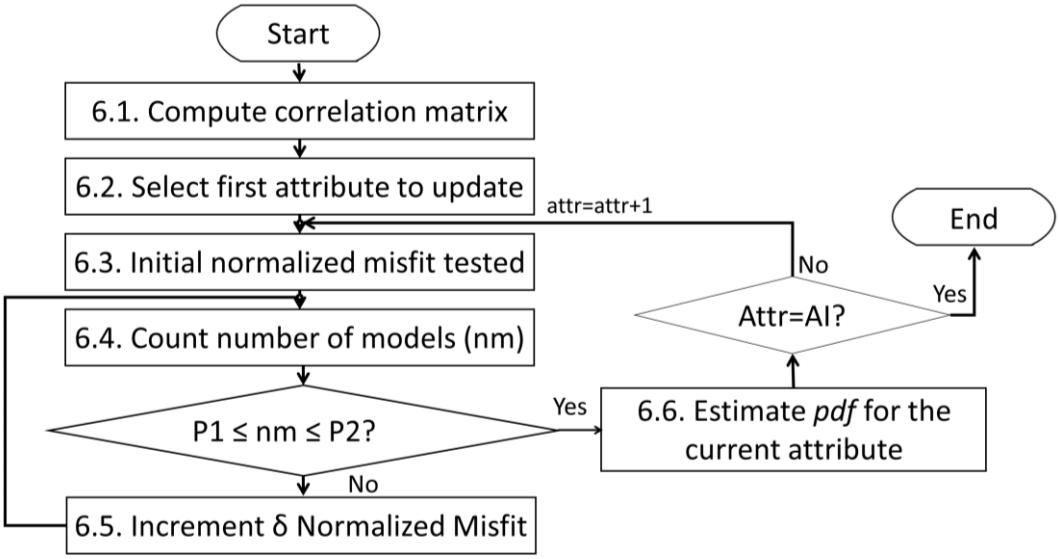


Figure 2.3. Flow chart from scenario selection, method 3 (Maschio and Schiozer, 2016).

A cut-off (R_c) applied to the coefficients of the correlation matrix (*STEP 6.1*) indicates the existence of a relationship between uncertain attributes and objective functions. The *AI* attributes considered correlated to at least one OF are updated. The updating routine starts in *STEP 6.2* with the first attribute to update, continuing until the last attribute (*AI*). The iterative process around *STEPS 6.4* to *6.5* guarantees two requirements: (a) a quantity of scenarios between a minimum (*P1*) and a maximum (*P2*) percentage of the scenarios sampled to avoid the collapse of the *pdf*, and (b) the selection of scenarios with smallest computed misfit.

Then, a nonparametric density estimation technique (*STEP 6.6*) leads to updating of uncertain attributes generating histograms representing the posterior distribution of each attribute. These posterior distributions are the prior distributions for the next iteration (*e.g.* prior in the sense that these are the distributions before the assimilation of the observed data in a given iteration). The iterative process of Figure 2.2 continues for the number of iterations predefined ($Iter_{max}$).

2.2.2 Normalised misfit as indicators of HM quality

In history-matching processes, indicators of quality for a scenario quantify the misfit between the simulation scenario and observed data. Four possible applications are to:

- Conduct the HM process, as objective functions to be minimised;
- Provide data to update the uncertain attributes;
- Diagnose scenarios revealing and guiding the review of reservoir characterisation;
- Support the evaluation of performance when comparing different methodologies.

We detail the two out of three normalised misfit groups applied in *STEP 5* of the HM methodology (Figure 2.2): NQDS and $NQDS_{BD}$ (NQDS of Breakthrough Deviation). In the methodology section, we present the third normalised misfit group $NQDS_{PD}$ (NQDS of Productivity Deviation) because it is subject of modification from previous work.

2.2.2.1 NQDS

NQDS (Avansi *et al.* 2016, modified) consolidates the misfit between history and temporal data series of production and injection wells. For example, $NQDS_{qw}$-Well 1 represents the misfit of water rate production for the Well 1 considering a time interval simulated for a given scenario. Similar notation applied to other data series, for example, oil production rate ($NQDS_{qo}$), production BHP ($NQDS_{ppbh}$), water injection rate ($NQDS_{iw}$) and injector BHP ($NQDS_{pibh}$).

Equation 2.1 computes the NQDS:

$$NQDS = \frac{\left(\sum_{j=1}^{n} Sim_j - Obs_j\right)}{\left|\sum_{j=1}^{n} Sim_j - Obs_j\right|} * \frac{\sum_{j=1}^{n}(Sim_j - Obs_j)^2}{\sum_{j=1}^{n}(Tol * Obs_j + C)^2} \quad (2.1)$$

where Sim_j and Obs_j are the simulated and observed (historical) data measured at the time j. Tol is the tolerance value (%) defined by the user for each data series; C is a constant used to avoid null or excessively small divisor, in case the production rate is close to zero (for example, water production rate in a recently opened well). Physically, the constant C represents the minimal tolerance for a given data series.

2.2.2.2 Water Breakthrough Deviation ($NQDS_{BD}$)

Water breakthrough is the time when water first reaches the production well. In the field management, this measured time and subsequent Water-Oil Ratio (WOR) trends are usually key performance indicators that also can be indicative of channelling and bypassing problems in the field (Baker, 1998).

The historical data of water production in wells is source of two-combined information: (a) water production rate through time, and (b) breakthrough time. In this sense, Almeida *et al.* (2018) adapted the NQDS as a punctual normalised misfit for breakthrough time (Equation 2.2), the $NQDS_{BD}$:

$$NQDS_{BD} = \frac{(BT_{sim} - BT_{obs})}{|BT_{sim} - BT_{obs}|} * \frac{(BT_{sim} - BT_{obs})^2}{(AE)^2} \quad (2.2)$$

where BT is the Breakthrough Time, and *AE* is the Acceptable Tolerance, for example, the maximal time between two consecutive measures of water production. A water rate cut-off to consider water breakthrough time avoids erroneous capture of breakthrough time: smaller water production rates when compared to this cut-off value are treated as residual water production. Even if the water breakthrough has not yet occurred in a given well at the historical period, it may add information to the HM process if some simulation scenarios have earlier breakthrough time.

Figure 2.4-a exemplifies water production against time for history data and some scenarios. The grey lines represent scenarios with production rate and breakthrough time within the acceptance range [$-\gamma$, $+\gamma$]. Scenarios 1 and 2 (brown and red lines) have early and late breakthrough time, respectively. Dashed and solid lines correspond to scenarios with matched and non-matched water production rates. The diagnostic of the $NQDS_{qw}$ plot (Figure 2.4-b) only identifies mismatches in the water production rate, keeping the two dashed scenarios within the acceptance range. On the other hand, the $NQDS_{BD}$ plot (Figure 2.4-c) identifies the difference between water breakthrough time for Scenarios 1 and 2. In this graph, two scenarios superpose in the extreme values of $NQDS_{BD}$ because the breakthrough time is identical for dashed and solid lines.

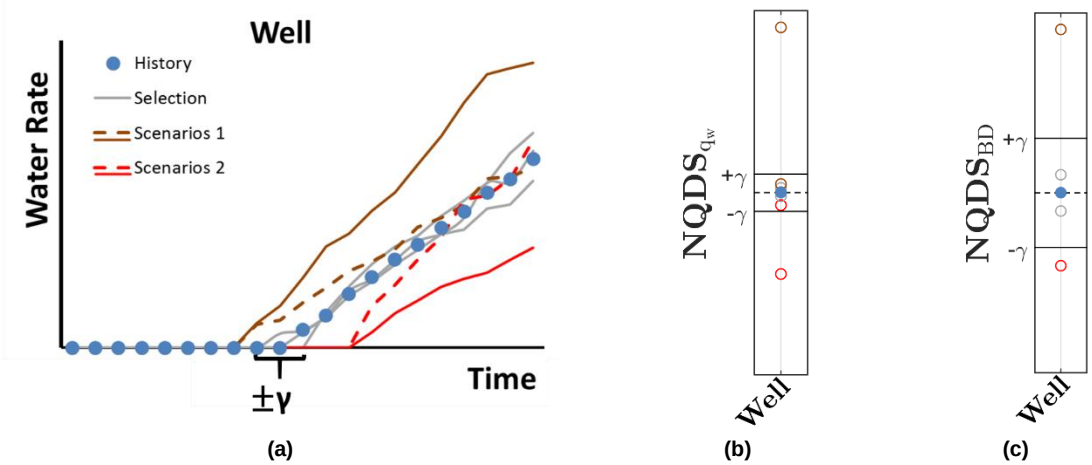


Figure 2.4. Breakthrough Deviation illustration - (a) Water production rate series for history data and several scenarios exemplifying differences between the information relative to water production rate and breakthrough time; (b) NQDSq_w plot summarising the production curves for the scenarios; (c) NQDS$_{BD}$ highlighting the mismatch in water breakthrough time for the scenarios.

2.3 Methodology: Productivity Deviation, case study, applications and assumptions

2.3.1 Productivity Deviation (NQDS$_{PD}$)

The transition between history and forecast period can cause fluid rate and bottom-hole-pressure fluctuations (Ranjan *et al.* 2014). In fact, at this point, the controls of the simulation scenario (boundary conditions) change: in the history period, liquid or oil production rates are treated as targets; during the forecast period, production restrictions are established (for example, minimal and maximal bottom-hole-pressure for producers and injectors and platform capacity). A possible cause of unconditioned reservoir scenarios is uncertain parameters, which can be wrongly defined or missing during the parameterisation.

As large fluctuations in the transition indicate non-realistic forecasted production rates, Almeida *et al.* (2018) defined an indicator related to the productivity of the well. The normalised misfit of Productivity Deviation (NQDS$_{PD}$) splits the historical dynamic data from wells into two parts simulated differently: (a) history controls, (b) forecast controls. This original implementation of the NQDS$_{PD}$ follows the simulation scenario by changing the control of the last history date from history control to forecast control.

In practical terms, history conditions usually include a target for liquid or oil production rate for the producer wells and forecast conditions apply operational conditions as minimal pressure for producers. Additionally, the simulation of the scenarios in the history period is not conditioned by platform and well restrictions, which is indispensable to perform the forecast simulation.

Two possible limitations may arise from the use of operational conditions to simulate the history period (as presented by Almeida *et al.*, 2018). Firstly, coupling operational conditions in the reservoir simulation requires information that may be uncertain, for example, a description of the multiphase flow in wells. Secondly, applying multiple restrictions simultaneously (*e.g.* well and platform capacities) potentially limit the identification of productivity mismatch.

Therefore, we propose an adaptation to the condition given to the last time step of the history from the one presented by Almeida *et al.* (2018). The measured BHP in the wells are the targets for production and injection wells, meaning that we change the data informed to the simulator. In this way, we limit the informed boundary condition to measured history data. This implementation of the PD indicator is generalisable and independent of other sources of data.

The modifications, in the last time step, of the simulation file are: (a) to reset non-restrictive maximal liquid production and injection for the wells (instead of non-restrictive maximal and minimal pressure applied to previous time steps, *i.e.* all-time steps except the last one); and (b) to inform the registered pressure for each well as new target condition (instead of informing well rates applied to the previous time steps).

Figure 2.5-a exemplifies, for a given producer well, the deviation for BHP informing the history pressure in the last time *t* of history. It illustrates most of the scenarios converging the target BHP condition because (1) liquid rate (Figure 2.5-b) has no production limit (q_{lmin}=0) and (2) a virtual maximal liquid rate is used to avoid simulation errors (q_{lmax}>>q_l).

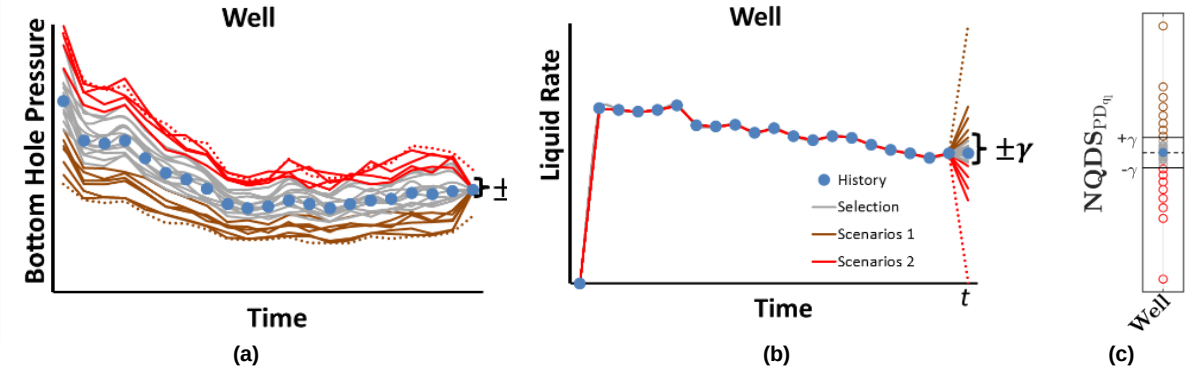


Figure 2.5. Productivity Deviation illustration - (a) BHP being informed only in the last time step of the history period; (b) Liquid production rate informed for all time steps except the last time steps, where non-restrictive conditions are reset; (c) Indicator of Productivity Deviation for liquid production.

The calculation of the productivity deviation applies to both production wells (*e.g.* for liquid rate - $NQDS_{PDql}$ - and BHP - $NQDS_{PDppbh}$) and injection wells (*e.g.* water rate - $NQDS_{PDiw}$ - and BHP - $NQDS_{PDpibh}$). Equation 2.3 computes the $NQDS_{PD}$:

$$NQDS_{PD} = \frac{(Sim_t - Obs_t)}{|Sim_t - Obs_t|} * \frac{(Sim_t - Obs_t)^2}{(tol * Obs_t + C)^2} \quad (2.3)$$

where Obs_t and Sim_t indicate the observed and simulation value in the last time (t) of the history data.

The $NQDS_{PDql}$ plot (Figure 2.5-c) indicates the deviation of simulated scenarios compared to the reference data. We consider that the scenarios in grey better present well productivity. Therefore, we expect that scenarios with smaller PD will provide better production predictions.

2.3.2 Case study

We applied the IDLHC methodology (Figure 2.2) in the UNISIM-I-H reservoir model (Avansi and Schiozer, 2015). This benchmark case is based on real data from the Namorado Field, a marine offshore turbidite reservoir in the Campos Basin – Brazil.

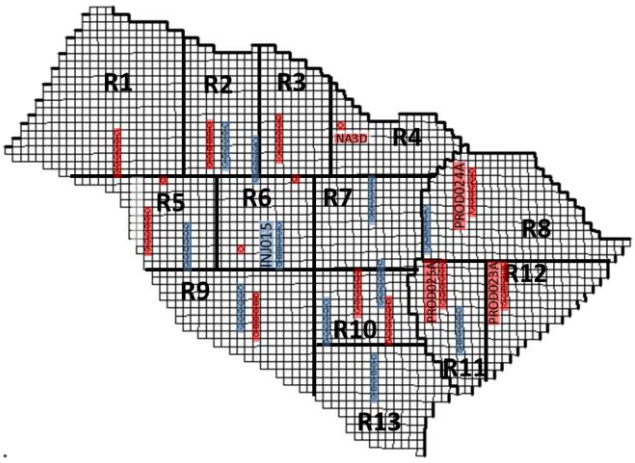


Figure 2.6. Bi-dimensional *x-y* view of the UNISIM-I-H with the position of the 13 regions defined by Maschio and Schiozer (2016). The production strategy contains 14 production wells (in red) and 11 injection wells (in green). Wells analysed in detail in the *Results and Discussion* section are identified: INJ015, NA3D, PROD025A, PROD023A and PROD024A.

The model UNISIM-I- H (Figure 2.6) has a production strategy with 14 producer wells and 11 injection wells and a production history of 11 years (4 018 days) available. The production forecast data for 19 years allows for the evaluation of methodologies in terms of predictability of the scenarios.

2.3.2.1 Initial parameterisation

The parameterisation defined in *STEP 2* (Figure 2.2) has 39 uncertain parameters as defined by Maschio and Schiozer (2016). Figure 2.6 retakes the 13 regions defined according to producer/injector pairs, attempting to capture the main drainage areas. Each region has multipliers of porosity (*mpor*), horizontal permeability (*mkx*) and vertical permeability (*mkz*). Isotropic permeability is taken for *x* and *y*-direction; initial *pdf* has uniform distribution for all levels. Table 2.1 summarises these uncertainties.

Table 2.1 - Uncertain attributes presented by Maschio and Schiozer (2016).

Uncertain attributes (for each region)	Minimum	Maximum	Number of levels	Initial pdf
Mpor	0.8	1.2	30	Uniform
Mkx	0.1	5.0	30	Uniform
Mkz	0.1	5.0	30	Uniform

2.3.3 Applications

Two applications performed in this study compute different groups of OFs:

- Application 1: 64 OFs, groups of $NQDS_{qo}$, $NQDS_{qw}$, $NQDS_{ppbh}$, for producer wells and $NQDS_{iw}$, $NQDS_{pibh}$ for injector wells;
- Application 2: 128 OFs resulting from adding the 64 OFs of Application 1, plus the additional OF groups ($NQDS_{BD}$, $NQDS_{PDql}$, $NQDS_{PDppbh}$, $NQDS_{PDiw}$ and $NQDS_{PDpibh}$).

In the *Results and Discussion* section, we compare their results for the field and wells in the history and forecast period.

2.3.4 Assumptions

Table 2.2 summarises the constants and tolerances for each OF applied in the calculation of the normalised misfit. Like Avansi *et al.* (2016), we defined 5% for controlled-data series ($NQDS_{iw}$); 10% for data series that are dependent on other series ($NQDS_{qo}$ and $NQDS_{qw}$, which are related to liquid rate, a target in the history period). Pressure related NQDS considers a tolerance of 5%. We applied a constant of 10 m^3/day for $NQDS_{qw}$ to moderate its impact on wells with low water rate production through a representative part of the history period. For example, the well NA3D production (Figure 2.7) reaches a maximum of 150 m^3/day and for this production, the tolerance adds up to 10+0.10*150=25 m^3/day. Higher constant would imply in smaller influence of the variations in q_w of this well in the updating process.

Table 2.2 - Constants used to calculate normalised misfit.

OF	C (unit of the variable)	Tol (%)
$NQDS_{qo}$	0	10
$NQDS_{qw}$	10	10
$NQDS_{ppbh}$	0	5
$NQDS_{iw}$	0	5
$NQDS_{pibh}$	0	5
$NQDS_{BD}$	AE=31	0
$NQDS_{PDql}$	10	10
$NQDS_{PDpbh}$	0	5
$NQDS_{PDiw}$	0	10
$NQDS_{PDpibh}$	0	5

$NQDS_{BD}$ has an AE of 31 days, the maximum interval between measurements. Productivity deviation is under forecast controls and under uncontrolled conditions. Therefore, we chose a tolerance of 10% for $NQDS_{PDql}$ and $NQDS_{PDiw}$, defining a minimal tolerance of 10 m^3/day for liquid production.

The cut-off applied to consider water breakthrough is 1 m^3/day for all the producers, except for NA3D with 6 m^3/day. Figure 2.7 shows the observed water production rate for this well, highlighting the portion of water rate considered residual. Applying 1 m^3/day cut-off for this well would mean to consider the breakthrough time of 669 days, which does not correspond to the effective breakthrough time of 3,226 days.

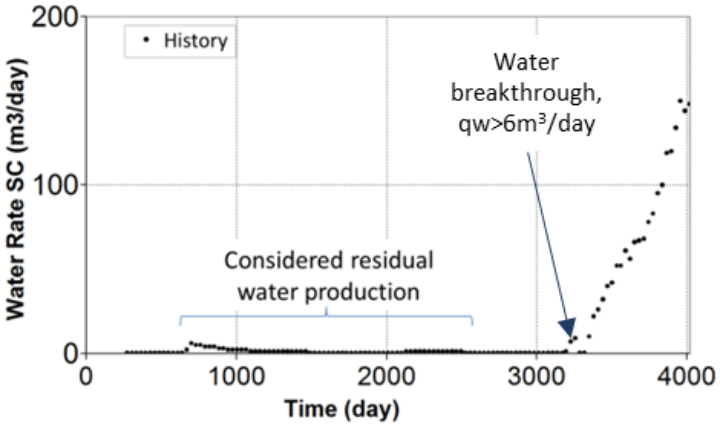


Figure 2.7. Water production rate for well NA3D in the history period.

Considering the recommendations proposed by Maschio and Schiozer (2016), the applications consider:

- 450 simulation scenarios per iteration in *STEPS 3* and *4*;
- A cut-off R_c=0.3 to the coefficients of the correlation matrix in *STEP 6.1*;
- An increment of the normalised misfit δ=1 in *STEP 6.5*;

- A minimum *P1*=5% and a maximum *P2*=15% of scenarios sampled to update the attributes;
- A maximal number of iterations $Iter_{max}$=8, set at the beginning of the process.

Moreover, to guarantee the reproducibility of the applications, the first run of the applications uses the same seed generated with Mersenne Twister (a random number generation algorithm).

2.4 Results and Discussions

To evaluate the assimilation of dynamic data series breaking down the conventional NQDS into more objective functions, we firstly exposed their impact with an overview of the indicators for the wells together with the field behaviour. Then, examples of additional OFs of some wells were used to complement the discussion. We decided on that approach because details for each of the 128 OFs individually were not feasible, with multiple relationships between OF and uncertain attributes.

The plots presented in this section consider the 450 scenarios of the 8th iteration in the HM process. In order to promote a clear visualisation of the impact in the forecast period and avoid fluctuations from changing boundary conditions, these final scenarios were simulated again with liquid production and water injection rate as a target during all the history period and the same operation conditions of the reference case in the forecast period.

2.4.1 History Period

The compilation of the results for the OFs allows for a broader evaluation of the general behaviour of the wells resulting from the implementation of the additional OFs. Figure 2.8 presents graphics for several OFs groups plotting the number of scenarios against the NQD[1] interval, from zero to the *x*-axis value. The higher the proportion of scenarios for a given NQD interval, the better. The *x*-axis is in logarithmic scale.

The assimilation of additional OFs (Application 2) reduces the mismatch of the OFs groups that have higher NQD values in Application 1 ($NQDS_{PDql}$ and $NQDS_{PDiw}$, Figure 2.8-a and -b). In contrast, the increased complexity of the HM through the assimilation of additional

[1] NQD (Normalized Quadratic Deviation) is the absolute value of the NQDS.

OFs leads to increasing the NQD values of traditional OF groups, exemplified by $NQDS_{qo}$ (Figure 2.8-c).

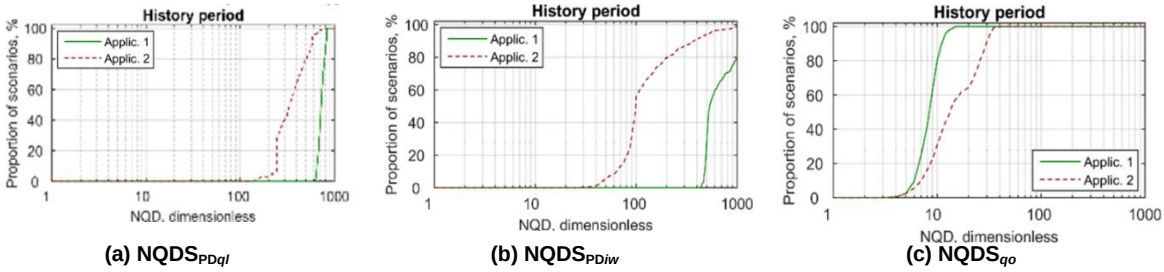


Figure 2.8. Proportion of scenarios against the NQD interval for OFs groups, semi-logarithmic scale: (a) $NQDS_{PDql}$ for 14 production wells; (b) $NQDS_{PDiw}$ for 11 injection wells; (c) $NQDS_{qo}$ for 14 production wells. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting in the traditional and proposed ones.

This analysis indicated that a comparison based only on the history period is insufficient. Therefore, in the next sections, we explore forecast data available for the benchmarking case.

2.4.2 Transition from history to forecast period

During the history period, the water injection rate is a target for the injection wells in the simulation. We expect scenarios very close to the reference data in this period. Nevertheless, the transition to the forecast period (Figure 2.9-a) shows fluctuations in the field rate when compared to the reference data. Application 2, including the additional OFs (in brown), provides less fluctuations and smother transition than Application 1.

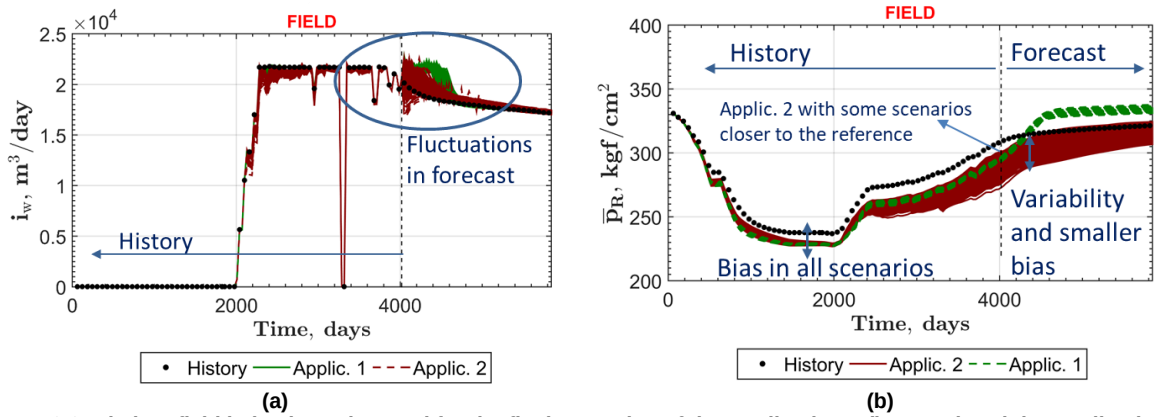


Figure 2.9. Distinct field behaviour observed for the final scenarios of the Application 1 (in green) and the Application 2 (in brown) including the history period (4 018 days) added to 5 years of production forecast: (a) Field water injection rate with smaller fluctuation in the transition for the final scenarios of the application that considers additional OFs; (b) Reservoir average pressure with a bias for both application in most of the history period, but Application 2 scenarios with better forecast and larger variability. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting of the traditional and proposed ones.

The average reservoir pressure (Figure 2.9-b) presents a bias for both applications in most of the history period: all the scenarios have reservoir pressure below the reference, and limited variability is observed. This is related to the fact that the initial liquid volume in place (oil and water) of the scenarios are smaller than the reference model (between 87-92% and 88-97% for Applications 1 and 2, respectively). Some scenarios of Application 2 are closer to the reference pressure at the end of the history period and it is closer to the reference in the 5-year forecast period (5 843 days of production). Note that the reservoir (and well) pressure is above the bubble point (around 210.03 kgf/cm^2), justifying the omission of the OFs related to gas production rate.

These results indicate that adding the OF groups related to Productivity Deviation and Breakthrough Deviation has the potential to limit oscillatory behaviour and improve the transition between history and forecast periods.

2.4.3 Forecast period

We use risk curves to evaluate the field forecast behaviour (Figure 2.10). In these curves, the cumulative oil production is plotted with the cumulative relative frequency observed in the 450 scenarios. Further than the two applications, we also plot the cumulative oil production for the first iteration (in grey) where all the uncertain attributes are in uniform prior distribution and the value of the reference model (black dotted line).

The three forecast period selected (one, five, and 19 years) show more scenarios closer to the reference value for Application 2. These graphs support that the inclusion of the new OFs has the potential to positively influence the predictability of field behaviour.

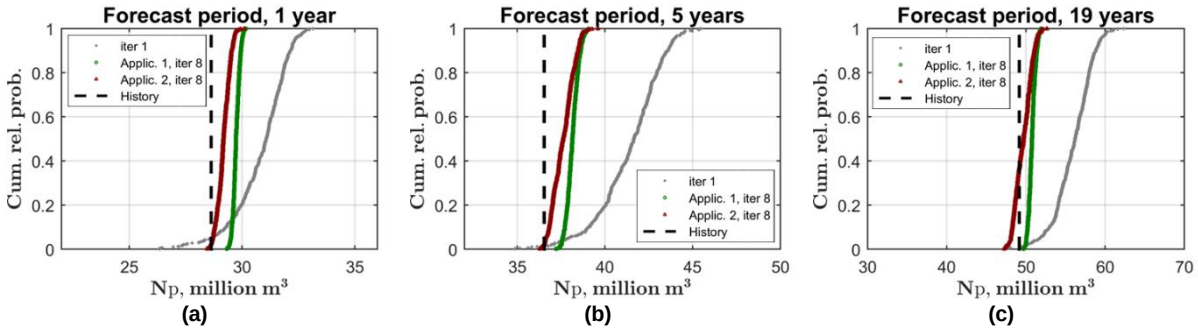


Figure 2.10. Forecast period, risk curves for the scenarios of iteration 1 and iteration 8 for the two Applications for: (a) 1 year; (b) 5 years and (c) 19 years. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting in the traditional and proposed ones.

In the next sections, some OFs illustrate the results in terms of well behaviour, individually.

2.4.4 Breakthrough Deviation

The assimilation of $NQDS_{BD}$ in Application 2 leads to the improvement of the breakthrough time of the scenarios for most wells. From the analysis of the importance of the OFs groups assimilated in the application (Appendix 2.A), Breakthrough Deviation was the additional OF group that contributed the most in the Application 2. Figure 2.11 shows the water production rate, $NQDS_{qw}$ and $NQDS_{BD}$ for the well PROD024A. Application 2 presents smaller breakthrough deviation than Application 1. In addition, the water rate of Application 2 is closer to the reference when compared to Application 1.

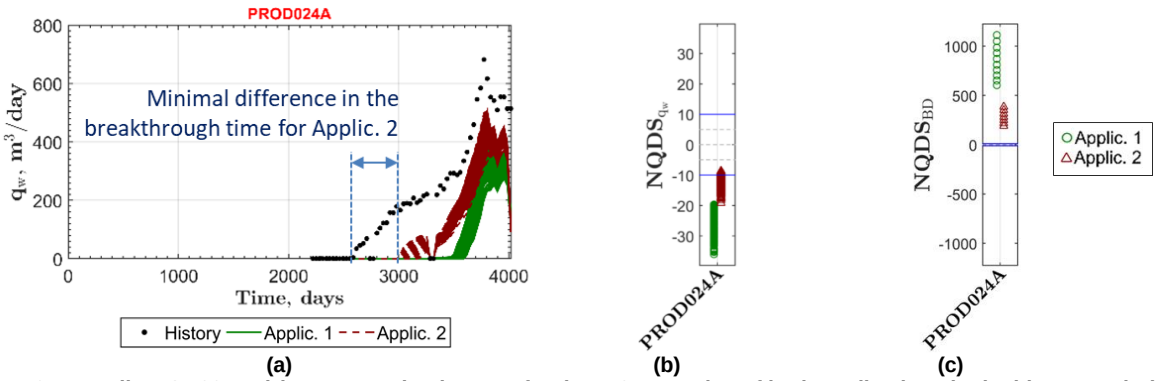


Figure 2.11. Well PROD024A: (a) Water production rate for the 450 scenarios of both applications in the history period; (b) Indicative of better $NQDS_{qw}$ for Application 2; (c) $NQDS_{BD}$ of the well PROD024A revealing an improvement in the BD, but still with a significant mismatch. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting of the traditional and proposed ones.

Water production of the well NA3D (Figure 2.12-a) indicates that neither water rate nor breakthrough time matches the historical data for both applications. The inclusion of the $NQDS_{BD}$ in the process was not sufficient to adjust the water breakthrough time (Figure 2.12-b) and, for some scenarios, lead to a worse water rate production (Figure 2.12-c). In fact, the parameterisation is limited to the regional multipliers and this result indicates the need for adding different uncertain parameters, for example, flow barriers with uncertain transmissibility.

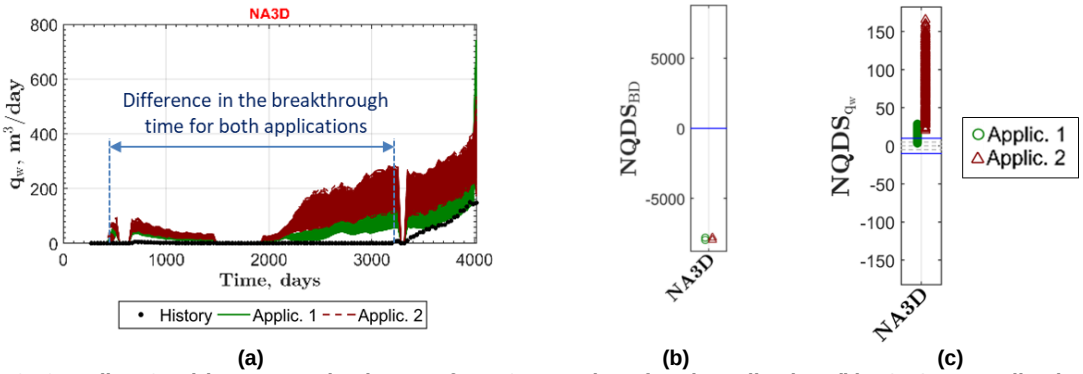


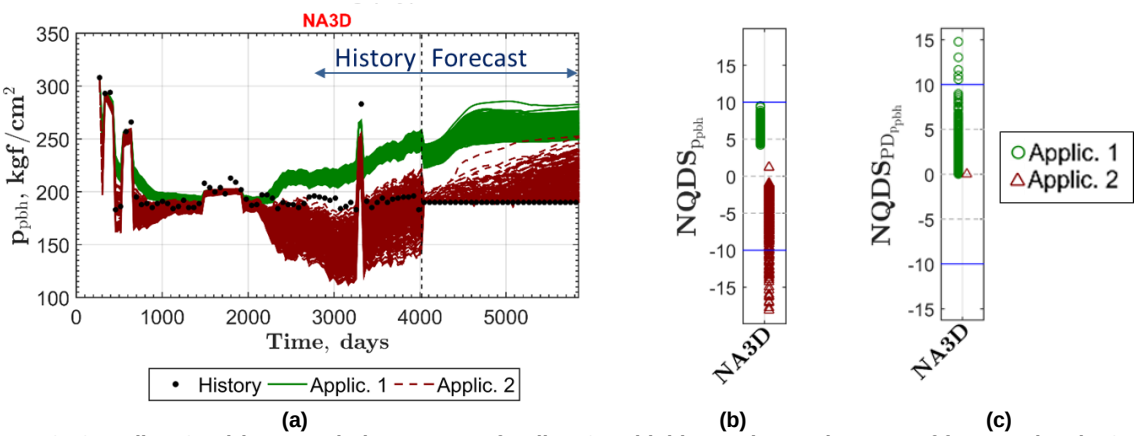
Figure 2.12. Well NA3D: (a) Water production rate for 450 scenarios of each application; (b) $NQDS_{BD}$ revealing large mismatch for all scenarios of both applications; (c) $NQDS_{qw}$ with some scenarios in the same range for both applications. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting of the traditional and proposed ones.

Therefore, a benefit of the additional OFs is to assist the identification of limitations in the reservoir parameterisation defined. The analysis of these extra indicators of reservoir quality can be useful when reviewing the reservoir parameterisation by supplying supplementary information to the scenarios' diagnostics, identifying unnoticed problems in the traditional approach.

2.4.5 Productivity Deviation

With the implementation of $NQDS_{PD}$, we observe an improvement in the transition from history to forecast periods for several wells as expected from the field results (Figure 2.9). The objective functions related to water injection rate and liquid production rate have higher impact in the history matching process. In Appendix 2.A, we show that these OFs groups are used to update a higher number of uncertain attributes when compared to $NQDS_{PDpbh}$ or $NQDS_{PDpibh}$. The justification for this behaviour refers to the definition of Productivity Deviation setup, which has BHP define as a boundary condition to the last time step (target informed to the simulator). We select as example production well (NA3D) and injection well (INJ015) to exemplify the positive impact of the assimilation of the additional OFs.

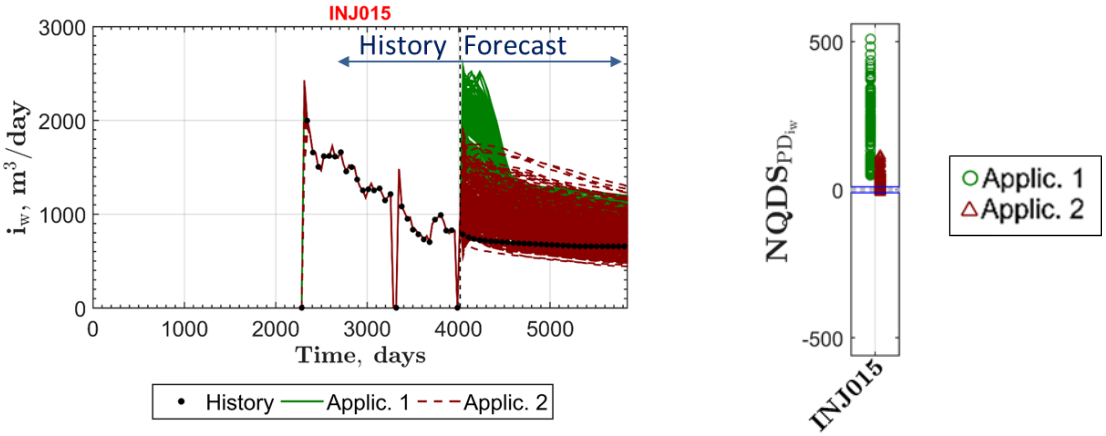
Figure 2.13-a presents BHP for the well NA3D during history and forecast periods with a total of 5 844 days (5 years of forecasting). The plots $NQDS_{ppbh}$ and $NQDS_{PDppbh}$ (Figure 2.13-b and -c) highlight pressure of the well closer to the reference (Application 2) data and with more variability around the history pressure than Application 1. In this sense, the scenarios of Application 2 are considered better conditioned than those in Application 1 for the OFs analysed. Jointly, these graphs provide evidence that scenarios with smaller indicators of Productivity Deviation provide better forecast behaviour.



(a) (b) (c)

Figure 2.13. Well NA3D: (a) Bottom hole pressure of well NA3D with history data and 5 years of forecast (total 5 844 days), (b) $NQDS_{ppbh}$ and (c) $NQDS_{PDppbh}$ highlighting the differences between the applications. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting of the traditional and proposed ones.

The transition of water injection between history and forecast period improved for several wells. The injection rate for well INJ015 (Figure 2.14-a) and its corresponding $NQDS_{PDiw}$ (Figure 2.14-b) are examples of better conditioning of scenarios in the transition.

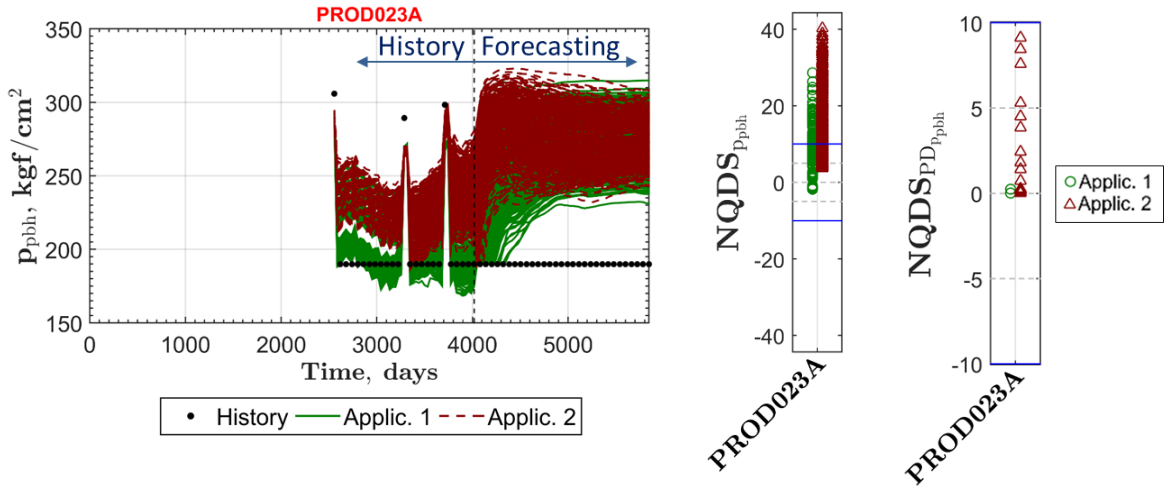


(a) (b)

Figure 2.14. Well INJ015: (a) Water injection rate of well with history data and 5 years of forecast (total 5 844 days), (b) $NQDS_{PDiw}$ highlighting the fluctuations in the last point of the history data simulated with forecast conditions. $NQDS_{iw}$ omitted because all scenarios matched the history data. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting in the traditional and proposed ones.

2.4.6 Detailing some OFs with poorer match

We also observe some objective functions with higher misfit for Application 2 than for Application 1. For these OFs, the addition of the unconventional OFs is not beneficial.



(a) (b) (c)

Figure 2.15. Well PROD023A – (a) Bottom hole pressure of well with history data and 5 years of forecast (total of 5 844 days); (b) $NQDS_{ppbh}$ showing the scenarios of Application 2 (in brown) limited to models with higher-pressure levels than the reference; meanwhile, Application 1 (in green) has more scenarios in the range [-10, +10]; (c) $NQDS_{PDppbh}$ showing that the assimilation of additional OFs is not beneficial for some OFs. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting of the traditional and proposed ones.

In our example, we explore the OFs of the well PROD023A. We detail this analysis from the bottom hole pressure for the history and 5-years forecast period (Figure 2.15-a). Highlighted by the NQDS plots (Figure 2.15-b and -c), the scenarios of Application 2 are limited to scenarios with higher-pressure levels than the reference. At the same time, Application 1 presents scenarios with higher variability, including scenarios with lower pressure values and closer to the reference.

The *mkz* of the region 12 influences only the $NQDS_{ppbh}$ well PROD023A in the Application 1 (Figure 2.16) but 6 OFS in the Application 2 ($NQDS_{ppbh}$, $NQDS_{PDppbh}$ of the well PRD023A and $NQDS_{PDql}$ of the wells PROD023A, PROD024A and PROD025A, Figure 2.17). For the second application, in order to provide a better match for $NQDS_{PDql}$ PROD025A, this uncertain attribute is updated in a detrimental manner from the perspective of the other OFs.

We investigate this effect through the correlation matrix, identifying the relationship between uncertain attributes and OFs. In the IDLHC methodology (Figure 2.3, *STEP 6.1*), the correlation matrix with the cut-off R_c captures this relationship for each of the 8 iterations. The number of iterations that a given OF is correlated to an uncertain attribute is added up and presented in two plots: Figure 2.16 and Figure 2.17 consider traditional and additional OFs, respectively. Each line corresponds to an uncertain attribute. In Figure 2.16, the R12 line corresponds to region 12. White colour means that the correlation coefficient is lower than the cut-off *Rc* in any iteration. Black colour means that the correlation is higher than the cut-off *Rc* in all the 8 iterations. The transitional colours correspond to intermediate values between 0 and 8 iterations.

The groups of the 64 conventional (Figure 2.16) and additional OFs ($NQDS_{BD}$ and $NQDS_{PD}$ – Figure 2.17) are plotted in the matrix with the uncertain attributes. Our focus is on the behaviour of the objective functions influenced by *mkz* (R12), marked with vertical lines in the plots. The analysis of the attribute *mkz* (R12) is direct because the only conventional OF correlated to it is the $NQDS_{ppbh}$-PROD023A. Figure 2.16 is built with data from Application 1. The attributes for vertical permeability multiplier (*mkz*) of region 12 are marked with a horizontal line because it influences the $NQDS_{ppbh}$-PROD023A. Because Application 2 has this same relationship, we do not present a correlation matrix computed for the additional OFs.

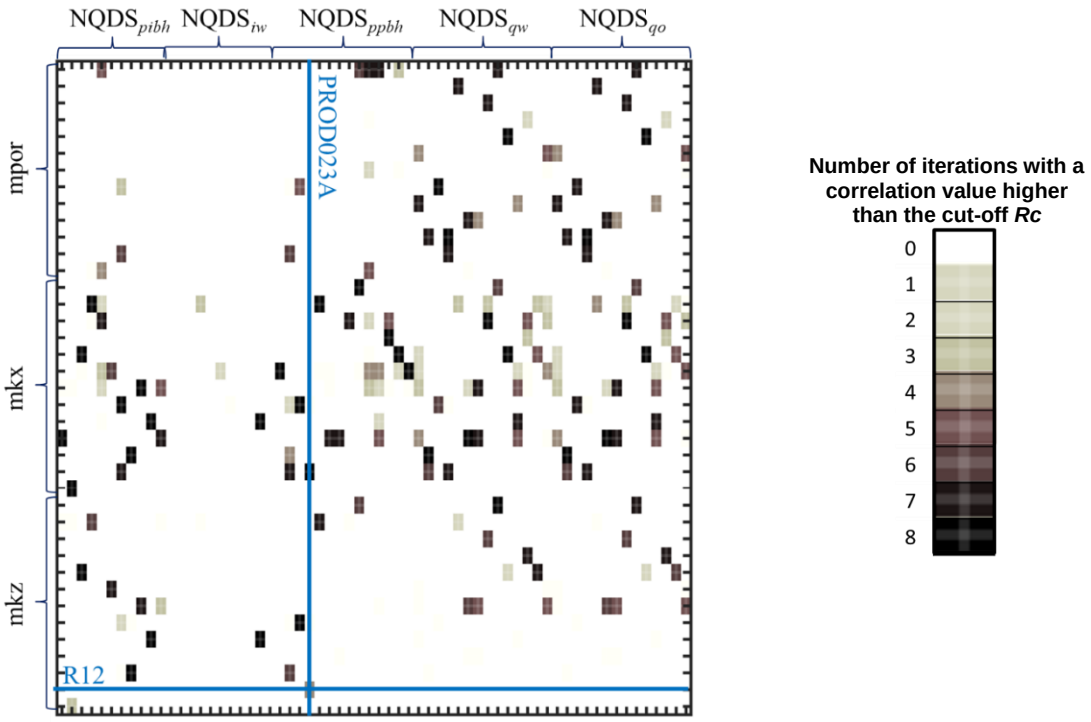


Figure 2.16. Matrix identifying the correlations captured in the 8 iterations for the group of 64 conventional OFs, Application 1. Black colour means that the correlation was of higher value than the cut-off R_c in all the 8 iterations. White color means that the correlation coefficient is lower than the cut-off R_c in any iteration. The transitional colours correspond to intermediate values between 0 and 8 iterations, as presented by the legend. The blue lines highlight the intersection between attributes and OFs mentioned in the text.

For Application 2, the NQDS$_{PDppbh}$ of the well PROD023A (Figure 2.15) is highlighted together with the other OFs influenced by this attribute (vertical lines).

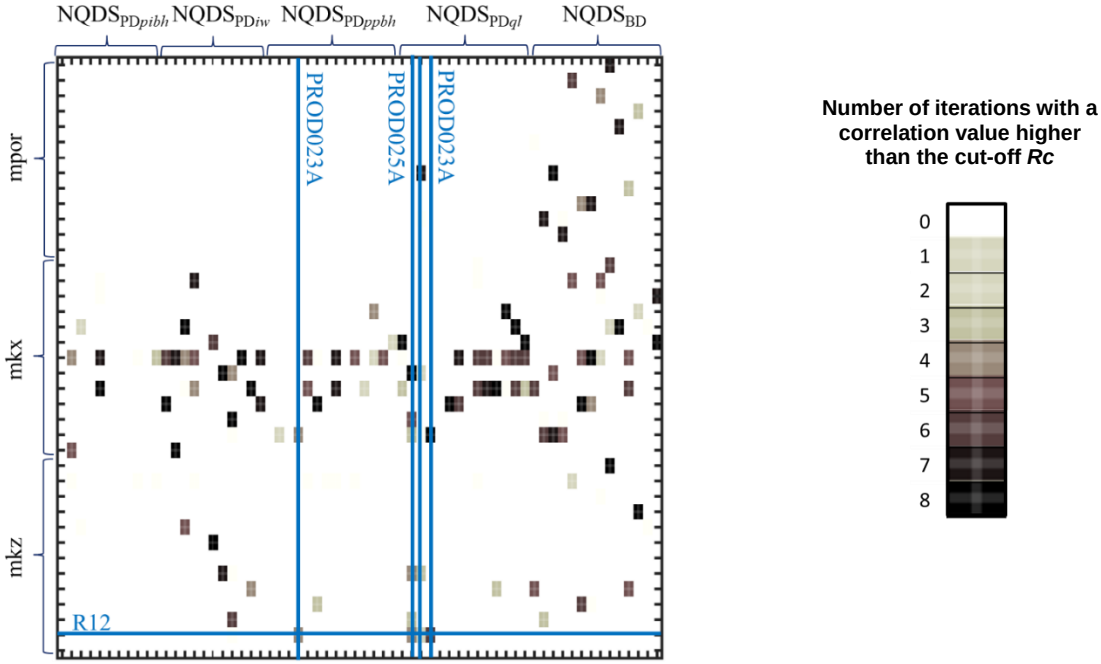
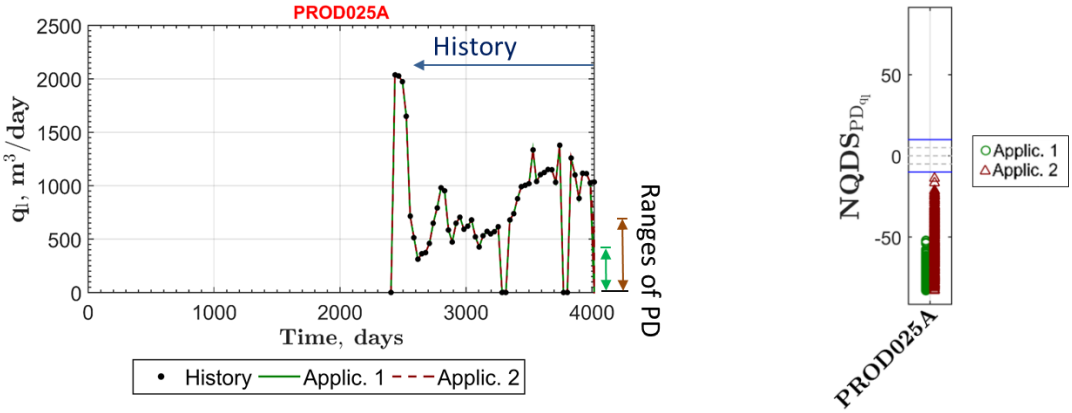


Figure 2.17. Matrix identifying the correlations captured in the 8 iterations for the NQDS$_{PD}$ and NQDS$_{BD}$ objective functions, Application 2. Black color means that the correlation was of higher value than the cut-off R_c in all the 8 iterations. White colour means that the correlation coefficient is lower than the cut-off R_c in any iteration. The transitional colours correspond to intermediate values between 0 and 8 iterations, as presented by the legend. The blue lines highlight the intersection between attributes and OFs mentioned in the text.

We observe that the $NQDS_{PDql}$ of the well PROD025A (Figure 2.18-a and -b) is closer to the reference in Application 2.



(a) (b)

Figure 2.18. The attribute *mkz* (R12) influences the $NQDS_{PDql}$ of the well PROD025A – (a) Liquid production rate in the history period for both applications highlighting the ranges of productivity deviation in the last history time step; (b) $NQDS_{PDql}$ of the well PROD025A highlighting smaller fluctuation in the transition between history and forecast period for Application 2 than for Application 1. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting of the traditional and proposed ones.

We also present the final distribution of the attribute *mkz* of region 12 (Figure 2.19). On the one hand, Application 1 (in green) presents a higher number of levels (variability) as well as higher multiplier values. On the other hand, Application 2 distribution (in brown) is concentrated to fewer levels and smaller multipliers (to the left of the *x*-axis).

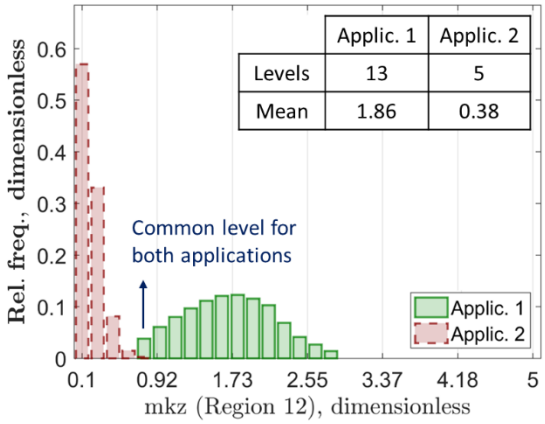


Figure 2.19. mkz of Region 12, an attribute correlated to the well PROD023A. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting in the traditional and proposed ones.

This attribute contributed to the behaviour described for this OF: smaller *kz* leads to a BHP closer to the reference for PROD025A (the scenarios in Application 1 have lower pressure when compared to Application 2 and the history data). Therefore, $NQDS_{ql}$ for this well

is smaller (Figure 2.18) because the liquid production rate of several scenarios does not diminish as much as in Application 1 to honour the informed pressure.

To summarise this example explaining why some OFs presented poorer match in Application 2, this uncertain attribute (*mkz* R12) influences traditional and additional OFs ($NQDS_{ppbh}$, $NQDS_{PDppbh}$ and $NQDS_{PDql}$). In order to provide a better match for the $NQDS_{PDql}$-PROD025A, the *pdf* concentrates in some levels but is detrimental to other OFs ($NQDS_{ppbh}$ and $NQDS_{PDppbh}$ of PROD023A).

This result indicates that with a large number of OFs assimilated, and a large number of uncertain attributes to update, the relationships between OFs and attributes increases the challenge to match the dynamic behaviour and all OFs assimilated.

2.5 Conclusions

We evaluated the impact of gathering and considering additional information from the dynamic data series in the History Matching (HM) performance. We presented a deep analysis of the assimilation of dynamic data series in an unconventional way, which is based on splitting the available historic time-series into more Objective Functions (OFs), detaching relevant events observed in the historical data. The OFs included measuring the Breakthrough Deviation (BD) and Productivity Deviation (PD).

We proposed an adaptation for the calculation of the additional objective function called Productivity Deviation (PD), which only uses information from the history data. It changes the information provided to the simulator from liquid production or water injection rate to bottom hole pressure.

Two applications show different field and well behaviour in the scenarios of the last iteration of the history matching process. The main identified advantages of the unconventional OFs in the HM matching process for this study case were:

- Smoother transition between history and forecast periods for field data;
- Water breakthrough time closer to the reference data for several wells and scenarios;
- Additional indicators of quality of the reservoir model to support the review of parameterisation: revealing problems in scenarios unnoticed by applying only the traditional OFs;

- Final scenarios with better predictability behaviour of the field in short (1-year), mid (5-years) and long (19-years) term.

Nevertheless, when considering the additional OFs, we observed a situation with traditional OF groups, presenting more distant scenarios from the history data. In fact, the HM problem becomes more complex to solve with the additional OFs because of the uncertain attributes considered influence more the OFs. In order to accommodate these additional OFs in the HM process, some traditional OFs result in a higher mismatch.

The improved predictability of the simulation scenarios indicates that a superior performance of HM process is possible by splitting the available dynamic data. At the same time, the evidence shown in this paper encourages the continuous improvement of HM methodologies and new approaches of data assimilation, which are able to accommodate a higher number of uncertain attributes and OFs.

Nomenclature – Article 1

BD	Breakthrough Deviation
BHP	Bottom Hole Pressure
DLHC	Discrete Latin Hypercube
HM	History Matching
IDLHC	Iterative Discrete Latin Hypercube
$Iter_{max}$	Maximal number of iterations in IDLHC
i_w	water injection rate
NQD	Normalised Quadratic Deviation
NQDS	Normalised Quadratic Deviation with Sign
OF	Objective Function
PD	Productivity Deviation
pdf	probability density function
p_{ibh}	Bottom hole pressure of injection wells
p_{pbh}	Bottom hole pressure of production wells
q_o	Oil production rate
q_w	Water production rate
R_c	Cut-off to the coefficients of the correlation matrix

Acknowledgement – Article 1

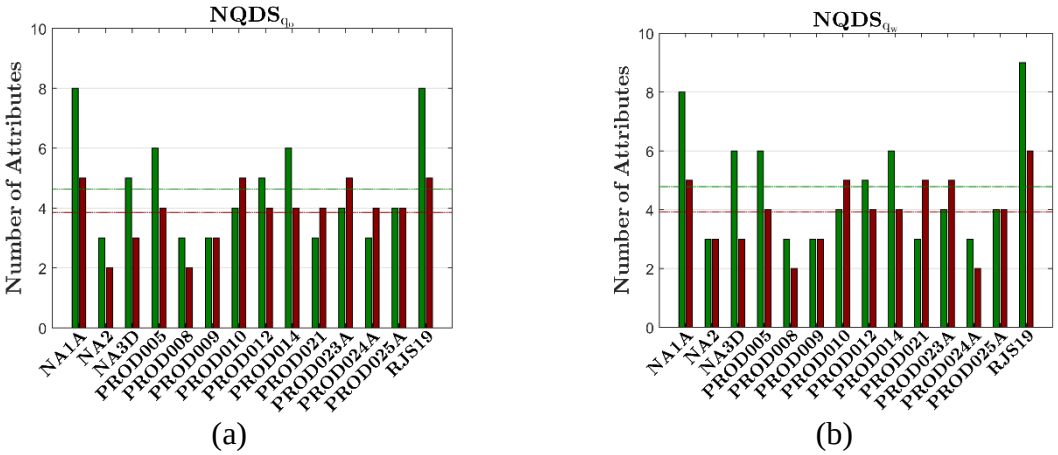
This work was conducted with the support of CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil, Energi Simulation and in association with the ongoing Project registered under ANP number 19708-7 as "BG-26 – Fomento à Formação de Recursos Humanos em Gestão de Incertezas e Tomada de Decisão: Um Programa BG

Fellowship" (UNICAMP/ Shell Brazil /ANP) funded by Shell Brazil, under the ANP R&D levy as "Compromisso de Investimentos com Pesquisa e Desenvolvimento". The authors also thank UNISIM, DE-FEM-UNICAMP, CEPETRO and Department of Mathematical Sciences (Durham University) for supporting this work and CMG, Emerson and Schlumberger for software licenses.

Appendix 2.A: Analysis of the importance of OF groups

The graphics below present all the objective functions displayed in groups according to the respective type of production data and application (Application 1 in green, Application 2 in brown). The bar's height represents the number of attributes that a given OF was selected to update uncertain attributes during all iterations. A horizontal line with the mean of all wells supports the differentiation between the two applications. Note that OFs from Figure 2.20, 2.21 and 2.22-*a* are assimilated in both Applications, but from Figure 2.22-*c*, 2.23 and 2.24, only in the Application 2.2. Also, the plots are on the same scale in the y-axis.

NQDS for oil and water rate (Figure 2.20-a and -b) have similar importance along with the wells, with a slight difference in the mean values. These plots evidence the complementarity between water and oil production when a simulation model is close to or meets the target values of the liquid production informed to the simulator.



(a) (b)

Figure 2.20. The number of attributes that a given OF was selected to update uncertain attributes by well: (a) NQDS$_{qo}$; (b) NQDS$_{qw}$. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting of the traditional and proposed ones.

Water injection rate is the boundary condition informed to the simulator in the history period, with exception to the last time which the target is set to be BHP. In Figure 2.21-

a, the mean number of attributes of $NQDS_{iw}$ is higher for Application 2 than for Application 1. Nevertheless, $NQDS_{iw}$ does not update more than two uncertain attributes for any well.

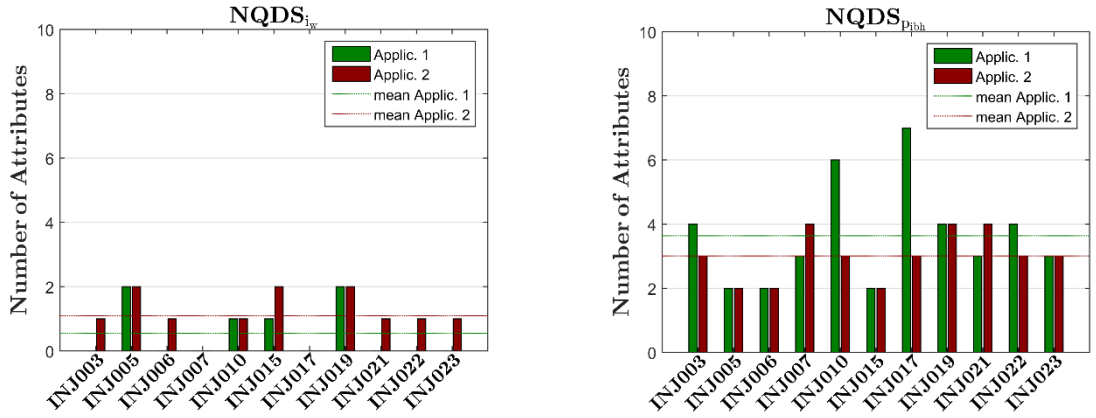


Figure 2.21. The number of attributes that a given OF was selected to update uncertain attributes by well: (a) $NQDS_{iw}$; (b) $NQDS_{pibh}$. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting of the traditional and proposed ones.

The mean number of attributes of $NQDS_{ppbh}$ is close to 4 for both applications (Figure 2.22-a), which indicates similar importance. Figure 2.22-b presents the NQDS of Breakthrough Deviation, which has a higher mean of uncertain attributes updated among the additional objective functions.

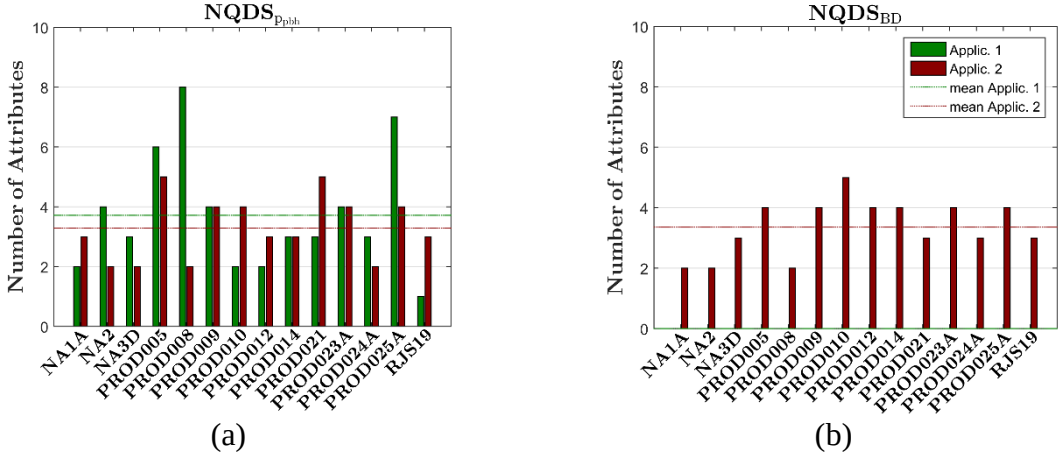
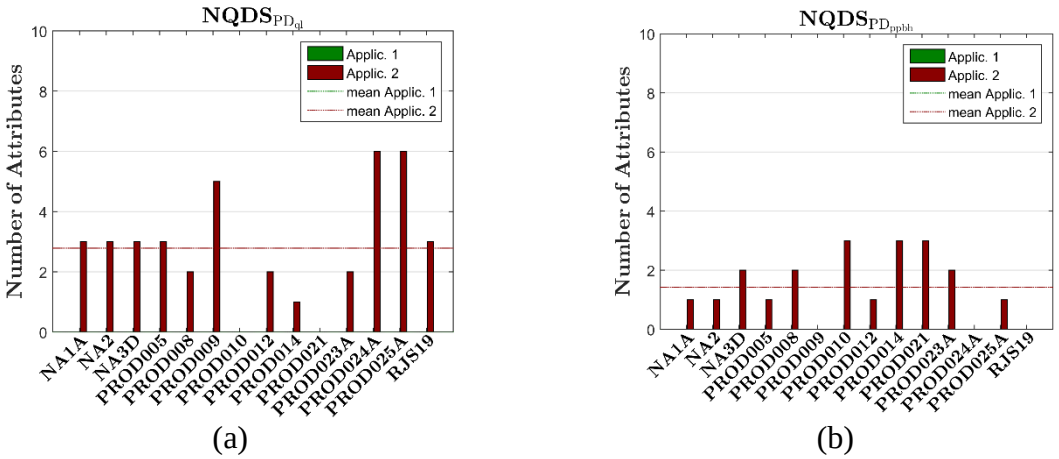


Figure 2.22. The number of attributes that a given OF was selected to update uncertain attributes by well: (a) $NQDS_{ppbh}$; (b) $NQDS_{BD}$. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting of the traditional and proposed ones.

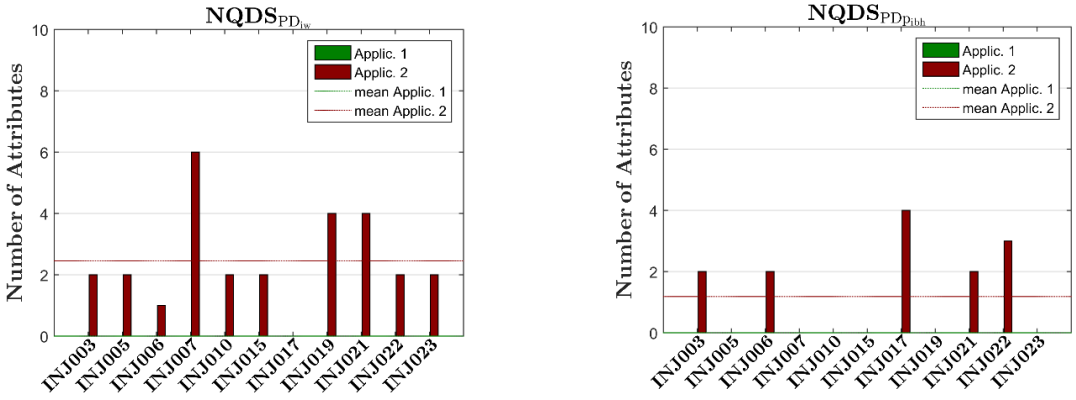
Because in the last time step the BHP is a target for the simulator, $NQDS_{PDql}$ group updates more uncertain attributes than $NQDS_{PDppbh}$, on average. Mismatches related to $NQDS_{PDpbh}$, have too small variability for some wells (for example, PROD024A, RJS019) or are uncorrelated with uncertain attributes (for example PROD010).



(a) (b)

Figure 2.23. The number of attributes that a given OF was selected to update uncertain attributes by well: (a) NQDS$_{PDql}$; (b) NQDS$_{PDppbh}$. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting of the traditional and proposed ones.

The same reasoning is applicable for PD of water injection and BHP of injectors. NQDS$_{iw}$ groups update more attributes than NQDS$_{PDpibh}$, on average.



(a) (b)

Figure 2.24. The number of attributes that a given OF was selected to update uncertain attributes by well: (a) NQDS$_{PDiw}$; (b) NQDS$_{PDpibh}$. Note: Application 1 assimilates 64 Objective Functions traditionally applied in the IDLHC methodology, and Application 2 considers 128 Objective Functions consisting of the traditional and proposed ones.

This analysis indicates that among the OFs groups added in the history matching process, the Breakthrough Deviation was more relevant in the process of updating uncertain attributes for the study case applied in this paper.

Appendix 2.B: Analysis of NQD used as cut

This section answers a discussion made after the publication of Formentin *et al.* (2019-a): Why does Application 2 (with more objective functions included in the process) has results with higher variability than Application 1 (Figure 2.9)? It is a fair question giving the intuition that more restrictions in the process lead to stronger restriction of variability.

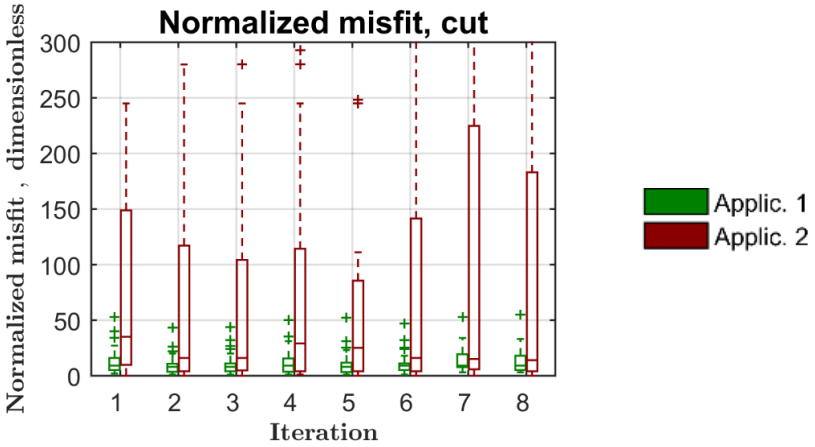


Figure 2.25. Normalised misfit cut used to update the pdf of each of the 39 attributes in each of the iterations for Application 1 (Boxplots in green) and Application 2 (Box plots in brown).

Figure 2.25 shows the normalised misfit selected in *STEP 6* of the methodology IDLHC (Figures 2.2 and 2.3). The box plots for Application 1 (green) and Application 2 (brown) are plotted considering the iterations of the process. We note that the medians for Application 2 (brown) are systematically higher the ones for Application 1, which provides more variability in the selected scenarios to update the pdfs of the attributes.

3 ARTICLE 2: SYSTEMATIC UNCERTAINTY REDUCTION FOR PETROLEUM RESERVOIRS COMBINING RESERVOIR SIMULATION AND BAYESIAN EMULATION TECHNIQUES: PART I

Helena Nandi Formentin, Ian Vernon, Guilherme Daniel Avansi, Camila Caiado, Célio Maschio, Michael Goldstein, Denis José Schiozer

To be submitted.

Contribution by the candidate Helena Nandi Formentin: Conceptualisation, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original + Edition, Visualization, Funding Acquisition, as described in the Statement of Contribution submitted to the SPE Europec (fully reproduced in Appendix B – Contribution Statements). The code in R was fully developed by Helena. CMG was selected as reservoir simulation software.

Abstract – Article 2

Reservoir simulation models incorporate physical laws, reservoir characteristics and production strategies. They represent our understanding of sub-surface structures based on the available information. Emulators are statistical representations of simulation models behaviour, offering fast evaluations of a sufficiently large number of reservoir scenarios, to enable a full uncertainty analysis. Bayesian History Matching for Uncertainty Reduction (BHMUR) aims to find the range of reservoir scenarios that are consistent with the historical data, to provide a comprehensive evaluation of reservoir performance and consistent, unbiased predictions incorporating realistic levels of uncertainty, required for full asset management. We proposed a systematic approach for uncertainty quantification that combines reservoir simulation and emulation techniques within a coherent Bayesian framework for uncertainty quantification. Our systematic procedure is an alternative and more rigorous tool for reservoir studies dealing with probabilistic uncertainty reduction. It comprises the design of sets of simulation scenarios to facilitate the construction of emulators, capable of accurately mimicking the simulator with known levels of uncertainty. Emulators can be used to accelerate the steps requiring large numbers of evaluations of the input space in order to be valid from a statistical perspective. Via implausibility measures, we compare emulated outputs with historical data incorporating major process uncertainties. Then, we iteratively identify regions of input parameter space unlikely to provide acceptable matches, performing more runs and reconstructing more accurate emulators at each wave, an approach that benefits from several efficiency improvements. The procedure was applied to reduce uncertainty in a complex reservoir case study with 25 wells, injectors and producers. The selection of one scenario as hypothetical reality allowed us to discuss analytical and theoretical aspects and to demonstrate the applicability of the procedure for complex petroleum reservoirs. The case study contains 26 uncertain attributes representing petrophysical, rock-fluid and fluid properties. We selected phases of evaluation considering specific events during the reservoir management, improving the efficiency of simulation resources use. With 15 waves and 115 valid emulators, we ruled out regions of the search space identified as implausible, and what remained was only a small proportion of the initial space judged as non-implausible (~10^{-11}%). The systematic procedure showed that uncertainty reduction using iterative Bayesian History Matching has the potential to be used in a large class of reservoir studies with a high number of uncertain parameters. In this paper, we advance the applicability of Bayesian History Matching for Uncertainty Reduction of reservoir studies with

two deliveries: (a) a general workflow for systematic BHMUR, and (b) the use of phases to progressively evaluate the historical data.

Keywords: Bayesian History Matching, Uncertainty Reduction, Emulation, Systematic Procedure.

3.1 Introduction

One of the biggest challenges for the energy industry is how to deal with many sources of uncertainty. Reservoir simulation models describe the understanding and interpretation of sub-surface structures, incorporating available data and technology. Reservoir model calibration is an inverse problem based on historical reservoir data: a high dimensional, ill-posed, non-linear problem.

The ultimate goal of a calibration process is to provide background for well informed and efficient decisions. Reservoir calibration can reduce, but not eliminate, the uncertainty in calibrated models. Finding the whole class of scenarios capable of representing the reservoir historical behaviour is essential as it adds value to an asset by giving a realistic evaluation of reservoir performance and consistent predictions with corresponding uncertainties. Calibrated models drive, for example, recovery strategies optimisation and risk quantification.

Multiple possible solutions are inherent in inverse problems for imperfect models possessing a large number of uncertain attributes (inputs) and outputs (production data observed with uncertainty). Advanced calibration techniques have been developed and applied in the energy industry referred to as History Matching (Oliver and Chen 2011, Oliver *et al.* 2008), Data Assimilation (Evensen 2009; Carrassi *et al.* 2018) and Uncertainty Quantification and Reduction (Smith 2014).

Challenges in the calibration of reservoir models that we address in this study are:

- **Preservation of variability** in calibrated models/scenarios while considering several sources of uncertainty involved in the calibration process;
- **High dimensional input space** implying that a large number of simulator evaluations would be required to provide a consistent representation of the uncertainties and for the maintenance of geological realism;

- **Time and resources** to evaluate a large number of simulations required in a calibration process (*i.e.* ideally an exhaustive assessment of all possible scenarios would be made);

- **High dimensional output space** requiring advanced analytic techniques to suitably capture patterns, trends, and associations of data with diverse characteristics.

Craig *et al.* (1995) made significant progress in this area by formulating an alternative calibration technique, combining simulation models and emulators under a Bayesian framework, referred to as Bayesian History Matching. The originality of this approach is seen in the identification of the whole range of solutions which are concurrently compatible with the historical reservoir performance, given major sources of uncertainty such as observation error and model discrepancy.

This class of technique was originally referred to Bayesian because Craig *et al.* (1995) used Bayes linear strategies to update emulator models using simulator output data. We qualify the procedures developed in this article as Bayesian mainly for (a) the uncertainty quantification, which is based on prior knowledge about the model and process characteristics, and (b) the uncertain nature of the reservoir parameters, attributed to our lack of knowledge about the subsurface (in contrast to a random process, as in a frequentist approach). Despite not using the Bayes theorem directly to update the attributes, we keep the traditional nomenclature for inverse problems in the petroleum industry with a brief reformulation: we highlight the main objective of this class of calibration process, building the term *Bayesian History Matching for Uncertainty Reduction*.

A central element of the BHMUR approach is the combination of simulation models and emulators. We summarise their main features in Table 3.1 and follow with additional definitions.

Table 3.1. Main characteristics of simulation models and emulators which are central elements of the Bayesian History Matching approach.

	Simulation models	**Emulators**
Definition	Numerical representation of the reservoir	Statistic approximations of simulation models
State	Physical laws, reservoir structure, sub-surface understanding	Mathematical models, statistical principles, data structure
Cost	Expensive evaluation of scenarios	Fast evaluations of large numbers of scenarios

- **Data set** consists of several possible reservoir model scenarios (inputs) and the corresponding quantities of interest (outputs); the scenarios are sampled in order to cover the input parameter space;

- **Training data set** is a data set used to construct emulators (*e.g.* select active variables and fit the statistical model); in this study, training data are plotted in red and orange;

- **Testing data set** is an independent data set applied to validate and select concurrent emulators constructed with the training set for a given quantity of interest; plotted in light and dark blue.

3.1.1 Bayesian History Matching

A fundamental characteristic distinguishes Bayesian History Matching for Uncertainty Reduction (BHMUR) approaches from other calibration techniques. To find the whole range of solutions for the calibration problem, the focus under a BHMUR is on the following question:

Which parts of the input space are unlikely to lead to acceptable fits between the model outputs and historical data?

From a practical perspective, the rationale for solving the inverse problem is reformulated. We look for what is not the solution to the problem in order to rule this part out of the input space and identify appropriate solutions. This process is far more effective than directly looking for good solutions as often large regions of the input space can be ruled out by considering only small numbers of outputs.

The core strategy of this approach relies on three elements (Figure 3.1). Firstly, a restricted number of scenarios is carefully sampled from the search space of interest (the input parameter space formed from the uncertain attributes). In Figure 3.1.a, a sample of scenarios from a search space of two dimensions (uncertain attributes named ϕ and k_r) is illustrated via a pair plot.

Using the reservoir simulator, we evaluate the sample set to acquire outcomes corresponding to measurable physical quantities (*e.g.* pressure and production rates). These outcomes are usually called quantities of interest or reservoir outcomes. We construct emulators (Figure 3.1.b) based on the data set from the simulated samples by applying appropriate

statistical techniques. Because emulators are fast, a large number of scenarios can then be evaluated, allowing exhaustive exploration of the search space.

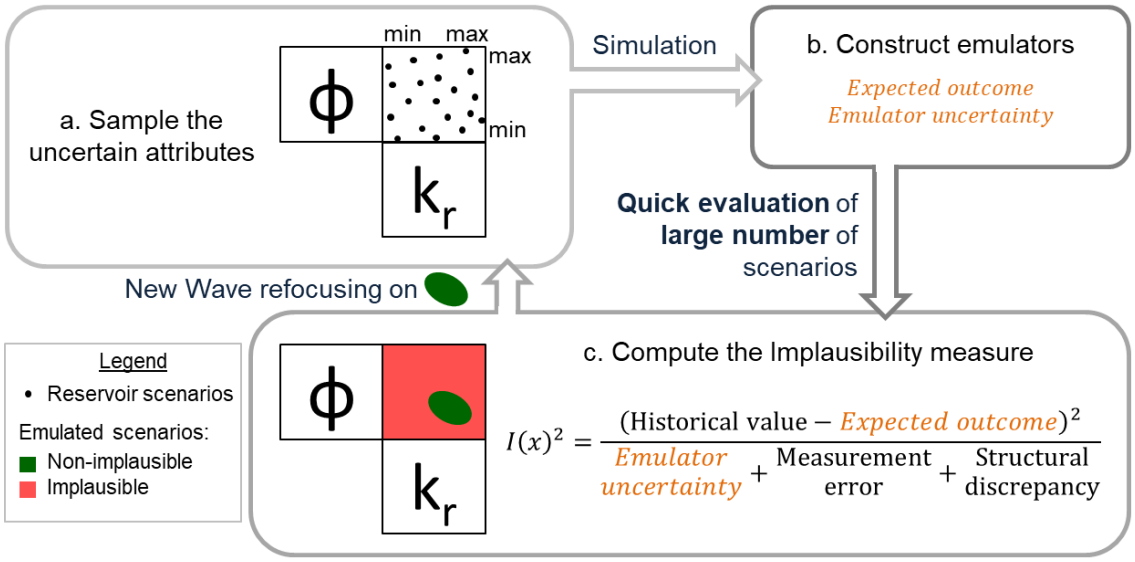


Figure 3.1. Three essential elements for model calibration through the Bayesian History Matching approach – (a) Sample the uncertain attributes evaluated with simulators; (b) Construct emulators able to compute expectation and uncertainty values and to quickly evaluate new scenarios; (c) Compute the implausibility measure to identify which parts of the input space are unlikely to lead to acceptable fits. Each iteration is a wave and refocuses the search space on non-implausible regions.

The implausibility measure (Figure 3.1.c) examines the difference between the historical value and the expected outcome from the emulator, compared to all the major uncertainties that arise: from using the emulator as a representation of the simulator, from using the measurement data to represent the real data and from using the reservoir model to represent the real physical reservoir. The implausibility measure is important because it identifies which parts of the input space are unlikely to lead to acceptable fits (*i.e.* the implausible regions) and answers the question stated above. We rule out regions of the search space identified as implausible, and what remains is only the proportion of the original space currently judged as non-implausible. Because a large number of scenarios are evaluated via the emulator, we can draw detailed images of the implausible regions projected into two-dimensional subspaces, referred to as implausibility pairs plots (red, Figure 3.1.c).

An iterative process is implemented in order to narrow the search space sequentially. A new wave refocuses the search space on the current non-implausible regions (green, Figure 3.1.c), from where a limited number of carefully designed scenarios is sampled. This data set is used to construct additional and more accurate emulators leading to further space reduction. The iterative process through waves sequentially discard regions of the input space, continually refocusing our search on the regions judged as non-implausible (Vernon *et*

al. 2018; Williamson *et al.* 2017). Jointly, the three elements of Figure 3.1 are able to mimic complex data structures with simple emulators in a reliable way.

The motivation for exploring this approach in the context of petroleum reservoirs includes: (a) efficient use of simulation and associated computational cost; (b) extensive exploration of reservoir scenarios to secure the representativeness of the whole class of models capable of describing the reservoir historical data; (c) integration of diverse sources of uncertainty coming from observed data and an imperfect simulation model; and (d) qualitative and quantitative insights about the reservoir description and performance.

BHMUR has been successfully employed across a diverse set of scientific domains, *e.g.* galaxy formation (Vernon *et al.* 2010), petroleum engineering (Moreno *et al.* 2018; Ferreira *et al.* 2014), climate modelling (Williamson *et al.* 2017) and system biology (Vernon *et al.* 2018). The main challenges for the application of BHMUR for reservoir models remain in (a) high dimensionality of inputs (*e.g.* spatial uncertain attributes linked to porosity, permeability and facies maps) and outputs (several observed measures to be used in the calibration process); (b) identification of structures, dependency and interdependency in data sets; (c) modelling structures of discrepancy between reservoir models and the real reservoir; and (d) time to evaluate scenarios through simulation.

3.1.2 Objective

We aim to advance the applicability of Bayesian History Matching for Uncertainty Reduction (BHMUR) in reservoir studies by offering:

- **A systematic workflow** to structure BHMUR techniques. The workflow is designed to (a) scale-up to high dimensional input and output data, (b) secure flexibility on combining diverse emulation techniques, and (c) perform stages automatically, centring users' focus on analysis and synthesis;

- **Phases of evaluation** which split historical data into physically meaningful periods to gradually introduce data into the analysis to take advantage of information from early time;

This work focuses on the development of methodology. In Part I, we focus on the general workflow and the results of its application. In Part II (Formentin *et al.* 2020-b), we

expand four additional parts of the work implemented in the general workflow and relevant for BHMUR applications.

We apply our procedures in a case study to discuss analytical aspects and to demonstrate its applicability for the analysis of petroleum reservoirs. This case study considers (a) 26 uncertain reservoir attributes, (b) 25 active wells and (c) 4018 days of historical data. The hypothetical reality is one of the combinations of the uncertain attributes considered. The aim is to test the potential of the procedure for a complex reservoir model under a controlled situation while illustrating the main steps of the methodologies developed. The consideration of a hypothetical reality is an important stage to ground more advanced studies, contemplating real observed data and inherent discrepancies between reservoir model and real physical reservoirs.

3.2 Statistical Methodology

We now introduce the standard form of Bayesian History Matching for Uncertainty Reduction (BHMUR), as described by Vernon *et al.* (2018). This framework is then extended by the incorporation of two-class emulators to address specific structures in reservoir data sets (developed in Formentin et al. 2020-b) and subsequently by the use of phases of evaluation. We provide background information for the applied statistical models and emulator validation.

3.2.1 Formulation of Bayesian History Matching for Uncertainty Reduction

Measurable quantities from the real reservoir are denoted by the vector z (*e.g.* well pressure in a given time). They result from the sum between the corresponding y quantities from the real physical reservoir and observational errors e (Equation 3.1). Sources of observational errors depend on field's well surveillance programme and the measurement process in place, including (a) equipment calibration (random and systematic types); (b) chemical analysis for gas-oil-ratio; (c) apportionment of field production to well production and production testing, and (d) data manipulation.

$$z = y + e \tag{3.1}$$

The reservoir simulation model computes a corresponding vector of quantities via a function f of x, where x is a vector of input parameter values representing a reservoir scenario

(*i.e.* a combination of the uncertain attributes). We represent the difference between the real reservoir and the reservoir model as the model discrepancy term ϵ. Equation 3.2 indicates that even evaluating the most appropriate scenario x^* through $f(x^*)$, there is a difference ϵ between the reservoir model and the real reservoir (Vernon *et al.* 2010).

$$y = f(x^*) + \epsilon \tag{3.2}$$

The model discrepancy arises, for example, from simplifications in physical laws modelling the phenomena in place (*e.g.* multi-phase flow in porous media), reservoir conditions and characteristics.

As an example, the case study described in the Section 3.3.1. Definition of the case study has x^* as the hypothetical reality, which is a vector of known uncertain attributes. In real applications, x^* is undefined: it is very unlikely that a simulation model perfectly represents the real field behaviour. Each element $i = 1, \dots, q$ of the vector $f(x)$ corresponds to measurable quantities of interest of the reservoir and is defined as $f_i(x)$.

Emulators $f^*(x)$ are statistical approximations of simulation models.

They offer fast evaluations of sufficiently large numbers of scenarios from the input space and enable a full uncertainty analysis (Vernon *et al.* 2010; Craig *et al.* 1997; Craig *et al.* 1995). For any input scenario, an emulator provides a mean and distribution describing how close it is likely to be to the simulator output. The expected outcome is a plausible interpolation (or extrapolation) of the training data. The distribution around the mean is a reasonable expression of emulator uncertainty (O'Hagan 2004). Distinct from simulators, emulators do not directly incorporate reservoir conditions, characteristics or physical laws. Table 3.1 summarised complementary characteristics between simulators and emulators.

We employ an implausibility measure to describe which parts of the input space are unlikely to lead to acceptable fits between the model output and observed data. Equation 3.3 standardises the difference between the emulator expectation $E\left(f_i^*(x)\right)$ conditioned to the observed data for the -*ith* considered output and corresponding historical data z_i by all uncertainties identified in the process. They are expressed in terms of the variance of the

emulator $Var\left(f_i^*(x)\right)$, the variance of model discrepancy $Var(\epsilon_i)$ and the variance of the observation error $Var(e_i)$.

$$I_i(x) = \sqrt{\frac{\left[E\left(f_i^*(x)\right) - z_i\right]^2}{Var\left(f_i^*(x)\right) + Var(\epsilon_i) + Var(e_i)}} \tag{3.3}$$

For each scenario x, the implausibility measures $I_i(x)$ calculated from each output emulated, $i \in [1, q]$, can be combined in various ways. Among the diverse possibilities, the first, second or third maximal implausibility can be selected, depending on the number of emulators and the understanding about the uncertainties in the calibration process. In this application, we use the maximal implausibility as in Equation 3.4. Alternative options are discussed by Vernon *et al.* (2010).

$$I_M(x) = \max_{i \in Q} I_i(x) \tag{3.4}$$

In parallel, the cut-off ω addresses an assessment about the appropriateness of the assumptions made to compute implausibility and the associated mean distribution of I_M. Pukelsheim (1994) states that for all continuous unimodal distributions (*e.g.* normal, double exponential, chi-squared, t, lognormal) with moments (*e.g.* expected value and variance), it holds that 95% coverage is obtained within 2.98 standard deviations from the mean. Vernon *et al.* (2010) applied the three-sigma rule for the individual univariate implausibility measure I_i. We defined for our application an implausibility cut-off $\omega = 3$ considering that the assumptions of continuity and unimodality are also reasonable for the distribution of I_M. The cut-off ω defines a boundary at each iteration or wave, to label regions of the search space as either:

- **Implausible scenarios** when $I_M(x) > \omega$, specifying that the combination of inputs x is unlike to lead to acceptable fits, *i.e.* the scenario response is sufficiently far from the historical data to be ruled out from the search space;

- **Non-implausible scenarios** when $I_M(x) \leq \omega$, which result from a good fit to historical data (the numerator in Equation 3.3), or high variance of $f_i(x)$, ϵ_i or/and e_i (denominator

of Equation 3.3). A high emulator variance can be resolved in later waves, where more accurate emulators are constructed.

Continuous quantities of interest are traditionally modelled using an emulator given in Equation 3.5. Active variables x_{A_i} are a subset of inputs selected as the most influential for a given quantity of interest i. We need to select a subset of variables as active because complex systems have a large number of uncertain attributes, but a smaller number of them are usually responsible for the larger variability of the simulator output. We discuss the choice of active variables in Section 3.3.4 (*see* STEP 10). They are used to construct the corresponding emulator:

$$f_i^*(x) = \sum_j \beta_{ij} g_{ij}(x_{A_i}) + u_i(x_{A_i}) + w_i(x) \tag{3.5}$$

The first term is a regression term, where g_{ij} are known deterministic functions of the active variables x_{A_i} and β_{ij} are unknown scalar regression coefficients. A common choice is low order polynomials, *e.g.* a regression with first and second order, and interaction terms (Vernon *et al.* 2018). The other terms, $u_i(x_{A_i})$ is a Gaussian process over x_{A_i} and its associated nugget $w_i(x)$, which is related to the fact that only a sub-set of uncertain attributes is included in the emulator as active variables.

Choosing appropriate statistical models to construct emulators is a strategical step while performing Bayesian History Matching for Uncertainty Reduction (BHMUR). The choice of emulator models can influence computational effort required, number of waves and, ultimately, the quality of the resulting calibrated model. Their performance to properly mimic simulators also depends on the size, dimensionality, quality and nature of the data used.

In our analysis, we identified two categories of quantities of interest: continuous and binary, the latter having not been previously employed in a BHMUR setting. Concurrently, we choose algorithms capable of modelling them: regression and two-class classification models. Here, we recall some features of statistical models for continuous quantities of interest; in Part II (Formentin et al. 2020-b), we develop the background for two class-emulation and two-class patterns identified in reservoir engineering.

Indicators for diagnostics support the analysis and validation of emulators. Table 3.2 summarises the main features of selected indicators. Combined, they assess emulators for

the continuous and binary quantities of interest. We expand the description for information index, credible interval diagnostics, positive and negative predictive value in Part II (Formentin et al. 2020-b). For further information about adjusted-R^2 and Normalised Root Mean Square Error (RMSE_n) for validation of emulators, we recommend Moreno *et al.* (2018). Other indicators can be used when simulating a test set is not affordable (*e.g.* leave-one-out diagnostics). For more details, *see* Bastos and O'Hagan (2009).

Table 3.2. Summary of selected indicators - their combination enables to evaluate and select emulators for continuous and two-class quantities of interest in a comprehensive way. Additional information in Formentin *et al.* (2020-b).

Indicator	Set	Quantity of interest	Description
Information index (D_{info})	Training & test	Continuous & binary	The proportion of scenarios expected to be implausible
Credible Interval Diagnostics (D_{CI})	Training & test	Continuous	The proportion of scenarios for which simulation outcome is covered by the emulator credible interval
Positive Predictive Value (PPV)	Training & test	Binary	The proportion of implausible scenarios correctly classified
Negative Predictive Value (NPV)	Training & test	Binary	The proportion of non-implausible scenarios correctly classified
Adjusted-R^2	Training	Continuous	The proportion of the variance of the output explained by the regression
RMSE_n	Test	Continuous	Normalised Root Mean square Error

At this point, we highlight the four safeguards in favour of an appropriate level of uncertainty reduction under the BHMUR formulation. Firstly, the implausibility measure accounts for all sources of uncertainties present in the calibration process, including uncertainty about the model to represent the real physical reservoir. Secondly, the emulators are diagnosed and validated under quality criteria that avoid overconfident emulators to be applied in the implausibility analysis. Thirdly, the combination of implausibility from different emulators for the same scenario considers the number of emulators and our understanding of the uncertainties in the calibration process. Finally, the cut-off (ω) addresses an assessment about the appropriateness of the assumptions made to compute implausibility and the associated mean distribution of I_M.

3.3 Description and Application of the Systematic Procedure

The proposed systematic workflow consists of a sequence of 20 steps collected in six groups where simulation and emulation techniques are combined to reduce uncertainty in a petroleum reservoir. The main features of the procedure are: (a) *repeatability*, due to the sequential nature of the steps, logically associated; (b) *flexibility*, since the steps of the high-level structure are adaptable to project requirements; (c) *scalability* to higher dimensions, as specific steps are planned to accommodate techniques for dimensionality reduction. Several

steps are automatable and enable the team to focus on the analysis and synthesis of the project. Eleven steps (1 to 4, 7, 8, 12, 15, 17, 18 and 20) concentrate such activities.

We overview these six groups before detailing the description of the systematic workflow (Figure 3.2):

- **Definition of case study**: we set the case study by combining our knowledge, information and data about the petroleum reservoir, models, uncertainties and observational error in historical data;

- **Definition of strategy for Bayesian History Matching for Uncertainty Reduction (BHMUR)**: We plan the data analysis in order to conduct the calibration process efficiently. Two objectives are explored: (a) we identify outputs to emulate considering data structure, and (b) we select phases of evaluation considering specific events during the reservoir management (*e.g.* wells and field behaviour);

- **Data preparation**: We build independent training and test sets of size m and n scenarios, respectively. The test set is used for the selection and validation of emulators; nevertheless, the workflow is adaptable to studies which an independent test set is not affordable (*e.g.* simulation cost). The estimation of model discrepancy incorporates the uncertainty due to the model being an imperfect representation of the real reservoir. Although a full description of the model discrepancy is not the main focus of this paper, we demonstrate the impact of a particular form of model inaccuracy related to simulation targets;

- **Construct and validate emulators**: We select outputs to emulate targeting the most informative ones. Complex or simple emulators are constructed as needed. Implausibility analysis only considers valid emulators;

- **Evaluation of scenarios and uncertainty reduction**: All valid emulators are applied simultaneously in order to evaluate scenarios through the implausibility measures;

- **Decision for phase and wave (definitions below):** We check the need for a new wave, a new phase or both. We anticipate our comments on the criteria that we applied. In *STEP 15*, a minimum number of scenarios in the training set $\lambda.m$ is required to construct new emulators (*e.g.* $0.5m$ scenarios); In *STEP 17* 'Criteria to change phase met?', we applied the minimum proportion of remaining space ruled out with the emulators constructed until

a given wave (*e.g.* $D_{info,M} \geq 70\%$). In *STEP 18* – 'Criteria to end calibration met', we defined the criteria that $phase$ is the last phase of evaluation. It indicates that our uncertainty reduction ends when the last phase is evaluated and when $D_{info,M} < 70\%$.

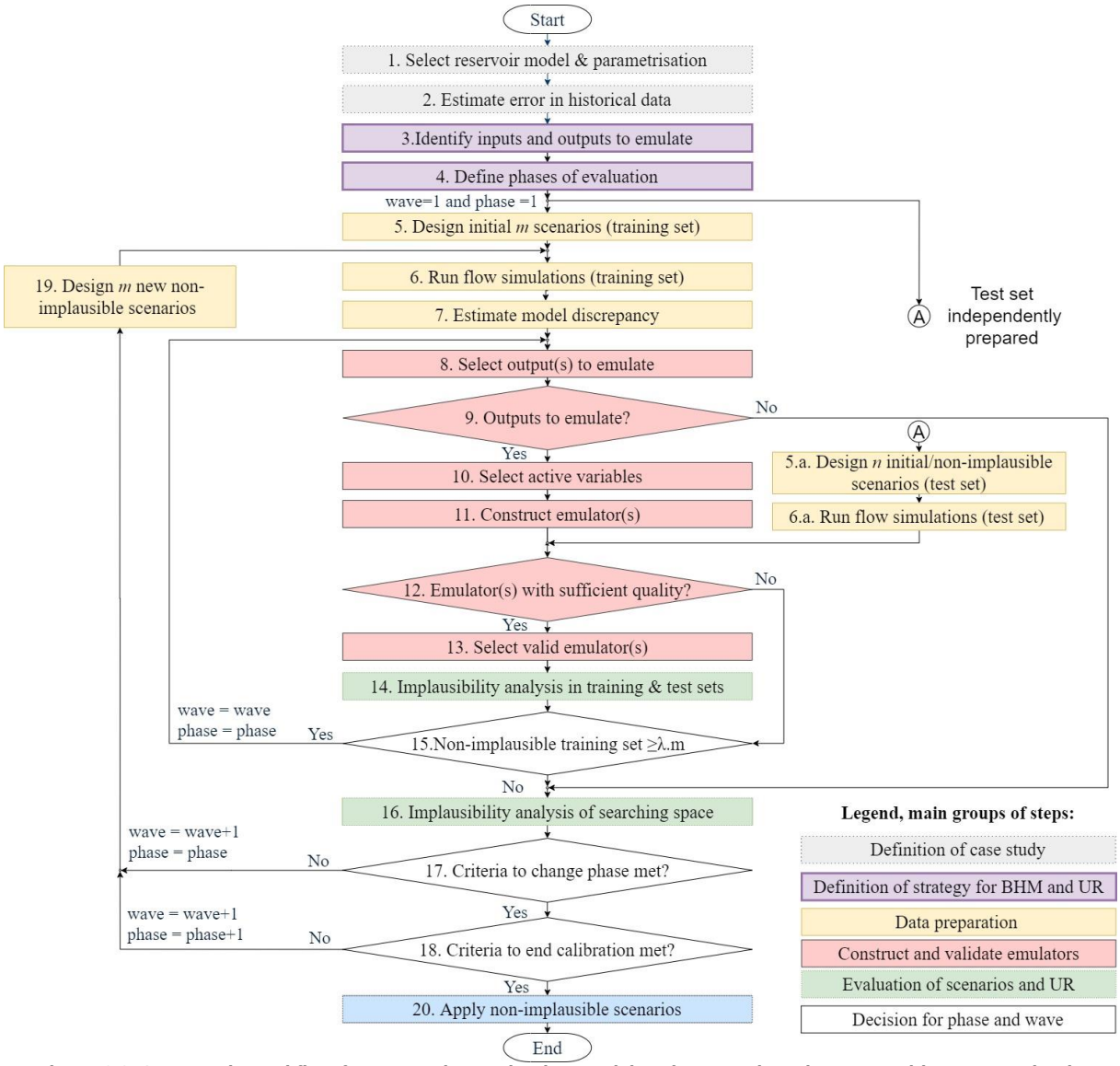


Figure 3.2. Systematic workflow for uncertainty reduction applying the Bayesian History Matching approach; six groups represent the 20 steps; *STEPS 1* to *4, 7, 8, 12, 15, 17, 18* and *20* concentrate activities for analysis and synthesis; the test set is independent and optional; each step of this high-level structure can be planned to answer requirements specific for a study; new waves require new simulations of scenarios; phases increment the amount of historical information considered in the process.

We highlight two core definitions in this systematic procedure:

- **Wave**: an iterative portion of the procedure whereby we simulate a limited number of scenarios. The corresponding quantities of interest are computed, emulated and used to reduce input space. To iterate through waves allows one to sequentially discard regions of the input parameter space, refocussing our search on the remaining non-implausible

inputs (Vernon *et al.* 2018). Waves greatly improve the efficiency of the uncertainty reduction process;

- **Phase of evaluation** (or phase): a distinct period in the historical period which is chosen considering reservoir behaviour and operational conditions. Calibrating models through phases allows one to (a) gradually incorporate the information available; (b) explore distinct relationships between uncertain attributes and reservoir outputs, for example, the physical relationships between simulation input and outputs are usually simpler in early time, while water breakthrough does not occur; and (c) reduce the computation effort spent in simulations. The definition of phases of evaluation require knowledge about the physical system to identify key behaviours related to reservoir management. On the extremes, a BHMUR strategy would consider each time step having data available or only the last time step of the history period. Expert judgement is expected to help compromising the arguments (a) to (c) highlighted in this paragraph. In Section 3.3.2, we exemplify relevant aspects of reservoir management to consider and Table 3.7 highlights the simulation time resulting from adopting phases of evaluation in the procedure.

To introduce the reader to the conceptual mechanism of waves and phases, we propose an illustrative example in Figure 3.3 for the training set. In Wave 1, we have 100 scenarios simulated until the end of the Phase of evaluation 1. Four valid emulators are constructed in this wave. The implausibility analysis of the set result in 18 non-implausible scenarios (smaller than $\lambda.m$=50). Since $D_{info,M}$ of Wave 1 is larger than 70%, the criteria in *STEP 17* is not met yet, we keep evaluating the same Phase 1 and move onto Wave 2.

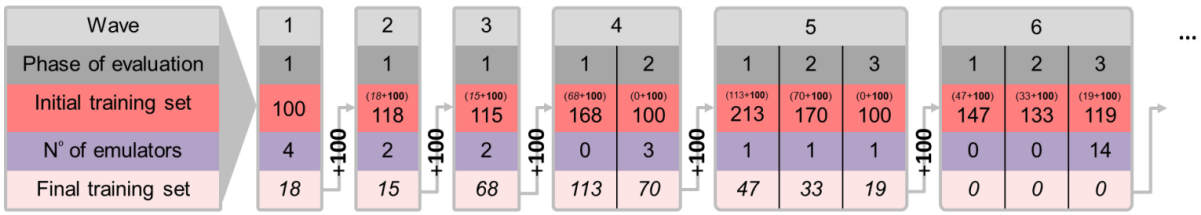


Figure 3.3. Conceptual mechanism of waves and phases in the systematic procedure – each wave run simulation models until a time which is characteristic from the corresponding phase; the size of the initial training set narrows to the final size after the consideration of the emulators constructed in each wave; each new wave increments m = 100 new simulated scenarios to the final training set of the previous wave; the need for a new phase is defined by the *STEP 17* criteria.

Wave 2 supplements the final training set of Wave 2 with 100 new non-implausible scenarios, resulting in an initial training set size of 118 scenarios. They enable the construction of two valid emulators, classifying as implausible 103 out of the 118 initial scenarios in Wave 2. Note that the row 'number of emulators' accounts only for emulators constructed in a given

wave. The emulators are evaluated cumulatively, meaning that in Wave 2, the total number of emulators evaluated is $4 + 2 = 6$.

The same logic of the transition between Wave 1 and 2 applies in the transition between Wave 2 and 3, which has an initial training set of 115 scenarios. The two emulators constructed in Wave 3 do not rule out a large proportion of the search space (*i.e.* are not informative enough). Then, *STEPS 17* and *18* call a new phase of evaluation, bringing in further (later in time) outputs for consideration. Wave 4 is the first wave that uses simulations that run until the end of Phase 2. Therefore, only 100 scenarios are available for this phase, but 168 is the size of the initial training set for Phase 1. The following waves keep a similar logic.

3.3.1 Definition of the case study

In *STEP 1*, the asset team defines a reservoir simulation model representing available information (*e.g.* seismic and well log data) and relevant uncertainties in the reservoir. Ranges and distributions define uncertain attributes which we expect to be informative about the reservoir behaviour. The uncertainty on these attributes arises, among other things, from our lack of knowledge of the sub-surface.

Uncertain attributes are classified in several ways: (a) uncertain physical properties themselves (*e.g.* rock compressibility, fluid viscosity) or pseudo-properties, which are not the actual properties but used to describe relationships between real-world quantities (*e.g.* effective permeability for phases, spatial properties of grid blocks); (b) as independent (*e.g.* fluid properties and rock compressibility) or dependent (*e.g.* porosity and permeability); (c) continuous and discrete (*e.g.* facies, PVT tables, fault existence).

Additionally, the asset team can set up the numerical simulator with tuning parameters and numerical approximations, among other things, which can have a significant impact in simulation time (Avansi *et al.* 2019), and give rise to additional uncertainties.

For the application of the proposed procedure, we established a synthetic black-oil reservoir model with a known hypothetical reality called HR-82 and based on the benchmarking case (Avansi and Schiozer, 2015). We describe the relevant features of HR-82 for this study in the main text. The full description is available in Appendix 3.A, where we detail how we defined the uncertain attributes.

An important feature is a sealing barrier separating the reservoir in two compartments (east and west blocks). The production and injection wells are represented by red and blue rectangles in Figure 3.4. The reservoir is spatially divided into 13 regions. Given the focus of this study, to describe and demonstrate methodological developments, we considered uncertain the parameters for regions 6, 8 and 10, leading to a total of 26 parameters, a sufficient number for our purposes. We highlight some steps that are planned to enable scalability on the number of dimensions (*see STEPS 3* and *10*).

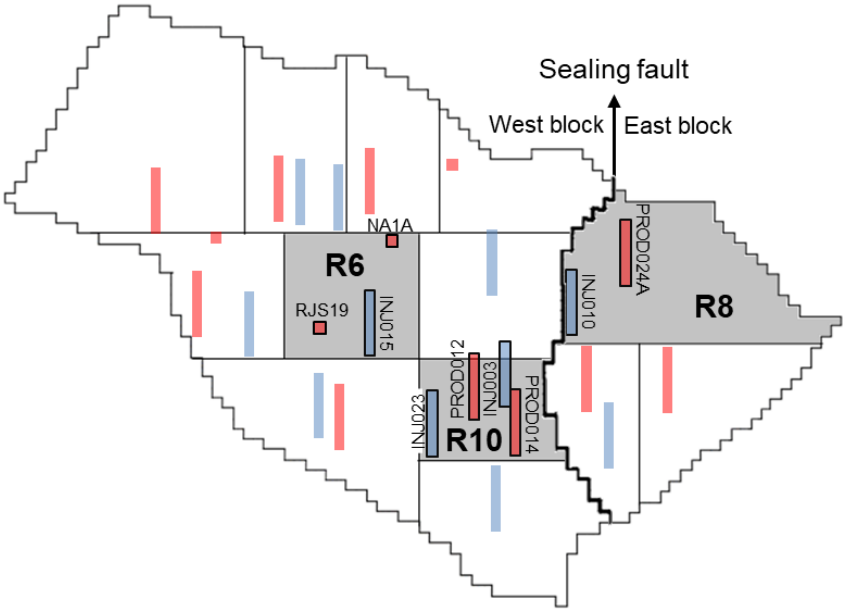


Figure 3.4. Two-dimensional aerial view of the reservoir model highlighting (1) the three regions considered uncertain, (2) the sealing fault diving the reservoir into two compartments, (3) the names of production and injection wells associated with local uncertain parameters called well index.

We defined all parameters as continuous and uniformly distributed, and we divide the 26 uncertain parameters of Table 3.3 into four classes:

- **Global parameters** (first two lines in blue, two uncertainties) influence the whole reservoir model;
- **Regional parameters** (four lines in green, twelve uncertainties) modify petrophysical properties of each the uncertain regions (regions 6, 8 and 10 in this study) with four parameters;
- **Sector parameters** (three lines in orange, three uncertainties) model attributes of the east block;
- **Local parameters** (last line in yellow, nine uncertainties) are independent multipliers of the Well Index (WI) for the nine wells located in the regions 6, 8 and 10.

Table 3.3. Uncertain parameters considered in the case study applied to demonstrate the systematic BHMUR procedure – parameters in blue have global influence; in green, regional influence (four attributes for each of the three regions); in orange, sector parameters influence only the east block; in yellow, local influence in nine wells. Appendix 3.A provides a complete description.

Parameter	Description	Ranges
c_p	Rock compressibility	[10, 96] E-6
c_{krw}	Relative permeability	[0.86, 1.28]
Mphi	Porosity multiplier	[0.75, 1.25]
A	kx, angular coefficient	[0.135, 0.175]
B	kx, linear coefficient	[-0.4, 1.1]
Mkz	kz multiplier	[0.1, 0.5]
$PVT_{co,EB}$	Oil compressibility	[1.40, 1.62] E-3
$PVT_{ai,EB}$	Oil viscosity related	[2.5, 50.0] E-4
WOC_{EB}	Water-Oil-Contact	[3169, 3179]
Wi_{ff}	Well index factor (9)	[0.6, 1.4]

In *STEP 2*, we select and estimate the error of the historical data. A critical analysis of the available data is performed in order to assess the levels of uncertainty in the information (*see* examples of sources of error in the measurement process in the Section 3.2.1. Formulation of Bayesian History Matching for Uncertainty Reduction).

In our application, the historical data of HR-82 is based on a hypothetical reality obtained from one of the simulated scenarios. The output from this scenario is noised up by adding both random and systematic errors. Table 3.4 presents the corresponding intervals of uncertainty in terms of $\pm 3\sigma$, where σ stands for the standard deviations of random or systematic portion of errors. Note that both contributions to the error are multiplicative in relation to the hypothetical measured data. We compute the corresponding variances of observational error $Var(e_i)$ to integrate into the implausibility measures.

Table 3.4. Error in the historical data. In our case study, it corresponds to the noise added to the production data of the hypothetical reality; random and systematic errors are defined in terms of three standard deviations and are multiplicative.

Observed data	Random error ($\pm 3\sigma_{z_i}^{ran}$)	Systematic error ($\pm 3\sigma_{z_i}^{sys}$)
Liquid production rate (q_l)	± 0.06	± 0.03
Water injection rate (i_w)	± 0.06	± 0.00
Water production rate (q_w)	± 0.05	± 0.05
Bottom-hole pressure of production and injection wells (p_{pbh}, p_{ibh})	± 0.02	± 0.02

3.3.2 Define the strategy for the BHMUR process

We dedicate *STEPS 3* and *4* to the definition of the strategy for uncertainty reduction. These activities are critical (a) to remain efficient in the use of computational resources, (b) to enable the dimensional scalability of our proposed workflow, and ultimately (c) to succeed in the quantification of uncertainty.

In *STEP 3*, we identify inputs and outputs to emulate, and we assess their relationship.

Firstly, our understanding of reservoir characteristics is used to perform informed judgements about physical relationships between outputs and uncertain attributes. This assessment avoids the consideration of spurious uncertain attributes as possible active variables for constructing emulators for some quantities of interest. For high dimensional case studies, we can also take advantage of (a) transformations in data, if coherent with physical laws and phenomena in place, background and examples available in Hoaglin *et al.* (1983), (b) supervised and unsupervised statistical learning techniques for dimensionality reduction.

Secondly, the selection of a sub-group of outputs to be considered during the iterative process makes sense for three reasons: (a) a high dimensional output space can be available (*e.g.* well data and seismic surveys) and trying to emulate all of them can be inefficient and unnecessary; (b) data quality can be diverse within the set available; (c) the assumption that all the quantities are independent is certainly not valid; therefore, the data structure can be considered viable for an effective dimensional reduction. For examples of dependency, *see* Fricker (2010).

We perform an initial evaluation of the appropriate outcomes to emulate. They can be of diverse natures, for example, single values (*e.g.* cumulative or rate at a given time); time series (*e.g.* coefficients of production curves); average values (*e.g.* average pressure on a given period); indicators of misfit between measured and simulated data (*e.g.* normalised quadratic deviation with sign, Almeida *et al.* (2014)).

For our case study, we identified inputs considering the sealing fault as the main reservoir feature, defining two non-communicating compartments, *e.g.* it is known that properties of the East block do not affect outputs of the West block. Uncertain attributes from the east block only influence outputs from wells in this compartment, similarly for west block inputs and outputs. Additionally, we assign the uncertain attributes with local impact to outputs from their respective wells. Therefore, emulators for outputs from the west block have a maximum of 11 active variables (two global, eight regional and one local attributes); in the east block, this maximum is ten active variables.

In *STEP 3*, as quantities of interest, we have identified average pressure (of production $\bar{p}_{pbh}$ and injection well $\bar{p}_{ibh}$), water breakthrough time and cumulative quantities

(liquid production L_p, water production W_p and oil production N_p, and water injection W_i) at the end of phases of evaluation. We approach the concept of phases in *STEP 4*. These quantities capture the reservoir behaviour related to material balance, communication between wells and the start of production of a new fluid phase in the wells (water). The average pressure is computed in the period between two consecutive phases of evaluation while the well is active (*e.g.* for Phase 2, the window is between 518 and 1,461 days). In parallel, we derive observational error from Table 3.4 corresponds appropriately to these quantities.

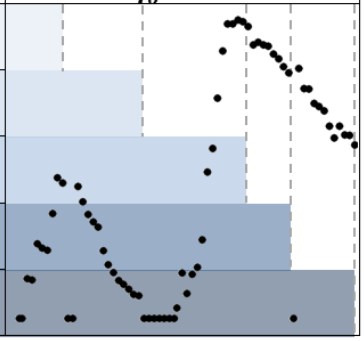
In Table 3.5, we summarise the pre-selected quantities of interest used to construct emulators. We explain why and how we emulate some quantities as binary outputs in the sections '4.2.1 Pattern 1: Simulator targets' and '4.2.2 Pattern 2: Breakthrough Time' in Part II (Formentin *et al.* 2020-b).

Table 3.5. Summary of quantities of interest identified in STEP 3, corresponding emulation techniques and models.

Quantities of interest emulated	Emulation technique	Emulation model
Cumulative quantities (L_p,W_i) Breakthrough Time (BT)	Two-class classification	Logistic regression and Linear discriminant analysis
Average pressure ($\bar{p}_{pbh}$, $\bar{p}_{ibh}$) Cumulative quantities (W_p, N_p)	Regression	Linear and quadratic regression

In *STEP 4*, we define the phases of evaluation. Phases of evaluation are windows in the historical period. They enable exploration of relationships between inputs and outputs while considering the evolution of time-based physics governing flow in porous media. Phases of evaluation formalise our understanding about temporal changes in the reservoir, for example, number of phases flowing, drainage area, recovery mechanism from early to late production stages.

Table 3.6. Characteristics of the five phases of evaluation defined for the case study; the window for each phase is from time zero to the one given in days; the plot of field oil production rate q_o *versus* time illustrates the time window.

Phase	Time	Characteristics	Field q_o *versus* time
1	518 days	Four active production wells in the WB, early stage, one well with water BT, before 1st maintenance stop	
2	1461 days	Last time with only 4 active production wells, two wells with water breakthrough (BT)	
3	2710 days	All 25 wells active, 21 wells opened for less than 700 days, 9 wells with BT, end of oil production plateau	
4	3256 days	Oil production plateau ends, 13 wells with water BT, before 2nd maintenance stop	
5	4018 days	Last time with historical data available, 14 wells with water breakthrough	

For the case study, the definition of phases of evaluation considers schedule of wells, water breakthrough, maintenance stops and field behaviour. Table 3.6 presents the characteristics of the phases and illustrates the corresponding time window in a plot for field oil production *versus* time.

Additionally, part of the strategy for BHMUR defines (a) how to combine the implausibility measure of the valid emulators; (b) the implausibility cut-off; and (c) all other design choices and decision statements in the workflow. The first two points depend on the stage of the reservoir modelling and calibration processes, the characteristics of the uncertain attributes (how much we know we do not know?) among others. For (a) and (b), we applied the choices discussed in the Section 3.2 - Statistical Methodology. We discuss the remaining choices within the corresponding steps.

3.3.3 Data preparation

In *STEP 5*, the objective is to sample a set of scenarios which is representative of the search space. The size of the training set (m) is a compromise between the affordability of simulations (*e.g.* a sample size as small as possible) and the accuracy of emulators (*e.g.* sufficiently large to enable the construction of informative emulators of sufficient accuracy in the current wave). In *STEP 5*.a, we design scenarios for the test set (sample size n). The test set is only used for the selection and validation of emulators. The workflow is adaptable to a situation where an independent test set is not affordable. Bastos (2010) describes several design possibilities, pointing differences between sampling strategies for training and test sets.

The specification of the design decision m depends on the number of uncertain attributes considered, the expected number of active inputs n_{Ai}, the complexity of emulated outcomes, and other practical aspects. To balance this decision, we can consider $\left(n_{A_i}+2\right)(n_{A_i}+1)/2$, which is the minimum number of points necessary to construct a quadratic response surface (Busby 2007). We highlight that the use of stepwise forward model selection to construct emulators with AIC or BIC criteria (Section 3.3.4) means that the limitation of runs never becomes an issue when fitting the linear model.

For our demonstration, we apply the space filling Latin Hypercube sampling technique (via lhs function in R) for generating samples from a multidimensional distribution. We defined $\mathrm{m} = \mathrm{n} = 100$ scenarios because we can run 100 simulations in parallel in the

cluster available. It is a reasonable size: the largest number of expected active variables in our application is 11, referred to in '*STEP 3* - Identify inputs and outputs to emulate'.

STEP 6 and *6.a* include the preparation of simulation files, simulation of scenarios and extraction of readable files in table format from the simulator.

In *STEP 7*, model discrepancy estimates the uncertainty about the simulation model in representing the real field (Equation 3.3). Goldstein *et al.* (2013) distinguish two types of model discrepancy:

- **Internal discrepancy**: This relates to any aspect of model discrepancy whose magnitude we may assess by experiments on the computer simulator. Internal discrepancy analysis gives a lower bound on the model discrepancy that we must introduce into our model analyses;

- **External discrepancy**: This relates to inherent limitations of the modelling process embodied in the simulator. There are no experiments on the simulator which may reveal this magnitude. It is determined by a combination of expert judgements and statistical estimation.

In reservoir engineering, the model discrepancy arises, for example, from lack of sufficient data or techniques to explore it, inaccuracy in reservoir size dimensions, in net-to-gross calculations, in reservoir architecture, spatial properties, upscaling models, fluid properties. After Ringrose and Bentley (2015), the aim in defining model uncertainties is to place our models within a framework that can overcome data limitations and personal bias and give us a useful way of quantifying forecast uncertainty. The need for considering model discrepancy and model error while performing history matching is studied in recent works related to petroleum reservoirs (Evensen 2018-a; Evensen 2018-b; Evensen and Eikrem 2018).

In our case study, we introduce a recurrent component of the model discrepancy due to the inherent errors in the liquid and injection rates used as simulation targets. These uncertainties arise in real case studies but are often neglected. Here we noised up the hypothetical reality outputs to mimic the real-world situation. Figure 3.5 represents this effect: *error-free* (brown line) is the direct outcome from the simulation of the hypothetical reality; *reference* (black dots) is the error-free outcome with independent random and systematic noise

added (with characteristics described in Table 3.4); *target=ref* (blue line) is the outcome of the simulation having the reference data as target.

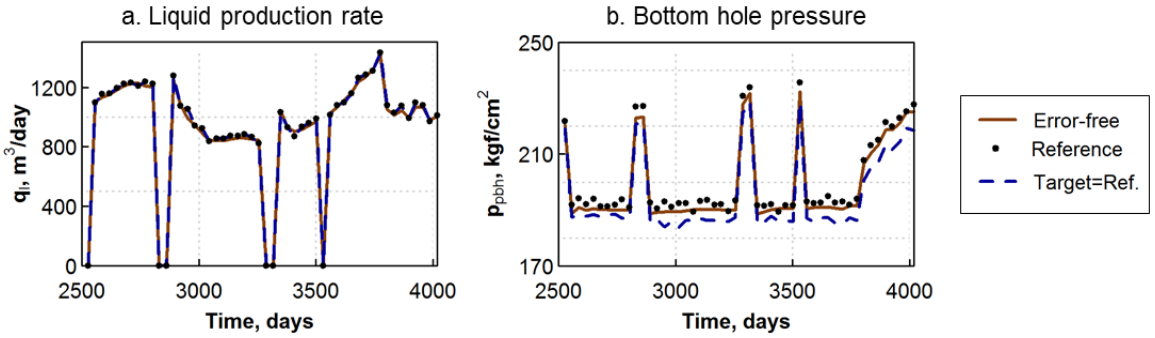


Figure 3.5. Representation of internal discrepancy due to noised liquid rate set as a simulation target, an example from the well PROD021. Liquid production rate is the simulation target. Error-free is the outcome from the hypothetical reality with no noise; Reference is the error-free data noised with random and systematic portions of error, Target=Ref is the simulation outcome having reference as target – (a) Liquid production rate of reference and Target=Ref are coincident; (b) Bottom-hole pressure reveals a source of model discrepancy with the bias of the blue curves in relationship with the other two curves.

In Figure 3.5.a for liquid production rate *versus* time, the reference and target=ref. are coincident, as the earlier is the simulation target. In Figure 3.5.b – bottom-hole pressure *versus* time, we verify differences between these three simulations. Target data is reached in Figure 3.5.a at the cost of an additional error on pressure and this additional error needs to be included in the calibration process.

Figure 3.5.b is a demonstration of internal discrepancy caused by the systematic error in the liquid production rate of HR-82: the model discrepancy is the difference between the reference data (black dots) and Target-Ref. (blue-dashed line) in the panel of Figure 3.5.b. In our application, we considered the discrepancy straightforwardly, but we emphasise the need to account for the model discrepancy in detail in reservoir history matching. Further discussions are available in Goldstein *et al.* (2013) and we leave a more advanced treatment to future work.

3.3.4 Construct and validate emulators

Selection of outputs to emulate (*STEP 8*) is critical for the effective use of information from a high dimensional output space and, ultimately, to the efficiency of the Bayesian History Matching for Uncertainty Reduction (BHMUR) process. We make a deliberate decision considering (a) the selection of all available quantities of interest, which would incorporate all the information available in the process; (b) the total time invested for emulation (construct, validate and evaluate new scenarios with the valid emulators), the most expensive decision; and (c) dependence between quantities of interest. By selecting a limited

number of outputs to emulate, we aim to identify the largest implausible region to rule out of our analysis while keeping the number of emulators low.

Because it involves some further subtleties, our workflow for *STEP 8* is fully described in Part II (Formentin *et al.* 2020-b). Firstly, we select the specific types of outputs to emulate. Secondly, we use the adapted implausibility measures $\tilde{I}_i$ and $\widetilde{D}_{info}$, presented in (Formentin *et al.* 2020-b), to estimate which combination of outputs has the highest potential to become the most informative. Each output selected in *STEP 8* is expected to respond more strongly to a sub-set of uncertain attributes – the active variables, and ideally, in a smooth way.

STEP 9 is a checkpoint, whether outputs are selected or not. It is mainly relevant when, in *STEP 8*, we try to select quantities of a phase φ which is earlier than the last phase simulated ($phase$), *see* Formentin et al. (2020-b) for further details of φ and $phase$.

For *STEP 10*, classical model fitting criteria such as AIC and BIC (using the "step" function in R) can be applied to select active variables for each selected output (Vernon *et al.* 2018, Vernon *et al.* 2010). The use of linear regression comes from our expectation that relevant relationships between attributes and outputs can be identified by linear operators as, if an input is at all active for a particular output, it will most likely induce some linear effect on that output, even if its actual functional dependence is far more complex and non-linear (Vernon *et al.* 2010).

With these active variables, *STEP 11* constructs several competitive emulators for each selected quantity. This step provides flexibility to our procedure. The time invested in constructing each competitive emulator can be estimated beforehand, generally depending on the statistical model and the number of scenarios in the training set.

Figure 3.6 details our workflow where: *STEP 11.1* defines the possible statistical models to be considered depending on the type of output (*see* Table 3.5). We can integrate multiple and diverse emulation techniques based on the requirements of the study. In our application, for continuous outcomes n_f*=2* encompassing linear regression with (1) linear terms and (2) with linear and quadratic terms. For two-class models $n = 3$ encompassing (1) linear discriminant analysis, (2) logistic regression with linear terms and (3) with linear and quadratic terms. We iterate in *STEPS 11.2 to 11.5* until all the competitive emulators are constructed.

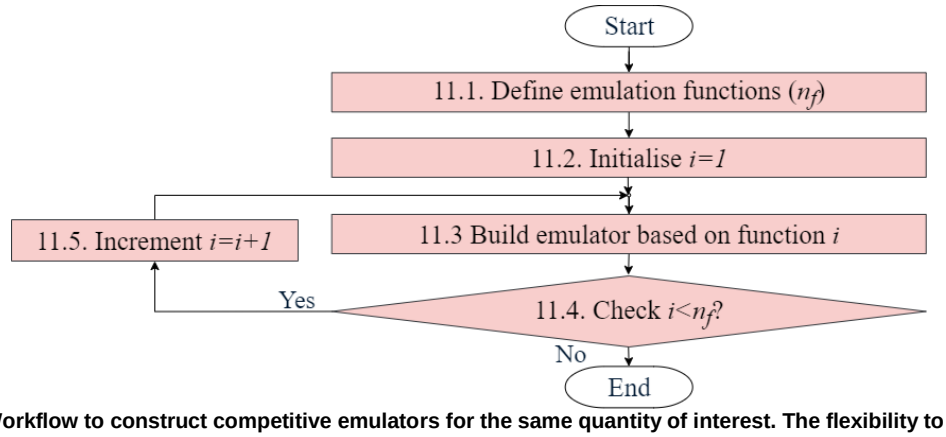


Figure 3.6. Workflow to construct competitive emulators for the same quantity of interest. The flexibility to incorporate diverse emulation functions brings an advantage to the systematic procedure.

In *STEP 12*, after constructing competitive emulators, we select the best emulator and validate them based on positive and negative predictive values (for two-class quantities) and on credible interval diagnostics and information index (for continuous quantities), detailed in Part II (Formentin *et al.* 2020-b). These indicators are calculated for training and test sets independently: while indicators for test sets are effectively used to select and validate emulators, we also monitor training sets. When test sets are not available, alternative criteria based on training sets can be applied to validate and select emulators (*e.g.* leave-one-out criteria).

3.3.5 Evaluation of scenarios and input space reduction

In *STEP 14*, all valid emulators are used to evaluate the implausibility for the training and test sets, checking how many non-implausible scenarios are available for each of these sets.

STEP 15 uses the number of non-implausible scenarios in the training set to identify the need for a new wave, which adds new non-implausible scenarios to the analysis. The value $\lambda.m$ – a minimum number of scenarios required to construct emulators - is a design option (*see* discussion related to *STEP 6*). For our case study, at least 50 scenarios in the training set are required to construct new emulators. If we have sufficient non-implausible scenarios in the training set (*STEP 15*), we have the possibility to construct new emulators.

In *STEP 16*, we follow with an implausibility analysis to evaluate the proportion of remaining space classified as implausible for all the emulators constructed up to this wave. As emulators are fast to evaluate, we can afford to evaluate a large number of new scenarios ($\gg 10^5$).

We described *STEPS 17* and *18* in Figure 3.2. The criteria to change phase (*STEP 17*) is set as $D_{info,M} < 70\%$ and the criteria to end calibration (*STEP 18*) is the last phase to be evaluated (jointly with the previous condition $D_{info,M} < 70\%$).

STEP 19 (design of new non-implausible scenarios) is performed in the situation where the workflow identifies the need for a new wave. We proceed as follow: (a) identify the range for each uncertain parameter to be sampled (*e.g.* maximal and minimal values for each parameter, which can be updated considering the current non-implausible space); (b) sample a large number of scenarios via Hypercube Latin sampling (same *lhs* function from R as the one used in *STEPS* 5 and 5.a); (c) evaluate the combined implausibility of those scenarios when considering all the emulators constructed until the current wave; (d) reject the implausible scenarios; (e) repeat the process until the required number of non-implausible scenarios is obtained; (f) check that the sets of scenarios fill the space appropriately via pairs plot; and (g) proceed the next steps with the non-implausible scenarios. These new scenarios are non-implausible for all the emulators constructed until the current stage of the analysis.

3.3.6 Apply non-implausible scenarios

When *STEP 18* indicates that we completed the uncertainty reduction, we proceed with the application of non-implausible scenarios. One possible application is for reservoir forecasting, which involves several additional concepts as described in Goldstein and Rougier (2006), Busby *et al.* (2007), Craig *et al.* (1997) and Craig *et al.* (2011).

3.4 Results and Discussions

Here we discuss important features resulting from the application of our systematic procedure to the case study HR-82. There are four figures of the process that we would like to highlight:

- We used 15 waves for 5 phases of evaluation;
- We evaluated 3,000 scenarios with the simulator with different stopping times; the simulation time would be equivalent to 1,637 full simulations without phasing;
- We obtained valid emulators for 115 quantities of interest out of a total of 198 selected quantities;

- We reduced the original space to 3.58e-11% (non-implausible space) by the end of the calibration process.

These records indicate that this procedure may be more efficient and far less expensive than traditional methods. We note that traditional methods often do not attempt to identify all input points consistent with historical data, which are essential for an accurate, unbiased prediction. We now look at a breakdown for each phase of evaluation in Table 3.7 and the performance of this process for different perspectives. It is worth noting that the cost-benefit between simulation time and space reduction in earlier phases is quite low, *i.e.* we considerably reduced the search space with short and cheap simulations.

Table 3.7. Breakdown of figures from our application based on phases of evaluation. For each phase of evaluation, we display the period simulated (a window between zero and the time given in seconds), the total number of scenarios simulated, the median time of the time invested in simulation, the number of emulators constructed in the phase, the mean of the time to evaluate 100,000 scenarios with emulators and the remaining proportion of the original space given as non-implausible.

Phase	Period simulated	Scenarios simulated	Median time simulation	Number of emulators	Average time emulators (100,000 scenarios)	Non-implausible space
1	518 days	600 (3 waves)	66 s (12% of the max)	8 (7%)	1.41 s	2.14%
2	1,461 days	200 (1 wave)	237 s (42% of the max)	3 (3%)	1.71 s	1.48%
3	2,710 days	1,600 (8 waves)	339 s (60% of the max)	73 (63%)	18.26 s	3.15e-9%
4	3,256 days	400 (2 waves)	456 s (81% of the max)	21 (18%)	21.83 s	8.68e-11%
5	4,018 days	200 (1 wave)	565 s (max)	10 (9%)	24.81 s	3.58e-11%

Phases of evaluation enable the efficient use of resources for reservoir simulation. We consider that the median of the time to simulate one scenario is representative of the process. The time to simulate scenarios in each phase is plotted in boxplots (Figure 3.7.a). From the total number of scenarios simulated, 20% were until the end of Phase 1 where the simulation time of each scenario corresponds to 12% of the median simulation time for scenarios until the end of Phase 5. The short simulations of Phase 1 allowed emulators to be constructed that ruled out 97.86% of the search space as implausible, *i.e.* very unlikely to match with the historical data.

The time spent to evaluate a new scenario with emulation is a fraction of the time spent to evaluate a new scenario through simulation. Figure 3.7.b presents the number of emulators constructed in each phase, highlighting that each phase considers all the emulators constructed since the beginning of the process (it is cumulative). Figure 3.7.c presents the time spent to evaluate scenarios with emulators: for Phase 5, in less than 25 seconds, we evaluate

100,000 scenarios with emulators, while it took 565 seconds to evaluate only one scenario through simulation, giving a speed increase of six orders of magnitude (note: we can run 100 scenarios in parallel in the cluster).

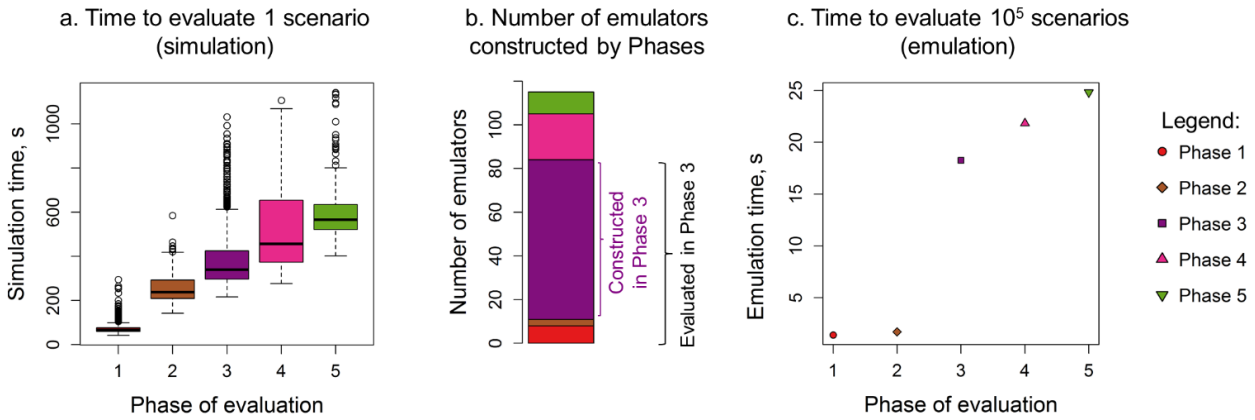


Figure 3.7. Analysis of time to evaluate scenarios by phases of evaluation – (a) Time required to evaluate one scenario through simulation, we report the median as the representative measure; (b) Number of emulators constructed in each phase, the example of interpretation for Phase 3 indicates that we constructed 73 valid emulators in this phase, but we evaluate the cumulative number of 84 emulators; (c) Time required to evaluate 100,000 scenarios through emulation, highlighting the large uncertainty reduction between Phases 2 and 3.

The potential of combining simulation and emulation techniques is evidenced: using simulation, we capture physical phenomena in place in the reservoir; with emulators, we speed up the parts of the history matching process requiring a sufficiently large number of evaluations to be consistent from a statistical perspective.

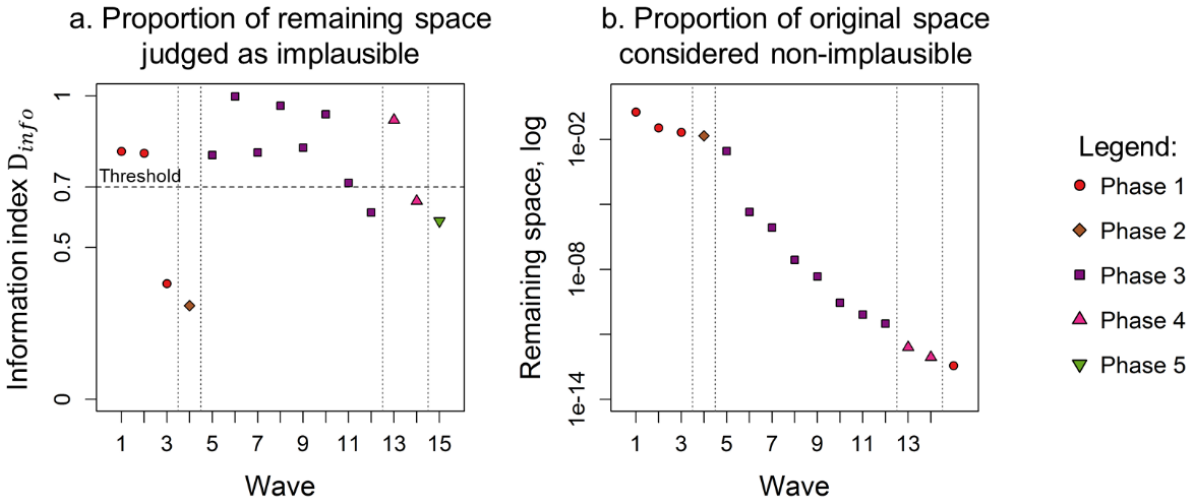


Figure 3.8. Overview of uncertainty reduction through waves – (a) Estimated proportion of the remaining space ruled out by the emulators; (b) Estimated proportion of the original space considered non-implausible.

For each wave, Figure 3.8 presents (a) information index D_{info} computed from a large number of scenarios during the implausibility analysis (*STEP 16*) and represents the proportion of remaining space judged as implausible, and (b) the proportion of initial space considered non-implausible, remaining in the analysis (note the log scale). In Figure 3.8.a, we

highlight the cut-off of 70% defined as criteria to change phase in *STEP 17*: each wave that was unable to achieve this threshold led to an increment in the phase of evaluation for the following wave. This happened in our application in Waves 3, 4, 12 and 14. Wave 15 evaluated Phase 5 and failed to reach 70%, leading to the end of the procedure by this wave.

Wave 4, evaluating Phase 2, had the lowest proportion of space cut-out among all the waves. Indeed, the same four wells of Phase 1 are active in Phase 2, leading this particular phase to a modest role in the uncertainty reduction. Phase 3 kept being highly informative for seven consecutive Waves (5 to 11). That level of uncertainty reduction is due to the 21 additional active wells in both east and west block of the reservoir. By the end of Wave 12, the proportion of the original space considered non-implausible is $3.15\text{x}10^{-9}\%$.

Phase 4 was highly informative only for Wave 13, which is coherent with the characteristics of this phase (*i.e.* the water breakthrough occurs for four additional wells). The fifth and last phase of the process did not offer new events to the reservoir historical data, leading to an unremarkable performance in terms of uncertainty reduction.

By the end of the calibration process, the hypothetical reality is also judged as non-implausible: a useful consistency check. The final proportion of input space remaining as non-implausible was 3.58e-11%. We see that this substantial uncertainty reduction is a direct result of the interplay between the simulator behaviour and the specification of all the required uncertainties, although we note that in high dimensional space such large reductions are expected. Note also that our approach should not suffer the problems of "ensemble collapse", as can be seen by the widely dispersed input points by the end of Wave 12, Phase 3, in Figure 3.9.

3.4.1 Uncertainty Reduction in search space and quantities of interest

The iterative process implemented narrowed the search space sequentially. Each new wave re-established the search space on non-implausible regions, from where a limited number of scenarios is sampled and used to construct additional emulators.

Figure 3.9 shows pairs plots for 18 uncertain attributes by the end of Wave 3 (left plots) and Wave 12 (right plots). The description of each attribute is presented in Table 3.3, but we highlight that they are more relevant to regions 6, 10 and 8, respectively. The implementation of the software to design those panels follows Vernon *et al.* (2010).

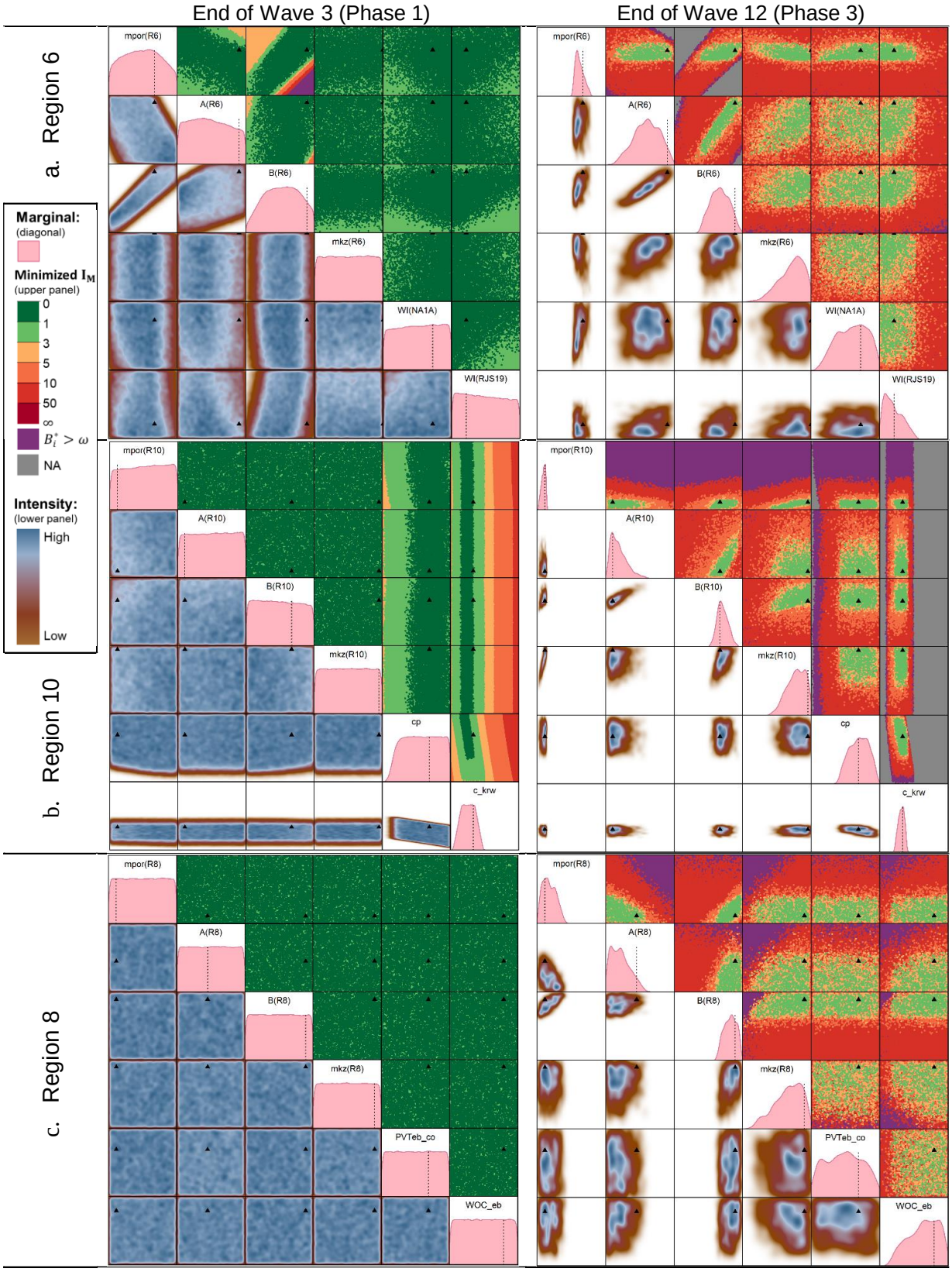


Figure 3.9. Non-implausible space for 18 uncertain attributes mostly relevant for regions 6 (top), 10 (middle) and 8 (bottom). The left side plots are by the end of Wave 3 (before the start of Phase 2); the right side plots by the end of Wave 12 (before the start of Phase 4); diagonals with the marginal density, upper panels the minimised implausibility and lower panels show the intensity of non-implausible scenarios. All plots indicate the hypothetical reality which correctly remains non-implausible. Note that the hypothetical reality of the attribute mkz(R6) for region 6 - fourth attribute in the diagonal – is the highest value on the right of the interval set for this attribute.

The graphics are based on a large number of scenarios applied in the implausibility analysis, which combined implausibility measure is evaluated. The structure of the pairs plot is a matrix of scatter plots (*e.g.* in R, we can use the commands *mfrow*, *mar*, *oma* and *pty* from *par* function to set this type of display), R Core Team (2018). We then iterate in the index of this matrix of graphics, plotting the data corresponding to the uncertain attributes using the appropriate visualisation tool.

For each pairs plot of Figure 3.9, the diagonal provides the marginal density of each attribute in the non-implausible space, providing a sense about the dispersion of the uncertain attributes compared to the initial distribution (uniform along the limits plotted). We highlight with dotted vertical lines the hypothetical reality. For some attributes, the hypothetical reality value is on (or very close to) the boundary of the initial range and superposed by the box of the diagonal plots (*e.g.* mkz(R6) in the plot for region 6). We implement the diagonal plots using the function *density* together with *polygon* of R (R Core Team, 2018) for the non-implausible scenarios of the corresponding uncertain parameter; *text* and *line* are applied to add the name of the uncertain attribute and the value of the Hypothetical Reality, respectively.

The upper panel (above diagonal) plots the minimised implausibility using an appropriate scale and show non-implausible regions in green (light and dark green represent implausibility smaller or equal to 3). In purple, we identified the regions ruled out by two-class emulators constructed in each phase. In red to yellow colours are the regions judged as implausible by emulators for continuous outputs. Grey (in the legend named NA) indicates regions ruled out up to Wave 3. Therefore, these regions do not make part of the search space of later phases. For both upper and lower panels, we plot the hypothetical reality as a black triangle.

To implement the graphics of minimised implausibility, we follow the procedure of Vernon *et al.* (2010). Firstly, we use *rev* and *order* functions of R (R Core Team, 2018) for the set of combined implausibility measure of the scenarios: in this way, we obtain the order which the software should plot the scenarios. The function *cut* is set with an appropriate number of breaks, which will be applied to define a limited number of colours in the graphic (*e.g.* the intuitive colour for non-implausible scenarios is green and for implausible scenarios is red). With those elements, we are able to plot a scatter graphic of the scenarios for the relevant uncertain parameters. Using the reversed order to plot the scenarios results in the visualisation of the minimised implausibility measure.

The lower panel plots (below diagonal) give the intensity (or optical depth) of non-implausible points for each pair of attributes (Vernon *et al.* 2010). Dark blue indicates that a high concentration of non-implausible scenarios is evaluated in the region. This plot provides a sense of the depth of the non-implausible space. To implement the lower panel, we compute the two-dimensional kernel density estimation (in R: *kd2d* function of the MASS package) for the non-implausible scenarios, selecting the appropriate uncertain attributes for a given graphic. We set the colours using *colorRampPalette* function in R and plot this density with the function *image*. Such pair plots allow us to visualise information of a high-dimensional space in an intuitive and informative way when performing an implausibility analysis.

Firstly, we analyse the plots from the end of Wave 3, which is the last wave evaluated until Phase 1. We highlight that in this phase (up to 528 days), only four vertical wells are active. We observe that most reduction of search space is on the attributes related to the porosity of Region 6 (mpor, A and B, left upper panel) and on global attributes (mainly c_{krw} related to water relative permeability, but also slightly rock compressibility, left middle panel). Intensity plots (lower panel) also reveal that the emulators captured the influence of other uncertain attributes (well index WI of RJS19, mpor and B of region 10).

Until Phase 1, two-class emulators are only applied for cumulative liquid production. The purple colour highlights that the regions ruled out by two-class emulators are far from the hypothetical reality (which is coherent with the arguments provided in the Section 4.2 Two-class quantities of interest and emulators, Formentin *et al.* 2020-b). The uncertain parameters of the east block (left lower panel) do not present any indication of uncertainty reduction, which is expected since no well of this region is active until 528 days.

Secondly, we focus on the plots for the end of Wave 12 (right), which is the last wave only evaluating until Phase 3 (2,710 days). Figure 3.8 showed that most of the uncertainty reduction occurs in the evaluation of Phase 3. Figure 3.9 highlights the regions of the search space remaining in the analysis. The diagonal reveals that the ranges of several attributes concentrate around the hypothetical reality.

Nevertheless, the concentration of points is not always symmetric around the hypothetical reality, nor should we expect them to be, especially when it falls in the corners of the non-implausible regions (*e.g.* A and B of region 6, in the upper plots) or in the limit of the ranges defined in *STEP 1* of the systematic procedure. The noise added to the target data, and

its corresponding model discrepancy will also contribute to shifting the cloud of non-implausible points around the true hypothetical reality.

The upper panels highlight several regions of the search space ruled out by the two-class emulators, especially noticeable for mpor(R10). Finally, the lower panels emphasise that the remaining non-implausible space is a fraction of the original space.

3.4.2 Successive waves and phases of evaluation

We dedicate this section to discuss the iterative principle of BHMUR applied in our systematic procedure. To iterate in waves allows the discarding of regions of the input parameter space sequentially and the application of simple emulators to mimic a complex simulator and its input space in a reliable way.

Figure 3.10 compares two emulators constructed for cumulative water production of the well RJS19 in Phase 3 (2,710 days). On the one hand, Figure 3.10.a plots a design of scenarios representing the original search space with five uncertain attributes in light blue. Note that this is only an illustrative example, supposing the case that we did not use phases of evaluation; this training set is not used in the calibration process (*i.e.* in Wave 1 we do not evaluate any scenario until Phase 3). This design of the original search space is used to construct an emulator, the cross-plot of which is presented in Figure 3.10.a in light blue. We observe a non-linear pattern, *i.e.* the simulated and emulated outcomes do not follow a straight line.

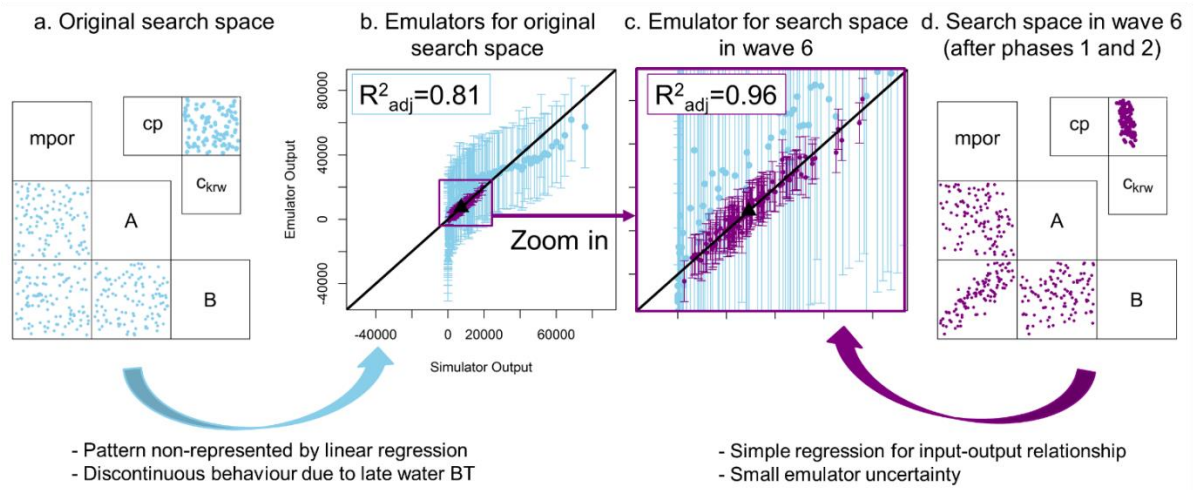


Figure 3.10. Cross-plot to compare two emulators for cumulative water production (W_p) of well RJS19 in Phase 3. Light blue is related to the emulator constructed from the original search space with illustrative purpose; purple is related to the emulator constructed during the application of the procedure, in Wave 6. The vertical bars show the uncertainty of the emulator within an interval of $\pm 3sd(f_i^*)$; (a) Five uncertain attributes illustrate the design of the scenarios sampled on the original space; (b) Emulator constructed with the sample on the original space, scaling the second emulator in purple; (c) Zoom in *b* to highlight the diagnostics for the emulator constructed from the search space in Wave 6; (d) Design of the sample on the search space in Wave 6, some regions already considered as implausible.

Moreover, the discontinuous behaviour due to a late water breakthrough time concentrates several points following the zero water production from the simulator output. Finally, the emulator uncertainty plotted in error bars $\pm 3sd\left(f_i^*(x_{A_i})\right)$ is very large. These diagnostics suggest that a more complex statistical model is required to accurately represent this level of complexity in the relationships between input and outputs existent in Figures 3.10.a and .b which would require many more simulator runs to train.

On the other hand, Figure 3.10.d plots a design of scenarios representing the search space in Wave 3 with five uncertain attributes in purple. Some regions of the space were already ruled out in previous waves, which is evident for mpor, B and c_{krw}. This design is used to construct an emulator, with the cross-plot presented in Figure 3.10.c in dark purple.

We compare this emulator with the first one through a zooming box relating Figures 3.10.b and c. We observe a linear pattern, *i.e.* the simulated and emulated outcomes follow a straight line, with a higher $R^2 adj$, which by definition indicates the proportion of the variance of the output explained by the regression (Table 3.2). The absence of simulated scenarios with no water breakthrough time ($W_p = 0$) is coherent with a limited range of the uncertain attribute related to water relative permeability c_{krw}. Finally, the emulator uncertainty plotted in bars is far narrower.

These diagnostics indicate that this simple statistical model is able to properly represent the relationships between input and outputs over the current non-implausible space, demonstrating the strength of the iterative approach: smaller volumes of input space in later waves are substantially easier to emulate as we have removed the erratically behaved and physically irrelevant regions from the search space.

Reducing the search space through the waves corresponds to reduce the uncertainty about the behaviour of the reservoir. In Figure 3.11, we summarise indicators for six emulators constructed for the average pressure of PROD024A ($\bar{p}_{pbh}$) in Phase 3 over Waves 6, and 8 to 12. Figure 3.11.a highlights the sequential reduction of the standard deviation (sd) of the quantity of interest, reduction of emulator uncertainty (residual standard deviation) and information index.

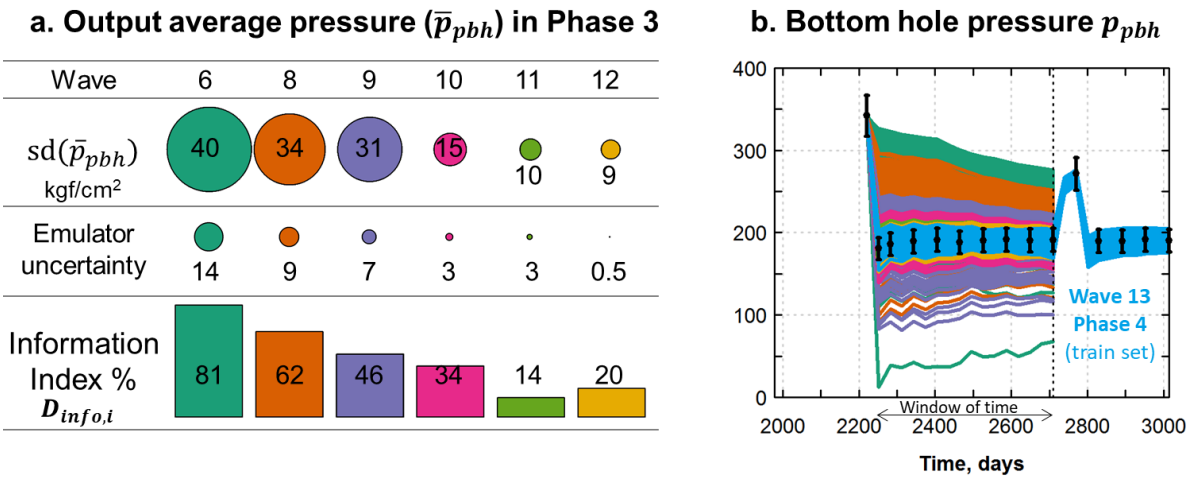


Figure 3.11. Uncertainty reduction with emulators – Production well PROD024: Example for through 6 waves, (a) indicators for emulators constructed for average pressure in Phase 3, based on the window of time presented with a horizontal arrow in *b*; (b) bottom-hole pressure *versus* time; the vertical line in 2,710 days highlights Phase 3; the colour of the curves corresponds to the scenarios used to construct emulators in the waves mentioned in *a*; the uncertainty reduction is highlighted by the pressure curves plotted by the end of Wave 12 and beginning of Wave 13 (in blue).

The corresponding pressure in time (Figure 3.11.b) corroborates the idea that we are narrowing down to the regions of interest in our input space and calibrating the model. Wave 6 is the first wave that this quantity of interest was selected as output to emulate (*STEP 8*). The high standard deviation of the average pressure (40 kgf/cm^2) is aligned with the spread of the corresponding pressure curves, in a range between the lower pressure limit and around 350 kgf/cm^2. The emulator uncertainty is sufficient to allow a reduction of 81% of the remaining search space (*e.g.* a very informative emulator).

In Wave 7, this quantity of interest is not chosen as output to emulate in *STEP 8* (Figure 3.2). The emulator of Wave 8 is constructed over a narrower search space, and the corresponding standard deviation of the average pressure is smaller, highlighting that no scenarios reached the lower limit boundary. The emulator uncertainty is smaller, and the information index keeps on a high level (62%).

The following emulators keep the trend. Comparing the figures for the emulator of Waves 11 and 12, we observe that the standard deviations of the average pressure are very similar (10 and 9), but the fact the emulator uncertainty is six times smaller in the emulator of Wave 12 leads to a higher information index for it (20% of the remaining space is ruled out).

3.4.3 Analysis of uncertainty reduction for production and injection wells

To show the potential of the systematic procedure proposed, we introduce the adapted implausibility measure $\tilde{I}_i$ (in boxplots) presented in Section 4.3.1 Information index

from Formentin et a. 2020-b; in a pragmatic way, $\tilde{I}_i$ is the implausibility measure (Equation 3.3) computing the variance of model discrepancy and of the observation error in consideration but with the variance of the emulator equals to zero (Formentin *et al.* 2020-b). This measure compares the simulation output against the historical data, a simplification of the traditional implausibility measure.

Figure 3.12 shows these results for production wells in Phase 5 (*i.e.* 4018 days). Plots from the original search space (Figure 3.12.a) are presented with their corresponding boxplots from the end of the calibration process (Figure 3.12.b). Note: we simulated scenarios from the original search space until Phase 5 for comparison purposes only, *i.e.* these scenarios were not applied in the analysis. The grey regions highlight the interval which scenarios are considered non-implausible, but also provide a sense of scale since the y-axis from the right plots are much smaller than the ones from the right side (*i.e.* the outputs are much closer than the historical data after the calibration process). The *x-axis* presents the names of the wells, which are the same for all production wells (labels provided on the last plot for production wells). We highlight specific wells in the box plots and present the corresponding well data.

The first pair of plots in Figure 3.12 is for cumulative oil production (N_p). Initially, these quantities of interest are distant from the historical data (*e.g.* several wells have the median above the threshold of three). Several outliers (mainly for wells NA1A and RJS19) indicates scenarios poorly representing the reservoir behaviour. We plot the well data oil production rate of NA1A. After the calibration process, all quantities of interest are within the grey region, have a very low median and the interquartile range describes much more calibrated scenarios.

The second pair of plots in Figure 3.12 is for cumulative water production (W_p). By the end of the calibration process, the medians are closer to the threshold of three. Nevertheless, the fact the median for PROD024A is above the grey region indicates that this well could be subject of further evaluations, possibly with new waves. For that, the criteria established in *STEP 17* and/or *18* could be replaced by the adapted implausibility measure of all quantities of interest evaluated, for example. We plot the well data water production rate of PROD024A. In our case study, higher medians W_p than N_p are related to the fact that both components of observational error are proportional to the production rate (Table 3.4). In this way, we have a smaller variance of the observational error for water production, which is inversely related to the adapted implausibility measure.

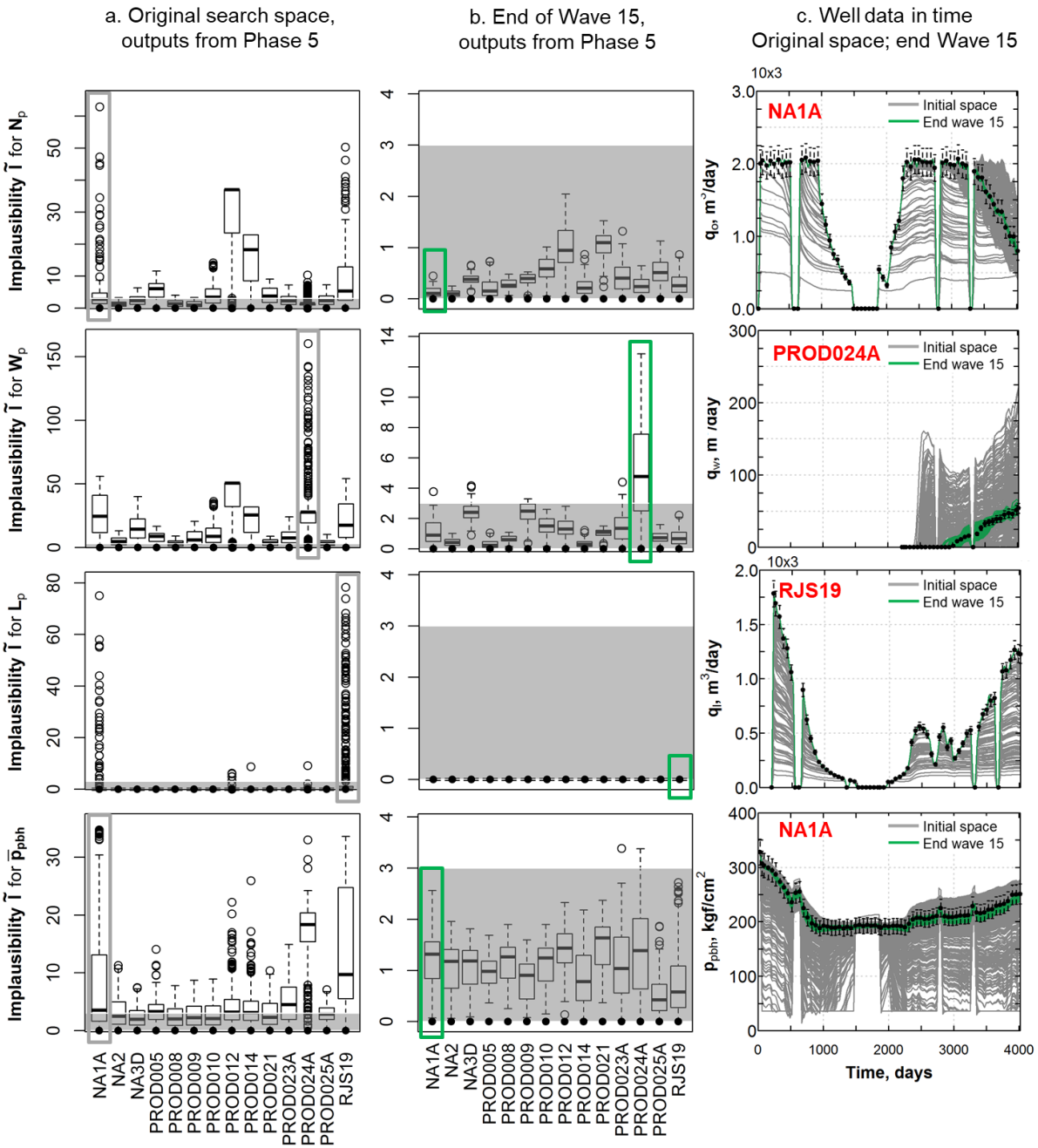


Figure 3.12. Adapted implausibility measure for simulated scenarios of all production wells for (a) initial search space and (b) after the end of the calibration process. The labels are the same for x and y-axis of figures (a) and (b). We select the production curve of a specific well for demostration purposes in (c), where we plot well data in time with curves simulated from the initial search space and by the end of the calibration process, labels in x-axis are the same for the plots of (c), error bars express observational error and discrepancy considered. Note that the third plots in the vertical are for liquid production which is a simulation target; therefore, adapted implausibility measure is expected to be close to zero by the wave 15 in (b).

The third pair of plots is for cumulative liquid production. Five wells (NA1A, PROD012, PROD014, PROD024A and RJS19) initially do not meet the liquid target informed to the simulator. By the end of the procedure, all wells meet the target informed to the simulator (*e.g.* no scenario reaches the lower limit pressure in the regions of the wells). We plot the liquid production rate of the well RJS19. The average pressure of production wells for the window

between Phases 4 and 5 shows a definite improvement in the representation of the behaviour of the wells by the end of the historical period. We plot the bottom-hole pressure for NA1A.

We observe the adapted implausibility measure for injection wells in Figure 3.13.

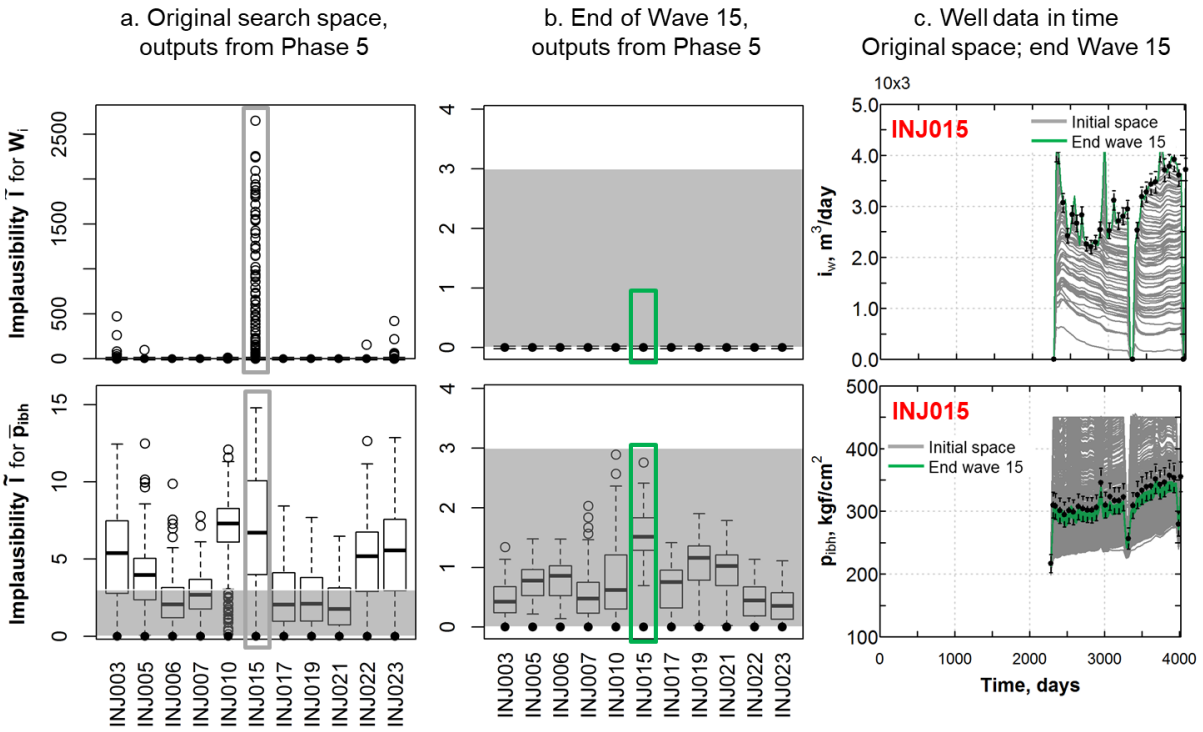


Figure 3.13. Adapted implausibility measure for simulated scenarios of all injection wells for (a) initial search space and (b) after the end of the calibration process. The labels are the same for x and y-axis of figures (a) and (b). We select the injection curve of a spectific well for demostration purpose in (c), where we plot well data in time with curves simulated from the initial search space and by the end of the calibration process, labels in x-axis are the same for the plots, error bars express observational error and discrepancy considered. Note that the first plots in the vertical are for water injection which is a simulation target; therefore, adapted implausibility measure is expected to be close to zero by the wave 15 in (b).

For the injection wells (Figure 3.13), initially, six wells (INJ003, INJ005, INJ010, INJ15, INJ22 and INJ23) could not inject as much as the simulation target. This lower volume injected is associated with scenarios reaching the upper-pressure limit, set in 450 kgf/cm^2. We illustrate the liquid production rate and bottom-hole pressure of the well INJ015 as examples. All the scenarios evaluated by the end of the calibration process meet the injection rate target. The average pressure of injection wells for the window between Phases 4 and 5 indicates that all scenarios have adapted implausibility measure below the threshold of 3.

We provide data for the well PROD021. Figure 3.14.a shows that liquid production rate (Figure 3.14.a) is the same as the simulator target during all the production time since the start of the calibration process. We observe bias in the bottom-hole pressure (Figure 3.14.b) of calibrated scenarios: all of the green lines are below the reference data. We emphasise that this

specific well presented an apparent internal discrepancy due to the noisy simulator target, as shown in Figure 3.14.c.

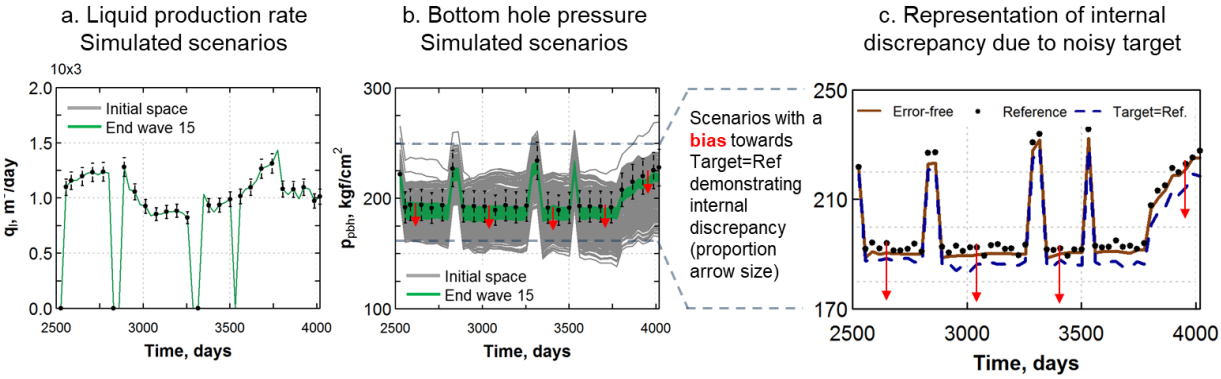


Figure 3.14. Well data for PROD021 - (a) liquid production rate for original space and scenarios after calibration, (b) bottom-hole pressure for original input space and scenarios after calibration highlighting a bias, (c) zoom in the representation of discrepancy caused by errors in target data.

3.4.4 Uncertainty Reduction in forecasting

In Figures 3.15 and 3.16, we plot scenarios from the initial search space (grey) and from the Wave 15 (green) in order to provide a visual confirmation about the quality of the systematic uncertainty reduction. The reference data from the scenario considered as the hypothetical reality is in black dots. The evaluation of forecasting is available because our application is a hypothetical reality. The forecasting information is not available in real cases. In this case, this analysis would require the limitation of historical data used in the calibration process to a fraction of the total historical data, and the remaining historical data could be explored to evaluate the results.

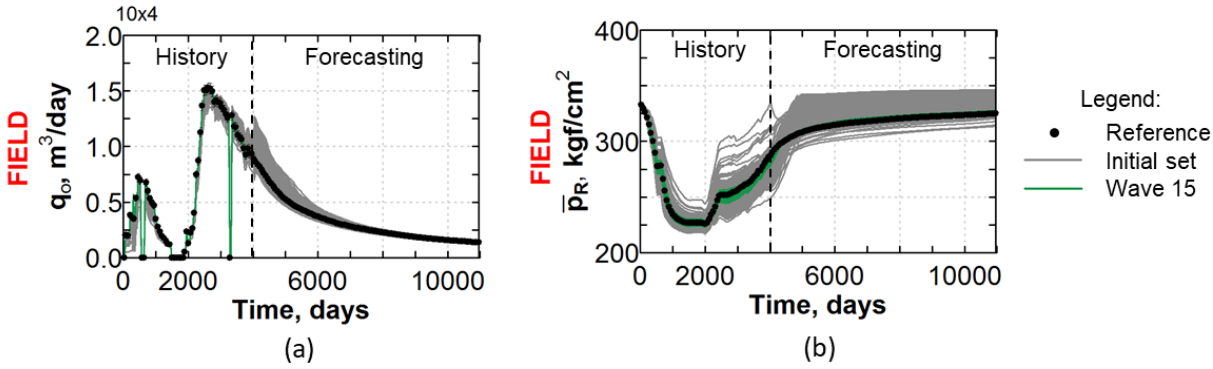


Figure 3.15. Time series for field data demonstrating the uncertainty reduction in forecasting from scenarios from the initial space in grey to non-implausible scenarios by the end of the wave 15, in green– (a) oil production rate, and (b) average reservoir pressure.

Figure 3.15 presents data for the field: (a) oil production rate and (b) average reservoir pressure during the historical period followed by 19 years of forecasting. The results

show a concentration of all scenarios around the reference data, with limited variability. This is expected from our case study:

(a) The uncertainty of the reservoir model to represent the hypothetical reality is tine (in more realistic studies, the uncertainty related to model discrepancy would be expected to be considerably larger);

(b) The observational errors of the historical data are well known and characterised in the process;

(c) A large period of historical data is available, *e.g.* we have enough information to cut out the parts of the search space that are implausible.

A relevant consistency check that our application resulted in an appropriate level of uncertainty reduction is that the scenario 'hypothetical reality' is kept as non-implausible by the end of the 15 waves.

Figure 3.16 presents (a) liquid production rate for the well RJS19 and (b) water injection rate of INJ003 during the historical period followed by 19 years of forecasting. Although these data are set as targets during the historical period, several scenarios from the initial set did not reach the reference data, especially in Figure 3.16a. The forecasting data demonstrated an apparent uncertainty reduction towards the reference data, although some bias is observed.

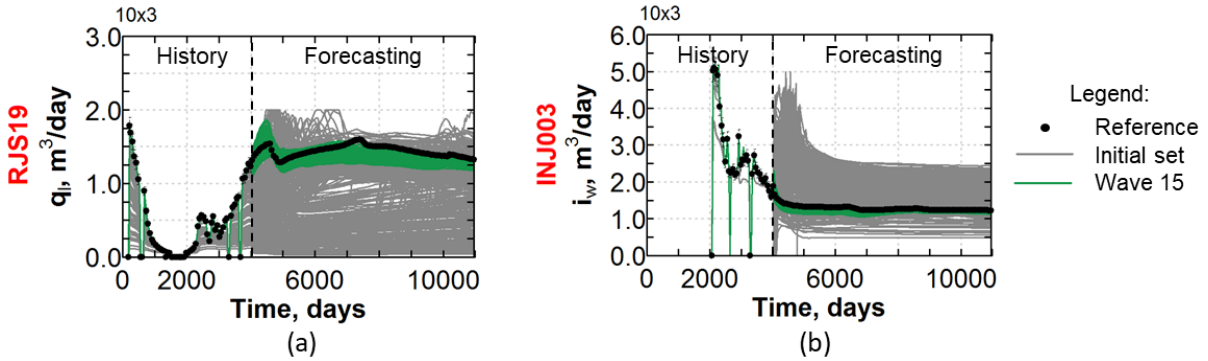


Figure 3.16. Time series for well data demonstrating the uncertainty reduction in forecasting from scenarios from the initial space in grey to non-implausible scenarios by the end of the wave 15, in green– (a) liquid production rate, and (b) water injection rate.

All the four safeguards against inappropriateness in the level of uncertainty reduction were carefully designed. Firstly, the implausibility measure accounted for all sources of uncertainties identified (observation error and model discrepancy). The uncertainty about the model to represent the real physical reservoir is not under consideration because the study case

applied is a hypothetical reality. Secondly, all the 115 emulators were validated, avoiding overconfident emulators. Thirdly, we applied a well-known way to combine the implausibility from different emulators. Finally, the definition of the cut-off ω considered an assumption about unimodality distribution of the implausibility measures, which is reasonable in this context.

3.5 Summary of control variables – Part I

In Table 3.7, we summarise the control variables from the procedures presented in this chapter with a short discussion of the choices made and likely sensitivity of the results to the choices.

Table 3.7. Summary of control variables summarising choices made and likely sensitivity.

Control variables	Choices and discussion
Cut-off of implausibility measure, ω	The choice of the cut-off of implausibility measure considers the degree of confidence that we have in the emulators constructed, the assumptions related to the distribution of the combined implausibility measure, among others. We have chosen $\omega = 3$ based on the three-sigma rule from Pukelshein (1994), considering that continuity and unimodality are reasonable assumptions for the distribution of I_M. In general, the larger ω: the smaller is the size of regions of the search space that are considered implausible in each wave, and the number of waves required to rule out as implausible proportions of the search space is higher.
Number of scenarios simulated for training set m	The specification of the design decision m depends on the number of uncertain attributes considered, the expected number of active inputs n_{Ai}, the complexity of emulated outcomes, and other practical aspects. To balance this decision, we can consider $(n_{A_i} + 2)(n_{A_i} + 1)/2$, which is the minimum number of points necessary to construct a quadratic response surface (Busby, 2007), where n_{A_i} is the number of active attributes to construct a specific emulator. Note that one important attribute of the training set is to be a representative sampling of the non-implausible space. A number of scenarios m=100 was chosen because of the number of simulations possible to run simultaneously in the cluster available and considering the number of uncertain attributes in the problem. As we construct emulators based on the training set, a smaller number of scenarios could lead to higher uncertainty around the emulator expectation or the impossibility of constructing new emulators, giving the number of active variables. A higher number m would lead to a higher computation cost for the simulation of scenarios.

Control variables	Choices and discussion
Number of scenarios simulated for test set n	The test set n is used to verify issues in the construction of the emulator and to validate emulators. Depending on the application, it can be unaffordable to have a test set, and $n = 0$ is needed. For these cases, we suggest alternative validation tools (Bastos, 2010). We have decided for a number of scenarios n=100 because of the number of simulations possible to simultaneously run in the cluster available and considering that this number sufficed to provide a representative sampling of the original search space.
The minimal proportion of scenarios λ required to construct new emulators	The number λ will depend on the number of scenarios simulated for training set m in a new wave and the number of uncertain parameters (and active variables for each emulator). A larger λ indicates that a new wave (with new simulations) will be more frequently required. A smaller λ would drive to smaller training sets being used to construct emulators, which by consequence could lead to more substantial uncertainty on the emulator output. We defined $\lambda = 50\%$, which indicates that a minimal of 50 scenarios is required to construct emulators. We considered the largest number of expected active variables in our application is 11 (see discussion in Section 3.3.3 related to STEP 5).
Number of emulation techniques considered in Step 11, n_f	We define the statistical models to be incorporated as emulators considering: (a) expert judgement on the types of relationships expected between inputs and outputs; (b) complexity of data structure; and (3) number of active variables to build emulators. For continuous outcomes n_f=*2* encompassing linear regression with (1) linear terms and (2) with linear and quadratic terms. For two-class models $n = 3$ encompassing (1) linear discriminant analysis, (2) logistic regression with linear terms and (3) with linear and quadratic terms. These models were selected considering that (a) linear and quadratic relationships are predominant, (b) data structure is sufficiently represented by binary and continuous outcomes, and (c) intermediate number of active variables allowing the construction of emulators with linear and quadratic terms with the training set size defined by m and λ.
Threshold to change phase of evaluation $\mathrm{D}_{info,M}$	We applied $\mathrm{D}_{info,M} \geq 70\%$ in our case study in a heuristic approach. Larger $\mathrm{D}_{info,M}$ values lead to faster incorporation of historical data (from new phases of evaluation) in the analysis. The simulation of scenarios for a larger period of times (in later phases) is costlier computationally, as demonstrated in Table 3.7. Conversely, smaller $\mathrm{D}_{info,M}$ values would slow down the integration of new historical data in the process, postponing expensive simulation of later phases for later waves (where a reduced non-implausible space is expected) but also the integration of additional (and possibly informative) data.

3.6 Conclusions

The ultimate goal of a model calibration process is to provide a background for well informed and efficient decisions. Finding the whole class of scenarios capable of representing

the reservoir historical behaviour is essential in order to give a realistic evaluation of reservoir performance and consistent, unbiased predictions incorporating realistic levels of uncertainty, required for full asset management.

We proposed a procedure for systematic uncertainty reduction for petroleum reservoirs combining reservoir simulation and Bayesian emulation techniques. We explored Bayesian History Matching techniques in order to provide an alternative and more rigorous tool for reservoir studies dealing with probabilistic uncertainty reduction. Challenges addressed by our systematic procedure are related to the consideration of several sources of uncertainty involved in a calibration process, efficient use of simulation time and high dimensional input and output spaces.

The procedure for systematic uncertainty reduction was applied to calibrate a case study with 26 uncertain attributes and 14 production and 11 injection wells. We defined the case study and took one of the scenarios to be the hypothetical reality (*i.e.* the known answer). The aim was to test the potential of the procedure for a complex reservoir model under a controlled situation while illustrating the main steps of the procedure. The consideration of a hypothetical reality is an important stage to ground more advanced studies, contemplating real observed data and inherent discrepancies between reservoir model and real physical reservoirs.

In total, we evaluated 4,018 days of historical data with five phases of evaluation; ran 15 waves sequentially; simulated 3,000 scenarios for training and test sets (the simulation time would be equivalent to 1,637 full simulations without phasing); and constructed valid emulators for 115 quantities of interest out of a total of 198 selected quantities. By the end of the calibration process and after removing the implausible regions, a small fraction in the order of 10^{-11}% of the input space remained as non-implausible, demonstrating a substantial uncertainty reduction. This substantial uncertainty reduction is a direct result of the interplay between the simulator behaviour and the specification of all the required uncertainties, although we note that in high dimensional space such large reductions are expected. The main novel contributions of this work (Part I) to the Bayesian History Matching framework for uncertainty analysis are as follows.

A **systematic workflow** to structure BHMUR techniques is presented. We deliver scalability to higher dimensions in input and outputs spaces by (a) using informed judgements of the asset team to identify inputs and outputs to emulate (*STEP 3*); (b) proposing a procedure

for output selection which considers the combination of quantities of interest considered as the most informative (*STEP 8*); and (c) planning the selection of active variables (*STEP 10*), where the sufficient technique currently applied (stepwise selection) can be substituted by techniques more appropriate for higher dimensions if required.

We secure flexibility for the integration of diverse emulation techniques. The workflow for the construction of concurrent emulators (*STEP 11*) for a given quantity of interest is capable of integrating diverse emulation techniques. In our application, we combined techniques for continuous and two-class quantities of interest.

We enable repeatability by developing a sequence of logically associated steps. The main workflow describes a high-level structure; for which individual steps can be designed in coherence with project requirements. It also allows the performing of several stages automatically, centring the users' focus on activities related to data analysis and synthesis.

Phases of evaluation split the historical data and progressively add information to the process. Phases allow us to explore more straightforward relationships from the early period of well production, and to exploit these relationships using emulation to rule out large regions of input space, based only on early time information. This promotes efficient use of resources for reservoir simulation. In our application, for example, Phase 1 (until 518 days) running time is estimated at 12% of the full historical period (4,018 days) running time. Three waves until Phase 1 rule out 97.86% of the input space (*i.e.* implausible regions), leaving only 2.14% of the space to evaluate in later phases, making subsequent emulation in later phases easier and more efficient.

The combination of each of these features in the systematic procedure is a novel approach to Bayesian History Matching. The application in the case study is a demonstration of the potential of the iterative nature of Bayesian History Matching combined with phases of evaluation and two-class emulators in addressing challenging problems of reservoir calibration, including the identification of the set of all inputs consistent with historical data, required for realistic predictions and asset management. The uncertainty reduction procedure was demonstrated as a powerful technique to search high dimensional space using substantially less computational time than using the complex simulation model alone.

Nomenclature – Article 2

B	=	indicator function of a transformation of simulation output
B^*	=	indicator function obtained from the emulator output
BHMUR	=	Bayesian History Matching for Uncertainty Reduction
BT	=	Breakthrough Time
CI	=	Credible Interval
D_{info}	=	information index
$\widetilde{D}_{info}$	=	adapted information index
D_{CI}	=	Credible Interval Diagnostics
ϵ	=	model discrepancy, the difference between the real reservoir and the reservoir model
E	=	expectation operator
e	=	vector of observational errors
f	=	function of reservoir simulation model that computes a vector of quantities of interest
f^*	=	emulator function
g	=	known deterministic function
I	=	implausibility measure
I_M	=	maximal implausibility measure
$\tilde{I}$	=	adapted implausibility measure
i_w	=	water injection rate
L_p	=	cumulative liquid production
m	=	number of reservoir scenarios simulated to increment the training set in each new wave
n	=	number of reservoir scenarios simulated to increment the test set in each new wave
n_A	=	number of active variables
n_f	=	number of statistical models used to build emulators
n_{sce}	=	total number of scenarios
N_p	=	cumulative oil production
NPV	=	Negative Predictive Value
$phase$	=	maximum phase of evaluation simulated in a given training set
p_{ibh}	=	bottom-hole pressure of injection wells
$\bar{p}_{ibh}$	=	average bottom-hole pressure of injection wells
p_{pbh}	=	bottom-hole pressure of production wells
$\bar{p}_{pbh}$	=	average bottom-hole pressure of production wells
PPV	=	Positive Predictive Value
q_l	=	liquid production rate
q_o	=	oil production rate
q_w	=	water production rate
RMSE_n	=	Normalised Root Mean Square Error
tol	=	tolerance applied to compute B
u	=	Gaussian process
Var	=	Variance
W_i	=	cumulative water injection
W_p	=	cumulative water production

w	=	nugget process
ω	=	implausibility cut-off
x	=	vector of input parameter values representing a reservoir scenario
x^*	=	most appropriate vector of uncertain attributes
y	=	vector of quantities from the real physical reservoir
z	=	vector of measurable quantities from the real reservoir
α	=	proportion covered by the credible interval
β	=	unknown scalar regression coefficients
φ	=	a phase of evaluation to select outputs to emulate $\varphi \in [1, phase]$
λ	=	the proportion of m providing a training set sufficiently large to construct emulators
ω^*	=	factor to rescale the indicator function of two-class emulators

Subscripts

A	=	active variables
i	=	a measurable quantity of interest of the reservoir, $i \in [1, q]$
j	=	the index corresponding to a regression term in the emulator equation

Superscripts

k	=	a scenario of the reservoir model, $k \in [1, n_{sce}]$

Acknowledgement

This research was carried out in association with the ongoing project registered as ANP 19708-7, “Fomento à Formação de Recursos Humanos em Gestão de Incertezas e Tomada de Decisão: Um Programa BG Fellowship" (UNICAMP/ Shell Brazil /ANP), co-funded by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brazil) through Science Without Borders Programme and Shell Brazil under the ANP R&D levy as “Compromisso de Investimentos com Pesquisa e Desenvolvimento”.

The authors thank ENERGI SIMULATION, UNISIM, DE-FEM-UNICAMP, CEPETRO and Department of Mathematical Sciences (Durham University) for supporting this work and CMG, Emerson and Schlumberger for software licenses. The authors also thank our colleagues who contributed to this paper: Alessandra Davolio and Forlan la Rosa Almeida (discussions about simulation targets), Carla Ferreira (discussions about visualisation), Daniel Carvalho (IT support), Gilson Moura (definition of well trajectory applied in the case study), Jonathan Owen (discussions about implausibility), Ricardo Moreno (discussions about emulation).

Appendix 3.A: Description of HR-82

The study case HR-82 derives from one of the geological realisations of the UNISIM-I-H model (Avansi and Schiozer 2015). A review of the inputs and outputs of the case study HR-82 differs it from the original benchmarking case. We firstly highlight the main characteristics of inputs and outputs from the HR-82 and then present additional details.

(a) 82 attributes model all uncertain parameters as continuous variables:

- Geological uncertainty: continuous variables represent geological uncertainties related to porosity and permeability (instead of geological maps in UNISIM-I-H);
- Tables of fluid properties and relative permeability: coefficients of equations model the relationships of the original tables in continuous attributes;
- Uncertain attributes added near the well: well index multipliers;
- All the considered uncertain parameters of the HR-82 have uniform distribution through an interval of values.

(b) The historical data is based on a hypothetical reality defined as a known combination of uncertain parameters:

- The original simulation outputs of the hypothetical reality (without noise) are added of random and systematic errors;
- The model discrepancy was assessed to incorporate errors in simulator target values during the historical period.

Table 3.A-1 summarises all the uncertain attributes of HR-82. All of them are uniformly distributed and have a value in their ranges which is considered the hypothetical reality.

Table 3.A-1. Uncertain parameters of HR-82. Parameters in blue have global influence; in green, regional influence (four attributes for each of the 13 regions); in orange, sector influence on the east block; in yellow, local influence in 25 wells.

Attribute	Description	Ranges	Quantity
c_p	Rock compressibility	[10, 96] E-6	1
c_{krw}	Relative permeability	[0.86, 1.28]	1
mphi	Porosity multiplier	[0.75, 1.25]	13
A(Rx)	kx, angular coefficient	[0.135, 0.175]	13
B(Rx)	kx, linear coefficient	[-0.4, 1.1]	13
Mkz(Rx)	kz multiplier	[0.1, 0.5]	13
$PVT_{co,EB}$	Oil compressibility	[1.40, 1.62] E-3	1
$PVT_{ai,EB}$	Oil viscosity related	[2.5, 50.0] E-4	1
WOC_{EB}	Water-Oil-Contact	[3169, 3179]	1
Wi_{ff}	Well index factor (9)	[0.6, 1.4]	25

The two first attributes of Table 3.A-1 have an impact on all the reservoir model. The parameter c_{krw} describes uncertainty related to relative permeability using a regression for three tables of the UNISIM-I-H. Figure 3.A-1 shows that water relative permeability grows as an exponential function of water saturation multiplied by c_{krw} varying on an interval.

Table 3.A-1 follows with four lines (in green) describing additional 52 (4 x 13 regions) uncertain parameters modelled. Geological properties of porosity and permeability are based on a map of porosity ϕ_{map}. The reservoir model is divided into the 13 regions – Figure 3.A-2 - as in Maschio and Schiozer (2016), each of them having four uncertain parameters associated: (a) a porosity multiplier $mphi^{region\ i}$; (b) two coefficients ($A^{region\ i}$ and $B^{region\ i}$) to model the horizontal permeability, and (c) one multiplier for vertical permeability $mkz^{region\ i}$.

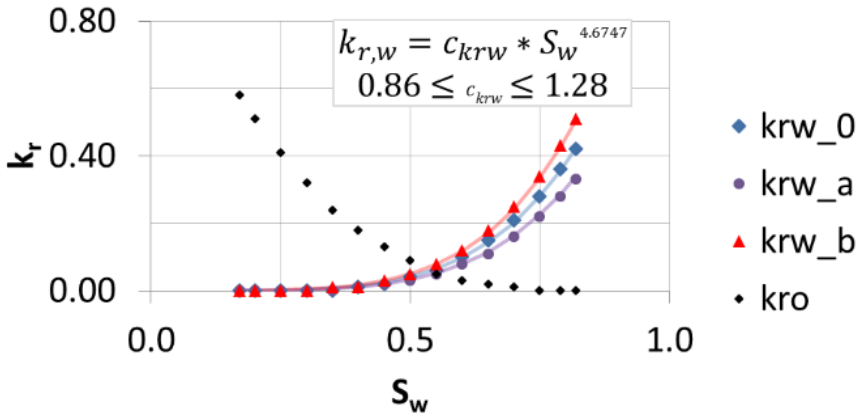


Figure 3.A-1. Relative permeability as a function of water saturation. Three tables from the benchmarking case UNISIM-I-H are plotted. A simple equation models the relationship with one uncertain attribute, the c_{krw}.

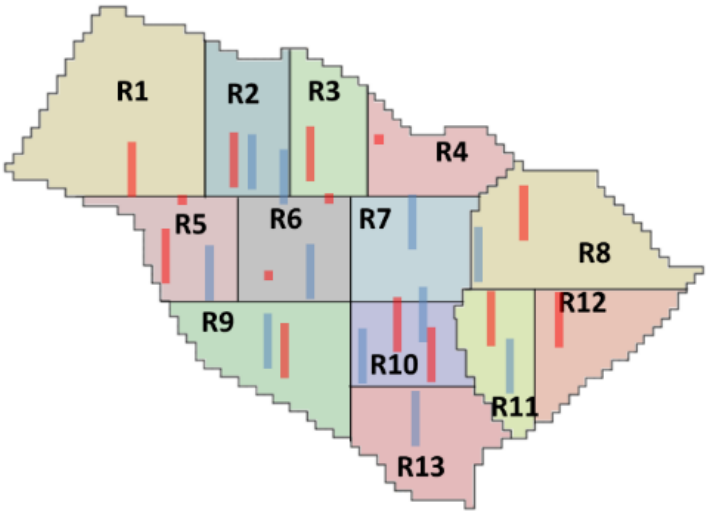


Figure 3.A-2. Two-dimensional aerial view of the reservoir model highlighting (a) the 13 regions composing the reservoir, (b) a sealing fault diving the reservoir into two compartments (east block with regions 8, 11 and 12, west block with the other regions), (c) the position of production and injection wells, in red and blue respectively.

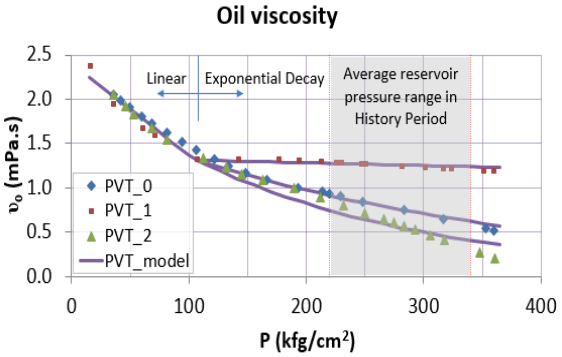
The Equations 3.16 to 3.18 represent the porosity of a region as a function of its multiplier, the horizontal permeability as a function of the porosity, and the vertical permeability as a function of horizontal permeability.

$$\phi^{region\ i} = mphi^{region\ i} * \phi_{map}^{region\ i} \tag{3.16}$$

$$kx^{region\ i} = 10^{100*A^{region\ i}\phi^{region\ i}-B^{region\ i}} \tag{3.17}$$

$$kz^{region\ i} = mkz^{region\ i} * kx^{region\ i} \tag{3.18}$$

Properties of three PVT tables were analysed and modelled as a function of two additional parameters. $PVT_{co,EB}$ directly represent oil compressibility; $PVT_{ai,EB}$ models oil viscosity through an equation with an exponential decay in the interval of interest (Figure 3.A-3).



$$v_o(p > 107) = 1.135 * e^{-ai*(P-107)};\ 2.5e-4 \leq ai \leq 50.0e-4$$

Figure 3.A-3. Oil viscosity dependence on pressure; the three PVT tables from the UNISIM-I-H benchmarking case are plotted. The PVT table is an uncertainty identifies in the east block only. When we consider the range of pressure under interest, we can model the oil viscosity by an equation with exponential decay and with one uncertain attribute a_i.

Remaining properties defined in the PVT tables were estimated based on a regression over the interval of pressure, with no uncertainty associated with them (Figure 3.A-4).

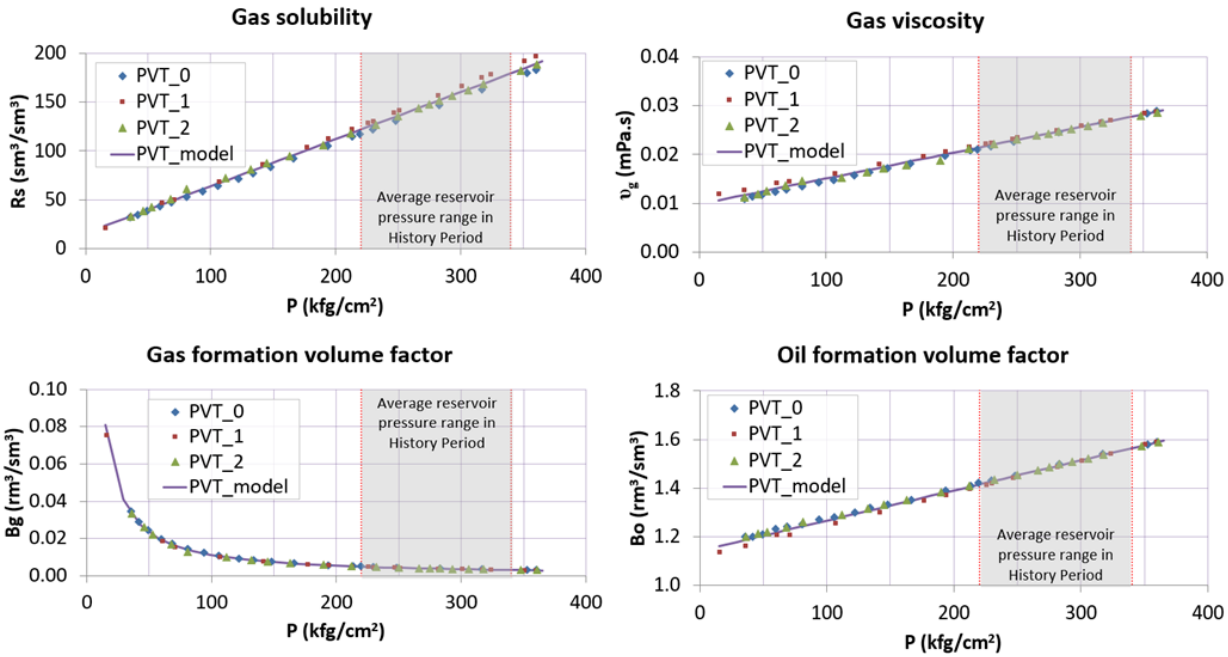


Figure 3.A-4. Models for properties with no uncertainty associated with the PVT tables.

Finally, each of the 25 production and injection wells has a multiplier of well index associated with representing local effects impacting pressure exclusively around the wells.

Other relevant aspects include the use of well trajectories to define the wells; pressure limits of 36-450 kgf/cm^2 during history period; operational conditions for forecasting are the same as the ones applied in UNISIM-I-H.

4 ARTICLE 3: SYSTEMATIC UNCERTAINTY REDUCTION FOR PETROLEUM RESERVOIRS COMBINING RESERVOIR SIMULATION AND BAYESIAN EMULATION TECHNIQUES: PART II

Helena Nandi Formentin, Ian Vernon, Camila Caiado, Michael Goldstein, Guilherme Daniel Avansi, Denis José Schiozer, Carla Janaína Ferreira

To be submitted.

Contribution by the candidate Helena Nandi Formentin: Conceptualisation, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original + Edition, Visualization, Funding Acquisition, as described in the Statement of Contribution submitted to the SPE Europec (fully reproduced in Appendix B – Contribution Statements). The code in R was fully developed by Helena. CMG was selected as reservoir simulation software.

Abstract – Article 3

Historical data from a producing reservoir support the calibration of the corresponding simulation model. A large number of uncertainties about the reservoir structure and physics combined with several sources of data places this inverse problem as one of the most complexes in the industry. Reservoir simulation models are the instrument to consolidate knowledge and data, incorporating the best of our understanding about the reservoir: from mathematical and physical models to seismic data, from laboratory tests to operations in the field. Bayesian History Matching for Uncertainty Reduction (BHMUR) is a calibration technique able to consider all sources of uncertainty in the process, aiming to identify the whole class of scenarios from reservoir simulation that are simultaneously consistent with the observed data and the uncertainties in the problem. This paper advances the applicability of BHMUR by exploring and proposing solutions for four critical steps in its implementation. Firstly, we extend the statistical methodology for BHMUR by incorporating two-class emulators. We address specific structures in data sets characterised by having distinct behaviour across two different regions of the input space. This section takes charge of the recognition of simulator behaviours untracked in previous reservoir studies. We characterise how simulator targets, *e.g.* liquid production, and water breakthrough lead to discontinuities in relationships between outputs and inputs of the reservoir simulation model. It is appropriate to emulate such simulator behaviours using two-class models. A fundamental feature of the BHMUR technique is the combination of evaluations from emulators and the numerical simulator. Emulators enable the assessment of new scenarios quickly, orders of magnitude faster than numerical simulators. The second point of this paper deals with a critical stage for the safe use of emulators: the validation process that indicates whether an emulator appropriately estimates the expected values for the simulation model and uncertainty around it. For this purpose, we combine complementary indicators of quality. The validation of emulators is one of the safeguards against the inappropriate exclusion of parts of the search space, and therefore, fundamental in a BHMUR approach. Thirdly, we propose a methodology to select outputs to emulate. This procedure systematically chooses a combination of quantities of interest able to balance the computational effort needed to (a) construct emulators and evaluate new scenarios, and (b) reduce the uncertainty about the reservoir behaviour through the implausibility analysis. In this way, we focus on the construction of informative emulators for a progressive and efficient analysis. The fourth topic addressed is the consideration of observational errors in the calibration process. We demonstrate a coherent approach to estimate the errors of cumulative and averaged quantities

of interest which are based on evaluations of error in rate and single-time data. The contribution of this paper is methodological: we advance the applicability of Bayesian History Matching for Uncertainty Reduction by offering (1) an extension of the BHMUR to behaviours in physical data which can be labelled as binary; (2) indicators of quality that, combined, are capable of validating emulators statistically and selecting one among concurrent emulators for the same quantity of interest; (3) a systematic procedure to choose a combination of quantities of interest to be emulated; and (4) a consistent demonstration of the assessment of random and correlated errors, using a straightforward approach that can be easily extended to more complex case studies. We illustrate the topics of this paper with the application presented in Formentin *et al.* (2020-a).

Keywords: Bayesian History Matching, Uncertainty Reduction, binary behaviour, simulation targets, breakthrough time, dimensionality reduction of outputs, validation of emulators, observation errors.

4.1 Introduction

Historical data from a petroleum reservoir reveal its actual performance, providing information to evolve the understanding of field characteristics and behaviour. The reservoir simulator, which closely reflects our knowledge about the sub-surface, takes advantage of historical information to be calibrated and is critically applied for forecasting, recovery strategy optimisation and, ultimately, decision-making.

Bayesian History Matching for Uncertainty Reduction (BHMUR) is an approach designed to identify the whole class of reservoir scenarios capable of honouring the historical data while considering all sources of uncertainties in the process. The formulation originally proposed by Craig et al. (1995) explicitly sets the problem and projects how this technique approaches and finds a coherent uncertainty reduction. We offer a brief overview about the statistical formulation, referring Craig *et al.* (1995), Vernon *et al.* (2010) and Vernon *et al.* (2018) for a deep understanding about the methodology.

The vectors y and z (Equation 4.1) contain, respectively, quantities from the physical reservoir itself and the corresponding historical data which are observed from the field (*e.g.* oil production rate in a given time). The difference between them is a random quantity, named observational error e.

$$z = y + e \tag{4.1}$$

Although an essential piece of the problem, the description of the observational error e is often overlooked in reservoir calibration process in several ways such as:

- They are frequently under-evaluated by considering only the calibration level of the equipment. Nonetheless, the observational error also includes errors in chemical analysis, in the apportionment of field production to production wells, and data manipulation;
- They are usually characterised as a simple percentage of z (Avansi *et al.* 2016) or a Gaussian process uncorrelated in time and space (Emerick and Reynolds 2011). These approaches are not in alignment with our understanding about the reservoir system and measurement process: we may expect to have random and systematic portions of error, correlated in different degrees in time, space (*e.g.* different wells) and across observed quantities (*e.g.* cumulative quantities, rates and pressure).

A possible opportunity for improvement is to analyse the successive stages of a field's well surveillance programme, evaluating the way which the uncertainty in the observations propagates through the process. However, how we characterise e and integrate it into the calibration process remains as a question. The structure of the observation error relies on a statistical description, the complexity of which ultimately depends on the objective of the study, information and statistical-skilled personal available. In this paper, we describe a simple, but coherent, formulation to account for random and systematic portions in observational errors for quantities of interest derived from measurable data (*e.g.* cumulative quantities).

In the BHMUR formulation, if x is a vector from the input space defining a reservoir scenario (*e.g.* properties related to porosity and permeability of the reservoir), the function f of x represents the corresponding vector of quantities computed via reservoir simulation (*e.g.* phase production rates, bottom hole pressure), with same dimensions as y. The level of complexity of the function $f(x)$ in Equation 4.2 includes mathematical and physical models and a numerical solver; such system is taken as a black-box, and we account only for inputs and outputs of the simulator:

$$y = f(x^*) + \epsilon \tag{4.2}$$

Acknowledging the quote from the statistician George Box ‘All models are wrong, but some are useful’, we define the model discrepancy ϵ: even when x is the most appropriate

reservoir scenario x^*, we expect a discrepancy ϵ between the reservoir simulation output and the physical reservoir itself, again represented as a random quantity (Vernon *et al.* 2010).

Two properties of the BHMUR differentiates this approach from other calibration techniques: (a) its focus on identifying regions of the input space x that are unlikely to emerge as an appropriate representation of the physical reservoir itself (opposed to attempting to directly find the best fitting solution), and (b) a dramatic increase in the speed of evaluation of new reservoir scenarios thanks to the construction and employment of emulators.

Emulators are statistical models representing the simulator. They provide quick evaluations of new scenarios before we simulate them, providing the expected outcome and uncertainty around it, and often lead to a speed increase of several orders of magnitude. While the traditional statistical model employed as emulator contains regression terms, Gaussian and nugget processes (Craig et al. 1995, Vernon et al. 2010), new behaviours in the data structure may require different or/and more complex statistical models. In parallel to all practical methods to recognise new patterns (*e.g.* cross-plots, pairs plots, correlation indicators), a good practice is to endorse such statistical refinement through physical justification or expert knowledge. We explored and recognised new data behaviours arising from the simulator, illustrating diagnostics techniques supported by physical interpretation.

BHMUR technique integrates the emulator results into the implausible measure (Equation 4.3), which uses a metric based on the number of standard deviations between z_i and $E(f_i^*(x))$. The implausibility measure relates the difference between the emulator expectation $E(f_i^*(x))$ for the -*ith* considered output and the corresponding observed value z_i, with all sources of uncertainty in the process, taking the variance of (a) the emulator $Var(f_i^*(x))$, (b) the model discrepancy $Var(\epsilon_i)$, and (c) the observation error $Var(e_i)$ (Craig et al. 1995, Vernon et al. 2010):

$$I_i^2(x) = \frac{\left[E\left(f_i^*(x)\right) - z_i\right]^2}{Var\left(f_i^*(x)\right) + Var(\epsilon_i) + Var(e_i)} \tag{4.3}$$

The implausibility measure $I_i(x)$ is calculated from each scenario x and output emulated $i \in [1, q]$. But, which set of outputs should be emulated? Among all possible outputs (*e.g.* rates and pressure in each time, averaged and combined index), the strategy for output selection stands as a decisive choice for the efficiency of the process, especially in contexts

with correlated outputs. Emulating correlated outputs individually can be both time-consuming and unnecessary (*e.g.* non-informative about implausible regions). Ferreira *et al.* (2020) discussed in-depth the considerations involved when selecting outputs to emulate. They proposed and demonstrated a methodology for output selection. Here, we approach this stage in an alternative way and broaden a discussion in the appropriate section.

Note that, independently on the method applied to select outputs to emulate, only emulators considered valid (*e.g.* with appropriate confidence level) are allowed to be considered in the implausibility analysis. The importance of this topic is detailed further in a dedicated section of this paper.

Several formulations to combine multiple implausibility measures from all valid emulators are possible. For example, as Vernon *et al.* (2010), one can select the first maximal implausibility by applying $I_M(x) = \max_{i \in Q} I_i(x)$, Q being the set of outputs currently considered A complete version of the implausibility measure is the Multivariate Implausibility Measure (Craig *et al.* 1995), a formulation rooted in the Mahalanobis distance.

On the one hand, reservoir scenarios with combined implausibility measure higher than a cut-off ω are labelled as implausible, indicating that these scenarios are unlikely to lead as an acceptable fit, given current observed data: we normally search for all scenarios that are consisted with current observed data, our physical insight and the uncertainties in the problem; some of these scenarios then may subsequently be found to be appropriate representations of the physical reservoir after much more data is collected, but some of these will not. On the other hand, non-implausible scenarios - that have low implausibility - are due to (a) an appropriate representation of physical reservoir itself, or (b) an uncertainty around the emulator estimation $Var\left(f_i^*(x)\right)$ being large.

The practice of the BHMUR technique enchains three core elements (Formentin *et al.* 2020-a):

1. Sampling and simulating of a relatively small number of scenarios from the input space. These scenarios must be carefully defined to be representative of the current non-implausible input space;

2. Constructing of emulators able to provide an expected outcome $E\left(f_i^*(x)\right)$ and the emulator uncertainty $Var\left(f_i^*(x)\right)$. The emulators allow us to evaluate a large number of

new scenarios very quickly, several orders of magnitude faster than with a numerical simulator;

3. Computation and combination of the implausibility measure to identify the regions of the space that are implausible, called implausibility analysis.

We name as a *wave* each iteration containing these three elements. By the end of each wave, we rule out implausible regions, preparing to refocus the sample of the next wave only in the smaller non-implausible regions. In each wave, we are able to construct more detailed emulators with smaller uncertainty because we are focusing our analysis on a smaller search space. A wave-based approach allows emulating complex surfaces with simple statistical structures.

4.1.1 Objectives

While implementing the BHMUR approach to complex systems such as petroleum reservoirs, some practical questions determine the efficiency, coherence and credibility of the solution. We advance the applicability of Bayesian History Matching for Uncertainty Reduction (BHMUR) for reservoir studies in four ways:

- Extending the traditional BHMUR formulation by including emulators for two-class quantities of interest, which mimic outputs from the physical system suitably labelled as binary;

- Describing a process for validation of emulators and selection of possible emulators for a given output considering whether its uncertainty is appropriate and whether it is informative enough. We review and establish indicators for this purpose;

- Proposing a methodology for selection of outputs which systematically decide on a combination of outputs to be emulated, considering the potential to reveal implausible parts of the input space;

- Modelling observational errors in historical data, to provide a coherent transformation of errors observed in time series measurement (*e.g.* rate, pressure) to cumulative and averaged quantities of interest.

We demonstrated the results in a stand-alone format, appropriate to be retaken individually in BHMUR approaches. Each of these objectives is explored and illustrated

through the application performed in Formentin *et al.* (2020-a), and are relevant for the *STEPS 2, 3, 8* and *12* the workflow of that paper.

4.2 Two-class quantities of interest and emulators

The standard form of Bayesian History Matching described in Craig *et al.* (1995), Vernon *et al.* (2010) and Vernon *et al.* (2018) treat as smooth the output functions on the inputs. Caiado and Goldstein (2015) studied the situation where simulator outputs in time series have more than one form of limiting behaviour (*e.g.* collapse or non-collapse of the physical system), which divides the input space into regions. The simulator output is smooth within each region, but discontinuous across boundaries. In this work, we extend the standard form of the Bayesian History Matching for Uncertainty Reduction (BHMUR) by incorporating two-class emulators to address specific behaviour in data sets.

Binary quantities of interest have two distinct classes of outcomes and present discontinuous behaviour across two discrete regions of input parameter space separated by a boundary. Classifiers are statistical models to deal with binary quantities. We provide background information for the application of these statistical models. We can apply the following extended form of BHMUR formulation to any problem with continuous and binary outputs and, here, we demonstrate a typical example from reservoir engineering. Moreover, the flexibility to integrate diverse statistical models is a strength of the procedure presented in Formentin *et al.* (2020-a), where we describe the incorporation of this extended form of BHMUR in a systematic procedure.

In reservoir studies, binary outcomes result from the classification of continuous outcomes from simulations into binary data. We identified a two-class pattern in well data such as cumulative liquid and water production for a given time and average bottom hole pressure within a time window. Therefore, we introduce classification models into our statistical formulation. Two-class emulators recognise to which of these regions a new scenario belongs to, before simulating the new scenario.

Three stages allow the integration of two-class emulators in BHMUR studies. The first stage classifies the simulator outcome in two classes, obtaining a binary outcome. In Figure 4.1.a, a pairs-plot for three illustrative uncertain attributes (ϕ, k_r and k_z) displays two distinct regions: the green region is classified as 0; the red one is classified as 1.

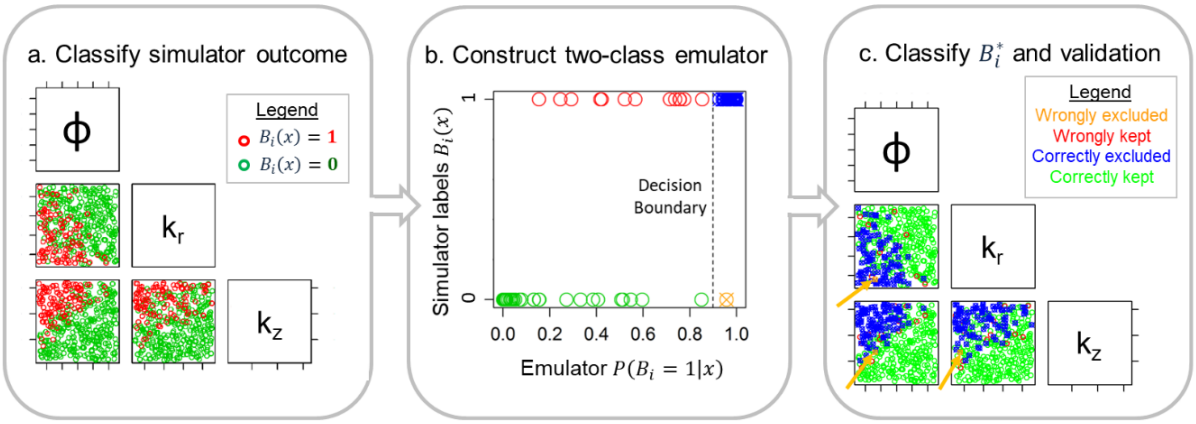


Figure 4.1. Two-class emulators recognise scenarios belonging to distinct regions – (a) Pairs plot for three uncertain attributes displaying two distinct regions, obtained from the classification of simulator outcomes; (b) Cross plot for a two-class emulator, the vertical line indicates the decision boundary, the legend is the same as figure c; (c) Pairs plot displaying scenarios classified by the two-class emulator.

To obtain these two regions, a transformation function maps f_i, the simulator output, onto a binary outcome B_i. In our application, we define in Equation 4.4 the binary quantity B_i by considering the difference between $f_i(x)$ and observed data z_i with respect to a tolerance tol_i. An alternative notation is given using an indicator function $\mathbf{1}(\cdot)$. The magnitude of the tolerance denotes the uncertainty due to the observational error e_i and the model discrepancy ϵ_i. Note that our convention labels as 0 the non-implausible and as 1 the implausible regions. We give explicit examples of binary outcomes in a reservoir simulator context in the next section.

$$B_i(x) = \begin{cases} 0, & if\ |z_i - f_i(x)| \leq tol_i \\ 1, & otherwise \end{cases} = \mathbf{1}(\,|z_i - f_i(x)| > tol_i) \tag{4.4}$$

We focus on removing one of the discrete regions, the non-implausible one. Nevertheless, we emphasise that a similar approach is adaptable for other data patterns; for example, a data structure that requires the identification of both discrete regions because it is convenient to emulate them separately. Additional discussion and examples are described by Caiado and Goldstein (2015).

In the second stage, Figure 4.1.b, we apply classification models (also called classifiers) to construct emulators. Classifiers provide the probability of a given new scenario x to be labelled as 1. Therefore, before simulating new scenarios, we can derive this probability. Logistic regression and linear discriminant analysis are classical two-class classifiers, and we adopted them with the implementation in R (R Core Team, 2018) for generalised linear models (*glm*, family *binomial*) and linear discriminant analysis (*lda*, Venables and Ripley, 2002).

Generalised linear model is a comprehensive class of statistical models. They have broader applicability than linear models because *glm* generalise the possible distributions of the residuals to the exponential family and use maximum likelihood for the estimation of coefficients. *Glm* includes linear models since the normal distribution is a particular case of the exponential family and, under normality conditions, the standard least-squares produces maximum likelihood estimation of coefficients (Myers, 1990).

For binary quantities of interest, a logistic model is considered through the logit link function in the left-hand side the log-odds or logit of the Equation 4.5. On the right-hand side a formulation equivalent to linear regression models. Therefore, we model the conditional distribution of the response $\mathrm{B_i}$, given the predictors $x_{\mathrm{A,i}}$as:

$$\log\left(\frac{\mathrm{P(B_i = 1|x)}}{1-\mathrm{P(B_i = 1|x)}}\right) = \sum_j \beta_{ij} g_{ij}\left(x_{A_i}\right) \tag{4.5}$$

Inverting Equation 4.5, we obtain the logistic function outcome (Equation 4.6) which predicts the probability of a given scenario to be classified as 1. This outcome is in the interval between 0 and 1.

$$\mathrm{P(B_i = 1|x)} = \frac{e^{\sum_j \beta_{ij} g_{ij}\left(x_{A_i}\right)}}{1 + e^{\sum_j \beta_{ij} g_{ij}\left(x_{A_i}\right)}} \tag{4.6}$$

The coefficients β_{ij} are determined through maximum likelihood, which for logistic regression takes the following form (James *et al.* 2013):

$$\ell(\beta_{i1}, \ldots, \beta_{ij}) = \prod_{a:B_{i,a}=1} p(x_{A_{i,a}}) \cdot \prod_{a':B_{i,a'}=0} \left(1 - p\left(x_{A_{i,a'}}\right)\right) \tag{4.7}$$

The estimates $\hat{\beta}_{ij}$ for a given quantity of interest i are chosen to maximise this likelihood ℓ, a multiplier function based on the calculated probabilities for scenarios a with $B_{i,a} = 1$ and scenarios a' with $B_{i,a'} = 0$. As the likelihood is a multiplicative function, applying the log simultaneously (a) transforms it in a summation operation, simpler to derivate and (b) provides the arguments that maximise the original function, as log is a monotonic transformation.

Another classifier considered is linear discriminant analysis (*lda*), and we explain the principle of this algorithm with the formulation from James *et al.* (2013). The distribution

of the predictors $x_{\mathrm{A,i}}$ is modelled separately in each of the response classes (*e.g.* for B_i, the possible number of classes k is 0 or 1). Bayes' theorem (Equation 4.8) is used to flip these around into estimates for $p_k(x_{\mathrm{A}_i}) = P(B_i = k|X = x_{\mathrm{A}_i})$:

$$\Pr(B_i = k|X = x_{\mathrm{A}_i}) = \frac{\pi_k f_k(x_{\mathrm{A}_i})}{\sum_{l=1}^{K} \pi_l f_l(x_{\mathrm{A}_i})} \tag{4.8}$$

Where π_k is the prior probability for the class k and the density function of X for an observation from the k class is defined by $f_k \equiv \Pr(X = x_{\mathrm{A}_i}|B_i = k)$. The *lda* algorithm needs to make some assumptions to define the form of f_k. With multiple predictors (*i.e.* x_{A_i} has more than one column), f_k can be drawn from a multivariate normal distribution $x_{\mathrm{A}_i} \sim N(\mu, \sum)$ with expectation $E(x_{\mathrm{A}_i}) = \mu$ and covariance matrix $cov(x_{\mathrm{A}_i}) = \sum$. The multivariate normal density is defined in Equation 4.9:

$$f(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right) \tag{4.9}$$

Plugging the multivariate normal distribution in Equation 4.8, the *lda* classifier assigns a probability to an observation $X = x_{\mathrm{A}_i}$ to be of class $B_i = 1$. When these distributions are assumed to be normal, the linear discriminant model is very similar in form to logistic regression.

From James *et al.* (2013), reasons to consider a linear discriminant analysis as an alternative to logistic regression include:

- When the classes are well-separated, the parameter estimates for the logistic regression model are surprisingly unstable. Linear discriminant analysis does not suffer from this problem;
- If the size of the training set is small and the distribution of the predictors x_{A} is approximately normal in each of the classes, the linear discriminant model is again more stable than the logistic regression model.

The cross-plot of Figure 4.1.b compares binary data from simulated scenarios with the corresponding probabilities predicted by the two-class emulator. A threshold probability called the decision boundary establishes a connection between Figures 4.1.b and 4.1.c. The

labels B_i^* for all the scenarios to the left of the decision boundary are 0; for all the scenarios to the right of the decision boundary are 1.

In Equation 4.5, this boundary links the probability estimated by the emulator $\mathrm{P}(\mathrm{B_i} = 1|\mathrm{x})$ with the labels B_i^*. We define ω^* larger than the implausibility cut-off ω, and this operation rescales the outcome from the emulator defining a new implausibility measure for the two-class emulators.

$$I_i(x) = B_i^*(x) = \begin{cases} 1 \cdot \omega^*, \text{if} \;\; \mathrm{P}(\mathrm{B_i} = 1|\mathrm{x}) \geq \text{decision boundary} \\ 0 \cdot \omega^*, \text{if} \;\; \mathrm{P}(\mathrm{B_i} = 1|\mathrm{x}) < \text{decision boundary} \end{cases} \tag{4.10}$$

In the third stage of Figure 4.1, we can classify training and test scenarios with the emulator. Figure 4.1.c presents the classification of the training set by the emulator, highlighting only one scenario wrongly excluded. The discussion about predictive values in the next section emphasises that among the scenarios ‘wrongly excluded’, ‘wrongly kept’, ‘correctly kept’ and ‘correctly excluded’, we are mainly concerned with scenarios wrongly excluded. This formulation for binary quantities of interest integrates uncertainties of the calibration process by the specification of a tolerance $\mathrm{tol_i}$ (incorporating observational error and model discrepancy) and a decision boundary (emulator uncertainty).

With quantitative and qualitative diagnostics, we check the validity of a given two-class emulator and, in parallel, we can select the best among concurrent (and valid) emulators for the same quantity of interest. Specificities of the validation process are described in the appropriate section.

Next, we explain two data behaviours (Simulator targets and Breakthrough Time) identified in historical data that are appropriately emulated as binary outputs in reservoir studies.

4.2.1 Pattern 1: Simulator targets

We usually specify observed production and injection rates as simulation targets for wells while simulating the historical period. Additionally, upper and lower pressure limits are set as boundary conditions for producers and injectors. The main objectives are to avoid unphysical fluid behaviour and to extrapolate the PVT table. Usually, pressure limits are much wider than the operational window, and not expected to be observed in the real field. That is, the lower pressure limit is much lower than the lower pressure required to produce fluid from the reservoir to the surface; the maximal pressure limit is much higher than rock fracture

pressure. Figure 4.2 illustrates these settings in plots for liquid production rate q_l and bottom hole pressure p_{pbh} *versus* time.

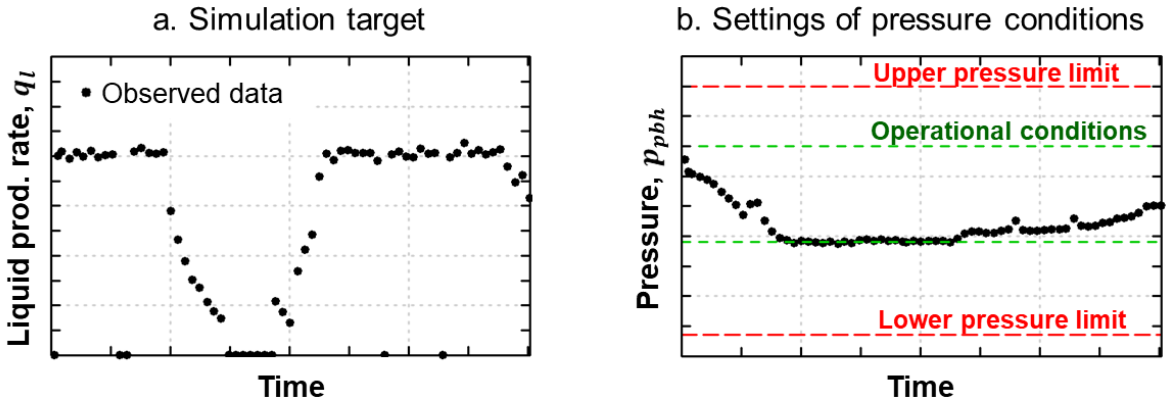


Figure 4.2. Settings for reservoir simulation in historical period – (a) the observed liquid production rate of the well is the simulation target; (b) limit pressures are much wider than the operational window.

The simulator calculates the bottom hole pressure required for producing or injecting at the given simulation target (*i.e.* the target is the primary control). The production rate of other flowing phases is dependent on the reservoir conditions surrounding the well. We usually perform the calibration process by a joint analysis of the ratios of fluid phases and bottom hole pressure. Nevertheless, when the simulation reaches a pressure limit in an attempt to deliver the targets, the simulation control of the well changes from target to the attained pressure limit.

This mechanism impacts the simulation outcomes in a particular manner described by two situations:

1. **Producers**: while a production well attains the lower pressure limit during the simulation, the production rate of the simulator is smaller than the specified target;

2. **Injectors**: while an injection well attains the upper pressure limit during the simulation, the injection rate is smaller than the specified target.

We recognised that when one of these situations occurs, it induces a change of behaviour in the simulation outcomes. The two distinct regions of behaviour characterise different relationships between inputs and outputs. Figure 4.3 illustrates these regions and relationships with a univariate example, *i.e.* only a porosity attribute is considered uncertain for the simulation runs. Figure 4.3.a shows whether the simulator reaches the simulation target during the whole time window (region 1) or not (region 2). This evaluation is made by the cumulative liquid production at the end of the time window. This binary behaviour is distinctly

driven by the condition of the bottom hole pressure (Figure 4.3.b): in region 2, the bottom hole pressure has stagnated at the lower pressure limit, which again, is much lower than the lower operational condition. We will capture this binary behaviour using two-class classification emulation techniques described.

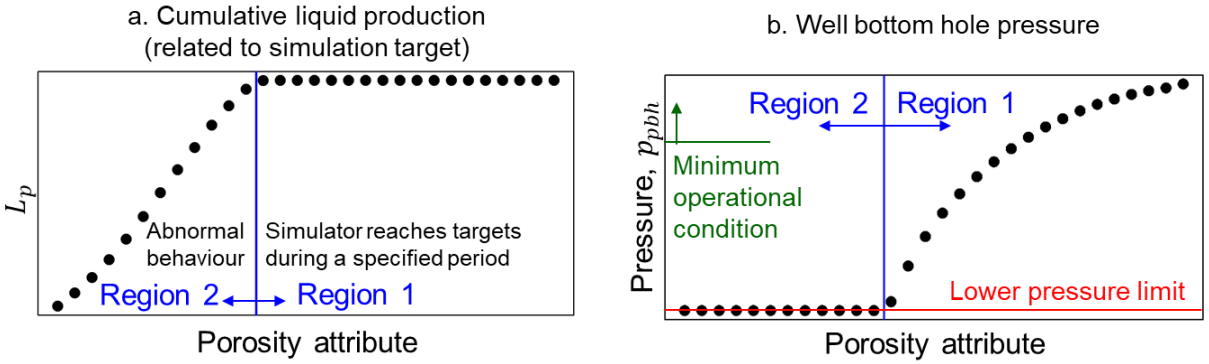


Figure 4.3. Conceptual description of the two distinct regions in the simulation outcome; one uncertain attribute related to porosity is plotted against (a) cumulative liquid production in a time window, and (b) bottom hole pressure. Region 2 presents an abnormal behaviour because the pressure is stagnated on the lower pressure limit. Note that the lower pressure limit is much lower than the lower operational condition.

The mechanism of simulation control change is also relevant when considering multiple uncertain attributes. Before exploring an example, we introduce a type of graphic frequently used in our analysis.

Cross-plots provide qualitative diagnostics of multivariate emulators. Simulation outputs and emulation outputs are respectively plotted in the x and y-axis. The axes usually have the same scale. For continuous quantities of interest, a 45-degree line shows where simulation output is equal to emulation output. Cross-plots usually contains a scatter plot of one or two data sets (in the last case, training and test sets can be compared, for example). With these elements, we can evaluate patterns in the residuals between emulator and simulator output. Alternatively, vertical bars can demonstrate additional information related to the emulator, for example, emulator uncertainty given by the standard deviation of the emulator. In this case, we indicate in the caption of the figure what the bars represent.

Figure 4.4 illustrates cross-plots of emulators for two continuous quantities of interest. These emulators are constructed with a regression model of first-order terms only and the training set of our first wave. We recognise a straightforward pattern in Figure 4.4.a: the cumulative liquid production of simulations is limited to the simulation target, but some scenarios produce much less than it. In Figure 4.4.b, the lower limit pressure is reached by these same scenarios with abnormal behaviour related to liquid production. Two regions are distinguishable in each of the figures and comparable with the regions described in Figure 4.3.

These cross plots show what would go wrong if we were to use a single standard emulator: the simulator behaviour would not capture and these regions would not be distinguished. Extending BHMUR with two-class emulators enable us to identify these regions, having emulators for continuous outputs that perform much better than these cross-plots.

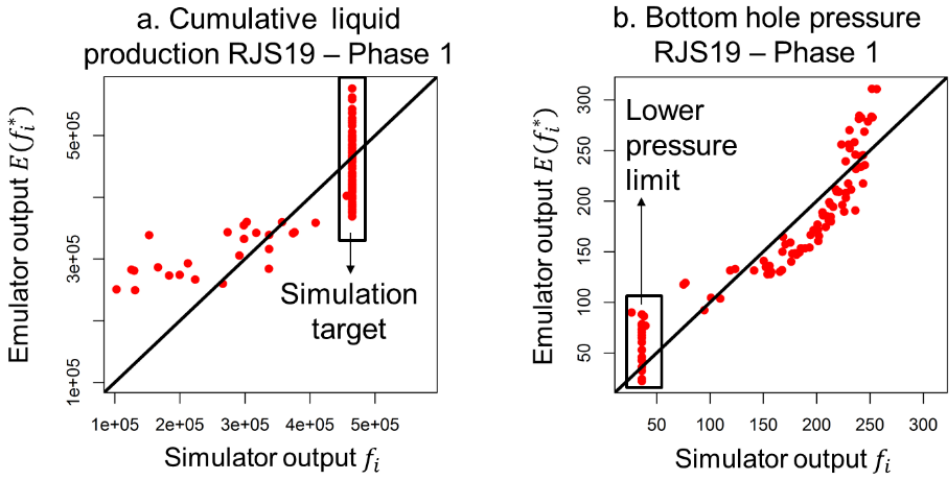


Figure 4.4. Cross-plot of emulators constructed for two continuous quantities of interest based on a training set of Wave 1, Phase 1. This a multi-variate example where: (a) the cumulative liquid production reaches the simulation target for the scenarios vertically aligned but fails for the others, which is deemed abnormal behaviour; (b) the bottom hole pressure stagnates at the lower pressure limit for the scenarios with abnormal behaviour. These plots show that a single standard emulator is not able to capture the structure of the simulator behaviour, demonstrating the importance of extending BHMUR technique with two-class emulators.

We recall Equation 4.4, which defines binary quantities from continuous outputs of the reservoir. When sufficiently large pressure limits are set in the simulator, we apply $tol_i = 0$ or a very small value (for example, considering possible effects in rounding or significant figures in the simulator output). When calibrating reservoir models for real fields, upscaling in the vicinity of the well may have a relevant role for the discrepancy in the injectivity and productivity of the well. An overview about model discrepancy is offered in Formentin *et al.* 2020-a. This local discrepancy can justify the pressure reaching the lower pressure limit for a short period when opening the well. In this case, we suggest estimating tol_i in order to encompass this short period that targets are not met due to a local, punctual problem.

Figure 4.5 illustrates the classification of simulated continuous quantities into binary quantities for $tol_i = 0$. On the one hand, on the top of Figure 4.5, all the scenarios classified as non-implausible scenarios produce the simulation target during the whole history, and $B_i(x) = 0$. For these scenarios, the corresponding pressure is not adjusted (top of Figure 4.5.b), but at least kept higher than the lower pressure limit at all times. On the other hand, all the scenarios classified as implausible do not produce the simulation target over the whole history period because they reached the lower pressure limit (bottom of Figure 4.5.b), $B_i(x) =$

1. Because the lower pressure limit is much lower than the lower operating pressure limit, we also want to rule out this part of the input space which gives rise to scenarios with abnormal behaviour.

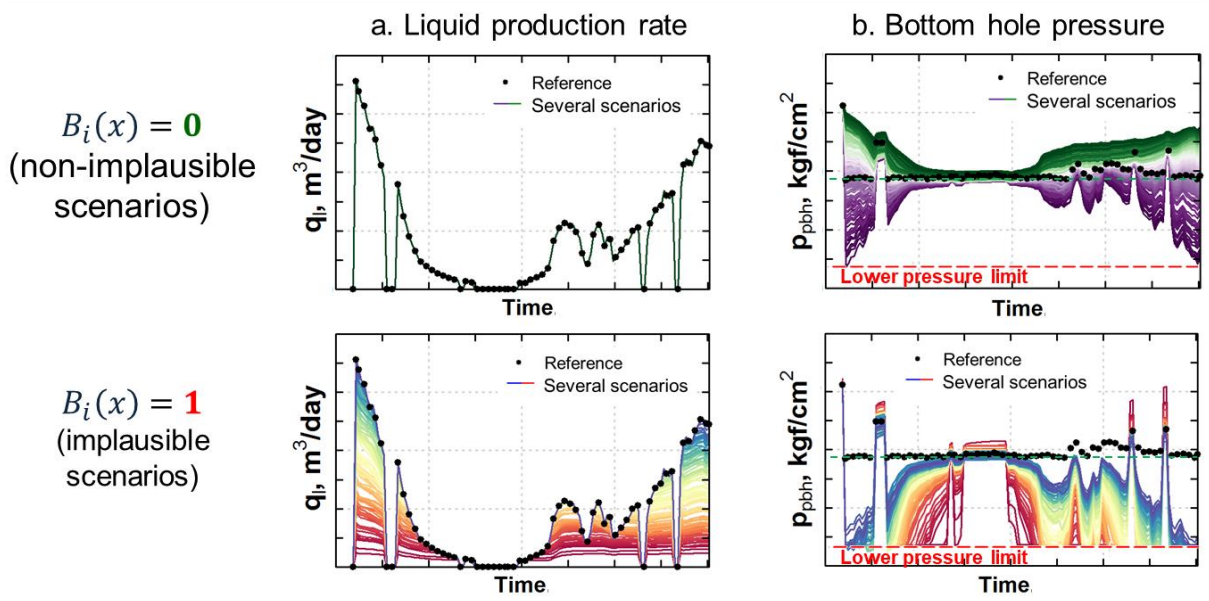


Figure 4.5. Simulated continuous quantities of interest classified in binary quantities; scenarios reaching the simulation target are currently classified as non-implausible $B_i(x) = 0$, note that the variability in the pressures is high; emulators for continuous quantities can later evaluate the pressure in the region with smooth behaviour. Scenarios which do not reach the simulation target are classified as implausible $B_i(x) = 1$ (red), it is due to pressures at the level of the lower pressure limit.

4.2.2 Pattern 2: Breakthrough Time

Water breakthrough time (BT) is the first time when water reaches the production well. In field management, this measured time and subsequent water-oil ratio trends are usually key performance indicators. They can also be indicative of channelling and bypassing problems in the field (Baker 1998).

Besides being a critical aspect for reservoir management, BT leads to a discontinuity in simulation outputs which are related to the water production. In a three-phase system (oil, water and gas), it defines two distinct regions:

- **Region 1**: BT did not occur; the cumulative oil production is equal to cumulative liquid production;
- **Region 2**: BT occurred; cumulative oil and water production add up to cumulative liquid.

Figure 4.6.a is a conceptual description of two distinct regions in cumulative water production given a unique uncertain attribute (related to porosity). To create one statistical

model to represent each one of the trends would be more efficient than to construct a single model representative of both regions. When we have the historical data indicating whether the BT occurred in the real well, we can define one of the regions to consider. Figure 4.6.b provides an example of a cross-plot for the well PROD024 highlighting the physical boundary ($W_p = 0$) for scenarios in which no BT occured.

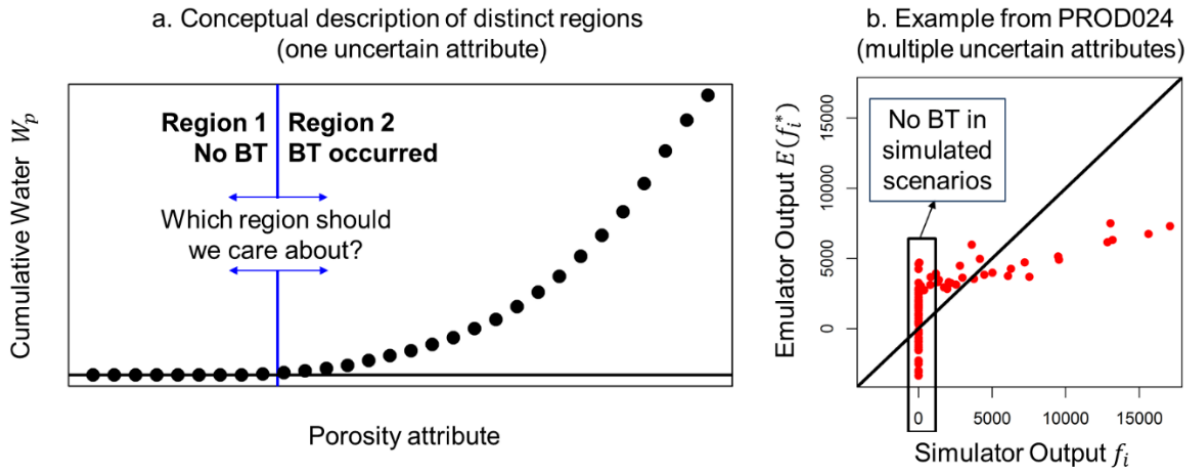


Figure 4.6. (a) Conceptual description of the two distinct regions in the simulation outcome; one uncertain attribute related to porosity is plotted against cumulative water production in a time window; the question raised is answered by the historical data of the well; (b) Cross-plot simulator output *versus* emulator output for cumulative water production of PROD024A before the application of a two-class emulator; a physical boundary draws a pattern for scenarios with no breakthrough. These plots show that a single standard emulator is not able to capture the structure of the simulator behaviour, demonstrating the importance of extending BHMUR technique with two-class emulators.

Our objective is to classify the two distinct regions highlighted in Figure 4.6.a. Therefore, instead of selecting the BT itself (*i.e.* a date) as a quantity of interest, we identify scenarios with similar behaviour as the historical data (did the breakthrough occur or not in historical data?). The proposed approach is suitable because of the iterative nature of BHMUR. Firstly, we identify Regions 1 and 2 ("No BT" or "BT occurred"). Secondly, we construct emulators able to identify these distinct regions and rule out the one which is dissimilar from the historical data. Finally, we remain with a region having smooth behaviour related to cumulative water production, which a far simpler emulator can model (when compared to the emulator required to represent both Regions 1 and 2) for further uncertainty reduction of the input parameter space.

We demonstrate the conceptual description of the labelling process in Figure 4.7 and illustrate it with examples from Phase 3 of our application. The labelling process is coherent with Equation 4.4 presented at the beginning of this section. The legend of Figure 4.7 shows the elements considered in the classification process: *timeline* indicates a time scale; *observed BT* is the historical data (z_i in the statistical formulation introduced in section 4.1); *BT tolerance*

considers our evaluation about the observational error e_i and model discrepancy ϵ_i, describing our uncertainty about the observed BT; tol_i is the tolerance applied in Equation 4.4, derived from BT tolerance; *simulated time* is the last time that scenarios were evaluated with the simulator.

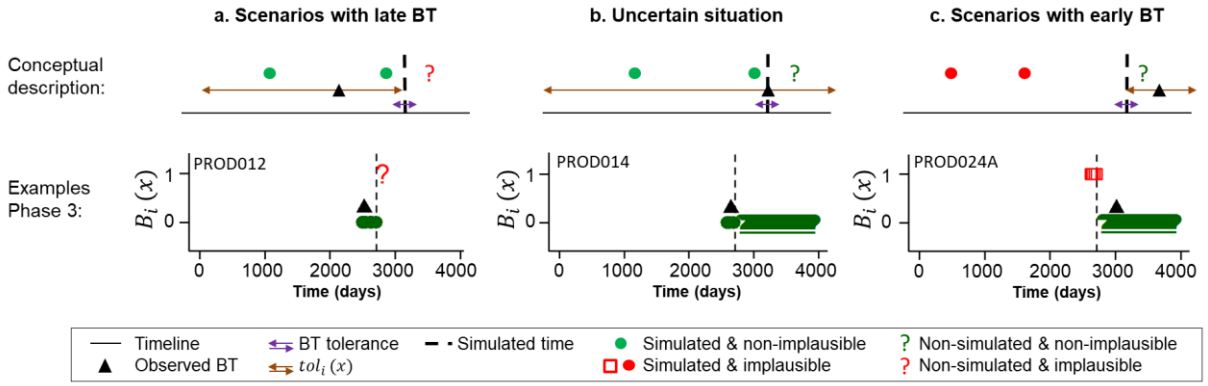


Figure 4.7. Conceptual description and examples for the classification of water Breakthrough Time (BT) into a binary quantity of interest. We aim to define scenarios in the two distinct regions 'NO BT' or 'BT occurred' - (a) Scenarios with late BT are represented by the red question marks. For these scenarios, we do not know when BT occurs; we only know that is sufficiently later than the observed BT; therefore they are labelled as implausible; (b) When the observed BT is near the latest simulated time, we have an uncertain situation about the region of interest, and all scenarios are non-implausible; (c) Implausible are the scenarios with BT occurring earlier than the simulated time, when the latter insufficiently earlier than the observed BT.

The BT of the simulated scenarios - $f_i(x)$ in the statistical formulation introduced in Section 4.1 - is represented in four distinct groups, a combination between the labels: green for non-implausible, red for implausible scenario; circle for scenarios which BT is identified during the simulated period; question mark for scenarios which BT is later than the simulated period (*i.e.* we do not know the exact BT of these scenarios, we *only* know that it is later than the simulated time).

Within this framework, we highlight three distinct cases to systematically classify simulation scenarios as implausible or non-implausible (*see* Figure 4.7):

(a) **Scenarios with late BT**: we classify as implausible the scenarios for which the water breakthrough did not occur during the simulated period, while the observed breakthrough occurred before the simulated time minus the breakthrough tolerance. These implausible scenarios are in Region 1 "No BT" in Figure 4.6. They lead to a discontinuity in the evaluation of cumulative water production. In this case, the value of the tolerance tol_i in Equation 4.4 is between zero and the simulated time and can be written as:

$$tol_i(x) = \begin{cases} z_i, & if\ z_i - f_i(x) \geq 0 \\ simulated\ time - z_i, & otherwise \end{cases} \quad (4.11)$$

(b) **Uncertain situation:** we classify as non-implausible all the simulated scenarios when the observed water breakthrough of a well occurred within the tolerance around simulated time (purple arrows in Figure 4.7), that is, it is uncertain which of the two regions in Figure 4.6 we should care. In this case, the value of the tolerance tol_i in Equation 4.4 can take any non-negative value, and $\text{tol}_\text{i} \rightarrow \infty$;

(c) **Scenarios with early BT:** we classify as implausible the scenarios for which the water breakthrough occurred during the simulated time, while the observed breakthrough occurred later than the simulated time plus the breakthrough tolerance. These implausible scenarios are in the region labelled as "BT occurred" in Figure 4.6. They lead to a spurious evaluation of cumulative water production since no water has been produced until the simulated time. In this case, the value of the tolerance tol_i in Equation 4.4 is greater than the simulated time and can be written as:

$$tol_i = z_i - simulated\ time \tag{4.12}$$

Once we set our tolerance tol_i for each case above, we define conditional tolerances. The definition of the water BT for observed data and simulated scenarios is discussed by Formentin *et al.* (2019-a), Almeida *et al.* (2018) and Lawal *et al.* (2007).

4.3 Validation of emulators and selection of concurrent emulators

One important safeguard in favour of an appropriate level of uncertainty reduction under the BHMUR formulation is the validation of emulators. Emulators diagnosis and validation indicate that predefined quality criteria are met, avoiding overconfident emulators to be applied in the implausibility analysis (and potentially wrongly exclude non-implausible regions of the input space). We can build several emulators for a given quantity of interest and check if they are valid (*i.e.* safe). Among concurrent (and safe) emulators, we also need to select one emulator among several emulators for a given quantity of interest. This process is called selection of concurrent emulators, and the chosen emulator (which is safe and informative) is used in the implausibility analysis.

In this section, we describe several indicators applied in the validation of emulators and selection of concurrent emulators (summary in Table 4.1). We exemplify how the combination of these indicators provide a framework to quantify the quality and performance of emulators.

Table 4.1. Summary of indicators applied to the validation of emulators and selection of concurrent emulators. Some indicators are appropriate for both continuous and binary quantities of interest.

Indicator	Type of emulator		Main application	
	Continuous	Binary	Validation of Emulators	Selection of concurrent emulators
Information index	•	•		•
Credible interval diagnostics	•		•	
Positive Predictive Values		•	•	
Negative Predictive Values		•	•	

4.3.1 Information index

Information index D_{info} (Equation 4.13) is an estimation of the proportion of the remaining input space that can be ruled out (as implausible) in the process. It is computed through the implausibility measure and implausibility cut-off ω, where $\mathbf{1}(\cdot)$ is an indicator function. The higher the information index, the better: it means that we are able to identify a large proportion of implausible scenarios.

$$D_{info} = \frac{1}{n_{sce}} \sum_{k=1}^{n_{sce}} \mathbf{1}(I(x^k) > \omega) \tag{4.13}$$

The set of size $k = 1, \ldots, n_{sce}$ input locations can be a training set, test set, or a collection of new scenarios (for implausibility analysis, for example). Therefore, we define the information index (a) $\mathrm{D}_{info,i}$ for an emulator of the output i, based on the implausibility $I_i(x^k)$, and (b) $\mathrm{D}_{info,M}$ for a set of emulators, based on the implausibility $I_M(x^k)$. In Equation 4.13, both $I_i(x^k)$ or $I_M(x^k)$ are based on the emulator expected outcome and emulator variance, the traditional implausibility measure (Equation 4.3).

Alternatively, we can estimate an adapted information index $\widetilde{D}_{info}$ based on outcomes from simulated scenarios. We define this indicator by making two changes to the original measure:

1. The emulator expectation $E\left(f_i^*(x)\right)$ is replaced by $f_i(x)$, the simulation outcome;

2. The emulator variance $Var\left(f_i^*(x)\right) = 0$.

In this way, we have:

$$\tilde{I}_i^2(x) = \frac{[f_i(x) - z_i]^2}{Var(\epsilon_i) + Var(e_i)}. \quad (4.14)$$

The corresponding adapted information index $\widetilde{D}_{info} = 1/n_{sce} \sum_1^{n_{sce}} \mathbf{1}(\tilde{I}(x^k) > \omega)$ is particularly relevant in steps of a BHMUR procedure where emulators were not yet constructed. For example, we applied this indicator to (a) select outputs to emulate (next section) and (b) evaluate the level of uncertainty reduction by the end of the BHMUR procedure (*see* the Results and Discussions section of Formentin *et al.* 2020-a).

4.3.2 Continuous quantities of interest

The statistical model that we apply to construct emulators for continuous quantities of interest (*e.g.* oil production, pressure) is a low order polynomial regression. Multiple indicators and qualitative analysis of diagnostics plots aid the validation of emulators. Besides the information index, we assess the Credible Interval Diagnostics (Bastos, 2010) and monitor adjusted-R^2 and RMSEn (Moreno *et al.* 2018).

Casella (2008) and Hair *et al.* (1998) are relevant references for specialised background in statistical inference and multivariate regression. Barber (2012), Bastos (2010), Busby (2009), O'Hagan (2004) and Oakley (1999) provide background for emulation techniques with Gaussian process.

4.3.2.1 Credible Interval Diagnostics.

Based on Bastos and O'Hagan (2009), we denote by $CI^k(\alpha)$ a $100\alpha\%$ credible interval for the simulator output $\mathrm{f_i}\left(\mathrm{x_{A_i}^k}\right)$ at $k = 1, \dots, n_{sce}$ input locations denoted $\mathrm{x_{A_i}^k}$, with $0 \leq \alpha \leq 1$. The Credible Interval Diagnostics is defined in Equation 4.15, where $\mathbf{1}(\cdot)$ is an indicator function. It computes the proportion of scenarios for which the simulation outcome is covered by the credible interval provided by the emulator.

$$D_{CI}(f_i(x)) = \frac{1}{n_{sce}} \sum_{k=1}^{n_{sce}} \mathbf{1}\left(f_i\left(x_{A_i}^k\right) \in CI^k(\alpha)\right) \quad (4.15)$$

Credible Interval Diagnostics is an appropriate indicator to determine whether the uncertainty estimation for an emulator corresponds to its actual uncertainty. We would expect the observed value for $D_{CI}(f_i(x))$ to be close to α when the emulator uncertainty is appropriately estimated. An appropriate emulator is neither (a) under-confident, with too large uncertainty on expected outcomes, nor (b) overconfident, with too small uncertainty on expected outcomes.

We are cautious about parts of the input space which are wrongly excluded based on the emulator's expectation. Therefore, the evaluation of an emulator is critical to guard against the possibility of being overconfident: it could lead to wrong conclusions about the implausible space. A cut-off value describes the lower D_{CI} limit to validate an emulator (*e.g.* $D_{CI} > 0.85$ for $\alpha = 0.95$) and we also select concurrent emulators based on this indicator.

4.3.3 Binary quantities of interest

4.3.3.1 Positive and Negative Predictive Value.

Predictive values estimate the probability of the emulator labels being correct. For a data set, the proportion of scenarios correctly classified as (a) implausible - label 1 - is the Positive Predictive Value (PPV); (b) non-implausible – label 0 - is the Negative Predictive Value (NPV) (Altman and Bland 1994), as shown in Equations 4.16 and 4.17, respectively:

$$PPV = \frac{True\ 1}{True\ 1 + False\ 1} \tag{4.16}$$

$$NPV = \frac{True\ 0}{True\ 0 + False\ 0} \tag{4.17}$$

The elements of these indicators are possible combinations between emulator and simulator labels. In Figure 4.8.a, we have: *True 0* (or True Negative) are the scenarios correctly kept by the emulator; *True 1* (True Positive) are the ones correctly excluded; *False 0* (False Negative) are the scenarios wrongly kept, and *False 1* (False Positive) are the ones wrongly excluded. While PPV judges the proportion of scenarios correctly excluded by the emulator, the NPV controls the emulator efficiency to keep suitable scenarios. Importantly, the iterative nature of BHMUR enables us to rule out in later waves scenarios that are wrongly kept. Nevertheless, regions of search space wrongly excluded are critical. Therefore, we must be extra cautious about PPV.

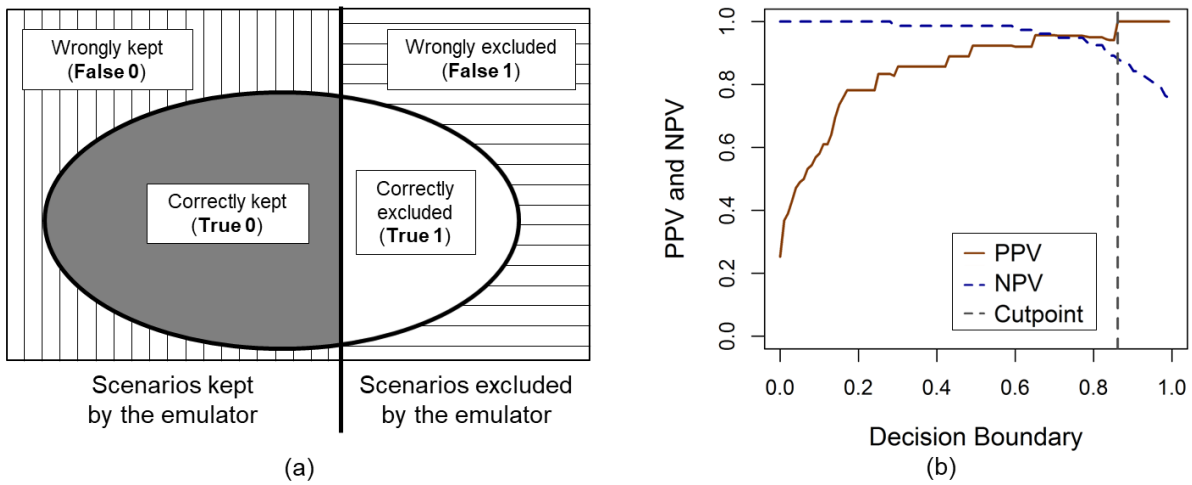


Figure 4.8. Positive and Negative Predictive Values – (a) Possible combinations between emulator and simulator classification, we are careful about scenarios wrongly excluded; (b) PPV and NPV *versus* all possible decision boundaries for an emulator, the cutpoint maximises the PPV and is a reference to define the decision boundary.

Our objective is to choose an optimal decision boundary that maximises the PPV while keeping NPV as high as possible. The graph in Figure 4.8.b facilitates this choice. It plots PPV and NPV *versus* each possible decision boundary. For this example, an optimal decision boundary named cut point is approximately 0.85, where PPV is the maximal.

We set a decision boundary linked to the training data set. Our aversion to wrongly exclude scenarios is implemented via two strategies: (a) increment the training set cutpoint to accommodate small dissimilarities between training and test sets characteristics; (b) choose a fixed unbalanced decision boundary setting that a very high emulator probability (*e.g.* 0.95) is required to exclude a scenario.

4.3.4 Application - Combining indicators to validate emulators and select among concurrent emulators

After constructing competitive emulators, we validate them and select the best emulator based on positive and negative predictive values (for two-class quantities) and on credible interval diagnostics and information index (for continuous quantities). This process constitutes the *STEP 12* of the systematic procedure from Formentin *et al.* (2020-a). These indicators are calculated for training and test sets independently: while indicators for test sets are effectively used to select and validate emulators, we also monitor training sets.

We now detail the choices made for binary quantities of interest in our application. Logistic regression models have a fixed boundary decision on 0.95; for linear discriminant models, we define the decision boundary from the training set (*e.g.* a larger value than the cut point defined in Figure 4.8).

As validation criteria, binary quantities of interest need a minimum Positive Predictive Value of 0.90, *e.g.* at least 90% of the emulator's implausible scenarios correspond to the simulator's implausible scenarios. This unbalanced PPV threshold ($PPV_{thres} \gg 0.50$) represent our aversion to rule out wrongly regions of the search space. Simultaneously, we consider a smaller threshold for the Negative Predictive Value ($NPV_{thres} = 0.50$) since bad scenarios kept in the analysis can be ruled out at later waves. When two concurrent emulators meet the validation criteria, the one with higher NPV is selected.

We show an example of the selection of concurrent emulators for a binary quantity. Emulator 1 is based on logistic regression (Figures 4.9.a and b); Emulator 2, a linear discriminant analysis (Figures 4.9.c and d). Training set is in red and orange and test set is in light and dark blue, respectively. Figures 4.9.a and c are traditional diagnostics plots. They compare binary simulation outputs with the emulator probability $\mathrm{P}(\mathrm{B_i} = 1|\mathrm{x})$. The grey, vertical lines are the boundary decision applied.

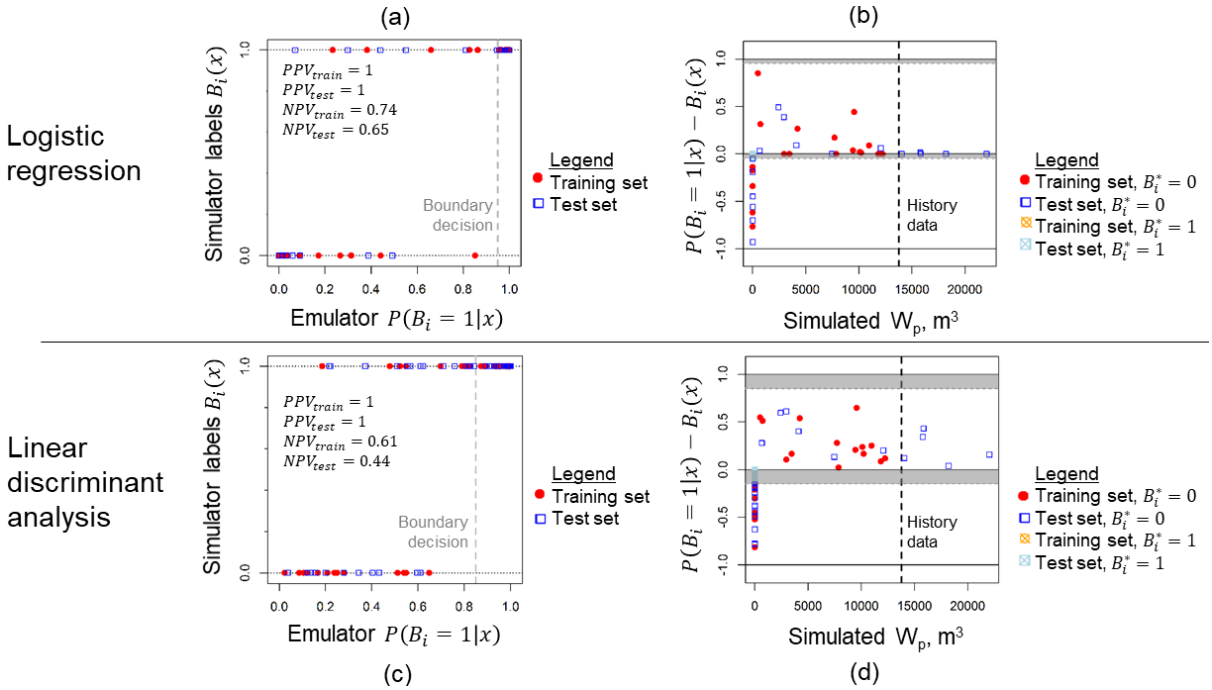


Figure 4.9. Validation and selection of concurrent two-class emulators – logistic regression and linear discriminant analysis; (a) and (c) cross plots for emulator outcomes and simulation labels, they highlight the decision boundary; (b) and (d) cumulative liquid production *versus* the difference between emulation probability and simulation labels, they highlight the volume of water produced in the history data, points in the grey regions are judged as implausible. Both emulators meet the validation criteria, but logistic regression is chosen because of its higher NPV.

Figures 4.9.b and d are alternative plots for diagnostics, relating the continuous output for the simulator in the *x-axis* (*i.e.* cumulative water production in 2,710 days, Phase 3) with the difference between the probability calculated by the emulator $\mathrm{P}(\mathrm{B_i} = 1|\mathrm{x})$ and the simulator binary outcome $\mathrm{B_i}\,(x)$ in the *y-axis*. The vertical line places the historical data. In this

example, W_p of history data larger than zero indicates that we should rule out the region 'No BT' and care about the region 'BT occurred'. Provided the decision boundary of the two-class emulator, we can highlight intervals in the *y-axis* with scenarios to be ruled out by the emulator; these scenarios are coloured in orange and light blue to enable the distinction. For example, in Figure 4.9.d, the decision boundary is at 0.95, *i.e.* $P(B_i = 1|x) > 0.95$ consists of the exclusion of the scenario. If $P(B_i = 1|x) - B_i(x)$ is in the interval [0.00; -0.05] or is equal to [0.95; 1.00] (grey regions), this means that $P(B_i = 1|x) > 0.95$ and the scenario will be excluded by the emulator (either a correct or wrong exclusion). Scenarios in these regions are plotted in orange and light blue; the absence of scenarios in the grey region for both emulators is in accordance with PPV=1 for training and test sets.

Both scenarios have PPV equal to 1, indicating that neither of them wrongly exclude scenarios based on the emulator evaluation. The drawback of these emulators is that they keep several scenarios in the 'No BT' region (*e.g.* NPV < 1). In this situation, we choose logistic regression, the more informative emulator, *i.e.* the one with higher NPV.

For continuous quantities of interest, we define a threshold for the information index $D_{info,thres} = 0.05$, *e.g.* at least 5% of remaining search space needs to be ruled out by a valid emulator. An emulator is valid when $D_{info,i} \geq 0.05$ and the Credible Interval Diagnosis with $\alpha = 95\%$ is larger than 0.85, *i.e.* the 95% credible interval defined by the emulator covers at least 85% of the simulation outcomes. These criteria combined validate an emulator as informative and with accurate uncertainty estimation. If more than one concurrent emulators are valid, the emulator with larger D_{CI} is selected, justified by our caution about wrongly ruling out non-implausible scenarios.

We expose two examples of cross-plot diagnostics of concurrent emulators for continuous quantities of interest (Figure 4.10 and 4.11). The plots are for the training (a and c) and test (b and d) sets. The concurrent emulators are regression models with (a) terms of first-order only and (b) terms of first and second order. Note that we omit labels for the *y-axis* of the cross plots for the test sets: they are the same as for the training sets.

The first example (Figure 4.10) highlights the potential of the information index as a selection indicator. The vertical bars account for $\pm\omega\left[Var\left(f_i^*(x)\right) + Var(\epsilon_i) + Var(e_i)\right]^{1/2}$, which is the implausibility cut-off ω multiplied by the denominator of the implausibility measure. The cross-plots of the first emulator (Figures 4.10.a and .b) have emulator expectation

with a curved trend, which indicates, for example, that it missed some quadratic relationship. Both emulators have D_{CI} of training and test sets higher than 90%, indicating that the level of uncertainty is appropriately estimated. Nevertheless, the information index D_{info} of training and test sets are lower than 5% expressing the large uncertainty related to this emulator.

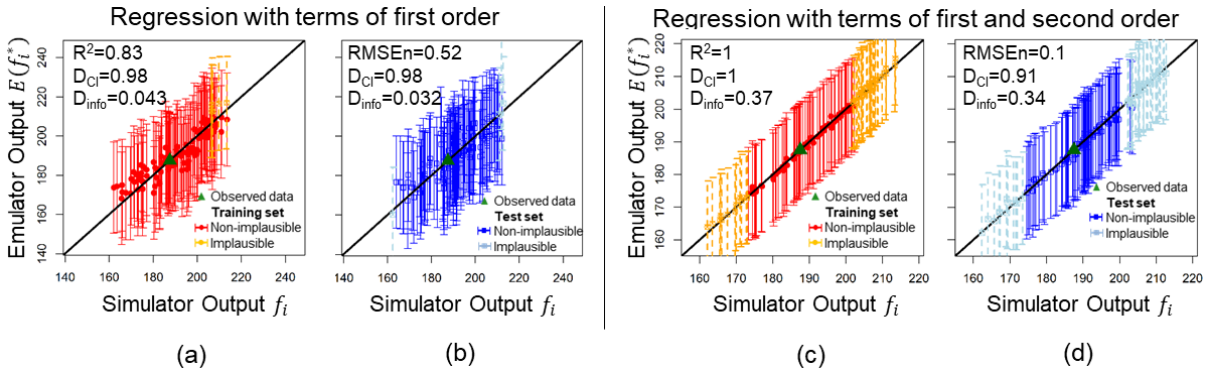


Figure 4.10. Concurrent emulators for cumulative water production of the well RJS19 in Wave 6, Phase 3. Examples of Information Index for training and test sets, vertical bars are $\pm\omega[Var(f_i^*(x)) + Var(\epsilon_i) + Var(e_i)]^{1/2}$ for each scenario emulated. Orange and light blue are implausible scenarios, red and blue are non-implausible. The first emulator is not informative enough to be selected. The second emulator is informative and estimates its uncertainty correctly.

The cross-plots of the second emulator (Figures 4.10.c and .d) show emulator expected outputs very close to the 45-degree line, indicating that the second-order terms more accurately captured relationships between active variables and the average pressure of the well RJS19 with coherent uncertainty estimation. The information index is above 30%; therefore we select the regression model containing terms of first and second order.

With Figure 4.11, we compare (a) an emulator with an appropriate estimation of uncertainty with (b) an over-confident emulator. For each plot, the vertical bars represent the 95% credible interval estimated around the expected outcome $E(f_i^*)$. Both emulators of Figure 4.11 are highly informative ($D_{info,1} = 0.46$ and $D_{info,2} = 0.67$). Nevertheless, a D_{CI} of the test set close to our threshold 0.85 indicates that the second emulator tends to estimate a small uncertainty, in this case, on the limit of over-confidence. The D_{CI} of the training set is very high (99%), and this difference with the D_{CI} of the test set also reveals this over-confidence characteristic, likely due to overfitting of the linear model. Considering the principles of BHMUR, we select the emulator with terms of first-order only, which has higher D_{CI} and is sufficiently informative.

These examples demonstrate the complementarity of the indicators of quality of an emulator: it is important to evaluate under diverse criteria in order to select and use the better

emulators, able to represent the reservoir simulator with a proper uncertainty and be sufficiently informative, justifying their consideration in the implausibility analysis.

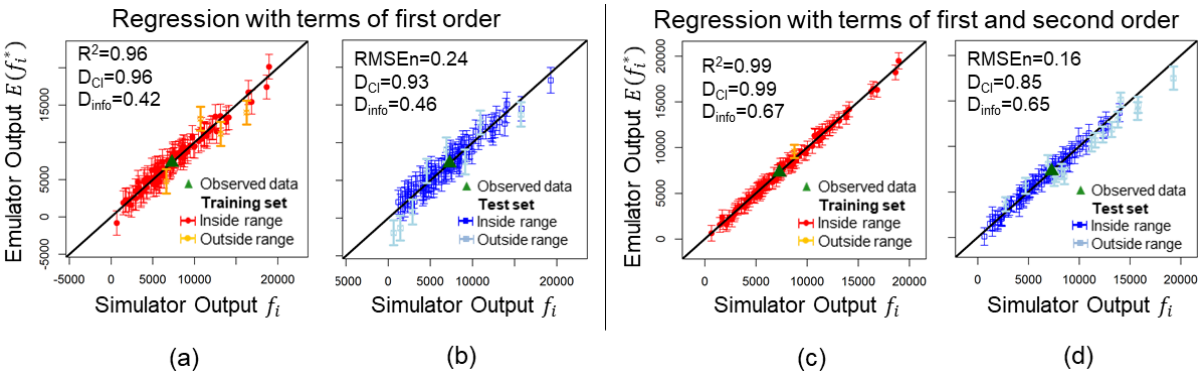


Figure 4.11. Concurrent emulators for cumulative water production of the well RJS19 in Wave 6, Phase 3. Examples of Credible Interval Diagnostics for training and test sets, vertical bars are the range estimated in the 95% credible interval for each scenario emulated. Orange and light blue are scenarios which simulation output is out of the range. Both emulators reach the selection criteria. Nevertheless, we select the emulator with terms of first-order because it estimates better the uncertainty related to the emulator expectation. The regression model with terms of first- and second-order tends to be overconfident $D_{CI} = 0.85$.

4.4 Selection of outputs to emulate

The selection of outputs to emulate is critical for the effective use of information from a high dimensional output space and, ultimately, to the efficiency of the Bayesian History Matching for Uncertainty Reduction (BHMUR) process. We aim to select a combination of outputs to emulate in order to balance the computational effort invested in constructing and evaluating new scenarios through emulators. In this way, we try to construct informative emulators to be considered in the analysis progressively. We propose a procedure for output selection, which considers the combination of quantities of interest to estimate which one is the most informative.

Two main points differentiates this procedure from the one presented by Ferreira *et al.* (2020): (1) we consider that all quantities of interest are similarly difficult to emulate, *i.e.* we do not invest resources constructing simple emulators that provide estimates of the possible full emulator uncertainty; (2) we look for a combination of quantities of interest that are the most informative, instead of a sequential approach that finds the most informative quantity of interest, then the 2nd most informative and so on.

The proposed systematic procedure is presented in Figure 4.12, and we illustrate its application considering it as *STEP 8* of Formentin *et al.* (2020-a). An important definition to retake is the concept of phases of evaluation, which split historical data into physically meaningful periods to gradually introduce data into the analysis to take advantage of

information from early time. We can construct emulators for quantities of interest from all phases already evaluated, *i.e.* the phase under evaluation φ starts from 1 to the last phase simulated (named here as $phase$). The starting point of the selection process enables this flexibility: *STEP 8.1* is a conditional statement to decide whether we should evaluate quantities of interest of Phase 1 ($\varphi = 1$) or more advanced phases already simulated. Note that in the diagram of Figure 4.12, $phase$ is the latest phase evaluated through simulation, and φ is the phase from which quantities of interest can be selected. If a new wave was simulated, we set $\varphi = 1$ with *STEP 8.2*; otherwise, we choose a more advanced wave in *STEP 8.3* (*i.e.* a successive increment in φ until the latest phase evaluated with simulators).

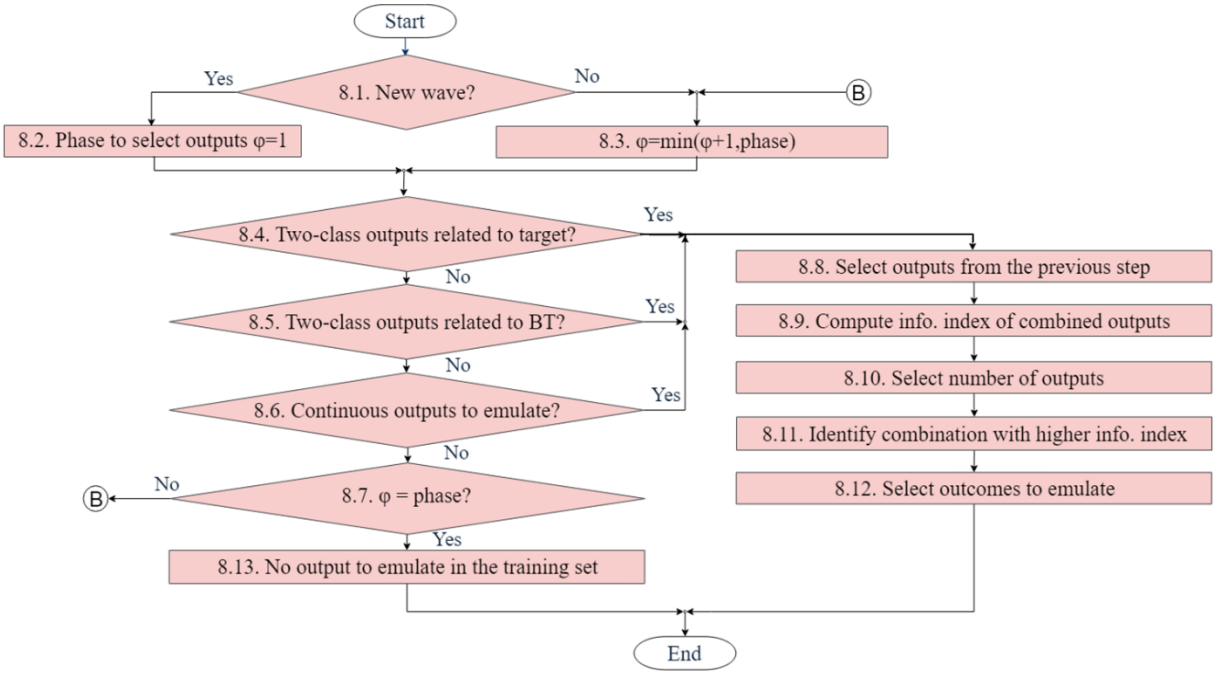


Figure 4.12. Workflow to select outputs to emulate in *STEP 8*. In the first three steps, the phase of evaluation φ is defined. *STEPS 8.4* to *8.6* identify the class of outputs to be considered in the selection. *STEPS 8.8* to *8.12* select a combination of outputs with the potential to have the highest information index.

In *STEPS 8.4* to *8.6*, we check specific classes of outputs to emulate. Two-class outputs allow the identification of scenarios reaching simulation targets or physical boundaries (water breakthrough). We recommend firstly to construct emulators for these binary quantities (*STEPS 8.4* and *8.5*). They can rule out regions of the input space that lead to discontinuous behaviour, which is beneficial to emulate continuous outputs (*STEP 8.6*). For continuous outputs, we use the formulation of adapted implausibility measure $\tilde{I}_i$ and $\widetilde{D}_{info}$, Equation 4.14 because *STEP 8* occurs previously than the construction of emulators in the main workflow (Formentin *et al.*, 2020-a). We highlight that this formulation implies an imprecision in the output selection, which consists of all outputs being considered equally challenging to emulate.

Although this consideration is not expected to be true, it allows us to speed up the process compared to a method requiring a preliminary emulation (Ferreira *et al.* 2020).

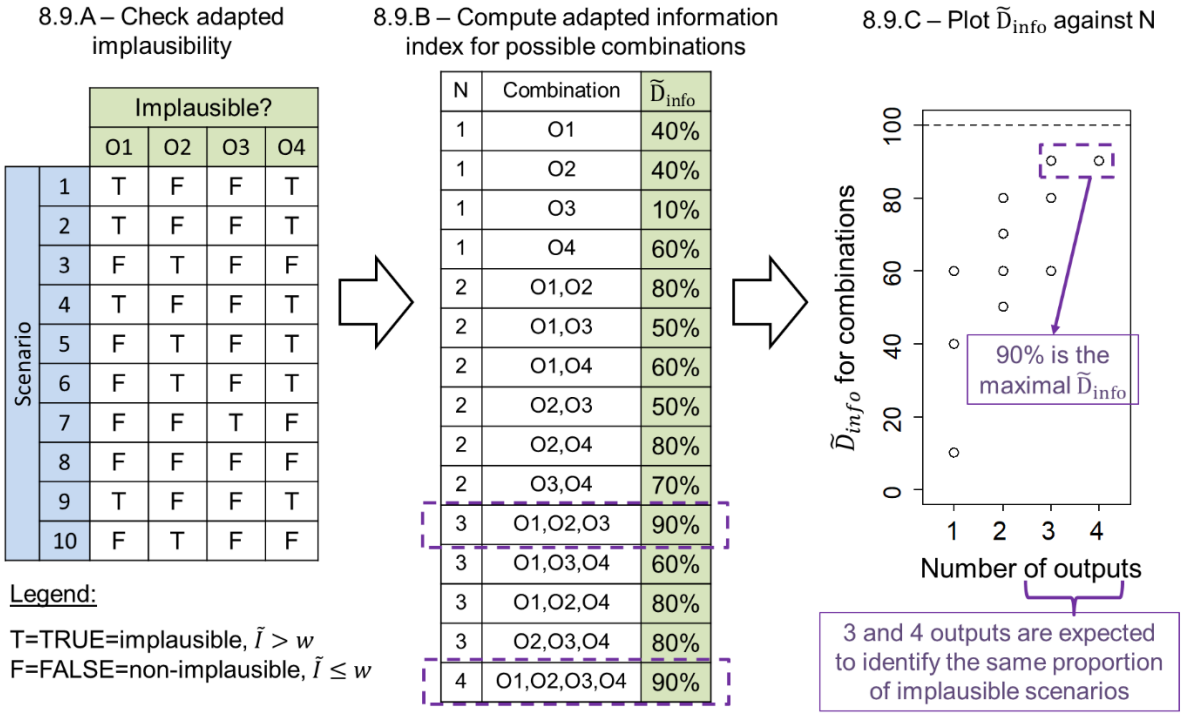


Figure 4.13. Illustration of *STEP 8.9* in the workflow to select outputs to emulate: (a) Compute the adapted implausibility, which is based on the simulations, and classify the scenarios based on a pre-defined implausibility cut-off; (b) Aggregate the adapted implausibility classification via adapted implausibility index, indicating the most informative combination of outputs; (c) Display the adapted implausibility for combinations of outputs with the corresponding number of outputs to support the choice.

When no quantity of interest offers (a) enough variability and (b) potential to rule out implausible regions of the space (*i.e.* $\widetilde{D}_{info} \geq D_{info,thres}$, in our application $D_{info,thres} = 5\%$), no quantity of interest is selected. In *STEP 8.7*, we consider the case where $\varphi = phase$, a condition that would terminate our selection process with no output to emulate based on the training set available (*STEP 8.13*). If there is no output offering variability and being informative, but $\varphi < phase$, we increment φ and evaluate the next phase that we have data available.

In *STEP 8.8*, we select relevant outputs following *STEPS 8.4* to *8.6*. In *STEP 8.9*, the adapted information index for all possible combinations of outputs is calculated. Figure 4.13 illustrates this procedure with ten scenarios. In *STEP* 8.9.A, for each output considered (O1 to O4), we classify scenarios from the training set based on the adapted implausibility: scenarios with $\tilde{I}_i$ larger than our implausibility cut-off w is set as TRUE (implausible). In *STEP* 8.9.B, we compute the adapted information index $\widetilde{D}_{info}$ for each possible combination (including outputs standing alone). We can visualise the $\widetilde{D}_{info}$ dependent on the number of outputs and

several possible combinations. In this illustrative example, two combinations of outputs are expected to deliver the same proportion of implausible scenario. Based on this information, we can, for example, decide to invest resources only emulating O1, O2 and O3 because this combination is expected to have a similar result as by emulating the four outputs.

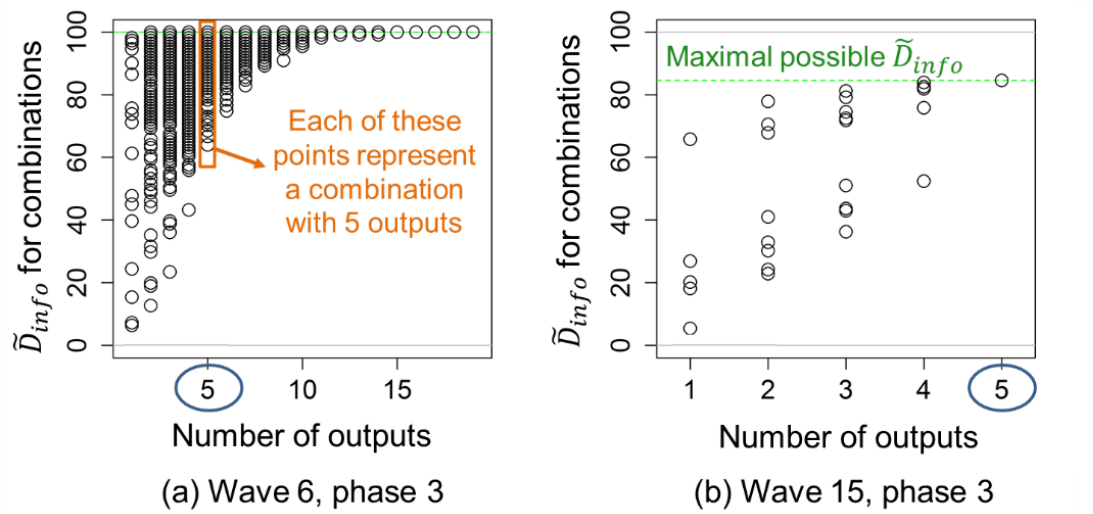


Figure 4.14. Adapted information index for combinations of continuous outputs of Phase 3 (STEP 8.9) – (a) in Wave 6, 19 outputs are available, the combinations of five outputs are highlighted; with 15 outputs the adapted information index reaches the limit of 100%; (b) in Wave 15, five outputs are available; the maximum adapted information index is approximately 85%. The comparison between the two waves emphasises the uncertainty reduction along the waves.

In Figure 4.14, we demonstrate the application of this step in the case study. The combinations with five outputs (Figure 4.14.a) are highlighted: each circle is a possible combination of 5 outputs among the 19 possible outputs. These plots were constructed to select outputs from Phase 3 in the Waves 6 and 15. They show a trend to increase the adapted information index as the number of outputs in the combinations increase. In Figure 4.14.a, the maximal possible $\widetilde{D}_{info}$ is reached by all combinations of 15 outputs, for example.

In *STEP 8.10*, we select a number of outputs (*e.g.* three in Figure 4.13). A design option stands for the number of outputs to be emulated. One should account for the possibility to miss informative outputs in the stage of the process, or for increasing the computation cost demanded to construct emulators and evaluate additional outputs. In *STEP 8.11*, we identify the combinations for the selected number of outputs with the highest information index ($\widetilde{D}_{info}$). In Figure 4.13, we would have only one combination with three outputs among the four possible combinations. Nevertheless, in Figure 4.14.a, all combinations of 15 outputs provide the maximal possible $\widetilde{D}_{info}$. A coherent way to select one of these combinations and the corresponding outputs to select (*STEP 8.12*) is to choose the combination that offers the highest combined adapted implausibility measure, in our application $\tilde{I}_M$.

We add Figure 4.14.b to emphasise that, in later waves, both (a) the maximal possible $\widetilde{D}_{info}$ reduces within the waves (approximately 100% in Wave 6 *versus* 85% in Wave 15), and (b) the number of outputs possible to emulate tends to decrease - we advanced nine waves in the process, and the maximal number of outputs with enough variability and $\widetilde{D}_{info} \geq D_{info,thres}$ is five outputs, the others do not offer enough variability.

4.5 Estimation of observational errors in historical data

We estimate the size of observational errors to be included in the BHMUR process. We follow the formulation stated from Equation 4.1, where y is a vector of dimensions $i \in R^q$ of quantities of interest from the physical reservoir. In our application, the physical reservoir itself is equivalent to the simulation of the hypothetical reality without any noisy. The vector z is a vector of measurable quantities (in our application, the hypothetical reality with noisy) and e^{total} is a vector of the total observational error.

$$z_i = y_i + e_i^{total} \tag{4.18}$$

We expand the total error into two components - random (ran) and systematic (sys):

$$z_i = y_i + e_{z_i}^{ran} + e_{z_i}^{sys} \tag{4.19}$$

In our analysis, we define these errors by univariate normal distributions with standard deviation σ with the following properties:

- Random error is uncorrelated and independent for each quantity i observed for a well w at any time t of the historical period, with distribution $e_{z_i}^{ran} \sim N\left(0, \sigma_{z_i}^{ran\,2}\right)$, $E\left[e_{z_i}^{ran}\right] = 0$ and $var\left[e_{z_i}^{ran}\right] = \sigma_{z_i}^{ran\,2}$;

- Systematic error is correlated in all the times of the historical period for each type of quantity *i* made for a well *w*; $e_{z_i}^{sys} \sim N\left(0, \sigma_{z_i}^{sys\,2}\right)$, $E\left[e_{z_i}^{sys}\right] = 0$ and $var\left[e_{z_i}^{sys}\right] = \sigma_{z_i}^{sys\,2}$. This means that we assumed that the systematic portion of error is the same in all measurements of the time series.

Statistical independence is assumed, *i.e.* $cov\left(e_{z_i}^{ran}, e_{z_i}^{sys}\right) = 0$. Therefore, the sum of errors associated with each measurement is:

$$\left(\sigma_{z_i}^{total}\right)^2 = \left(\sigma_{z_i}^{ran}\right)^2 + \left(\sigma_{z_i}^{sys}\right)^2 \quad (4.20)$$

Table 4.2 lists quantities of interest usually measured in the field with the respective errors applied in our case study. We define $\sigma_{z_i}^{ran} = prop_{z_i}^{ran} \cdot z_i$ and $\sigma_{z_i}^{sys} = prop_{z_i}^{sys} \cdot z_i$. In a univariate normal distribution, we have a 99% probability that the measurement falls in the interval $\pm 3\sigma$ from the mean.

A consistency check of the total volume of liquid produced implies that $q_o = q_l - q_w$. If liquid and water production rates are independent measurements, *i.e.* $cov\left(e_{q_l}^{total}, e_{q_w}^{total}\right) = 0$, we derive Equation 4.21 to compute the error associated with q_o:

$$\left(\sigma_{q_o}^{total}\right)^2 = \left(\sigma_{q_l}^{total}\right)^2 + \left(\sigma_{q_w}^{total}\right)^2 \quad (4.21)$$

For an under-saturated reservoir (*i.e.* reservoir pressure above the bubble point), gas production is proportional to oil production. Therefore, $\frac{\sigma_{qg}^{total}}{q_g} = \frac{\sigma_{q_o}^{total}}{q_o}$.

Table 4.2. Maximal errors considered in the hypothetical reality of the HR-82.

Observed data	Random error ($\pm 3\sigma_{z_i}^{ran}$)	$prop_{z_i}^{ran}$	Systematic error ($\pm 3\sigma_{z_i}^{sys}$)	$prop_{z_i}^{sys}$
Liquid production rate	$\pm 0.06 \cdot q_l$	0.02	$\pm 0.03 \cdot q_l$	0.01
Water injection rate	$\pm 0.06 \cdot i_w$	0.02	$\pm 0.00 \cdot i_w$	0.00
Water production rate	$\pm 0.05 \cdot q_w$	0.05/3	$\pm 0.05 \cdot q_w$	0.05/3
Bottom hole pressure of production and injection wells	$\pm 0.02 \cdot p_{pbh}$ or $\pm 0.02 \cdot p_{ibh}$	0.02/3	$\pm 0.02 \cdot p_{pbh}$ or $\pm 0.02 \cdot p_{ibh}$	0.02/3

We follow with an analysis to compute variances and expected errors for quantities derived from measured data in the field (*e.g.* rates and pressure). The analysis results in:

- Equation 4.40 for Cumulative quantities (Liquid and Water Production);
- Equation 4.43 for Cumulative Oil Production (N_p);
- Equation 4.61 for averaged quantities of interest (*e.g.* bottom hole pressure).

4.5.1 Cumulative quantities of wells

In this section, we propose an approach to account for errors in cumulative quantities of interest, when the actual measurements are made for rates. Corresponding cumulative quantities *Z* and *Y* are defined for a given series of measurements from the same

type i (*e.g.* time series of liquid production rate). N is the number of measurements considered until a specific time step t ($t \in [1, N]$), where we compute the cumulative quantity:

$$Z_{i,N} = \sum_{s=1}^{N} z_{i,s} \cdot (t_s - t_{s-1}) \tag{4.22}$$

From Equation 4.19, we have cumulative quantities:

$$Z_{i,N} = Y_{i,N} + e^{ran}_{Z_{i,N}} + e^{sys}_{Z_{i,N}} \tag{4.23}$$

In our application, the time between measurements varies from 28-31 days, but for simplification we consider $\Delta t = \Delta t_s = 30\ days$, or $Z_{i,N} = 30 \sum_{s=1}^{N} z_{i,s}$. Therefore:

$$E\left[e^{ran}_{Z_{i,N}}\right] = E\left[\Delta t \sum_{s=1}^{N} e^{ran}_{z_{i,s}}\right] = 0 \tag{4.24}$$

And

$$Var\left[e^{ran}_{Z_{i,N}}\right] = Var\left[\Delta t \sum_{s=1}^{N} e^{ran}_{z_{i,s}}\right] \tag{4. 25}$$

Because $e^{ran}_{Z_{i,N}}$ is an independent error and $prop^{ran}_{z_i}$ is a constant value for a given z_i (check Table 4.2):

$$Var\left[e^{ran}_{Z_{i,N}}\right] = \Delta t^2 \cdot Var\left[\sum_{s=1}^{N} e^{ran}_{z_{i,s}}\right] = \Delta t^2 \cdot \sum_{s=1}^{N} Var\left[e^{ran}_{z_{i,s}}\right] \tag{4.26}$$

$$Var\left[e^{ran}_{Z_{i,N}}\right] = \Delta t^2 \cdot \sum_{s=1}^{N} (prop^{ran}_{z_i} \cdot z_{i,s})^2 \tag{4.27}$$

$$Var\left[e^{ran}_{Z_{i,N}}\right] = (prop^{ran}_{z_i} \cdot \Delta t)^2 \cdot \sum_{s=1}^{N} (z_{i,s})^2 \tag{4.28}$$

$$\sigma^{ran}_{Z_{i,N}} = \sigma\left[e^{ran}_{Z_{i,N}}\right] = \Delta t \cdot prop^{ran}_{z_i} \cdot \sqrt{\sum_{s=1}^{N} (z_{i,s})^2} \tag{4.29}$$

$$\frac{\sigma_{Z_{i,N}}^{ran}}{Z_{i,N}} = \frac{\Delta t \cdot prop_{z_i}^{ran}}{Z_{i,N}} \cdot \sqrt{\sum_{s=1}^{N} {z_{i,s}}^2} \tag{4.30}$$

For the systematic portion of error:

$$E\left[e_{Z_{i,N}}^{sys}\right] = E\left[\Delta t \sum_{s=1}^{N} e_{z_{i,s}}^{sys}\right] = 0 \tag{4.31}$$

And

$$Var\left[e_{Z_{i,N}}^{sys}\right] = Var\left[\Delta t \sum_{s=1}^{N} e_{Z_{i,s}}^{sys}\right] \tag{4.32}$$

Taking $e_{z_i}^{sys}$ as $e_{st} \cdot \sigma_{z_i}^{sys}$, where e_{st} is originated from a normal distribution with $E[e_{st}] = 0$ and $Var[e_{st}] = 1$, we can formulate as:

$$Var\left[e_{Z_{i,N}}^{sys}\right] = \Delta t^2 Var\left[\sum_{s=1}^{N} e_{st} \cdot \sigma_{z_{i,s}}^{sys}\right] = \Delta t^2 Var\left[e_{st} \cdot \sum_{s=1}^{N} \sigma_{z_{i,s}}^{sys}\right] \tag{4.33}$$

$$Var\left[e_{Z_{i,N}}^{sys}\right] = \Delta t^2 \left(\sum_{s=1}^{N} \sigma_{z_{i,s}}^{sys}\right)^2 \cdot Var[e_{st}] \tag{4.34}$$

$$Var\left[e_{Z_{i,N}}^{sys}\right] = \Delta t^2 \cdot \left(\sum_{s=1}^{N} \sigma_{z_{i,s}}^{sys}\right)^2 \tag{4.35}$$

$$\sigma_{Z_{i,N}}^{sys} = \sigma\left[e_{Z_{i,N}}^{sys}\right] = \Delta t \cdot \sum_{s=1}^{N} \sigma_{z_{i,s}}^{sys} = \Delta t \cdot \sum_{s=1}^{N} prop_{z_i}^{sys} \cdot z_{i,s} \tag{4.36}$$

$$\sigma_{Z_{i,N}}^{sys} = prop_{z_i}^{sys} \cdot Z_{i,N} \tag{4.37}$$

$$\frac{\sigma_{Z_{i,N}}^{sys}}{Z_{i,N}} = prop_{z_i}^{sys} \tag{4.38}$$

$$\left(\frac{\sigma_{Z_{i,N}}^{total}}{Z_{i,N}}\right)^2 = \left(\frac{\sigma_{Z_{i,N}}^{ran}}{Z_{i,N}}\right)^2 + \left(\frac{\sigma_{Z_{i,N}}^{sys}}{Z_{i,N}}\right)^2 \tag{4.39}$$

$$\left(\frac{\sigma_{Z_{i,N}}^{total}}{Z_{i,N}}\right)^2 = \left(\frac{\Delta t \cdot prop_{Z_i}^{ran}}{Z_{i,N}}\right)^2 \cdot \sum_{s=1}^{N} (z_{i,s})^2 + \left(prop_{z_i}^{sys}\right)^2 \quad (4.40)$$

Since the ratio $(\sum_{s=1}^{N} {z_{i,s}}^2)/\,(Z_{i,N})^2$ tend to be very small, the portion of systematic error tends to be more important than the part of random error.

For cumulative oil production (Np):

$$\left(\sigma_{N_{p,N}}^{total}\right)^2 = \left(\sigma_{L_{p,N}}^{total}\right)^2 + \left(\sigma_{W_{p,N}}^{total}\right)^2 \quad (4.41)$$

$$\left(\sigma_{N_{p,N}}^{total}\right)^2 = {L_{p,N}}^2 \cdot \left[\left(\frac{\Delta t \cdot prop_{q_l}^{ran}}{L_{p,N}}\right)^2 \cdot \sum_{s=1}^{N} {q_{l,s}}^2 + \left(prop_{q_l}^{sys}\right)^2\right] + {W_{p,N}}^2 \cdot \left[\left(\frac{\Delta t \cdot prop_{q_w}^{ran}}{W_{p,N}}\right)^2 \cdot \sum_{s=1}^{N} {q_{w,s}}^2 + \left(prop_{q_w}^{sys}\right)^2\right] \quad (4.42)$$

$$\frac{\sigma_{N_{p,N}}^{total}}{N_{p,N}} = \frac{\sqrt{\Delta t \cdot \left[\left(prop_{q_l}^{ran}\right)^2 \cdot \sum_{s=1}^{N} {q_{l,s}}^2 + \left(prop_{q_w}^{ran}\right)^2 \cdot \sum_{s=1}^{N} {q_{w,s}}^2\right] + \left(prop_{q_l}^{sys}\right)^2 \cdot {L_{p,N}}^2 + \left(prop_{q_w}^{sys}\right)^2 \cdot {W_{p,N}}^2}}{N_{p,N}} \quad (4.43)$$

4.5.2 Average quantities of wells

We perform similar calculations for averaged quantities $\bar{z}$ and $\bar{y}$ which are defined for a given series of measurements from the same type i (*e.g.* bottom hole pressure); N corresponds to the number of measurements to be averaged (*e.g.* a time window). From Equation 4.18, we have:

$$\overline{z_{\iota,N}} = \overline{y_{\iota,N}} + e_{\overline{z_{\iota,N}}}^{ran} + e_{\overline{z_{\iota,N}}}^{sys} \quad (4.44)$$

Where we define:

$$\overline{z_{\iota,N}} = \frac{\sum_{s=1}^{N} z_{i,s}}{N} \quad (4.45)$$

$$\overline{y_{\iota,N}} = \frac{\sum_{s=1}^{N} y_{i,s}}{N} \tag{4.46}$$

$$E\left[e_{\overline{z_{\iota,N}}}^{ran}\right] = \frac{E\left[\sum_{s=i}^{N} e_{z_{i,s}}^{ran}\right]}{N} = 0 \tag{4.47}$$

These equations lead to:

$$\sigma_{\overline{z_{\iota,N}}}^{ran} = \frac{prop_{z_i}^{ran}}{N} \cdot \sqrt{\sum_{s=1}^{N} \left(z_{i,s}\right)^2} \tag{4.48}$$

In the case of bottom hole pressure (BHP), the measurement z_i is considered a smooth series, usually more precise than other production data because it is a direct measurement (*e.g.* compared to water and oil rates usually requiring indirect evaluations from the separator). We expect similar BHP measurements while operating in the same conditions, *e.g.* no maintenance stop or significant variation of production or injection rates. Considering $z_{i,s} \cong \overline{z_{\iota,N}}$ and $\sum_{s=1}^{N}(z_{i,s})^2 \cong N \cdot \overline{z_{\iota,N}}^2$, we have the mean of the random error smaller as N increases:

$$\sigma_{\overline{z_{\iota,N}}}^{ran} \cong \frac{prop_{z_i}^{ran}}{N} \cdot \sqrt{N \cdot \overline{z_{\iota,N}}^2} = \frac{prop_{z,i}^{ran} \cdot \overline{z_{\iota,N}} \cdot \sqrt{N}}{N} \tag{4.49}$$

For the systematic portion of the measurement error, we have:

$$E\left[e_{\overline{z_{\iota,N}}}^{sys}\right] = \frac{E\left[\sum_{s=1}^{N} e_{z_{i,s}}^{sys}\right]}{N} \tag{4.50}$$

Perfectly correlated errors in each time steps s imply that errors have similar size for all measurements i. The assumption of perfect correlation is taken for demonstration purposes and the its adoption for real applications should consider the observational errors verified on field's well surveillance programme and the measurement process in place, including (a) equipment calibration (random and systematic types); (b) chemical analysis for gas-oil-ratio; (c) apportionment of field production to well production and production testing, and (d) data manipulation.

$$e_{z_{i,s}}^{sys} = e_{z_i}^{sys} \tag{4.51}$$

And

$$\sigma_{z_{i,s}}^{sys} = \sigma_{z_i}^{sys} \tag{4.52}$$

$$Var\left[e_{\overline{z_{\iota,N}}}^{sys}\right] = Var\left[\frac{\sum_{s=1}^{N} e_{z_{i,s}}^{sys}}{N}\right] = \frac{1}{N^2} \cdot Var\left[\sum_{s=1}^{N} e_{st} \cdot \sigma_{z_{i,s}}^{sys}\right] \tag{4.53}$$

$$e_i^{sys} = e_{st} \cdot \sigma_{z_i}^{sys} \tag{4.54}$$

$$Var\left[e_{\overline{z_{\iota,N}}}^{sys}\right] = \frac{\left(\sum_{s=1}^{N} \sigma_{z_{i,s}}^{sys}\right)^2}{N^2} \cdot Var[e_{st}] \tag{4.55}$$

$$Var\left[e_{\overline{z_{\iota,N}}}^{sys}\right] = \frac{\left(\sum_{s=1}^{N} prop_{z_i}^{sys} \cdot z_{i,s}\right)^2}{N^2} \tag{4.56}$$

$$\sigma_{\overline{z_{\iota,N}}}^{sys} = \frac{prop_{z_i}^{sys} \cdot \sum_{s=1}^{N} z_{i,s}}{N} \tag{4.57}$$

$$\sigma_{\overline{z_{\iota,N}}}^{sys} = \overline{z_{\iota,N}} \cdot prop_{z_i}^{sys} \tag{4.58}$$

Summing random and systematic errors, we have:

$$\sigma_{\overline{z_{\iota,N}}}^{total} \cong \sqrt{\left(\frac{prop_{z,i}^{ran} \cdot \overline{z_{\iota,N}} \cdot \sqrt{N}}{N}\right)^2 + \left(\overline{z_{\iota,N}} \cdot prop_{z_i}^{sys}\right)^2} \tag{4.59}$$

$$\frac{\sigma_{\bar{z},N}^{total}}{\overline{z_{\iota,N}}} \cong \sqrt{\frac{\left(prop_{z_i}^{ran}\right)^2}{N} + \left(prop_{z_i}^{sys}\right)^2} \tag{4.60}$$

Because the number of measurements averaged is usually large (*e.g.* $N > 10$), we approximate:

$$\frac{\sigma_{\bar{z},N}^{total}}{\overline{z_{\iota,N}}} \cong prop_{z_i}^{sys} \tag{4.61}$$

This analysis allows us to emulate additional quantities of interest derived from measured data and incorporate the respective uncertainty in the implausibility measure, described in Formentin *et al.* (2020-a).

4.6 Summary of control variables – Part II

In Table 4.3, we summarise the control variables from the procedures presented in this chapter with a short discussion of the choices made and likely sensitivity of the results to the choices.

Table 4.3. Summary of control variables summarising choices made and likely sensitivity.

Control variables	Choices and discussion
Credible interval diagnostics D_{CI} for α	Credible Interval Diagnostics is an appropriate indicator to determine whether the uncertainty estimation for an emulator corresponds to its actual uncertainty. We would expect the observed value for D_{CI} to be close to α when the emulator uncertainty is appropriately estimated. An appropriate emulator is neither (a) under-confident, with too large uncertainty on expected outcomes, nor (b) overconfident, with too small uncertainty on expected outcomes. We defined $D_{CI} > 0.85$ for $\alpha = 0.95$, which indicates that valid emulators may have a slight overconfident response. A larger D_{CI} for the same α turns more selective the validation of emulators and a smaller D_{CI} would allow to cut-off regions of the space as implausible earlier (a larger number of valid emulators are constructed).
Decision boundary for binary emulators	Our objective is to choose an optimal decision boundary that maximises the PPV while keeping NPV as high as possible. We decided for a decision boundary fixed in 0.95 for the emulators based on logistic regression to express our aversion to exclude scenarios wrongly. Choosing a larger value leads to the construction of less informative emulators (smaller proportions of scenarios would be rule-out); smaller values indicate that we are less averse to the possibility of wrongly ruling out parts of the input space as implausible.
Negative Predictive Value (NPV) and Positive Predictive Value (PPV)	As validation criteria, binary quantities of interest need a minimum Positive Predictive Value of 0.90, e.g. at least 90% of the emulator's implausible scenarios correspond to the simulator's implausible scenarios. This unbalanced PPV threshold ($PPV_{thres} \gg 0.50$) represent our aversion to rule out wrongly regions of the search space. Simultaneously, we consider a smaller threshold for the Negative Predictive Value ($NPV_{thres} = 0.50$) since bad scenarios kept in the analysis can be ruled out at later waves. When two concurrent emulators meet the validation criteria, the one with higher NPV is selected.
Threshold for the information index $D_{info,thres}$	We define a threshold for the information index $D_{info,thres} = 0.05$, *e.g.* at least 5% of remaining search space needs to be ruled out by a valid emulator. The threshold for the information index considers: (a) cost to construct emulators; (b) cost of evaluation of scenarios via the emulator; (c) uncertainties in the process. A smaller $D_{info,thres}$ allows all emulators constructed and validated by the quality criteria (e.g. D_{CI}) to be considered in the analysis; more valid emulators have an impact of performing the implausibility analysis, to be justified by the D_{info}. A larger value of $D_{info,thres}$ would lead to a smaller number of emulators to be used in the implausibility analysis, possibly leading to a slower reduction of the uncertainty.

Control variables	Choices and discussion
Scaling choice for implausibility measure of binary quantities ω^*	We choose ω^* as a very large value (e.g. 150) for practical reasons: we would like to highlight the regions of the input space ruled out by binary emulators (see the purple region in Figure 3.9 of Formentin et al. (2020-a). This number does not have any further impact because regions considered implausible by binary emulators are imperatively ruled out of the input space.
Number of outputs to emulate	In STEP 8.10, we select a number of outputs (e.g. three in Figure 4.13). A design option stands for the number of outputs to be emulated. One should account for the possibility to miss informative outputs in the stage of the process, or for increasing the computation cost demanded to construct emulators and evaluate additional outputs. In the application of Formentin et al. (2020-a and 2020-b), this was a discretionary decision in each wave.

4.7 Conclusions

The ultimate goal of a model calibration process is to provide a background for well informed and efficient decisions. Finding the whole class of scenarios capable of representing the reservoir historical behaviour is essential to give a realistic evaluation of reservoir performance and consistent, unbiased predictions incorporating realistic levels of uncertainty, required for full asset management. Bayesian History Matching for Uncertainty Reduction (BHMUR) technique approaches this objective by incorporating all sources of uncertainties in the calibration process and combining evaluations from emulators and numerical simulators.

In this present work, we advanced the applicability of BHMUR approach by exploring and developing four topics, critical in the implementation of BHMUR for complex problems:

- **Emulators for two-class quantities of interest,** that appropriately model discontinuous behaviour identified for some quantities of interest typically found in reservoir simulations. We formulated a version of the implausibility measure for two-class emulators, integrating binary outcomes. In our application, we recognised data behaviours related to simulation targets and water breakthrough that required two-class emulators to construct appropriate statistical representations. This class of emulators allows one to rule out parts of the input space that lead to discontinuities, leaving regions with smoother relationships between inputs and outputs that can be subsequently analysed using standard emulators for continuous and smooth quantities;

- **Validation of emulators and selection of competitive emulators for a given quantity of interest** which is based on indicators defined to verify if an emulator is an appropriate estimator of the reservoir simulation model and to rank competitive emulators;

- **Methodology for selection of outputs,** allowing to efficiently select a combination of outputs to emulate, that is expected to be the most informative and efficient choice;

- **Estimation of observational errors in historical data,** where we developed a naïve, but consistent, form to compute the cumulative and averaged quantities of interest derived from measured time series (*e.g.* rate, pressure).

We demonstrated the application of methodologies proposed for these four topics with examples from Formentin *et al.* (2020-a). Each of these topics consists of significant contributions for uncertainty reduction procedures that cover a large number of outputs and require efficient use of reservoir simulations while applying emulators to represent simulation outputs.

Nomenclature – Article 3

B = indicator function of a transformation of simulation output
B^* = indicator function obtained from the emulator output
BHMUR = Bayesian History Matching for Uncertainty Reduction
BT = Breakthrough Time
CI = Credible Interval
D_{info} = information index
$\widetilde{D}_{info}$ = adapted information index
D_{CI} = Credible Interval Diagnostics
ϵ = model discrepancy, the difference between the real reservoir and the reservoir model
E = expectation operator
e = vector of observational errors
f = function of reservoir simulation model that computes a vector of quantities of interest
f^* = emulator function
g = known deterministic function
I = implausibility measure
I_M = maximal implausibility measure
$\tilde{I}$ = adapted implausibility measure
i_w = water injection rate
L_p = cumulative liquid production
m = number of reservoir scenarios simulated to increment the training set in each new wave

n = number of reservoir scenarios simulated to increment the test set in each new wave
N = number of time steps or measurements to be averaged t
n_A = number of active variables
n_{sce} = total number of scenarios
N_p = cumulative oil production
NPV = Negative Predictive Value
$phase$ = maximum phase of evaluation simulated in a given training set
p_{ibh} = Bottom hole pressure of injection wells
$\bar{p}_{ibh}$ = average bottom hole pressure of injection wells
p_{pbh} = Bottom hole pressure of production wells
$\bar{p}_{pbh}$ = average bottom hole pressure of production wells
PPV = Positive Predictive Value
q_l = liquid production rate
q_o = oil production rate
q_w = water production rate
RMSE_n = Normalised Root Mean Square Error
tol = tolerance applied to compute B
t = specific time from a time series given the subscript s
u = Gaussian process
Var = Variance
W_i = cumulative water injection
W_p = cumulative water production
w = nugget process
ω = implausibility cut-off
x = vector of input parameter values representing a reservoir scenario
x^* = most appropriate vector of uncertain attributes
y = vector of quantities from the real physical reservoir
z = vector of measurable quantities from the real reservoir
α = proportion covered by the credible interval
β = unknown scalar regression coefficients
φ = a phase of evaluation to select outputs to emulate $\varphi \in [1, phase]$
λ = proportion of m providing a training set sufficiently large to construct emulators
ω^* = factor to rescale the indicator function of two-class emulators

Subscripts

A = active variables
i = a measurable quantity of interest of the reservoir, $i \in [1, q]$
j = index corresponding to a regression term in the emulator equation
s = index corresponding to a time step of the historical period

Superscripts

k = a scenario of the reservoir model, $k \in [1, n_{sce}]$

Acknowledgement – Article 3

This research was carried out in association with the ongoing project registered as ANP 19708-7, “Fomento à Formação de Recursos Humanos em Gestão de Incertezas e Tomada de Decisão: Um Programa BG Fellowship" (UNICAMP/ Shell Brazil /ANP), co-funded by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brazil) through Science Without Borders Programme and Shell Brazil under the ANP R&D levy as “Compromisso de Investimentos com Pesquisa e Desenvolvimento”.

The authors thank ENERGI SIMULATION, UNISIM, DE-FEM-UNICAMP, CEPETRO and Department of Mathematical Sciences (Durham University) for supporting this work and CMG, Emerson and Schlumberger for software licenses. The authors also thank our colleagues who contributed to this paper: Alessandra Davolio and Forlan la Rosa Almeida (discussions about simulation targets), Daniel Carvalho (IT support), Gilson Moura (definition of well trajectory applied in the case study), Jonathan Owen (discussions about implausibility), Ricardo Moreno (discussions about emulation).

5 ARTICLE 4: ACCOUNTING FOR MODEL DISCREPANCY IN UNCERTAINTY ANALYSIS BY COMBINING NUMERICAL SIMULATION AND BAYESIAN EMULATION TECHNIQUES

Helena Nandi Formentin, Ian Vernon, Michael Goldstein, Camila Caiado, Guilherme Daniel Avansi, Denis José Schiozer

To be submitted

Contribution by the candidate Helena Nandi Formentin: Conceptualisation, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original + Edition, Visualization, Funding Acquisition. Michael Goldstein and Ian Vernon together with the PhD candidate developed the algorithm. The code in R was fully developed by Helena, with exception to the function to construct Gaussian emulators (inherited from a course by Michael Goldstein, Ian Vernon and Camila Caiado in Unicamp, 2015). CMG was selected as reservoir simulation software.

Abstract – Article 4

Model discrepancy specifies unavoidable differences between a physical system and its corresponding computer model. Incomplete information, simplifications and lack of knowledge about the physical state originate model discrepancy. Misevaluation of model discrepancy exposes decision-makers to overconfident and biased forecasting, a risky situation. We describe a methodology to account for one type of model discrepancy in the Bayesian History Matching for Uncertainty Reduction (BHMUR), an approach that combines reservoir simulation and emulation techniques to find all reservoir scenarios consistent with observed data and uncertainties in the problem. Our methodology is an alternative and more rigorous tool to account for the model discrepancy caused by errors in target data while performing uncertainty analysis. Target data used in historical period contain observational errors that propagate through the simulator, causing one type of model discrepancy. We follow a systematic procedure for uncertainty reduction previously presented by the authors, expanding the step dedicated to the model discrepancy. Our methodology: (1) obtains a training set by evaluating model discrepancy in multiple scenarios of the search space, an expensive simulation-based process; (2) characterises the model discrepancy across the entire search space via Bayesian emulators; and (3) integrates the model discrepancy in the BHMUR via bias and covariance structures. The methodology is demonstrated in a case study: 27 valid emulators for model discrepancy were constructed and integrated into the implausibility analysis and uncertainty reduction process. Two perspectives showed the impact of this type of model discrepancy. Firstly, neglecting model discrepancy resulted in all the search space being implausible –an indicator to review the problem characterisation and uncertainties; by contrast, when considering the model discrepancy, the non-implausible region consists of 8% of the search space. Secondly, we demonstrated the uncertainty reduction in the historical and forecasting periods. A key finding is that the error in target data results in a substantial model discrepancy over many other simulation outputs, being both time and location dependent. We advance the applicability of BHMUR by proposing a statistically consistent tool to account for one type of model discrepancy in the uncertainty quantification process. We showed that errors in target data cause model discrepancy with a complex structure. Appropriate consideration of model discrepancy is vital to (a) identify the whole class of solutions consistent with historical data and uncertainties in the problem, (b) appropriately represent the physical system; (c) avoid making decisions based on over-confident and biased information while enabling more reliable production forecast.

Keywords: Uncertainty Quantification, Model Discrepancy, Imperfect models, Bayesian History Matching for Uncertainty Reduction, Reservoir Simulation.

5.1 Introduction

Reservoir simulation models are complex computer models that incorporate physical laws, data available, and an interpretation of the sub-surface structure; it formalises our understanding of the real physical system. Reservoir models routinely support reserves estimation, reservoir behaviour analysis, field development optimisation and production forecasting, having a direct influence in the asset value, exploration and production phases. The irresolvable difference between the output of the model and the performance of the physical system is often termed as model discrepancy (Goldstein *et al.* 2013).

Schiozer *et al.* (2019) described a workflow for integrated decision analysis in the development and management of petroleum fields considering reservoir simulation, risk analysis, history matching, uncertainty reduction, representative models, and production strategy selection under uncertainties. The physical system in the sub-surface is unique and our lack of knowledge about it turns the physical state uncertain. By recognising and characterising their limitations, we can use these imperfect representations of the physical system to gain insights, forecast relevant information and make timely decisions with an appropriate degree of uncertainty.

Restricted, imperfect and incomplete information imposes challenges while building reservoir models, accounting for uncertainties and model discrepancy, which includes (Cosentino, 2001; Zee Ma *et al.*, 2011): (a) indirect information; (b) availability of small support volume (except seismic); (c) information available from various methods and techniques with different scales, *e.g.* porosity measured through core plugs, well logs or well testing; (d) accuracy and precision of measured data; and (e) uncertainties quantification in geologic interpretations, *e.g.* structural and stratigraphic models often built under a deterministic approach. Finally, propagation of uncertainties in complex, multi-stepped workflows based on siloed data is expensive and demanding in terms of time, computational resources and skills required. In practice, the difficulties of uncertainty propagation often lead to simplifications in the processes, possibly steaming an incomplete evaluation of uncertainties in forecasting.

A powerful technique to calibrate reservoir simulation models based on production data is the Bayesian History Matching approach (Craig *et al.* 1995; Vernon *et al.* 2018). The combination of simulation and emulation enables to identify the whole classes of simulations that are compatible with observed data and uncertainties specific to the problem. The set of

simulations identified as appropriate is applied for forecasting and risk analysis; this objective was reinforced by extending the name to Bayesian History Matching for Uncertainty Reduction – BHMUR (Ferreira *et al.* 2020, Formentin *et al.*, 2019-b). Ferreira *et al.* 2020 demonstrated a conceptual schema comparing BHMUR with optimisation approaches, showing that the evaluation of the entire input space via emulators is powerful to full uncertainty analysis.

Formentin *et al.* (2019-b) presented a systematic procedure for uncertainty reduction that combines simulation and emulation techniques under the BHMUR approach. Among the steps, it is critical to identify and characterise the uncertainties originated from diverse sources, including (a) observed data to represent the data from the physical system; and (b) computer model to describe the physical system which is the model discrepancy (Goldstein et al. 2013).

Model discrepancy accounts for approximations in the state of the physical system. Brynjarsdóttir & O'Hagan (2014) highlighted that models will virtually always be wrong in unknown ways and degrees with important consequences of ignoring the presence of model discrepancy:

- The precision of parameter estimates is likely to be over-estimated;
- Estimates of model fit may be overly optimistic;
- The degree of uncertainty in model forecasting may be under-estimated.

The interpretation and meaningfulness of parameters and their estimates become awkward (Brynjarsdóttir & O'Hagan, 2014): a parameter has clear meaning when a model is correct, but we do not know the real mean of a parameter when we acknowledge that a model is wrong.

Approaches to account for the model discrepancy in history matching and uncertainty reduction procedures gained space in the recent literature: Alfonzo and Oliver (2019), Evensen (2019), Rammay *et al.* (2019), Arnold *et al.* (2019), Wutzler (2018), Oliver and Alfonzo (2017), Brynjarsdóttir and O'Hagan (2014), Goldstein *et al.* (2013).

5.1.1 The problem

Identifying sources of model discrepancy and characterising them are challenges for fields that use computer models as predictive tools for physical systems, including reservoir

engineering and energy industry. Reservoir simulation usually applies some observed data in the historical period as boundary conditions for the simulations which are usually called simulation targets. Typical data used as target are liquid production rate or oil production rate (for production wells) and water injection rate (for injection wells).

Errors in simulation target can propagate for other quantities of interest through the computer model, representing a source of model discrepancy, as highlighted in Formentin et al. (2019-b) by using a hypothetical reality. By simulating the same scenario of a model (*e.g.* the hypothetical reality) with three conditions, we state the problem of interest of this paper.

In Figure 5.1, the brown line characterises data from the physical system (*e.g.* represents the reservoir in the subsurface). In practice, we simulate the reservoir scenario with operational conditions (*e.g.* maximal and minimal operation pressures and flow capacity of well), which we name as *forecasting mode*. A measurement system captures the physical behaviour; we noise the brown line to mimic the observational error and set as reference (or historical data in black dots). One of the time series of the reference data is set as simulation target, a set-up named a simulation in *history mode*; for example, the blue lines result from a simulation of the hypothetical reality with the oil production rate as target and primary boundary condition. When we evaluate this simulation, we observe that the error in the target data (oil production rate, on the left) propagates through other outputs of the simulator (bottom hole pressure, on the right panel), causing model discrepancy.

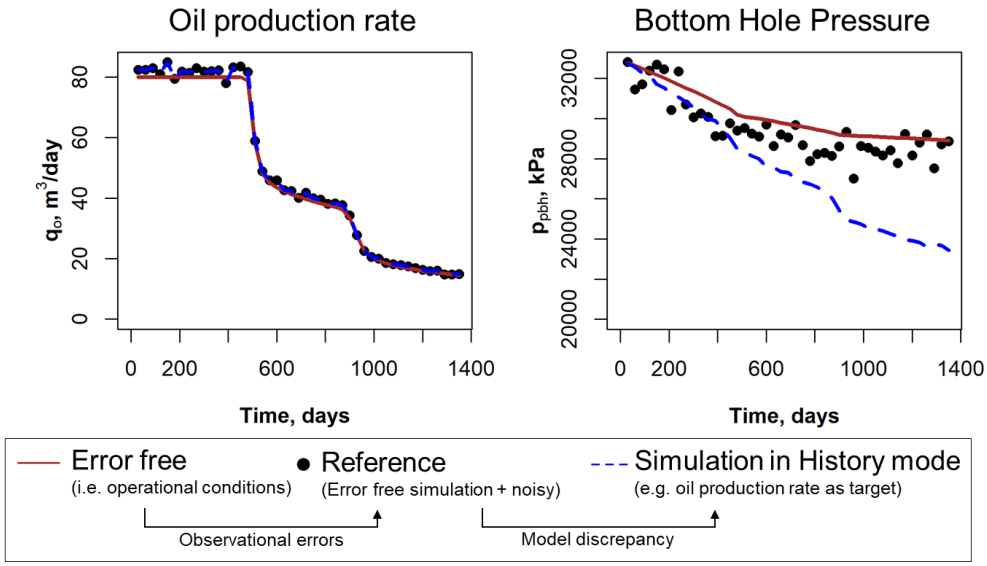


Figure 5.1: One scenario simulated under three different simulation conditions to describe the model discrepancy originated from error in target data: (a) oil production rate, (b) Bottom Hole Pressure. The brown line is the simulation with operational conditions, *e.g.* maximal and minimal operation pressures, flow capacity in wells; black dots represent the reference (or historical) data, which is the brown line added with noisy; the blue line results from the simulation in history mode, *e.g.* oil production rate set as target to the simulator.

The problem that we deal with in this paper is the model discrepancy originated by errors in target data (*i.e.* the difference between the black dots and the dashed-blue line in the panel on the right of Figure 5.1).

5.1.2 Objectives

We aim to contribute to the rapidly growing research area of uncertainty quantification, characterisation and integration of model discrepancy into the uncertainty analysis, by offering:

- A methodology to account for model discrepancy originated from errors in target data that can be integrated into the Bayesian History Matching for Uncertainty Reduction approach;
- The use of advanced emulation techniques to increase the speed of the evaluation of model discrepancy across the input space;
- The application of the methodology in a case study to demonstrate the impact of a specific source of model discrepancy in the uncertainty reduction process.

Our approach starts by presenting the statistical methodology, followed by the description of the application, results and conclusions.

5.2 Statistical methodology

A brief description of the standard form of Bayesian History Matching (BHM) is followed by the introduction of a systematic procedure for uncertainty reduction (Formentin *et al.* 2019-b). This sequence of 20 linked steps turn the practice of BHMUR more standardised across applications, and we highlight the steps expanded in the section 'Application of the Systematic Procedure for Uncertainty Reduction'.

5.2.1 Formulation of Bayesian History Matching

Craig *et al.* (1997) made a significant contribution to the uncertainty quantification field by stating the principles of Bayesian History Matching (extended to BHMUR). A vector of measurable quantities z from the physical system, *e.g.* water production rate or bottom hole pressure, results from the sum of the corresponding vector in the physical system y and observational errors e, giving $z = y + e$. Each element $i=1, \ldots, q$ of this vector has a corresponding quantity calculated by the computer model via an unknown function f of x, where

x is a vector of uncertain parameter values or scenario belonging to the input space (*e.g.* permeability, porosity parameters).

We account for imperfections in the computer model by incorporating the model discrepancy ϵ and defining $y = f(x^*) + \epsilon$: even the most appropriate scenario x^* presents outputs with a level of discrepancy (Vernon *et al.* 2010). Synthetic case studies with a Hypothetical Reality (HR) allow interpreting the HR as the most suitable scenario; nevertheless, for real applications to gain this insight is more challenging or even unaccomplishable. A bottleneck in statistical studies is the evaluation of scenarios via numerical models, which can be intensive in terms of time, computational and other resources. BHMUR combines simulation and emulation to enable the assessment of the whole input space affordably.

Emulators $f^*(x)$ are statistical models built on structured data (*i.e.* inputs and corresponding outputs) which is obtained from the simulation of a relatively small number of scenarios in the input space. Emulators are a powerful tool because they enable fast evaluations of sufficiently large numbers of scenarios, supporting a full uncertainty quantification (Craig *et al.* 1995; Craig *et al.* 1997; Vernon *et al.* 2010). For any combination of uncertain attributes, an emulator computes an expectation (*e.g.* mean) and uncertainty (*e.g.* variance) expressing its capability to mimic the simulator output. O'Hagan (2004) offers established arguments about emulators characteristics.

Figure 5.2 highlights the features of a statistical model for the output *f(x)* which is traditionally used as the basis for constructing an emulator of the same output. It is composed of: (a) a regression term to capture the global trend of active attributes x_A – a subset of inputs identified as the most influential for each quantity i; (b) a Gaussian process u, which adds flexibility to local behaviour; and (c) a nugget term which is a function of the scenario x and independent on u. The nugget acknowledges for simplifications that may lead to hidden effects in the emulator structure, such as the inclusion of only a sub-set of uncertain attributes in the two first terms of the emulator as active variables. The sum $v = u + w$ is useful for demonstration purposes later in this section.

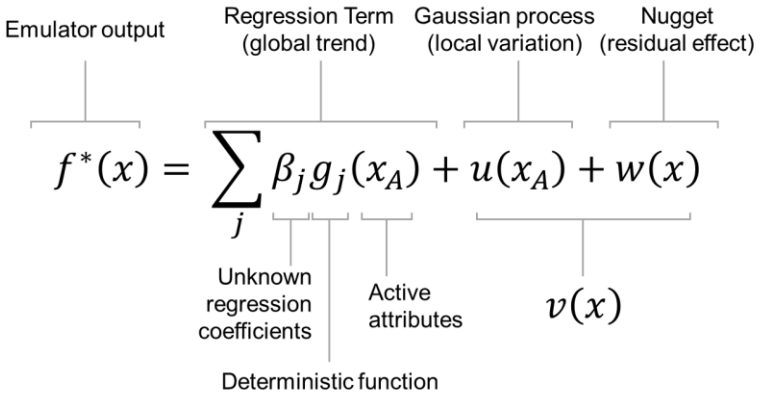


Figure 5.2: Features of a traditional form of an emulator, a sum of the global trend, local variation and residual effect.

A covariance matrix is symmetric and has the variance in the diagonal; the correlation matrix is the covariance matrix standardised by variances, resulting in the diagonal equal to one. Covariance structure refers to patterns in the covariance matrices.

A usual form for the Gaussian emulator has $E[u(x_A)]$ equals to a constant (or zero) and a base function built on an exponential covariance structure (Equation 5.1). This continuous function has two hyper-parameters to be defined: (1) the amplitude of the covariance function, σ_u^2, and (2) the correlation length θ, which scales the influence between points in the training set.

$$\mathrm{Cov}[u(x_A), u(x'_A)] = \sigma_u^2 \exp\left(\frac{-\|x_A - x'_A\|^2}{\theta^2}\right) \tag{5.1}$$

A nugget can be included in the process with the structure of Equation 5.2, where $\mathbf{1}(\cdot)$ is an indicator function, *i.e.* the covariance is equal to σ_w^2 if $x = x'$, and 0 otherwise.

$$\mathrm{Cov}[w(x_A), w(x'_A)] = \sigma_w^2\, \mathbf{1}(x = x') \tag{5.2}$$

Because the structure of u and w are taken as independent, we have $\sigma_v^2 = \sigma_u^2 + \sigma_w^2$ and define $\sigma_u^2 = (1 - \delta)\sigma_v^2$ to integrate in Equation 5.3:

$$\mathrm{Cov}[v(x_A), v(x'_A)] = \sigma_u^2 \exp\left(\frac{-\|x_A - x'_A\|^2}{\theta^2}\right) + \sigma_w^2\, \mathbf{1}(x = x') \tag{5.3}$$

Equation 5.4 summarises the covariance structure of both Gaussian process and nugget. Note that the nugget adds a fractional variance δ to the diagonal of the covariance matrix.

$$\mathrm{Cov}[v(x_A), v(x'_A)] = (1-\delta)\sigma_v^2 \exp\left(\frac{-\|x_A - x'_A\|^2}{\theta^2}\right) + \delta\,\sigma_v^2\,\mathbf{1}(x = x') \tag{5.4}$$

Advanced methods using cross-validation and maximum likelihood can be used to define the hyper-parameters σ_v^2 and θ (Bachoc, 2013; Bastos, 2010; Rasmussen and Willians. 2006). Alternatively, we can assess educated guesses such as (a) σ_v^2 equals to the variance of the outputs of simulator used as training set or the residual variance from the regression term, and (b) θ being an intermediate value between the maximal and minimal Euclidian distance between two scenarios – advised to scale the input internals in the same range, which was our approach in this article. More in-depth discussions and sensitivity analysis can be examined in Rasmussen and Williams (2006) and Bastos (2010).

As in Vernon *et al.* 2010, we follow the Bayes Linear approach (Goldstein and Wooff, 2007); the updating rules for expectations E and variances Var for a vector B, given a vector D, are in Equations 5.5 and 5.6, where $E_D[B]$ and $Var_D[B]$ are termed the adjusted mean and variance of B given D.

$$E_D[B] = E[B] + Cov(B,D)Var(D)^{-1}(D - E(D)) \tag{5.5}$$

$$Var_D[B] = Var[B] - Cov(B,D)Var(D)^{-1}(D - B) \tag{5.6}$$

Jackson (2018) - Chapter 2 provides a full derivation of emulator adjustments. We adopt $\sum_j \beta_j g_j(x_A) = \beta_0$; a constant and known value, implying in $E(\beta_j) = \beta_j$ and $Var(\beta_j) = 0$. Therefore:

- $E[B]$ is a vector in the dimension of new scenarios n_x and equals to β_0;
- $Cov(B,D)$ comes from the covariance matrix $\mathrm{Cov}[v(x_A), v(x'_A)]$ with dimensions n_x and n_l;
- $Var(D)$ is a squared n_l matrix;
- D is a vector with n_l with the quantities simulated; and
- $E(D)$ is a β_0-vector with n_l elements.

Because emulators are very fast, we can evaluate a large number n_x of new scenarios quickly and integrate them into an implausibility measure. The Mahalanobis Distance

is a standardised version of the Euclidian distance that centres individual points around the mean and scales on the standard deviation (Mardia *et al.* 1982; Morrison, 2005). It can be thought of as an appropriate statistical distance for use in sample spaces where there exist differential variances and correlations between variables (Krzanowski, 2000).

Craig *et al.* (1997) reread the Mahalanobis Distance as a Multi-Variate Implausibility (MVI or $\mathcal{I}$) measure in Equation 5.7. The MVI measure centres outputs from the emulator around the historical data defining $\mu_D(x) = E\big(f^*(x)\big) - z$, and scales on the sum $\Sigma_D(x) = \Sigma_{f^*}(x) + \Sigma_e(x) + \Sigma_\epsilon(x)$, which accounts for all uncertainties in the process. The covariance matrix Σ_D expresses the uncertainties:

- From the emulator to mimic the simulator Σ_{f^*};
- From the observed data to represent the physical system Σ_e and
- From the computer model to describe the physical system Σ_ϵ.

$$\mathcal{I}^2(x) = \mu_D(x)^T \Sigma_D(x)^{-1} \mu_D(x) \tag{5.7}$$

If the MVI is assumed to follow a chi-squared distribution with dimension of μ_D minus one degrees of freedom; we can then define a cut-off based on the critical value of chosen confidence level from a chi-squared distribution with the number of observations as degrees of freedom (Ferreira et. al. 2019).

Exact calculation of Σ_D requires the specification of the full covariance structure between all components (*e.g.* all outputs considered) for any vector of x values. One alternative is to compute univariate implausibility measure I (Equation 5.8), which simplifies by requiring only the specification of the respective variances – diagonal of the covariance matrix – for each of the three uncertainties in the process. Several possible formulations allow to combine univariate implausibility measure from multiple outputs; one frequent choice is the maximal implausibility measure $I_M(x) = \max_{i \in Q} I_i(x)$. When the univariate implausibility measure is expected to be a continuous unimodal distribution, a common choice is to define the implausibility measure based on the Pukelsheim's three-sigma rule: it has the powerful property of covering 95% within ±2.98 standard deviations from the mean (Pukelsheim, 1994; Vernon *et al.* 2010).

$$I_i^2(x) = \frac{\left[E\left(f_i^*(x)\right) - z_i\right]^2}{Var\left(f_i^*(x)\right) + Var(e_i) + Var(\epsilon_i)} \tag{5.8}$$

The adapted implausibility measure (Formentin *et al.*, 2019-b) has a similar formulation except for the use of simulator output instead of emulator expectation, consequently implying in $\Sigma_{f^*}(x) = 0$. The adapted implausibility measure was applied for the selection of outputs to emulate and to the evaluation of scenarios through the uncertainty reduction process. In this paper, we apply the concept of adapted implausibility measure because we decide to focus on one component of uncertainty, which is a source of model discrepancy ϵ.

5.2.2 Systematic Procedure for Uncertainty Reduction

Formentin *et al.* (2020-a) proposed a systematic procedure for uncertainty reduction that combines reservoir simulation and emulation techniques under the BHMUR approach (Figure 5.3).

The main features of this sequence of 20 linked steps are repeatability, flexibility and scalability, which turn the practice of the BHMUR more standardised and less manual across applications. Each step of this high-level structure can be planned to answer requirements specific for a study. Formentin *et al.* (2020-a) describes each of the steps of the workflow, and important definitions to retake here are: (a) waves are each iteration with the simulations of new scenarios; (b) phases increment the amount of historical information considered in the process.

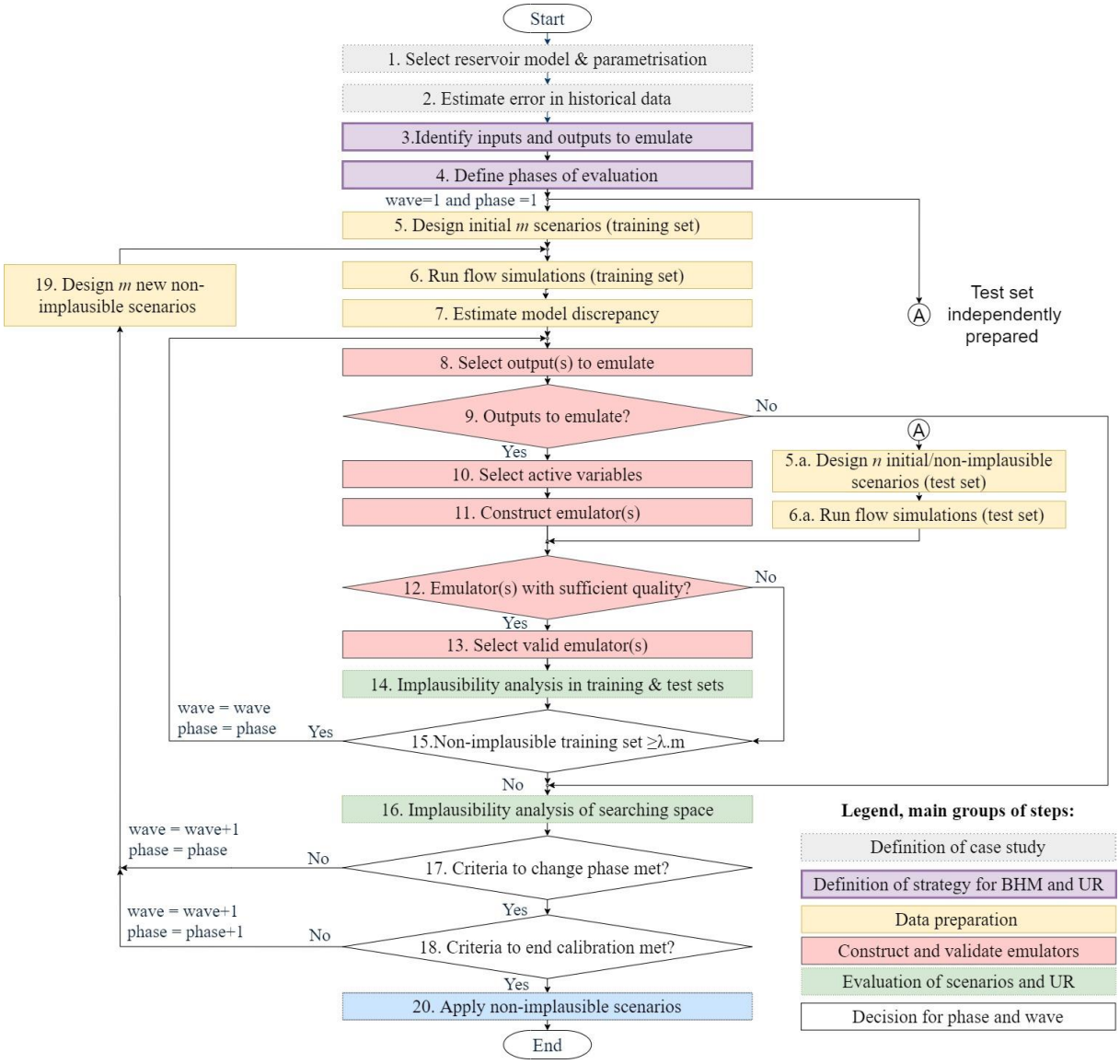


Figure 5.3: Systematic procedure for uncertainty reduction combining reservoir simulation and emulation techniques under the Bayesian History Matching for Uncertainty Reduction approach (Formentin *et al.*, 2020-a).; six groups represent the 20 steps; each step of this high-level structure can be planned to answer requirements specific for a study. This paper focuses on a methodology to account for one form of model discrepancy (*STEP 7*) and shows results mainly of *STEPS 16* and *20*.

In the next sub-sections, we focus on the expansion of *STEP 7* of the systematic procedure, where the model discrepancy is estimated. *STEP 16* is reviewed to accommodate an additional term to describe the model discrepancy in the implausibility measure. We recommend the consultation of the reference papers for additional details.

5.2.3 Estimation of Model Discrepancy caused by Errors in Target Data

Characterising the observational error e and model discrepancy ϵ accompanies the increasing interest in uncertainty quantification, to use computer models appropriately, gain

knowledge about the physical system and quantify uncertainties. Vernon *et al.* (2010), Kennedy and O'Hagan (2011), Mirams *et al.* (2016) and Goldstein (2017) proposed different classifications for the various sources of uncertainties in computer models with relevant discussions about the topic. One possible cause of model discrepancy in reservoir models is the use of target data, *e.g.* liquid or oil production rates, with inherent observational errors.

Specific regulations limit errors in measurement of oil production, especially for fiscal and custody transfer purposes, *e.g.* ANP and INMETRO (2013). For those applications, the measurement equipment is inspected and calibrated regularly. Frøysa *et al.* (2018) highlighted perspectives from diverse stakeholders involved: (a) authorities have requirements concerning maximum uncertainty to secure the national interests; (b) the partners selling the oil have interests to secure their incomes; (c) buyers of oil have interest in ensuring that they are not getting a lower amount of oil than what they pay for.

Fiscal measurement is usually followed by apportionment and reconciliation – the processes to allocate grouped production for individual wells. In many fields, those processes rely on proportions specified by test separator, which is a vessel used to separate and meter oil and gas from the wells. NFOGM (2005) presents several concepts of installations for well surveillance and monitoring with test separators and multiphase flow meters (MPFM). In several cases, observational errors, both random and correlated between wells, may be expected for measurements in production rates of flowing phases (*e.g.* oil, water, gas and liquid).

The simulator calculates certain outputs assuming that the target data is precise. In fact, due to the characteristic of the measurement process, it is not so. The observed target data is a measurement made with random and systematic errors on the true data values. These measurements are usually set as boundary conditions (targets) to the simulator while simulating in the historical period. Therefore, for a given choice of inputs, all of the simulator outputs will contain an error, *e.g.* different errors in the target data will lead to different results for all outputs. These errors must be accounted for in the implausibility calculations.

The inherent uncertainty of error in target data is propagated to other outputs as a cause of model discrepancy, *e.g.* if the liquid rate is the target, its observation error leads to uncertainty propagation to bottom hole pressure. Formentin et al. (2020-a) exemplified it: even if we had a reservoir model perfectly representing the physical system (*e.g.* hypothetical reality in the proposed case study), when a production target larger or smaller than the non-noised

production was set, the bottom hole pressure tends to be smaller or larger than the outcome of the hypothetical reality, respectively. We also exemplified in Figure 5.1.

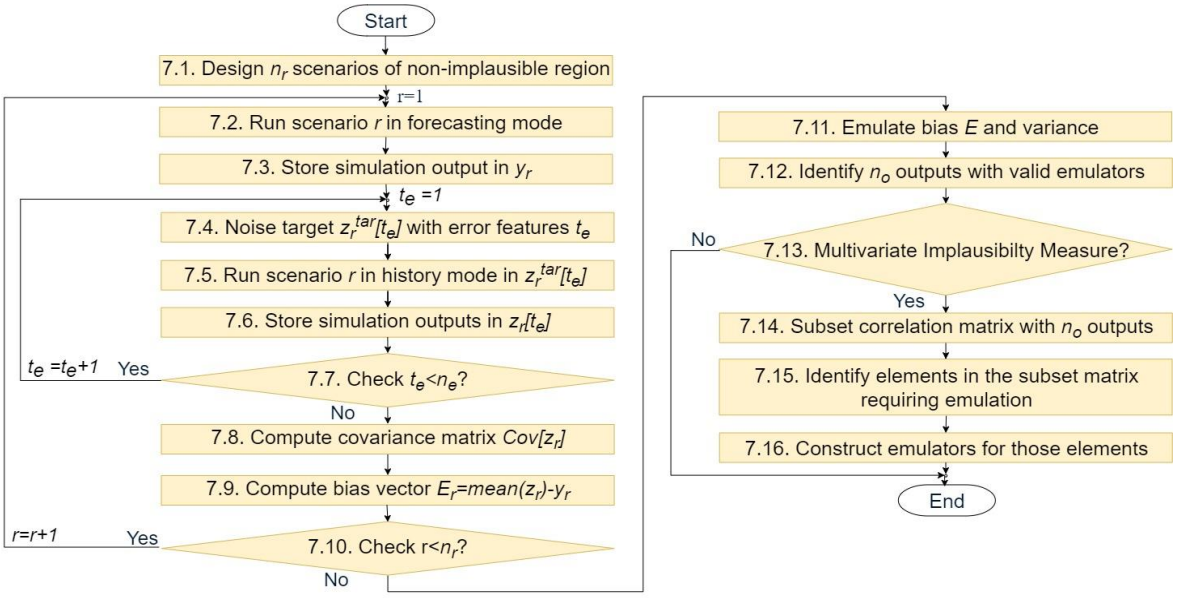


Figure 5.4: Methodology for estimating model discrepancy due to error in target data.

We propose a methodology to evaluate the model discrepancy originated by errors in target data (the methodology in Figure 5.4 with an illustration in Figure 5.5).

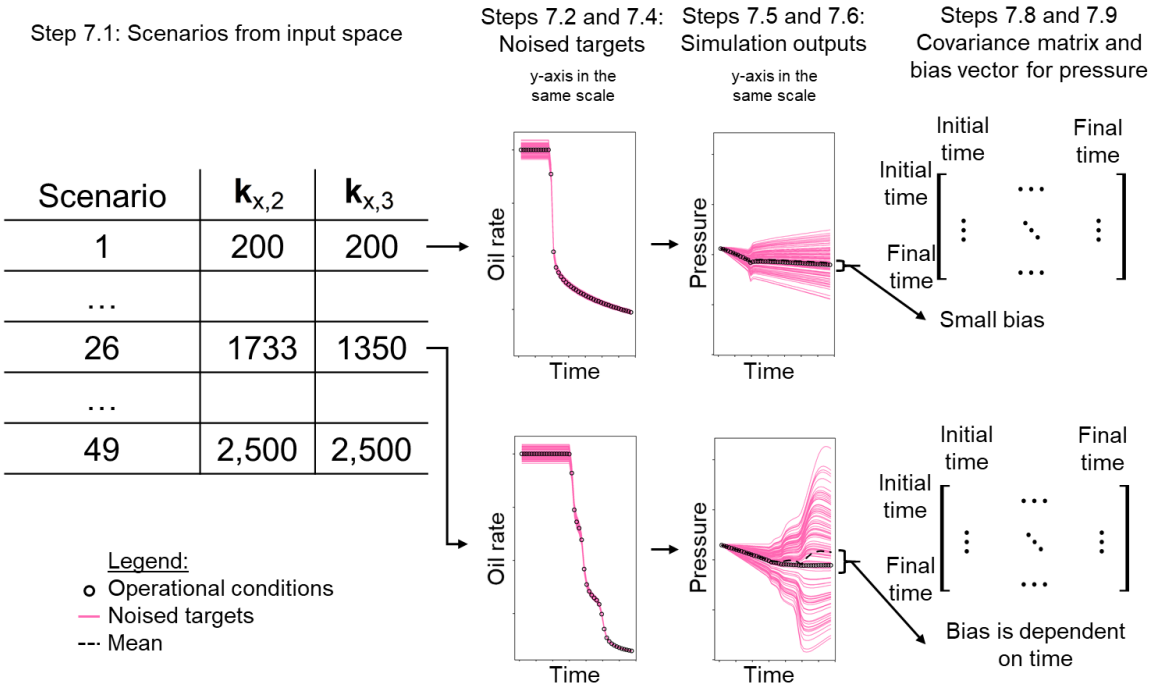


Figure 5.5: Illustration of the methodology for estimating the model discrepancy due to error in target data.

It integrates into *STEP 7* of the systematic procedure presented in Figure 5.3 (Formentin *et al.*, 2020-a) and can also be used independently, in other procedures based on

BHMUR. We follow with its integration in the BHMUR approach and application in a case study. In *STEP 7.1*, we design a training set with n_r scenarios in the input space; sampling methods include uniform, space-filling and Latin hypercube designs (Bastos, 2010; Fang *et al.* 2006; Johnson *et al.* 1990; McKay *et al.* 1979).

After initialising the variable r to iterate in the n_r scenarios, the scenario r is simulated in forecasting mode in *STEP 7.2* (*i.e.* having the operational conditions such as liquid capacity in the well and platform, and maximal and minimal bottom hole pressures as boundary conditions).

We store the outputs of this simulation (*e.g.* phases rate and pressure for each well and time step) in the vector y_r, *STEP 7.3*; it presents the physical behaviour in the unlikely case that all the information in the reservoir model were perfect stated. We follow with the initialisation of t_e, which is associated with the total number of target realisations n_e, where each realisation represents a new target that has a sample from the observational error. The vectors z_r^{tar} and y_r^{tar} are sub-sets of z and y corresponding to the target data (usually oil production rate or liquid production rate).

In *STEP 7.4*, the subset y_r^{tar} corresponding to the target data used in the historical period (*e.g.* liquid or oil production rate) is noised with the covariance structure of the observational error e, which results in the vector $z_r^{tar}[t_e]$; the covariance structure of e may contain random and correlated errors.

We use the target vector $z_r^{tar}[t_e]$ to simulate the scenario r in history mode (*i.e.* with target data as boundary condition), *STEP 7.5*.

In *STEP 7.6*, the vector of simulation outputs (*e.g.* phases rate and pressure for each well and time step) are stored in $z_r[t_e]$.

STEP 7.7 incurs the increment t_e until it equals the pre-stablished number of realisations of the observational error n_e. Note that each scenario r is associated with a matrix z_r of n_e rows and number of outputs as the number of columns.

In *STEP 7.8*, we compute the covariance matrix $Cov[z_r]$ for the scenario r: a square matrix with the number of outputs as the number of columns and rows. *STEP 7.9* determines the bias vector E_r which results from the mean of each column in z_r minus the corresponding value in y_r. Note that we may have random and systematic portions of error in the target data

which generates errors in the simulator outputs. These outputs are compared to the simulator outputs which would arise if the *true* target data were incorporated, the average difference being the bias, and the variance just being the variance of the outputs over variation in the observed target data. Note that this bias and variance depends on both the choice of input parameters at which we perform the calculation and also on the true values of the target data.

STEP 7.10 incurs the increment of r until it equals the pre-stablished number of scenarios n_r.

The covariance matrix and bias vector that characterise the model discrepancy can be calculated for each scenario. As Figure 5.5 illustrates, both the expectation and variance of the discrepancy caused by the error in the target data is dependent on the scenario and time. The calculations involved to define the uncertainty of each scenario tend to be computationally costly, what evidences the need of emulators able to mimic the behaviour of this kind of model discrepancy across the whole input space. The *STEPS* 7.1 to 7.10 provide a training set in the input space that enables the emulation of the model discrepancy across the entire region in *STEPS* 7.11 to 7.16.

In *STEP 7.11*, we emulate the relevant quantities of interest in the vector E and the corresponding values in the diagonal of the Cov. We develop elements for the emulation of these quantities in the next section, which complement the traditional formulation to construct emulations that were explained in the section "5.2.1 Formulation of Bayesian History Matching".

STEP 7.12 is fundamental as we determine a subset of outputs n_o under the validation criteria. Several validation methods are available: in Formentin *et al.* (2019-b), we applied validation based on training and test sets; in the section 5.2.3.1, we explain leave-one-out method, which is based on training set only and is appropriate when the evaluation of a test set may be excessively expense in terms of computational time.

Until this point, we characterised the model discrepancy term used in the univariate implausibility measure (Equation 5.8), which is the focus of this paper.

If the uncertainty quantification requires the use of multivariate implausibility measure (Equation 5.7 and *STEP 7.13*), the covariance matrix needs to be further explored and

we provide some leads on how to do it. The correlation matrix is the covariance matrix standardised and having diagonal equals to one.

In *STEP 7.14*, we subset the correlation matrix with the quantities of interest represented in n_o. The diagonals of the n_r correlation matrices are one, and our attention is now in the off-diagonal elements – try to identify structures in the data that limit the need to emulate.

In *STEP 7.15*, we suggest identifying the off-diagonal elements of the correlation matrix that have variability across the n_r training set, and, therefore, need to be emulated in *STEP 7.16*. Off-diagonal elements that do not vary across the input space can be estimated by the multiplication between the appropriate variances (estimated in *STEP 7.11* via emulation) and correlation value.

The high computational cost involved in the estimation of the model discrepancy of each scenario of the input space justifies the application of emulation techniques in *STEP 7.11* (and 7.16 in case of Multivariate Implausibility Measure). Therefore, we proposed an efficient methodology to account for model discrepancy originated from errors in target data in the BHMUR approach. In the next sub-sections, we an analysis specific to construct emulators considering the training set obtained from the methodology of Figure 5.4.

5.2.3.1 Estimating the nugget to emulate model discrepancy originated from errors in the target

Figure 5.2 and Equations 5.1 to 5.6 described a traditional formulation to construct emulators. Our objective is to emulate the bias and variance caused by noise in target data. The nugget term can be estimated with an educated guess, expert judgment, statistical model, or a combination of these approaches. In this section, we propose a statistical model to estimate the nugget defined by δ in Equation 5.4. For example, if we define the target error through a normal distribution and consider that this pattern propagates to other outputs such as bottom hole pressure. We could also assume that the population is independent and identically distributed, and normally distributed $N(\mu, \sigma^2)$. The sample variance for a given scenario r is given by ${s_r}^2 = (n_e - 1)^{-1} \sum_{e=1}^{n_e} (z_r^{(e)} - \bar{z_r})^2$, where $\bar{z}_r = {n_e}^{-1} \sum_{e=1}^{n_e} z_r^{(e)}$, and z_r is the simulated output for a given scenario following Figure 5.4.

The distribution of the sample variance ${s_r}^2 \sim \frac{{\sigma_r}^2}{(n_e-1)} \chi^2_{n_e-1}$, a chi-squared distribution with $n_e - 1$ degrees of freedom and $Var({s_r}^2)$ can be written as:

$$Var({s_r}^2) = \frac{{\sigma_r}^4}{(n_e - 1)^2} Var\left(\chi^2_{n_e - 1}\right) = \frac{2{\sigma_r}^4}{n_e - 1} \tag{5.9}$$

And we approximate to define the quantity δ:

$$\sigma_w^2 = \delta_{Var}\, \sigma_v^2 \sim \frac{2{s_r}^4}{n_e - 1} \tag{5.10}$$

The bias is computed through the mean, and its nugget can be approximated through:

$$Var(\bar{Z}_r) = \delta_{bias}\, \sigma_v^2 = \frac{\sigma_r^2}{n_e} \sim \frac{{s_r}^2}{n_e} \tag{5.11}$$

With Equations 5.10 and 5.11, we can estimate the magnitude of δ_{Var} and δ_{bias} to apply for the construction of emulators for the quantities that characterise the model discrepancy examined, e.g. the quantities $\delta\, \sigma_v^2$ in Equation 5.4 for emulators constructed to represent the variance and bias.

5.2.3.2 Emulator Diagnostics

In order to identify over-confident (and undesired) emulators, the diagnostic is a critical step in the BHMUR approach. Leave-one-out (Bastos, 2010; Rougier *et al.*, 2009) is a cross-validation diagnostic which consists on (a) the removal of one scenario of the training set; (b) the construction of the emulator with $n_r - 1$ scenarios; (c) prediction of the removed scenario; (d) computation of error standardised by the simulated value and standard deviation from the emulator; and (e) analyse the standardised errors, which is assumed as a continuous unimodal distribution and can be interpreted through the Three-Sigma Rule (Pukelsheim, 1994). We use leave-one-out diagnostics in our application and show some diagnostics plot to illustrate in the results section.

5.2.4 Accounting for Model Discrepancy in the BHMUR approach

We review the implausibility measure (Equation 5.8) to integrate the bias from the model discrepancy calculation as in Equation 5.12:

$$I_i^2(x) = \frac{\left[E\left(f_i^*(x)\right) + bias - z_i\right]^2}{Var\left(f_i^*(x)\right) + Var(e_i) + Var(\epsilon_i)} \tag{5.12}$$

We have a corresponding Adapted Implausibility Measure in Equation 5.13:

$$\tilde{I}_i^2(x) = \frac{[f_i(x) + bias - z_i]^2}{Var(e_i) + Var(\epsilon_i)} \tag{5.13}$$

For multivariate implausibility measure of Equation 5.7, we have $\mu_D(x) = E\big(f^*(x)\big) + bias - z$.

5.3 Application of the Systematic Procedure for Uncertainty Reduction

In this section, we describe the application of the systematic procedure for uncertainty reduction in Figure 5.3 in terms of groups of steps, highlighting the application of the methodology for accounting the model discrepancy caused by errors in target data. We emphasise that we apply emulation techniques for the computation of model discrepancy only. The conventional use of emulation for outputs themselves (*e.g.* bottom hole pressure) is deliberately avoided in this application to maintain the focus on the emulation of the model discrepancy.

5.3.1 Definition of Case Study and Strategy of BHMUR

The case study (*STEP 1*) is a three-layer cross-sectional vertical model (x and z directions) with one injector in one extremity and one producer in another (Maschio and Schiozer, 2018). Figure 5.6 illustrates that (a) the permeability in the *x*-direction is set as equal for all blocks of the same layer; (b) the vector defining the reference scenario is $\overrightarrow{k_x^*}$=(1,000; 600; 2,000) mD; and (c) the permeabilities of the two inferior layers are considered independent and uncertain in the interval from 200 to 2,500 mD. The error in the historical data for phase rates (oil and water) and bottom hole pressure is characterised by 2% of random and 2% of systematic errors (*STEP 2*).

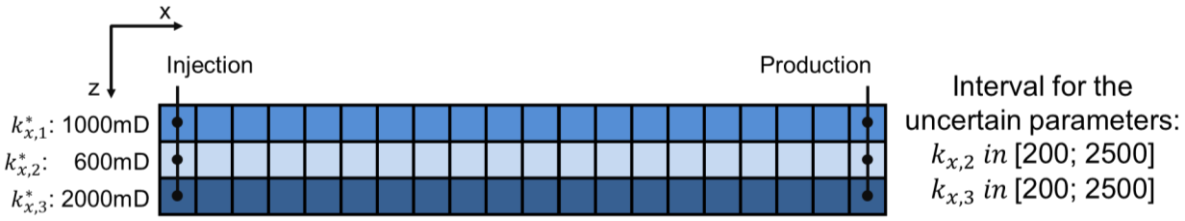


Figure 5.6: Permeability of the reference model (adapted from Maschio and Schiozer, 2018).

Historical period is set from 0 to 1,350 days and follows with forecasting until 2,700 days, and only one phase of evaluation is established. An additional injection well is placed after the historical period as in Maschio and Schiozer (2018). Figure 5.7.a to 5.7.d shows water saturation of hypothetical reality at four different times. To have historical data, we noised the

quantities of interest with one standard deviation from the simulation of the hypothetical reality, as shown in Figure 5.1. Our focus is on the emulation of the terms related to the model discrepancy, not the simulation outputs themselves; therefore, *STEPS 3* and *4* of the systematic procedure are applicable in this paper.

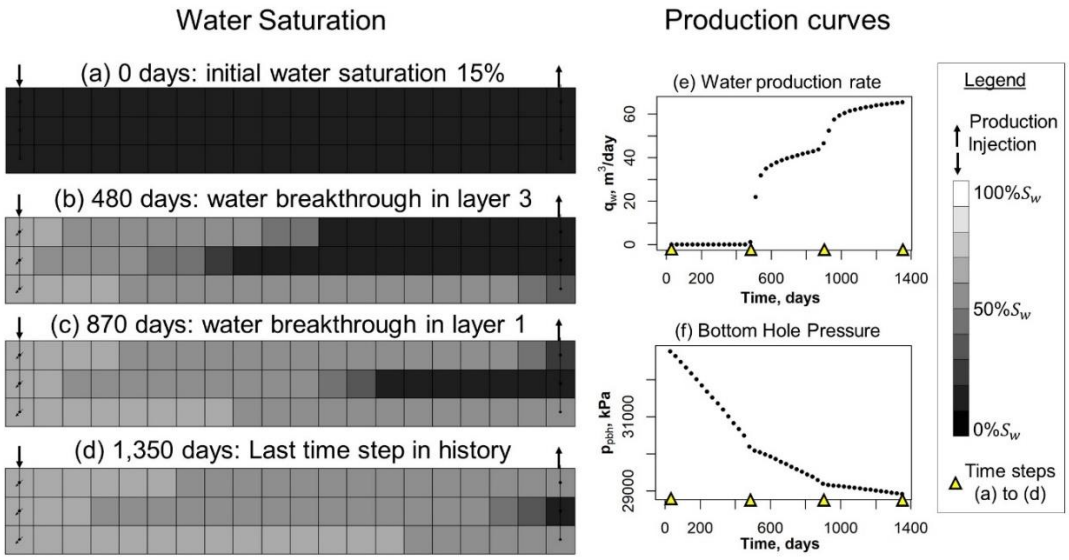


Figure 5.7: Hypothetical Reality, water saturation in 0, 480, 870 and 1350 days; corresponding production plots for (e) water production rate and (f) bottom hole pressure, marking the time steps (a) to (d).

The initial water saturation is 15% of irreducible water, the remaining proportion of the porous volume is saturated with oil; 480 days after water injection starts, layer 3 is the first to have water breakthrough. Figures 5.7.e. and 5.7.f. show the time which water production starts with a steep positive trend in water production rate and an abrupt change in the pressure gradient. A similar effect is observed when water production starts in layer 1. At the final time step of the historical period (Figure 5.7.d), water breakthrough still did not reach layer 2.

Although simple in its definition, this case study with two uncertain attributes enables to demonstrate the proposed methodology. The fast speed to run simulations is valuable in this experimental research and evidences potential impacts in the uncertainty reduction process. Variations in permeability offer non-linear relationships between inputs and outputs of the simulator and Maschio and Schiozer (2018) observed multiple modes in the posterior distributions calibrated scenarios, *i.e.* unconnected regions in the input space provide scenarios that are coherent with observed data and uncertainties in the process.

5.3.2 Data Preparation

We explore a source of model discrepancy in the 2-dimensional input space, which even for a fast simulation model can take a significant time to sample the form of the error structure of the target data. Because our experiment is low dimensional, we opt for sampling a grid of scenarios equally spaced.

Our focus is on the emulation of the terms related to the model discrepancy, that means, we are not focusing on the emulation of outputs themselves. Therefore, we designed a large set of simulations (Evaluation set in Figure 5.8) to evaluate the proposed methodology in *STEPS 5* and *6* of the systematic procedure (Figure 5.3). The evaluation set forms a grid composed of 50 equally spaced values for each uncertain attribute, totalling 2,500 scenarios and scenarios simulated with the noised target. The reference scenario defines a Hypothetical Reality (HR) that allows generating synthetic historical data.

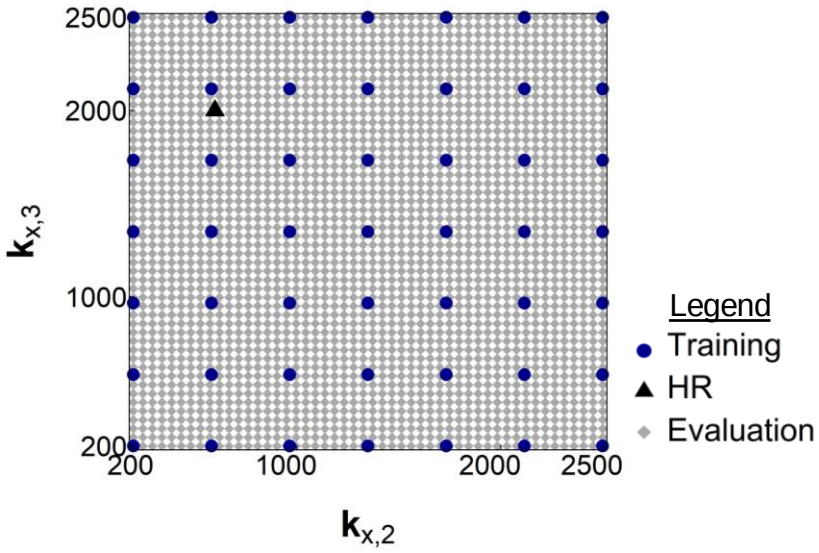


Figure 5.8: Set of data used in the analysis of this paper, where $k_{x,2}$ and $k_{x,3}$ are permeability of layers 2 and 3, respectively: the training set with 49 scenarios; Hypothetical Reality (HR) and evaluation set with 2,500 scenarios.

5.3.2.1 Estimation of model discrepancy originated from error in target data

The training set designed for *STEP 7.1* is a grid with seven equally spaced values, and we initialise r to be incremented until $n_r = 49$ scenarios. For each scenario, we follow the methodology proposed. We set forecasting mode (*STEP 7.2*) by defining as operational conditions the following:

- Maximal liquid rate in wells: 80 m^3/day;
- Maximal bottom hole pressure (for injection well): 45,000 kPa;

- Minimal bottom hole pressure (for production well): 15,000 kPa;

We highlight that simulation in the forecasting period do not reach the minimal bottom hole pressure (for example, Figure 5.1) and the primary operational condition restricting the system is the maximal liquid rate. The simulation outputs are stored as a vector y_r (*STEP 7.3*), and we select the oil production rate in each time step until 1350 days as the subset y_r^{tar}.

We initialise t_e, predefined to iterate until $n_e = 100$. For each t_e, the oil production rate in y_r^{tar} is noised with a systematic error sampled from a normal distribution having 2% as standard deviation and mean zero, which is added to the vector y_r^{tar} resulting in the z_r^{tar} (*STEP 7.4*). The multiplicative form of error was defined for demonstration purposes. For real applications, more complex forms with the combination of random and systematic errors can be considered. We do not noise water injection rate for this case study. In *STEP 7.5*, the noised oil production rate is set as the boundary condition for the simulation, which outputs are stored in the vector z_r (*STEP 7.6*). *STEP 7.7* ensures that the loop runs until $n_e = 100$.

For each scenario r, the covariance matrix and bias vector is computed (*STEPS 7.8* and *7.9*). We construct emulators for the bias and variance (*STEP 7.11*) with the elements described in Figure 5.2. Via leave-one-out diagnostics, we select the n_o outputs with valid emulators (*STEP* 7.12). As the focus of our analysis is on univariate implausibility measure, *STEPS 7.13* to *7.16* are suggestive about how to proceed. A total of 4,900 simulations were run to the training set of the model discrepancy, a demanding computational power that was affordable in our simple experiment. For more complex models, we highlight the need to estimate the computational cost in advance.

5.3.3 Construct and Validate Emulators

STEPS 8 to *13* of the systematic procedure (Figure 5.3) are dedicated to the emulation of the simulation outputs, which is not the focus of this paper. Instead, we use the results of the evaluation set (Figure 5.8) to proceed with the uncertainty analysis.

5.3.4 Evaluation of scenarios and UR

We performed two experiments to compare scenarios resulting from the BHMUR approach, and both are composed by one single wave: Experiment 1 does not account for the model discrepancy; Experiment 2 takes in consideration model discrepancy via the proposed

procedure. We apply Equation 5.13 to account for the Adapted Implausibility Measure in the implausibility analysis steps.

We compare these experiments in terms of uncertainty reduction in the input and output spaces and analyse forecasting to demonstrate the possible impact of neglecting model discrepancy in the BHMUR process.

We opted to demonstrate our evaluation based on results for water production rate and bottom hole pressure in different time steps. The univariate implausibility measure of outputs that have valid emulators were combined though maximal implausibility and used to compare the results of experiments 1 and 2. We apply the results of the Uncertainty Reduction process (*STEP 20*) by analysing forecasting of the initial search space with the non-implausible scenarios from these experiments.

5.3.5 Decision for phase and wave

This group of steps is not applicable: the simplicity of the case study does not justify the use of phases of evaluation; our objective is accomplished with the performance of one single wave.

5.4 Results and Discussions

In this section, we explore results for (1) emulation of variance and bias of model discrepancy with corresponding emulators diagnostics; (2) comparison of implausibility analysis between experiments 1 and 2; and (3) uncertainty reduction of reservoir behaviour in the historical and forecasting period. We discuss essential features resulting from the application of the BHMUR approach to the case study, comparing Experiment 1 (neglecting model discrepancy) and Experiment 2 (accounting for model discrepancy).

5.4.1 Emulation of variance and bias for the model discrepancy

To be applicable, both emulators of bias and variance of the model discrepancy need to satisfy diagnostics criteria. Out of a total of 61 evaluated outputs (water rate and pressure in different time steps), we constructed valid emulators for 27 quantities: 22 for bottom hole pressure and five for water production rate. We select a series of plots for the (a) expectation and (b) standard deviation of emulators of the model discrepancy, with their corresponding (c) diagnostics plots for leave-one-out. An advantageous characteristic of

BHMUR is that we do not need to analyse every possible output to complete an uncertainty analysis because a subset of outputs enables us to remove parts of the input space carefully.

We address the question of why model discrepancy changes through the input space, *i.e.* different scenarios have different discrepancy characteristics. Figure 5.9 presents saturation images for three scenarios at 990 days, the water breakthrough time for Scenario 1. The analysis is complementary to Figure 5.5, where scenarios 1 and 26 can be compared. The discrepancy demonstrated in the application is also dependent on time, as can be checked in Figure 5.5. This helps to understand some insights from the emulator expectations – plots (a) in Figures 5.9, 5.11 to 5.12– that indicate that the discrepancy changes depending on the position that a scenario takes in the input space, the output and the time-step considered.

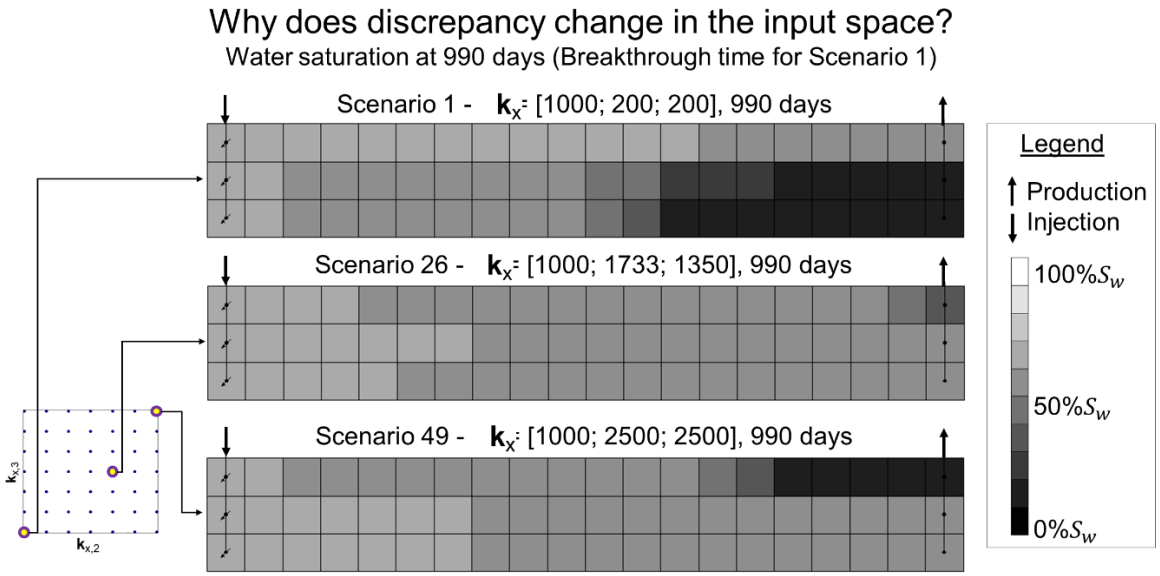


Figure 5.9: Water saturation at 990 days from different scenarios demonstrating physical reasons why the model discrepancy is dependent on the scenario evaluated.

Scenario 1 has the lowest permeability for layers 2 and 3, and the water breakthrough happens in layer 1. At 990 days, Scenario 1 still has a relatively high volume of oil to be drained compared with Scenarios 26 and 49. A variation in the target data with the structure of its observational error would require a less dramatic change in the bottom hole pressure in Scenario 1 than in the other two. Simulations in Figure 5.5 corroborate this intuition. Note that Scenario 26 presents high water saturation in all blocks of the reservoir model; recovering the residual oil in the pores to meet a higher target in oil production demands an increase in the water injection rate which leads to higher pressure.

In Figures 5.10, 5.12 and 5.13, we show results for several emulators. The vertical sequence of three plots per output emulated presents in (a) the emulator expectation for the

output calculated from Equation 5.7; (b) the emulator standard deviation for the output from Equation 5.8; (c) the outcome from the leave-one-out diagnostic of the emulator, counting how many samples are out of the interval [-3, +3] which was defined by considering the 3-sigma rule (Pukelsheim, 1994). We considered that a maximal of 2 out of the 49 samples could be out of this interval to make an emulator valid (~4%).

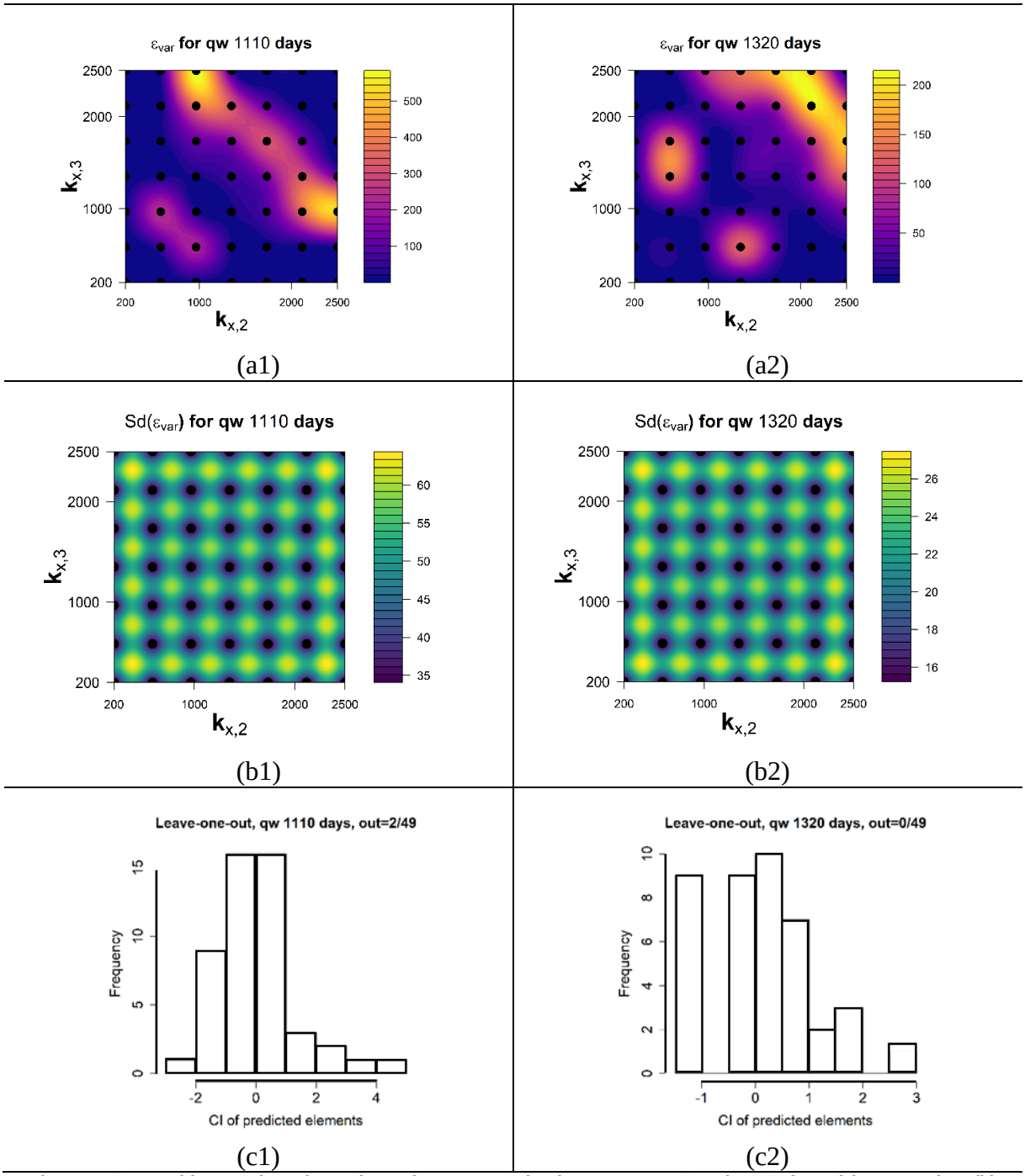


Figure 5.10: Resulting emulator for variance for water production rate at 1,110 and 1,320 days: (a) expectation, (b) standard deviation, (c) corresponding diagnostics plots.

Figure 5.10 shows emulators built for the variance of the model discrepancy – $\mathrm{Var}(\epsilon)$ in Equations 5.12 and 5.13 – for the water production rate at 1,110 and 1,320 days, respectively. The expectation (Figure 5.10.a1 and 5.10.a2) follows a symmetric pattern towards the secondary diagonal.

In fact, a quasi-symmetric behaviour is explained by similar effects that a lower permeability has being in the second or third layers, not perfectly symmetric due to gravity, *i.e.* water density is higher than oil density, leading to a non-uniform sweep even if the three layers had the same permeability. Figure 5.11 shows this effect: scenarios 19 and 31 are symmetric in the input space. At the water breakthrough time, the layers with higher permeability (layers 2 and 3 respectively) have a similar distribution of water saturation, except that gravity forces lead to slightly higher water saturation in the production well: 38% for layer 2 of scenario 19 and 40% for layer 3 of scenario 31.

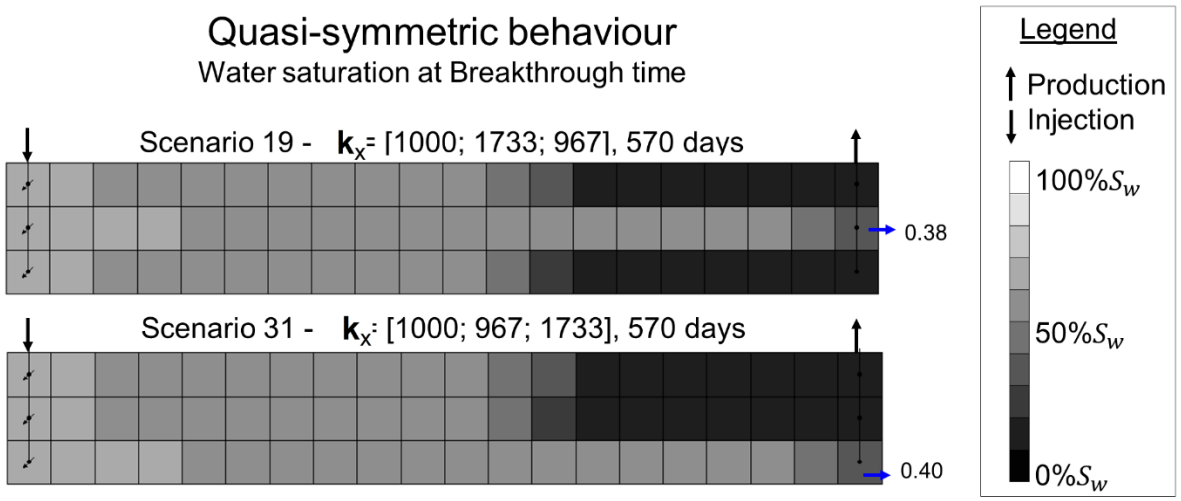


Figure 5.11: Water saturation at 570 days from symmetric scenarios demonstrating the quasi-symmetric behaviour expected in the case study.

The standard deviation of the variance in Figure 5.10.b1 and 5.10.b2 shows that the emulator uncertainty near the measured points (training set) is relatively low. The diagnostics plots in Figure 5.10.c1 and 5.10.c2 indicates only 2 and 0 out of 49 scenarios of the training set have the standardised error out that are out of the interval [-3, +3], which leads us to accept the emulators as valid.

Figure 5.12 shows emulators built for the bias of the model discrepancy – $bias$ in Equations 5.12 and 5.13 – for the water production rate at 1,110 and 1,320 days. The expectation (Figures 5.12.a1 and 5.12.a2) follows a symmetric pattern towards the secondary diagonal and has a similar pattern as the variance of the model discrepancy in Figures 5.10.a. There is a

difference in the scale: the bias tends to be much smaller than the variance of the model discrepancy: the maximum absolute value is around 8 and 6 for the expectation of the bias, while it is 600 and 200 for the expectation of the variance. The difference in the scale is also notable in the standard deviation of the emulator: around 2 for the bias and 65 and 28 for the variance. Finally, most of the standardised errors in Figures 5.12.c1 and 5.12.c2 are concentrated in the interval [-1, +1].

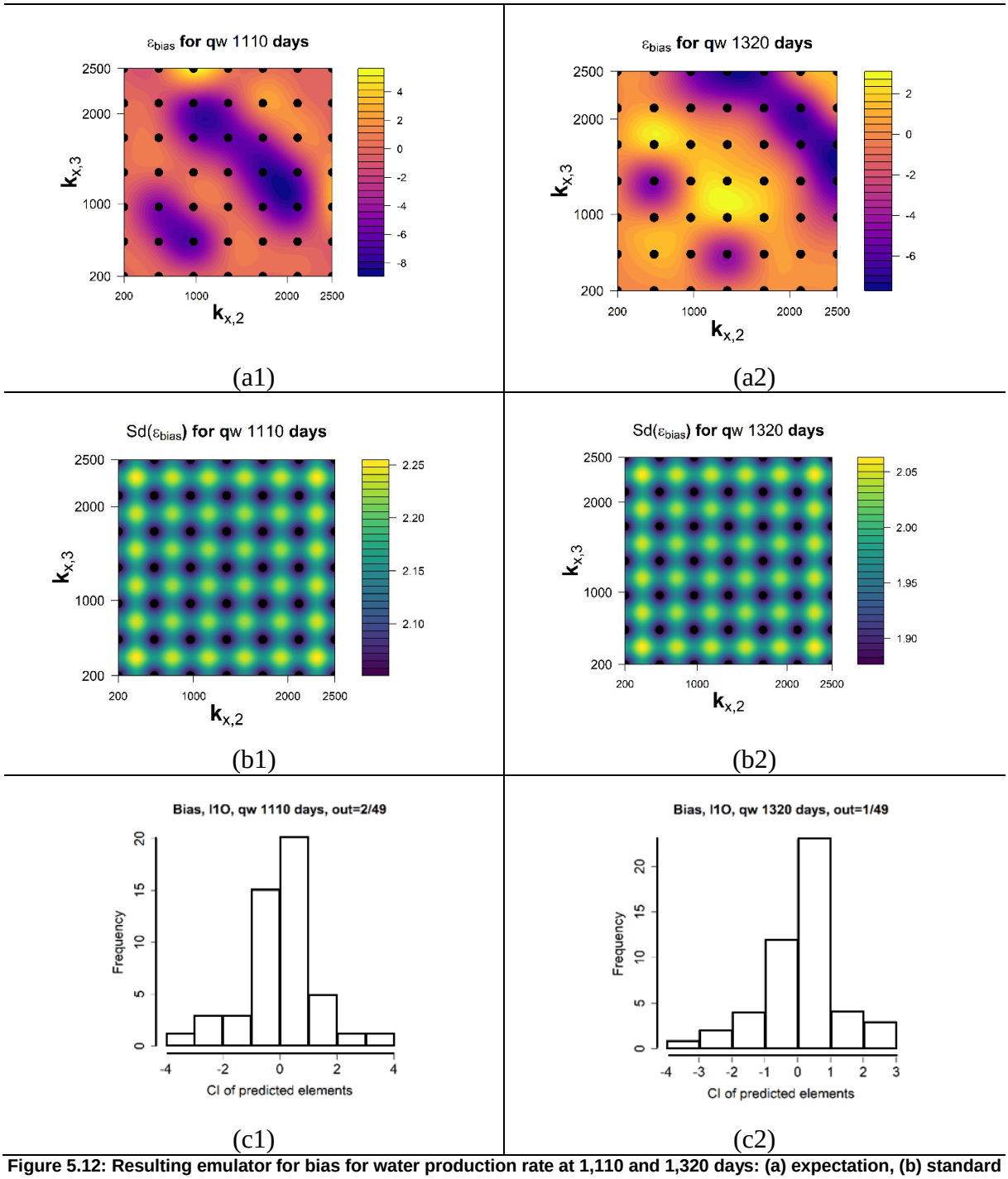


Figure 5.12: Resulting emulator for bias for water production rate at 1,110 and 1,320 days: (a) expectation, (b) standard deviation, (c) corresponding diagnostics plots.

We plot the results of emulators for bottom hole pressure at 630 and 1,350 days in Figure 5.13. In the expectation of the variance of the model discrepancy, we note the similar symmetry in the second diagonal and dependence on time as in the emulators for water production rate. Comparing bottom hole pressure at 630 and 1,350 days, the range for the expectation of the variance increases significantly, coherent with the plots in Figure 5.5.

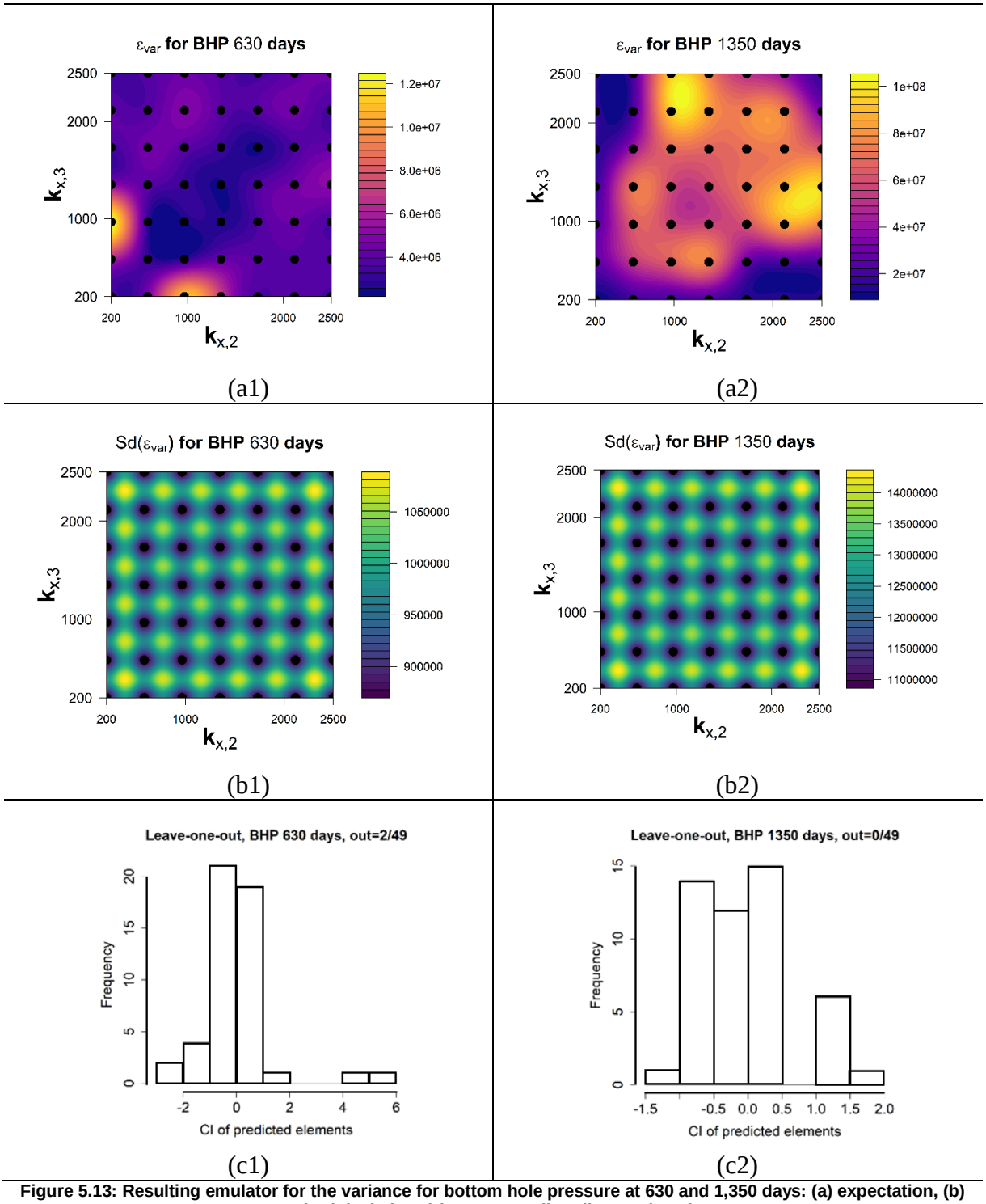


Figure 5.13: Resulting emulator for the variance for bottom hole pressure at 630 and 1,350 days: (a) expectation, (b) standard deviation, (c) corresponding diagnostics plots.

The structure of the standard deviations in Figures 5.13.b1 and 5.13.b2 have a similar interpretation as in Figures 5.10.b and 5.12.b, due to the use of the same correlation length. Finally, the diagnostics of leave-one-out in Figure 5.13.c1 has only two points larger than 3, which are the scenarios close to permeability [1,000; 200] and [200; 1,000] with very high expectation of variance.

5.4.2 Implausibility analysis and impact of model discrepancy

We computed the adapted implausibility measure of Equation 5.13 for each of the 27 valid emulators using the simulations of the evaluation set (Figure 5.8). We combine these results through the maximal implausibility, which are presented in Figure 5.14: the plot (a) presents the implausibility measure when we do not account for the model discrepancy and (b) when we account for it. The scales of the implausibility plots are selected to make green non-implausible scenarios and a scale from yellow to red the implausible regions.

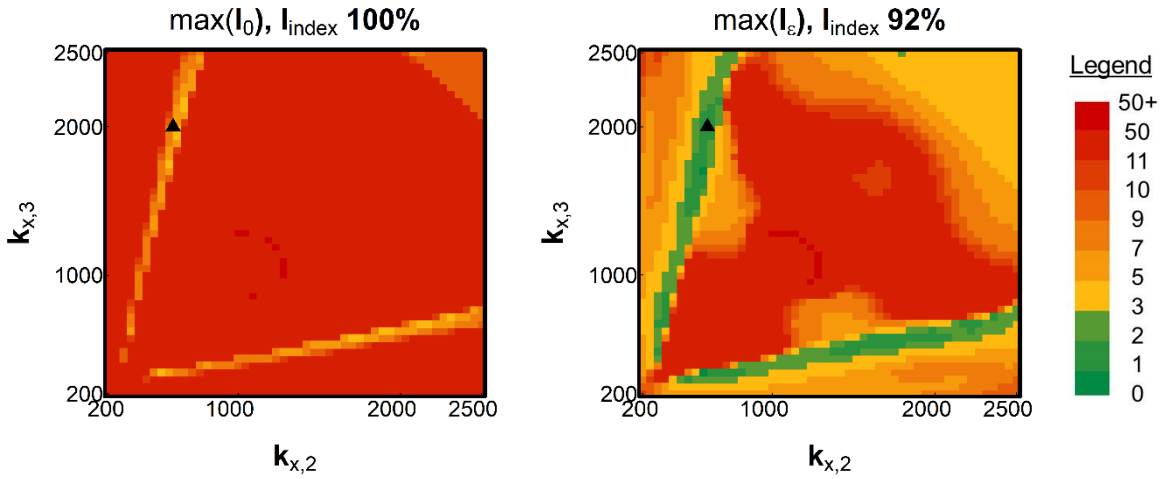


Figure 5.14: Combined adapted implausibility measure for (a) experiment 1 and (b) experiment 2, where the model discrepancy is not considered and considered respectively.

Without accounting for the model discrepancy (Figure 5.14.a, experiment 1), all the input space was deemed implausible, indicating that the problem characterisation should be reviewed (*e.g.* in this case, the uncertainties of the problem). This consists of one of the strengths of the BHMUR approach.

When accounting for the model discrepancy (Figure 5.14.b, experiment 2), 8% of the input space is deemed as non-implausible. A useful consistency check is that the hypothetical reality is considered non-implausible. In the next section, we plot the production curves of the non-implausible space, showing the uncertainty reduction in the output space.

5.4.3 Uncertainty Reduction in Historical and Forecasting Periods

The production curves of Figure 5.15 present the reduction of uncertainty in the behaviour of several outputs for non-implausible scenarios of Experiment 2 (green region of Figure 5.14-b). In grey, we have the simulation of scenarios of all evaluation set, showing the behaviour across the complete input space; the non-implausible region is plotted in green. We also have the historical data in dots and the observational error in the form of error bars.

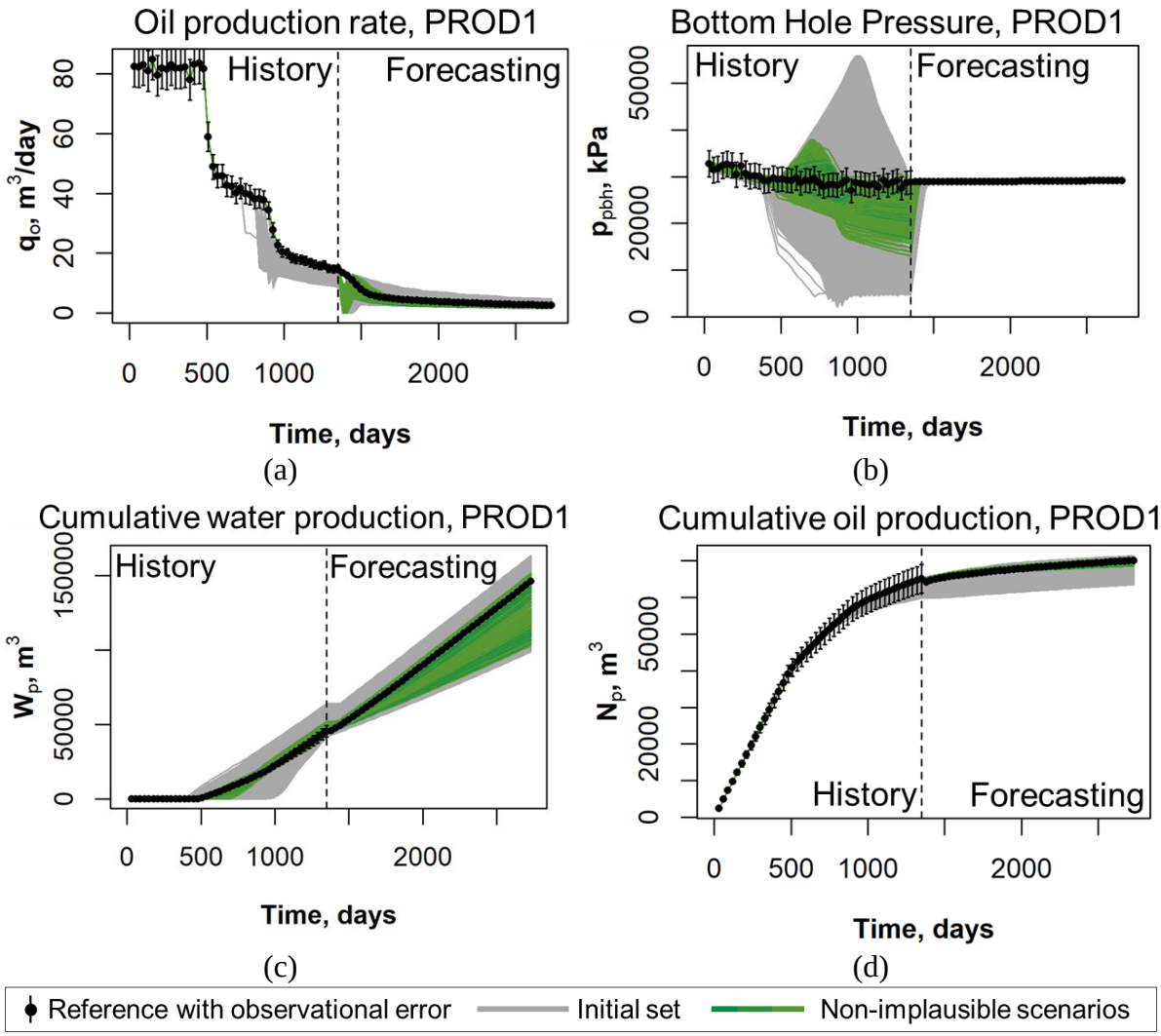


Figure 5.15: Production curves of original input space (grey) and non-implausible scenarios after performing the BHMUR approach (green region in Figure 5.14-b), dots represent the historical data, and the error bars consider the observed error: (a) oil production rate; (b) bottom hole pressure; (c) Cumulative Water production; (d) Cumulative Oil Production. Note that the green tones used in this plot match with the implausibility colours plotted in Figure 5.14-b.

We analyse the uncertainty reduction of the outputs:

(a) Oil production rate: the target data of the simulations had not all the initial input space (grey) reaching the targets because of the minimal pressure set in the simulator (which

was set as 5,000 kPa). The non-implausible space (green lines) shows a clear improvement by having all scenarios reaching the target data;

(b) Bottom hole pressure: the uncertainty in the historical period was reduced and not all green lines are within the error bars because we considered the model discrepancy. In the forecasting period, the uncertainty is kept in similar levels for grey and green curves, mainly because this case study did not have high variability in the forecasting period;

(c) Cumulative water production: a consistent uncertainty reduction in the historical period is propagated to the forecasting period. We believe that this output would benefit from another wave of BHMUR to reduce further the uncertainty in the forecasting period;

(d) Cumulative oil production: the uncertainty in this output is a consequence of the uncertainty reduction of the oil production rate, and this curve evidences the positive impact in the forecasting period.

We demonstrated the robustness of the proposed procedure by presenting the results for the emulation of model discrepancy for different quantities of interest, integrating these emulators in the implausibility analysis and showing the uncertainty reduction in the behaviour of the reservoir.

5.5 Conclusions

Model calibration, history matching, Bayesian History Matching for Uncertainty Reduction (BHMUR) and data assimilation are model-based techniques of reservoir management, usually time and resources consuming; BHMUR emphasises on identifying all the possible scenarios that are coherent with data available and uncertainty in the problem. A critical stage is to characterise the uncertainties inherent from a numerical model which is an imperfect representation of the physical system. The model discrepancy is referred to as the difference between the computer model and the physical system, influencing procedures for model calibration.

As an example, we focus on one particular source of model discrepancy which is generated by setting, as a boundary condition, the target data with observational errors (*e.g.* oil production). Observational error in target data is expected at some level given the monitoring plan in place, which includes errors from measurement, processing and production

reconciliation, among others. We explored how errors in target data are propagated through the simulation model as a source of discrepancy.

The systematic procedure for uncertainty reduction previously developed by the authors is used, and we propose a procedure for accounting for this source of model discrepancy in the BHMUR approach. We deliberately focused on the emulation of the model discrepancy, evaluating the outputs themselves via simulation – a simplification to be withdrawn in more complex studies. To account for the model discrepancy, we apply an iterative process to compute the bias vector and covariance matrix in several scenarios of the input space, characterising the model discrepancy in a training set. Because this iterative process tends to have high computational cost, we use Bayesian emulators to estimate the expectation of the model discrepancy in the entire input space. Finally, we integrate this source of model discrepancy in the implausibility analysis.

We applied the procedure in a case study with two uncertain attributes: a two-dimensional reservoir model with two out of three layers having horizontal permeability as uncertain. The case study allowed to demonstrate that: (1) error in target data propagates as a source of discrepancy to other reservoir outputs (*e.g.* bottom hole pressure), and (2) the procedure proposed is efficient for quantifying and integrating model discrepancy in the BHMUR approach.

The main numbers of the application are: valid emulators for 27 outputs were constructed and used to compute the implausibility measure in the input space, which was combined through the maximal implausibility measure. We compared two experiments. Firstly, when we did not consider the model discrepancy, all the input space was deemed as implausible, which in the BHMUR approach is an indication to review the model characterisation (including model discrepancy specification), uncertain attributes and uncertainties in the problem. When we considered the model discrepancy, in experiment 2, 8% of the input space was classified as implausible. A valid consistency check is that the Hypothetical Reality (our reference scenario) is deemed as non-implausible by the end of the uncertainty reduction process. We also presented the production curves comparing the initial input space with the non-implausible region, demonstrating that we consistently reduced uncertainty in the historical and forecast periods.

We demonstrated one cause of model discrepancy that can be characterised through numerical experiments, and we proposed and applied a procedure to describe and integrate model discrepancy in the BHMUR approach. The results of our application suggested that errors in target data can be a relevant source of model discrepancy, indicating careful consideration in more complex case studies is suitable.

Nomenclature – Article 4

BHM	=	Bayesian History Matching
BHMUR	=	Bayesian History Matching for Uncertainty Reduction
CI	=	Credible Interval
ϵ	=	model discrepancy, the difference between the real reservoir and the reservoir model
E	=	expectation operator
e	=	vector of observational errors
f	=	function of reservoir simulation model that computes a vector of quantities of interest
f^*	=	emulator function
g	=	known deterministic function
I	=	Univariate implausibility measure
I_M	=	maximal implausibility measure
$\tilde{I}$	=	Univariate adapted implausibility measure
$\mathcal{I}$	=	Multivariate implausibility measure
$\tilde{\mathcal{I}}$	=	Multivariate adapted implausibility measure
i_w	=	water injection rate
L_p	=	cumulative liquid production
n	=	number of something
n_A	=	number of active variables
N_p	=	cumulative oil production
p_{ibh}	=	bottom hole pressure of injection wells
p_{pbh}	=	bottom hole pressure of production wells
q_l	=	liquid production rate
q_o	=	oil production rate
q_w	=	water production rate
u	=	Gaussian process
Var	=	Variance
W_i	=	cumulative water injection
W_p	=	cumulative water production
w	=	nugget process
ω	=	implausibility cut-off
x	=	vector of input parameter values representing a reservoir scenario
x^*	=	most appropriate vector of uncertain attributes
y	=	vector of quantities from the real physical reservoir
z	=	vector of measurable quantities from the real reservoir
α	=	proportion covered by the credible interval

β = unknown scalar regression coefficients

Subscripts

A = active variables
i = a measurable quantity of interest of the reservoir, $i \in [1, q]$
j = index corresponding to a regression term in the emulator equation

6 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

This thesis addressed the problem of using observed data to gain knowledge about the physical behaviour of a petroleum reservoir field and provide a better basis for decision making. We delivered a general and systematic Bayesian methodology for uncertainty quantification combining simulation and emulation techniques with a petroleum reservoir application. To this end, this thesis has four complementary fronts of contribution.

Firstly, we explored the impact of adding misfit functions related to specific phenomena in the physical system: (a) water breakthrough time redefines the multi-phase flow in the porous media and has implications for the recovery factor of the field; (b) switching the boundary conditions defined as target for the simulations provide new information on the surrounding regions of the wells. The procedure using Iterative Discrete Latin Hypercube technique (Maschio and Schiozer, 2016) aptly allowed the comparison of two applications with a different number of misfit functions (64 and 128) and the same parameterisation for the UNISIM-I-H case study. The results of these applications demonstrated the beneficial potential of breakthrough deviation and productivity deviation to reduce uncertainty while keeping variability in the resulting collection of scenarios.

Founded in reservoir simulation acumen, statistical principles and an understanding of the case study, the general and systematic Bayesian methodology for uncertainty quantification combining simulation and emulation techniques builds on 20 linked steps. It combines implausibility measure, sequential waves, several emulation techniques and the concept of phases of evaluation. The methodology enables the practice of Bayesian History Matching for Uncertainty Reduction to be more standardised across applications and has three main features: (1) flexibility, since the steps of the high-level structure are adaptable to project requirements, (2) repeatability, because it is a systematic procedure and (3) scalability to higher dimensional spaces (inputs and outputs). The methodology was validated with a case study with hypothetical reality, and additionally, advances and extensions that are particular to a given step can be integrated into a general procedure.

The high-level structure to consolidate incremental developments. Four extensions were proposed in paper 3 of the thesis: (1) an adaptation of the BHMUR to patters in physical data which can

be labelled as binary; (2) presenting and combining indicators of quality capable of validating emulators statistically and selecting one among concurrent emulators for the same quantity of interest; (3) a systematic procedure to choose a combination of quantities of interest to be emulated; and (4) a consistent demonstration of assessment of random and correlated errors, where this naïve approach can be extended to more complex evaluations. They are stand-alone developments that were straightforwardly integrated into the high-level structure.

The fourth front addresses the challenge of characterising model discrepancy – the difference of behaviour between the physical system and the reservoir simulation model. In reservoir simulation, observed data such as oil or liquid production rates are usually used as boundary conditions, *i.e.* simulation targets. We observed that errors in target data propagate through the reservoir simulation model as a cause of model discrepancy affecting other outputs, such as bottom-hole-pressure. We explored this source of model discrepancy and proposed a procedure that allows evaluating covariance matrix and bias vector through the input space to characterise and quantify the model discrepancy. Then, we apply Bayesian emulators to quickly evaluate the model discrepancy in the complete input space. The application in a simple case study demonstrated that the model discrepancy could have an impact on the uncertainty reduction, suggesting that an analysis for other case studies may be important.

We contributed to the Uncertainty Quantification and Reservoir Engineering communities through (a) an understanding of the physical system and relevance of water breakthrough time and boundary conditions (targets) for reservoir simulation; (b) the proposition of the systematic procedure combining reservoir simulation and emulation techniques; (c) the implementation of statistical methods required for an appropriate data analysis when applying Bayesian History Matching for Uncertainty Reduction; and (d) a procedure to quantify and characterise the model discrepancy considering a specific, but recurrent, cause of error related to simulation target.

Certainly, a lot more research and work are needed to provide tools that are scalable, efficient, robust and applicable for real case studies.

The interdisciplinary nature of uncertainty quantification techniques claims for the integration of reservoir engineering and statistics experts. Future work on the usability of the systematic Bayesian methodology will play a central role to enable dissemination of the technique and straightforward application by professionals and for multiple projects/studies. In the energy industry, uncertainty quantification for petroleum reservoirs gained a place in corporate workflows and has a material impact in decision making. Investments in the usability of the systematic Bayesian methodology have the potential to place Bayesian History Matching as cutting-edge technology, to improve the understanding of the reservoir behaviour and to reduce time to the decision-makers.

In what follows, we outlined some of the prominent topics to be addressed and incrementally developed, and which can be subject to future work. We expose the topics as they integrate

into the groups of the systematic Bayesian methodology, marking them by alphabetic order while sharing some ideas and suggestions to approach them.

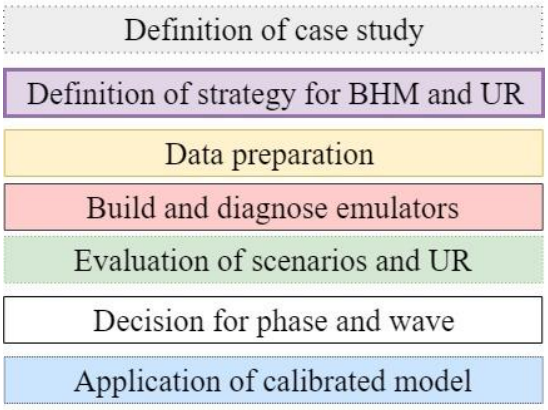


Figure 6.1: Groups of steps presented in the systematic procedure for uncertainty quantification.

6.1. Definition of case study

The parameterisation applied in the papers is built on uncertain attributes associated with regions of the reservoir. Regional multipliers and parameters represent a simplification of subsurface properties, rough and certainly lacking geological realism, but necessary for the current state of art and objectives of this thesis. The applicability of Bayesian History Matching for Uncertainty Reduction in realistic contexts is fundamentally dependent on advances for the treatment of high dimensional input space– *e.g.* porosity, permeability, facies.

Two directions particularly represent a breakthrough for the integration of geological maps in the methodology: (A) dimensionality reduction to extract the main features of input data in a reasonable number of uncertain attributes that enable the construction of valid emulators; (B) dimensionality expansion of attributes with reduced uncertainty, allowing the physical and geological interpretation of the results.

(C) Multiple levels of fidelity may be available for the numerical model used to describe the physical system, *e.g.* fine to coarse grids. Models with higher fidelity level (and higher resolution) are expected to better represent the physical behaviour and be more expensive to run than lower fidelity models. In such contexts, multi-level emulation techniques can be developed. A useful reference to this topic is Cumming and Goldstein (2007).

6.2. Definition of strategy from BHM and UR

We identified two sources of progress in this group of steps.

(D) This work concentrated on production data from wells as sources for objective functions (paper 1) and quantities of interest (articles 2, 3 and 4). Production data is sparse in the reservoir. We may have access and identify as relevant the integration of quantities of interest related to

4D-seismic data, which provide us with information about fluid saturation and pressure changes in time within the field scale. The progress in this area is naturally subsequent to the previous suggestions.

(E) A careful and comprehensive comparison between classes of quantities of interest would be useful to guide a first screening of the data. Besides of rate data, cumulative data, averaged data, misfit index and multi-variate implausibility measure derived from production data, we could assess and compare the efficiency of additional sources of quantity of interest such as productivity deviation presented in the article 1 of this thesis. The investigation would focus on finding if the simulation of the last time step of a given phase with a different target (*e.g.* pressure) could be informative for identifying implausible regions in initial, and cheap to simulate, phases of evaluation.

6.3. Data preparation

(F) Efficient sampling of non-implausible scenarios of the input space is challenging, even when using emulators to evaluate new scenarios. In situations with high dimensional space and/or small uncertainties in the process, non-implausible scenarios are only a tiny portion of the original search space. Further development in this area could recur to (1) methods inspired in Metropolis–Hastings procedure, substituting the rejection criteria by emulator evaluations plus implausibility measures, (2) dimensionality reduction of the non-implausible input space (with Principal Component Analysis, for example), sampling new scenarios within the limits of the reduced space, projecting samples to the original space and evaluating them with emulators and implausibility measure for a final verification. This second suggestion would be particularly efficient if the uncertain attributes have a relevant correlation in the non-implausible regions, which is dependent on an exploratory analysis of the data under consideration.

(G) Together with the treatment of high-dimensional spaces and despite the current efforts in this area, the characterisation of the model discrepancy remains a topic of rupture to the application of Bayesian History Matching for Uncertainty Reduction in realistic contexts. For all applications in Uncertainty Quantification, accounting for the differences between the simulation model and physical system is challenging. Article 4 of this thesis explored one source of model discrepancy, and the subject clearly would deserve more than one PhD thesis to reach the level of practical and consistent consideration of model discrepancy.

6.4. Construct and validate emulators

(H) We proposed a methodology for the efficient selection of a combination of outputs to emulate. The computational cost of this selection tends to increase with higher dimensional output spaces. Therefore, in parallel to the consideration of map information as outputs to emulate (*i.e.* 4D-seismic data), new techniques for output selection would gain space and increment the efficiency of the process.

(I) The emulators applied in this thesis are based on the regression model (articles 2 and 3), which capture the main trends in the input-output relationships. The Gaussian process (applied in article 4) is an alternative and complete approach to capture local variations. Moreover, the full emulator accounts for a nugget related to the inactive input variables. The investigation of the full version of an emulator in higher dimensions is a natural step for the consideration of complex relationships between inputs and outputs. The definition of hiper-parameters of the Gaussian process (*e.g.* correlation length) in high dimension is critical for the application of the full version of the emulator.

6.5. Evaluation of scenarios and UR

(J) In steps using implausibility measure to reject regions of the input space, investigation of criteria such as (1) the combination of the implausibility measure from several emulators by the first, second and third maximal implausibility, and (2) cut-off value for implausibility would be valuable. As most of the literature, our approach is straightforward and a deeper discussion around these criteria would be especially relevant in a more complex application, with imperfect models and observed data.

6.6. Decision for phase and wave

(K) The criteria applied to change phase and wave were sufficient to demonstrate our systematic methodology but empirical in their nature and subject of future work and consideration.

6.7. Application of non-implausible scenarios

(L) The application of non-implausible scenarios for forecasting production and strategy optimisation is a field by itself. A point of attention raised from this thesis is the transition between history and forecasting period and the consequent changes in the boundary conditions. As the usual procedure in reservoir simulation is to switch from target to operational conditions mode, we may lose important information related to the model discrepancy. For example, a well underproducing oil in historical period would keep this pattern when boundary conditions are modified? If so, we could, for example, verify the possibility to consider this discrepancy for forecasting.

.

7 REFERENCES

Alfonzo M. and Oliver, D. S. (2019) Seismic data assimilation with an imperfect model. Computational Geosciences. https://doi.org/10.1007/s10596-019-09849-0.

Almeida, F. L. R., Davolio, A. and Schiozer, D. J. 2014. A New Approach to Perform a Probabilistic and Multi-Objective History Matching. SPE Annual Technical Conference and Exhibition, Amsterdam, The Netherlands, 27-29 October. SPE-170623-MS. https://doi.org/10.2118/170623-MS.

Almeida, F. L. R., Formentin, H. N., Maschio, C. *et al.* 2018. Influence of additional objective functions in the history matching and uncertainty reduction. Accepted to the SPE Europec featured at 80th EAGE Annual Conference & Exhibition, Copenhagen, Denmark, 11-14 June. SPE-190804-MS.

Altman, D. G. and Bland, J. M. 1994. Diagnostic Tests 2: Predictive Values. BMJ Jul 9 309 (6947): 102. https://doi.org/10.1136/bmj.309.6947.102.

ANP and INMETRO 2013. Joint resolution ANP/INMETRO Number 1, dated June 10th 2013 published in the Official Journal of the Federal Government of Brazil, rectified in June 17th 2013 (in Portuguese: Resolução Conjunta ANP/INMETRO Nº 1, DE 10.6.2013 - DOU 12.6.2013 – RETIFICADA DOU 17.6.2013)

Arnold, D., Demyanov, V., Rojas, T. *et al.* 2019. Uncertainty Quantification in Reservoir Prediction: Part 1—Model Realism in History Matching Using Geological Prior Definitions. Mathematical Geosciences 51: 209. https://doi.org/10.1007/s11004-018-9774-6

Avansi, G. D., Schiozer, D. J. 2015. UNISIM-I: Synthetic Model for Reservoir Development and Management Applications. International Journal of Modeling and Simulation for the Petroleum Industry 9 (1): 21–30, April, Brazil.

Avansi, G., Rios, V., Schiozer, D.J. 2019. Numerical Tuning in Reservoir Simulation: it is Worth the Effort in Practical Petroleum Applications. Journal of the Brazilian Society of Mechanical Sciences and Engineering (01/2019): 41-59. https://doi.org/10.1007/s40430-018-1559-9.

Avansi, G.D., Maschio, C., Schiozer, D. J. 2016. Simultaneous History-Matching Approach by Use of Reservoir-Characterization and Reservoir-Simulation Studies. SPE Reservoir Evalution & Engineering 19 (04): 694–712. Paper No. SPE-179740-PA. https://doi.org/10.2118/179740-PA.

Bachoc, F. 2013. Parametric estimation of covariance function in Gaussian-process based Kriging models. Application to uncertainty quantification for computer experiments. Statistics [math.ST]. Université Paris-Diderot - Paris VII, 2013. English. fftel-00881002. https://tel.archives-ouvertes.fr/tel-00881002

Baehyun, M., Kang, J. M., Chung, S. *et al.* 2014. Pareto-Based Multi-Objective History Matching with Respect to Individual Production Performance in a Heterogeneous Reservoir. Journal of Petroleum Science and Engineering 122 (October) Elsevier: 551–66. doi: https://doi.org/10.1016/j.petrol.2014.08.023.

Baker, R. 1998. Reservoir Management for Waterfloods - Part II. Journal of Canadian Petroleum Technology 37 (1): 12–17. PETSOC-98-01-DA. https://doi.org/10.2118/98-01-DA.

Barber, D. 2012. Bayesian Reasoning and Machine Learning. Cambridge: Cambridge University Press. Available in http://www.cs.ucl.ac.uk/staff/d.barber/brml/, 10/01/2019. ISBN: 9780521518147.

Bastos, L. S. and O'Hagan, A. 2009. Diagnostics for Gaussian Process Emulators. Technometrics 51 (4): 425–438. doi:10.1198/TECH.2009.08019.

Bastos, L. S. 2010. Validating Gaussian Process Models in Computer Experiments. University of Sheffield. Thesis. Statistics. Available in http://etheses.whiterose.ac.uk/963/2/Bastos_thesis_final.pdf?origin=publication_detailA, accessed in 18/11/19.

Bertolini, A. C. and Schiozer, D. J. 2011. Influence of the Objective Function in the History Matching Process. Journal of Petroleum Science and Engineering 78 (1): 32–41. doi: https://doi.org/10.1016/j.petrol.2011.04.012.

Brynjarsdóttir, J. and O'Hagan, A. 2014. Learning about physical parameters: the importance of model discrepancy. Inverse Problems, Volume 30, Number 11. https://doi.org/10.1088/0266-5611/30/11/114007

Busby, D. 2009. Hierarchical Adaptive Experimental Design for Gaussian Process Emulators. Reliability Eng. and System Safety 94 (7): 1183–93. https://doi.org/10.1016/j.ress.2008.07.007.

Busby, D. G. M. 2007. Uncertainty Propagation and Reduction in Reservoir. PhD Thesis, University of Leicester, Leicester.

Busby, D., Farmer, C. L. and Iske, A. 2007. Uncertainty Evaluation in Reservoir Forecasting by Bayes Linear Methodology. Proc., 5th International Conference, Algorithms for Approximation, Chester, July 2005. https://doi.org/10.1007/978-3-540-46551-5_14

Caiado, C.C.S. & Goldstein, M. 2015. Bayesian uncertainty analysis for complex physical systems modelled by computer simulators with applications to tipping points. Communications in Nonlinear Science and Numerical Simulation 26(1-3): 123-136.

Carrassi, A., Bocquet, M., Bertino, L. *et al.* 2018. Data Assimilation in the Geosciences: An Overview of Methods, Issues, and Perspectives. Wiley Interdisciplinary Reviews: Climate Change 9 (5): 1–79. doi:10.1002/wcc.535.

Casella, G. and Berger, R.L., 2008. Statistical Inference, international student edition. Cengage Learning. ISBN13: 978-0-495-38187-6

Cosentino, L. 2001. Integrated Reservoir Studies (Fundamentals of Exploration and Production). Editions Technips, 310 pages. ISBN-13: 978-2710807971

Craig, P S, Goldstein M., Seheult A.H. *et al.* 1996. Bayes Linear Strategies for Matching Hydrocarbon Reservoir History. In Bayesian Statistics 5 - Proceedings of the Fifth Valencia International Meeting June 5-9, 1994, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, 69–95. Oxford University Press.

Craig, P. S., Goldstein M., Seheult A. H. *et al.* 1997. Pressure Matching for Hydrocarbon Reservoirs: A Case Study in the Use of Bayes Linear Strategies for Large Computer Experiments. In: Case Studies in Bayesian Statistics. Lecture Notes in Statistics, ed. C. Gatsonis, J.S. Hodges, R. E. Kass, R. McCulloch, P. Rossi, N.D. Singpurwalla, Vol. 121: 37–93. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-2290-3_2.

Craig, P. S., Goldstein, M., Rougier, J.C. *et al.* 2011. Bayesian Forecasting for Complex Systems Using Computer Simulations. Journal of the American Statistical Association 96: 717–729. https://doi.org/10.1198/016214501753168370

Craig, P. S., Smith, J.A., Goldstein, M. *et al.* 1995. Matching Hydrocarbon Reservoir History - a Bayes Linear Approach. Proc., the Third Applied Statistics in Industry Conference, 371–79.

Cumming, J. A. and Goldstein, M., 2007. Multilevel Emulation. MUCM report available in http://www.mucm.ac.uk/Pages/Downloads/Internal%20Reports/INT1.3.3%20JC%20Multilevel%20Emulation.pdf, accessed in 04/January/2020.

Emerick, A. and Reynolds, A. 2011. History Matching a Field Case Using the Ensemble Kalman Filter with Covariance Localization, SPE Reservoir Evaluation & Engineering-Reservoir Engineering, Vol. 14, Number 4, 2011. DOI: 10.2118/141216-PA.

Evensen, G. 2009. Data Assimilation: The Ensemble Kalman Filter, second edition. Berlin: Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-642-03711-5.

Evensen, G. 2018-a. Accounting for Model Errors in Iterative Ensemble Smoothers. ArXiv preprint:1806.00237 [Physics.Data-An], 1–17. Submitted on 1 Jun 2018. http://arxiv.org/abs/1806.00237.

Evensen, G. 2018-b. Introducing Stochastic Model Errors in Ensemble-Based History Matching. Proc., ECMOR XVI - 16th European Conference on the Mathematics of Oil Recovery, Barcelona, 3-6 September 2018. doi:10.3997/2214-4609.201802280.

Evensen, G. and Eikrem, K. S. 2018. Conditioning Reservoir Models on Rate Data Using Ensemble Smoothers. Computational Geosciences 22 (5): 1251–70. https://doi.org/10.1007/s10596-018-9750-8.

Evensen, G. 2019. Accounting for Model Errors in Iterative Ensemble Smoothers. Computational Geosciences (2019) 23:761–775. https://doi.org/10.1007/s10596-019-9819-z

Fang, K.-T., Li R. and Sudjianto A. 2006. Design and Modelling for Computer Experiments. Boca Raton: Chapman; Hall/CRC Press.

Ferreira, C. J., Vernon, I. R., Schiozer, D. J. *et al.* 2014. Use of Emulator Methodology for Uncertainty Reduction Quantification. Proc., SPE Latin America and Caribbean Petroleum Engineering Conference, Maracaibo, Venezuela, 21-23 May, SPE-169405-MS. https://doi.org/10.2118/169405-MS.

Ferreira, C. J., Vernon, I. R., Caiado, C., Nandi-Formentin, H., Avansi, G. D., Goldstein, M., Schiozer, D. J. 2020. Efficient Selection of Reservoir Model Outputs within an Emulation-Based Bayesian History Matching Uncertainty Analysis. SPE-201209-PA.

Formentin, H. N., Almeida, F. L., Avansi, G. D. *et al.* 2019-a. Gaining More Understanding About Reservoir Behavior Through Assimilation of Breakthrough Time and Productivity Deviation in the History Matching Process. Journal of Petroleum Science and Engineering 173 (February): 1080–96. https://doi.org/10.1016/j.petrol.2018.10.045.

Formentin, H. N., Vernon, I., Avansi, G. D. *et al.* 2019-b. Systematic uncertainty reduction for petroleum reservoirs combining reservoir simulation and Bayesian emulation techniques. https://doi.org/10.2118/195478-MS

Formentin, H. N., Vernon, I., Avansi, G. D. *et al.* 2020-a. Systematic uncertainty reduction for petroleum reservoirs combining reservoir simulation and Bayesian emulation techniques: Part I. To be submitted.

Formentin, H. N., Vernon, I., Avansi, G. D. *et al.* 2020-b. Systematic uncertainty reduction for petroleum reservoirs combining reservoir simulation and Bayesian emulation techniques: Part II. To be submitted.

Formentin, H. N., Vernon, I., Goldstein, M. *et al.* 2020-c. Accounting for Model Discrepancy in Uncertainty Analysis by Combining Numerical Simulation and Bayesian Emulation Techniques. To be submitted.

Fricker, T. 2010. Emulators for Multiple Output Computer Models. PhD Thesis. University of Sheffield, United Kingdom (January 2010).

Frøysa, K. E., Lunde, P., Lied, G. Ø. *et al.* 2018. Handbook of uncertainty calculations for ultrasonic, turbine and Coriolis oil flow metering stations - Documentation of uncertainty models and internet tool. Report CMR-17-A211576-RA-1 from the Norwegian Society for Oil and Gas Measurement, February 2018. Available in https://nfogm.no/wp-content/uploads/2018/02/CMR-17-A211576-RA-1-Rev01-Handbook-fiscal-ultrasonic-oil-metering-station-updated-2017_new-front-page.pdf, accessed on 18/11/2019.

Goldstein, M. and Rougier, J. 2006. Bayes Linear Calibrated Prediction for Complex Systems, Journal of the American Statistical Association, 101, 1132–1143

Goldstein, M. and Rougier, J. 2006. Bayes Linear Calibrated Prediction for Complex Systems. Journal of the American Statistical Association. 101 (475): 1132–43. https://doi.org/10.1198/016214506000000203.

Goldstein, M., Seheult, A. and Vernon, I. 2013. Assessing model adequacy. In Environmental Modelling: Finding Simplicity in Complexity, second edition, ed. J. Wainwright and M. Mulligan, Chap. 26, 435–49. Wiley-Blackwell. https://doi.org/10.1002/9781118351475.ch26.

Goldstein and Wooff, 2007. Bayes Linear Statistics: Theory and Methods. DOI: 10.1002/9780470065662.

Goldstein, M. 2017. Uncertainty analysis for computer models. https://www.ncl.ac.uk/media/wwwnclacuk/cesi/files/Uncertainty%20considerations%20Prof%20Goldstein%20-%20Durham.pdf, last acessed in 02/11/2019.

Goodwin, N. 2015. Bridging the Gap Between Deterministic and Probabilistic Uncertainty Quantification Using Advanced Proxy Based Methods. Presented at the SPE Reservoir Simulation Symposium, Houston, Texas, 23-25 February. SPE-173301-MS. https://doi.org/10.2118/173301-MS.

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L. *et al.* 1998. Multivariate Data Analysis, fifth edition. Prentice Hall. ISBN 0-13-84858-5.

Hoaglin, D. C., Mosteller, F., Tukey, J. W. 1983. Understanding Robust and Exploratory Data Analysis, first edition. Wiley Series in Probability and Mathematical Statistics. ISSN 0271-6356.

Hutahaean, J. J., Demyanov, V. and Christie, M. A. 2017. On Optimal Selection of Objective Grouping for Multi-objective History Matching. SPE Journal 22 (04). SPE-185957-PA. https://doi.org/10.2118/185957-PA.

Hutahaean, J. J., Demyanov, V. and Christie, M. A 2015. Impact of Model Parameterisation and Objective Choices on Assisted History Matching and Reservoir Forecasting. Presented at the SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition, Nusa Dua, Bali, Indonesia. 20-22 October. SPE-176389-MS. https://doi.org/10.2118/176389-MS.

Jackson, S. E. 2018. Design of Physical System Experiments Using Bayes Linear Emulation and History Matching Methodology with Application to Arabidopsis Thaliana, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/12826/, assessed in 07/November/2019.

James, G., Witten, D., Hastie, T. *et al.* 2013. An Introduction to Statistical Learning with Applications in R. Springer. Available in: http://www-bcf.usc.edu/~gareth/ISL/, 14/10/2018. ISBN 978-1-4614-7138-7 (eBook). doi:10.1007/978-1-4614-7138-7.

Jansen, J. D., Brouwer R. and Sippe G. D. 2009. Closed Loop Reservoir Management. Presented at the SPE Reservoir Simulation Symposium, The Woodlands, Texas. 2-4 February. SPE-119098-MS. https://doi.org/10.2118/119098-MS.

Johnson, M.E., Moore, L.M. and Ylvisaker, D. 1990. Minimax and maximin distance designs. Journal of Statistical Planning and Inference. Volume 26, Issue 2, October 1990, Pages 131-148. https://doi.org/10.1016/0378-3758(90)90122-B.

Kam, D., Han, J. and Datta-Gupta, A. 2017. Streamline-based history matching of bottomhole pressure and three-phase production data using a multiscale approach. Journal of Petroleum Science and Engineering 154 (June): 217–33. doi: https://doi.org/10.1016/j.petrol.2017.04.022.

Kennedy M. C. and O'Hagan A. 2011. Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology)Volume 63, Issue 3. https://doi.org/10.1111/1467-9868.00294

Krzanowski, W. J. 2000. Principles of Multivariate Analysis: A User's Perspective. ISBN-10: 0198507089. Oxford University Press, USA; 2 edition (28 Dec. 2000).

Lawal, K. A., Utin, E. and Langaas, K. 2007. A Didactic Analysis of Water Cut Trend During Exponential Oil-Decline. Proc., Nigeria Annual International Conference and Exhibition, Abuja, Nigeria, 6-8 August, SPE-111920-MS. https://doi.org/10.2118/111920-MS.

Mardia, K. V., Kent, J. T. and Bibby, J. M. 1982. Multivariate Analysis. Paperback ISBN: 9780124712522. Academic Press.

Maschio, C. and Schiozer, D. J. 2016. Probabilistic History Matching Using Discrete Latin Hypercube Sampling and Nonparametric Density Estimation. Journal of Petroleum Science and Engineering 147: 98–115 (November 2017). https://doi.org/10.1016/j.petrol.2016.05.011.

Maschio, C. and Schiozer, D. J. 2017. A New Methodology for Bayesian History Matching Using Parallel Interacting Markov Chain Monte Carlo. Inverse Problems in Science and Engineering 5977 (May). Taylor & Francis: 1–32. doi: https://doi.org/10.1080/17415977.2017.1322078.

Maschio, C. and Schiozer, D. J. 2018. A new methodology for history matching combining iterative discrete Latin Hypercube with multi-start simulated annealing. Journal of Petroleum Science and Engineering 169 (October). Elsevier: 560-577. doi: https://doi.org/10.1016/j.petrol.2018.06.004.

McKay, M. D., Beckman, R. J. and Conover, W. J.1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics. Vol. 21, No. 2 (May, 1979), pp. 239-245. https://www.jstor.org/stable/1268522.

Mirams, G. R.; Pathmanathan P., Gray R. A. *et al.* 2016. Uncertainty and variability in computational and mathematical models of cardiac physiology. The Journal of Physiology published by John Wiley & Sons Ltd on behalf of The Physiological Society; 594(23):6833-6847. doi: 10.1113/JP271671.

Moreno, R., Avansi, G. D., Schiozer, D. J. *et al.* 2018. Emulation of Reservoir Production Forecast Considering Variation in Petrophysical Properties. Journal of Petroleum Science and Engineering 165 (June 2017): 711–25. https://doi.org/10.1016/j.petrol.2018.02.056.

Morrison, F. D. 2005. Multivariate Statistical Methods. ISBN-10: 9780534387785. Cengage Learning; 4th Revised edition.

Myers, H. R., 1990. Classical and modern regression with applications. 2nd edition, PWS Publishers.

NFOGM, 2005. Handbook of Multiphase Flow Metering. ISBN 82-91341-89-3. Available in https://nfogm.no/wp-content/uploads/2014/02/MPFM_Handbook_Revision2_2005_ISBN-82-91341-89-3.pdf, accessed in 18/11/2019.

O'Hagan, A. 2004. Bayesian Analysis of Computer Code Outputs. Reliability Engineering & System Safety 91 (10-11): 1290-1300. http://dx.doi.org/10.1016/j.ress.2005.11.025.

Oakley, J. 1999. Bayesian Uncertainty Analysis for Complex Computer Codes. PhD Thesis, University of Sheffield, Sheffield (December 1999).

Oliver, D. S. and Alfonzo, M. 2017. Calibration of imperfect models to biased observations. Computational Geosciences, February 2018, Volume 22, Issue 1, pp 145–161. https://doi.org/10.1007/s10596-017-9678-4.

Oliver, D. and Chen, Y. 2011. Recent Progress on Reservoir History Matching: A Review. Computational Geosciences 15 (1): 185–221. https://doi.org/10.1007/s10596-010-9194-2.

Oliver, D., Reynolds A. C. and Liu N. 2008. Inverse Theory for Petroleum Reservoir Characterization and History Matching, first edition. Cambridge, United Kingdom: Cambridge University Press. ISBN: 9780521881517.

Pukelsheim, F. 1994. The Three Sigma Rule. The American Statistician 48 (2): 88–91. https://www.jstor.org/stable/2684253.

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rammay, M. H., Elsheikh, A. H. and Chen, Y. 2019. Identifiability of Model Discrepancy Parameters in History Matching. SPE-193838-MS. SPE Reservoir Simulation Conference held in Galveston, Texas, USA, 10–11 April 2019. https://doi.org/10.2118/193838-MS.

Ranjan, R., Masoudi, R., Karkooti, H. *et al.* 2014. History Matched Models Can Mis-Lead the Forecasting in the Brownfield Redevelopment Projects: HM to Prediction Transition. Presented in the International Petroleum Technology Conference, Kuala Lumpur, Malaysia, 10-12 December. IPTC-17744-MS. https://doi.org/10.2523/IPTC-17744-MS.

Rasmussen, C. E. and Williams C. K. I. 2006. Gaussian Processes for Machine Learning. The MIT Press. ISBN 026218253X, available in http://www.gaussianprocess.org/gpml/chapters/RW.pdf, last accessed in 04/November/2019.

Ringrose, P. and Bentley, M. 2015. Reservoir Model Design: A Practitioner's Guide, 2015 edition. Springer. ISBN 978-94-007-5497-3.

Rougier, J., Sexton, D. M. H., Murphy, J. M. *et al.* 2009. "Analyzing the Climate Sensitivity of the HadSM3 Climate Model Using Ensembles From Different but Related Experiments," Journal of Climate, 22, 3540–3557.

Rwechungura, R. W., Dadashpour, M. and Kleppe, J. 2011. Advanced History Matching Techniques Reviewed. Presented in the SPE Middle East Oil and Gas Show and Conference, Manama, Bahrain, 25-28 September. SPE-142497-MS. https://doi.org/10.2118/142497-MS.

Schiozer, D. J., Avansi, G. D. and Santos, A. A. S. 2017. Risk Quantification Combining Geostatistical Realizations and Discretized Latin Hypercube. Journal of the Brazilian Society of Mechanical Sciences and Engineering 39 (2) Springer Berlin Heidelberg: 575–87. doi: https://doi.org/10.1007/s40430-016-0576-9.

Schiozer, D. J., Santos, A. A. S. and Drumond, P. S. 2015. Integrated Model Based Decision Analysis in Twelve Steps Applied to Petroleum Fields Development and Management. Presented in the EUROPEC 2015, Madrid, Spain, 1-4 June. SPE-174370-MS. https://doi.org/10.2118/174370-MS.

Schiozer, D. J., Santos, A. A. A., Santos, S. M. G. *et al.*. 2019. Model-Based Decision Analysis Applied to Petroleum Field Development and Management, Oil & Gas Science and Technology, v. 74, pp. 1-20, May, 2019.

Silva, V. L. S., Emerick, A. A., Couto, P. *et al.* 2017. History Matching and Production Optimization under Uncertainties – Application of Closed-Loop Reservoir Management.

Journal of Petroleum Science and Engineering 157 (August): 860–74. https://doi.org/10.1016/j.petrol.2017.07.037.

Smith, R. C. 2014. Uncertainty Quantification: Theory, Implementation, and Applications. SIAM. ISBN 978-1-611973-21-1.

Tillier, E., Da Veiga, S. and Derfoul, R. 2013. Appropriate Formulation of the Objective Function for the History Matching of Seismic Attributes. Computers & Geosciences 51 (February): 64–73. https://doi.org/10.1016/j.cageo.2012.07.031.

UNISIM webpage, https://www.unisim.cepetro.unicamp.br/benchmarks/en/, accessed in January 4th, 2020.

Venables, W. N. Ripley, B. D. 2002. Modern Applied Statistics with S. Fourth Edition. Springer. ISBN 0-387-95457-0.

Vernon, I., Goldstein, M. and Bower, R. G. 2010. Galaxy Formation: A Bayesian Uncertainty Analysis. Bayesian Analysis 5 (4): 619–70. doi:10.1214/10-BA524.

Vernon, I., Liu, J., Goldstein, M. *et al.* 2018. Bayesian Uncertainty Analysis for Complex Systems Biology Models: Emulation, Global Parameter Searches and Evaluation of Gene Functions. BMC Systems Biology 12 (1). https://doi.org/10.1186/s12918-017-0484-3.

Wang, C., Gaoming, L. and Reynolds, A. C. 2009. Production Optimization in Closed-Loop Reservoir Management. SPE Journal 14 (03). SPE-109805-PA. https://doi.org/10.2118/109805-PA.

Williamson, D. B., Blaker, A. T. and Sinha, B. 2017. Tuning without Over-Tuning: Parametric Uncertainty Quantification for the NEMO Ocean Model. Geoscientific Model Development 10 (4): 1789–1816. https://doi.org/10.5194/gmd-10-1789-2017.

Wutzler, T. 2018. Efficient treatment of model discrepancy by Gaussian Processes - Importance for imbalanced multiple constraint inversions. Submitted to Inverse Problems. arXiv:1812.07801, accessed https://arxiv.org/pdf/1812.07801.pdf in 18/Nov/2019.

Zee Ma, Y. and La Pointe, P. 2011. Uncertainty Analysis and Reservoir Modeling. AAPG Memoir. American Association of Petroleum Geologists. https://doi.org/10.1306/M961330.

APPENDIX A – LICENSE AGREEMENTS FROM PUBLISHERS GRANTING PERMISSION TO REPRODUCE PUBLISHED ARTICLES IN THIS THESIS

Title:	SPE Europec featured at 81st EAGE Conference and Exhibition
Article ID:	978-1-61399-661-4
Publication:	Publication1
Publisher:	CCC Republication
Date:	Jan 1, 2019

Logged in as:
Helena Nandi Formentin
Account #:
3001518302

LOGOUT

Order Completed

Thank you for your order.

This Agreement between Helena Nandi Formentin ("You") and Society of Petroleum Engineers ("Society of Petroleum Engineers") consists of your order details and the terms and conditions provided by Society of Petroleum Engineers and Copyright Clearance Center.

License number	Reference confirmation email for license number
License date	Oct, 15 2019
Licensed content publisher	Society of Petroleum Engineers
Licensed content title	SPE Europec featured at 81st EAGE Conference and Exhibition
Licensed content date	Jan 1, 2019
Type of use	Thesis/Dissertation
Requestor type	Author of requested content
Format	Print, Electronic
Portion	chapter/article
Number of pages in chapter/article	46
The requesting person/organization	Helena Nandi Formentin
Title or numeric reference of the portion(s)	Complete paper, pages 1 to 46
Title of the article or chapter the portion is from	Systematic Uncertainty Reduction for Petroleum Reservoirs Combining Reservoir Simulation and Bayesian Emulation Techniques
Editor of portion(s)	N/A
Author of portion(s)	Helena Nandi Formentin (Durham University and University of Campinas) | Ian Vernon (Durham University) | Guilherme Daniel Avansi (University of Campinas) | Camila Caiado (Durham University) | Célio Maschio (University of Campinas) | Michael Goldstein (Durham University) | Denis José Schiozer (University of Campinas)
Volume of serial or monograph	N/A
Issue, if republishing an article from a serial	N/A
Page range of portion	1-46
Publication date of portion	SPE Europec featured at 81st EAGE Conference and Exhibition, 3-6 June 2019, London, England, UK
Rights for	Main product and other compilations/derivative products
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	yes
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	yes
Lifetime unit quantity of new product	Up to 1,000,000
Title	Uncertainty Quantification of Petroleum Reservoirs: a Structured Bayesian Methodology

	(possible minor reviews in the thesis tittle)
Institution name	Durham University and University of Campinas
Expected presentation date	Feb 2020
Requestor Location	Helena Nandi Formentin Mathematical Sciences Lower Mountjoy, Stockton Rd Address, other DH1 3LE United Kingdom Attn:
Billing Type	Invoice
Billing address	Helena Nandi Formentin Mathematical Sciences Lower Mountjoy, Stockton Rd Address, United Kingdom DH1 3LE Attn: Helena Nandi Formentin
Total (may include CCC user fee)	0.00 USD
Total	0.00 USD

CLOSE WINDOW

APPENDIX B – CONTRIBUTION STATEMENTS

Subject: **Statement of contribution of paper based on CRediT author statement**
Title: "**Gaining More Understanding About Reservoir Behavior Through Assimilation of Breakthrough Time and Productivity Deviation in History Matching Process"**

Period of development: **February/2017 to July/2018**

Submitted to: **JPSE**

- **Helena Nandi Formentin:** Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization
- **Forlan la Rosa Almeida:** Conceptualization, Methodology, Software, Formal Analysis, Investigation, Writing – Review & Editing
- **Guilherme Daniel Avansi:** Conceptualization, Writing – Review & Editing, Supervision, Project Administration, Funding Acquisition
- **Célio Maschio:** Conceptualization, Methodology, Software, Writing – Review & Editing
- **Denis Schiozer:** Conceptualization, Methodology, Writing – Review & Editing, Resources, Supervision, Project Administration, Funding Acquisition
- **Camila Caiado:** Conceptualization, Supervision, Project Administration, Funding Acquisition
- **Ian Vernon:** Conceptualization, Supervision, Project Administration, Funding Acquisition
- **Michael Goldstein:** Conceptualization, Supervision, Project Administration, Funding Acquisition

CRediT author statement (Source: https://www.elsevier.com/authors/journal-authors/policies-and-ethics)

Term	Definition
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims
Methodology	Development or design of methodology; creation of models
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components
Validation	Verification, whether as a part of the activity or separate, of the overall replication/ reproducibility of results/experiments and other research outputs
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data

Term	Definition
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse
Writing – Original Draft	Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)
Writing – Review & Editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre-or postpublication stages
Visualization	Preparation, creation and/or presentation of the published work, specifically visualization/ data presentation
Supervision	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team
Project Administration	Management and coordination responsibility for the research activity planning and execution
Funding Acquisition	Acquisition of the financial support for the project leading to this publication

SPE Paper Information Form

Society of Petroleum Engineers

Meeting Name and Year: SPE Europec featured at 81st EAGE Annual Conference 2019

Paper Number: SPE-195478-MS

Paper Title: Systematic uncertainty reduction for petroleum reservoirs combining reservoir simulation and Bayesian emulation techniques

☐ **AUTHOR 1:** Helena (First *(Forename)*) Nandi (Middle) Formentin (Last *(Family Name)*)

SPE Member Number: 4333473 Company Affiliation: Durham University and University of Campinas

Complete Address: Durham University - Department of Mathematical Sciences - DH1 3LE Durham UK

Telephone: +4407759466150 Fax: _____ E-mail: helena.n.formentin@durham.uk.ac

Briefly describe nature of author's contribution: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original + Edition, Vizualization, Funding Acquisition

☐ **AUTHOR 2:** Ian (First *(Forename)*) Richard (Middle) Vernon (Last *(Family Name)*)

SPE Member Number: _____ Company Affiliation: Durham University

Complete Address: Durham University - Department of Mathematical Sciences - DH1 3LE Durham UK

Telephone: _____ Fax: _____ E-mail: i.r.vernon@durham.ac.uk

Briefly describe nature of author's contribution: Conceptualization, Methodology (statistical formulation), Formal Analysis (noisy+review), Writing (structure+original+review), Supervision, Project Administration, Funding Acquisition

☐ **AUTHOR 3:** Guilherme (First *(Forename)*) Daniel (Middle) Avansi (Last *(Family Name)*)

SPE Member Number: 3252249 Company Affiliation: University of Campinas

Complete Address: University of Campinas: UNISIM, Rua Cora Coralina, 350, 13083-896, Campinas, Brazil

Telephone: _____ Fax: _____ E-mail: avansi@cepetro.unicamp.br

Briefly describe nature of author's contribution: Methodology (review), Resources (related to case study), Writing (structure + review), Visualization (review), Supervision, Project Adm, Funding Acquis.

☐ **AUTHOR 4:** Camila (First *(Forename)*) Caiado (Last *(Family Name)*)

SPE Member Number: _____ Company Affiliation: Durham University

Complete Address: Durham University - Department of Mathematical Sciences - DH1 3LE Durham UK

Telephone: _____ Fax: _____ E-mail: c.c.d.s.caiado@durham.ac.uk

Briefly describe nature of author's contribution: Conceptualization (BHM), Methodology (review), Formal Analysis (review), Writing (structure+review), Visualization (review), Supervision, Project Adm, Funding Acquis.

☐ **AUTHOR 5:** Célio (First *(Forename)*) Maschio (Last *(Family Name)*)

SPE Member Number: _____ Company Affiliation: University of Campinas

Complete Address: University of Campinas: UNISIM, Rua Cora Coralina, 350, 13083-896, Campinas, Brazil

Telephone: _____ Fax: _____ E-mail: celio@cepetro.unicamp.br

Briefly describe nature of author's contribution: Resources (support related to the case studied - parameterization - and simulation in cluster), Writing (Review)

☐ **AUTHOR 6:** Michael (First *(Forename)*) Goldstein (Last *(Family Name)*)

SPE Member Number: _____ Company Affiliation: Durham University

Complete Address: Durham University - Department of Mathematical Sciences - DH1 3LE Durham UK

Telephone: _____ Fax: _____ E-mail: michael.goldstein@durham.ac.uk

Briefly describe nature of author's contribution: Conceptualization (BHM, phases of evaluation), Formal Analysis (review), Methodology (review), Writing (structure + review), Visualization (review), Supervision, Funding Acquisition

SPE Paper Information Form (continued)

Society of Petroleum Engineers

Meeting Name and Year: SPE Europec featured at 81st EAGE Annual Conference 2019

Paper Number: SPE-195478-MS

Paper Title: Systematic uncertainty reduction for petroleum reservoirs combining reservoir simulation and Bayesian emulation techniques

☐ **AUTHOR 7:** Denis José (First *(Forename)*) (Middle) Schiozer (Last *(Family Name)*)

SPE Member Number: _____ Company Affiliation: University of Campinas

Complete Address: University of Campinas: UNISIM, Rua Cora Coralina, 350, 13083-896, Campinas, Brazil

Telephone: _____ Fax: _____ E-mail: denis@cepetro.unicamp.br

Briefly describe nature of author's contribution: Conceptualization (simulation targets), Visualization (review), Supervision, Project Administration, Funding Acquisition, Writing (Review)