UNIVERSIDADE ESTADUAL DE CAMPINAS

FACULDADE DE ENGENHARIA MECÂNICA

E INSTITUTO DE GEOCIÊNCIAS

DANIEL RODRIGUES DOS SANTOS

NEW PARAMETERIZATION AND MACHINE LEARNING APPROACHES FOR EFFICIENT OIL FIELD MANAGEMENT

NOVAS ABORDAGENS DE PARAMETRIZAÇÃO E APRENDIZADO DE MÁQUINA PARA O GERENCIAMENTO EFICIENTE DE CAMPOS DE PETRÓLEO

CAMPINAS

2023

DANIEL RODRIGUES DOS SANTOS

NEW PARAMETERIZATION AND MACHINE LEARNING APPROACHES FOR EFFICIENT OIL FIELD MANAGEMENT

NOVAS ABORDAGENS DE PARAMETRIZAÇÃO E APRENDIZADO DE MÁQUINA PARA O GERENCIAMENTO EFICIENTE DE CAMPOS DE PETRÓLEO

A thesis presented to the Mechanical Engineering Faculty and Geosciences Institute of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Petroleum Sciences and Engineering in the area of Reservoirs and Management.

Tese de Doutorado apresentada à Faculdade de Engenharia Mecânica e Instituto de Geociências da Universidade Estadual de Campinas como parte dos requisitos exigidos para obtenção do título de Doutor em Ciências e Engenharia de Petróleo na área de Reservatórios e Gestão.

Orientador: Prof. Dr. Denis José Schiozer

Este exemplar corresponde à versão final da Tese defendida pelo aluno Daniel Rodrigues dos Santos e orientada pelo Prof. Dr. Denis José Schiozer.

CAMPINAS

2023

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

Sa59n

Santos, Daniel Rodrigues dos, 1990-
New parametrization and machine learning approaches for efficient oil field management / Daniel Rodrigues dos Santos. – Campinas, SP : [s.n.], 2023.

Orientador: Denis José Schiozer.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Mecânica.

1. Reservatórios (Simulação). 2. Otimização. 3. Incerteza. 4. Aprendizado de máquina. 5. Engenharia de petróleo. I. Schiozer, Denis José, 1963-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Mecânica. III. Título.

Informações Complementares

Título em outro idioma: Novas abordagens de parametrização e aprendizado de máquina para o gerenciamento eficiente de campos de petróleo
Palavras-chave em inglês:
Reservoir simulation
Optimization
Uncertainty
Machine learning
Petroleum engineering
Área de concentração: Reservatórios e Gestão
Titulação: Doutor em Ciências e Engenharia de Petróleo
Banca examinadora:
Denis José Schiozer [Orientador]
João Carlos Von Hohendorff Filho
Guilherme Palermo Coelho
Eduardo Gildin
Leonardo José do Nascimento Guimarães
Data de defesa: 06-12-2023
Programa de Pós-Graduação: Ciências e Engenharia de Petróleo

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-3219-2890
- Currículo Lattes do autor: http://lattes.cnpq.br/3575190794602020

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA MECÂNICA
E INSTITUTO DE GEOCIÊNCIAS

Tese de Doutorado Acadêmico

NEW PARAMETERIZATION AND MACHINE LEARNING APPROACHES FOR EFFICIENT OIL FIELD MANAGEMENT

Autor: Daniel Rodrigues dos Santos
Advisor/Orientador: Prof. Dr. Denis José Schiozer

The Examining Committee composed of the members below approved this thesis:
A Banca Examinadora composta pelos membros abaixo aprovou esta tese:

Prof. Dr. Denis José Schiozer, President/Presidente
DEP / FEM / UNICAMP

Dr. João Carlos Von Hohendorff Filho
UNISIM / CEPETRO / UNICAMP

Prof. Dr. Guilherme Palermo Coelho
FT / UNICAMP

Prof. Dr. Eduardo Gildin
Harold Vance Department of Petroleum Engineering / College of Engineering / Texas A&M University

Prof. Dr. Leonardo José do Nascimento Guimarães
DEC / Universidade Federal de Pernambuco

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

December 6, 2023
Campinas

DEDICATION

In loving memory of my beloved grandparents, whom I have always admired.

ACKNOWLEDGEMENTS

I am deeply grateful to the individuals and organizations who have been instrumental in making this research journey possible and providing me with support and guidance throughout the process.

First and foremost, I express my gratitude to Professor Dr. Denis José Schiozer for the guidance and providing me this opportunity to work on an industry-relevant topic. His mentorship and encouragement have been pivotal in shaping the direction of this thesis.

I would like to express my sincere appreciation to Professor Dr. André Ricardo Fioravanti for his valuable insights and contributions to this research. His expertise and feedback have been very important in refining the ideas presented in this work.

My thankfulness also goes to Dr. Vinicius Eduardo Botechia for his contributions to this endeavor. Not only did he meticulously correct and fine-tune my research, but he also extended his assistance and expertise throughout this project.

I would like to acknowledge Antonio Alberto de Souza dos Santos for his co-authorship in two of the articles incorporated in this thesis. I also want to extend a special thanks to Guilherme Roberto Tonin for his logistical support, which ensured the smooth progression of this project

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

I am also grateful to Energi Simulation for the financial support, without which this research would not have been possible. Additionally, I extend my appreciation to UNICAMP, FEM, CEPETRO, and UNISIM for providing a conducive and stimulating environment that allowed me to enhance my skills and knowledge during this scientific exploration.

Finally, I must express my heartfelt gratitude to my family and friends. Their constant encouragement, understanding, and unwavering support have been a source of strength and motivation, guiding me through the challenges encountered on this journey.

RESUMO

Santos, Daniel Rodrigues dos, *Novas Abordagens de Parametrização e Aprendizado de Máquina para o Gerenciamento Eficiente de Campos de Petróleo*, Campinas, Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas, 2023. 171 p. Tese (Doutorado)

A seleção da estratégia de produção sob incertezas é uma etapa importante no desenvolvimento e gerenciamento de campos de petróleo, pois impacta diretamente o fator de recuperação de óleo e o retorno econômico do projeto. A abordagem mais confiável para determinar a estratégia ótima envolve a realização de inúmeras simulações usando modelos numéricos de reservatórios para otimizar um grande número de variáveis.

Apesar dos avanços computacionais, as simulações necessárias para determinar a estratégia ideal ainda requerem tempo significativo, especialmente em situações reais, que exigem uso de modelos de maior fidelidade para representar o campo, hidrocarbonetos com múltiplos componentes e incertezas. Além do alto esforço computacional, outra desvantagem comumente identificada nos procedimentos tradicionais é o fato de fornecerem estratégias excessivamente complexas de serem implementadas.

A fim de abordar os problemas mencionados, esta tese propõe procedimentos que aceleram o processo de otimização das variáveis de controle de poço ao longo do ciclo de vida do campo, ao mesmo tempo que geram estratégias bem-comportadas para implementação prática em casos de injeção de água e de injeção alternada de água e gás (WAG).

Este trabalho foi dividido em quatro artigos. O primeiro explora quatro regras de parametrização das variáveis de controle do ciclo de vida do campo buscando maximizar o retorno econômico do projeto. A parametrização inicial otimiza a distribuição das vazões e o tempo de fechamento dos poços. As outras três consistem em equações paramétricas, incluindo a logística e as polinomiais de primeira e segunda ordem, para definir a pressão de fundo de poço (BHP) ao longo do tempo. No segundo artigo, apresentamos o método IDLHC-ML, o qual integra técnicas de aprendizado de máquina ao algoritmo hipercubo latino discretizado iterativo (IDLHC). Esse método visa diminuir a quantidade de simulações necessárias no processo de otimização da melhor parametrização encontrada no trabalho anterior. No terceiro, apresentamos o IDLHC-MLR, uma extensão do IDLHC-ML, que visa incorporar incertezas

através da otimização robusta de modelos representativos. Expandindo o conceito de parametrização introduzido no primeiro artigo, o quarto estudo apresenta uma nova equação paramétrica para acelerar o processo de otimização de injeção WAG. Essa equação define a prioridade de injeção de água ou gás para cada poço em cada intervalo. Para isso, a equação considera dados de produção do reservatório, incluindo corte de água, razão gás óleo e produção acumulada de gás dos poços.

As parametrizações propostas para as variáveis de controle de poço do ciclo de vida do campo, seja para definição do BHP ou da injeção WAG ao longo do tempo, melhoraram a convergência do algoritmo de otimização para soluções melhores quando comparadas com as de abordagens tradicionais. Além disso, essas estratégias exibiram comportamentos favoráveis para aplicação prática. O método de aprendizado de máquina também foi bem-sucedido reduzindo o número de simulações necessárias sem comprometer o retorno econômico nos cenários nominal e probabilístico.

ABSTRACT

The selection of production strategy under uncertainties is a crucial step in the development and management of oil fields, as it directly impacts the oil recovery factor and the economic return of the project. The most reliable approach to determine the optimal strategy involves conducting numerous simulations using reservoir numerical models to optimize a large number of variables.

Despite computational advancements, the simulations required to ascertain the optimal strategy still demand a substantial amount of time, particularly in real-world scenarios that necessitate the use of higher-fidelity models to represent the field, hydrocarbons with multiple components, and uncertainties. In addition to the high computational effort, another drawback frequently observed in traditional procedures is their tendency to yield excessively intricate implementation strategies.

In order to address the mentioned issues, this thesis proposes procedures that accelerate the well control life-cycle optimization process while also generating well-behaved strategies for practical implementation in cases of water injection and water alternating gas injection (WAG).

This work has been divided into four articles. The first explores four parameterization rules for the well control life-cycle variables, aiming to maximize the project's economic return. The initial parameterization optimizes the distribution of flow rates and the well shut-in time. The other three consist of parametric equations, including logistic and first and second-order polynomial equations, to define the wells bottom-hole pressure (BHP) over time. In the second article, we present the IDLHC-ML method, which integrates machine learning techniques into the iterative discretized Latin hypercube (IDLHC) algorithm. This method aims to reduce the number of simulations required in the optimization process of the best parameterization found in the previous work. In the third article, we introduce the IDLHC-MLR, an extension of IDLHC-ML that aims to incorporate uncertainties through robust optimization of the representative models. Expanding on the parameterization concept introduced in the first article, the fourth study presents a new parametric equation to expedite the optimization process of WAG injection. This equation defines the priority of water or gas injection for each well in each interval. To achieve this, the equation considers reservoir production data, including water cut, gas-oil ratio, and the wells’ cumulative gas production.

The parameterizations proposed for the well control life-cycle variables, whether for defining BHP or WAG injection over time, improved the optimization algorithm's convergence towards better solutions when compared to those of traditional approaches. Furthermore, these strategies exhibited favorable behaviors for practical implementation. The machine learning method was also successful in reducing the number of required simulations without compromising the economic return in both nominal and probabilistic scenarios.

List of Figures

List of Tables

Table of Contents

1 Introduction

The development and management of oil field projects pose significant challenges, including high investments, extended payback periods, economic risks (Yanting and Liyun, 2011), uncertainties, along with a multitude of decision variables. Such projects require a thorough investigation conducted by a multidisciplinary team and the implementation of a systematic field development and management plan to improve decision making-process and profitability while mitigating risks stemming from high investments and uncertainties.

In an effort to enhance the decision-making process in petroleum field development and management, Schiozer et al. (2019) put forth a meticulously model-based methodology integrating reservoir simulation, risk analysis (Risso et al., 2011; Santos et al., 2017b; Santos et al., 2018a), history matching, uncertainty reduction, representative models, and production strategy selection under uncertainties. These steps are iteratively executed utilizing accumulated information over time in order to maximize the project's objective function in a process known as closed-loop field development and management (CLFDM).

The step of selecting a production strategy under uncertainties is important for maximizing the economic return and minimizing the risks (Santos et al., 2018b) of the project. To achieve an optimal strategy is necessary to optimize a multitude of variables. According to Gaspar et al. (2016) these variables can be separated into three distinct groups based on their impact on economic returns, the stage of the field's life cycle during which they are implemented, and the need for investment. These groups are named as follows: group 1 (G1) includes design variables, group 2 (G2) comprises control variables, and group 3 (G3) consists of revitalization variables.

The G1 variables relate to the pre-development implementation of the field's complete production system. They have a significant impact on the project's economic return and involve substantial investments. Examples of G1 variables include well number, type, placement, platform capacity, and the recovery method selected (e.g., water flooding, water alternating gas). The G2 variables encompass equipment operation specifications during field management and can be easily adjusted throughout the field's lifespan with minimal costs. They have a smaller, yet significant impact on the economic return compared to G1. The G2 variables can be further categorized into three subcategories as proposed by Schiozer et al. (2022): (1) life-cycle control rules (G2L) representing control operations over the field's entire life, (2)

closed-loop cycle control rules (G2C) referring to the control within the current closed-loop cycle, (3) short-term controls (G2S) involving real-time or short-term equipment control in the field. Examples of G2 variables include control of choke at both the platform and well region level, alongside WAG cycles. The G3 variables consist of the reconfiguration of the production system with additional investments during the field management. Their purpose is to provide future alternatives to G1 as the uncertainties decrease over the management period. Some examples include wells conversion, infill drilling and recompletion.

This work focuses on planning the G2L strategy at the beginning of field management, so costs incurred in the development stage cannot be reversed even if the G2L strategy suggests G1 modifications. Nevertheless, it is crucial to emphasize that a preliminary assessment of G2 should be conducted during the developmental phase. This is because it wields influence over field design variables (Li and Jafarpour, 2012; Zandvliet et al., 2008), and neglecting these interdependence could yield suboptimal outcomes.

The management of wellbore valves is typically executed indirectly by controlling the bottom-hole pressure (BHP) or flow rate overtime (Isebor, 2009; Isebor et al., 2014; Van Essen et al., 2009). This entails creating rules to govern the operational mode of the wells. These rules are generally based on monitoring variables (e.g., BHP, water cut, and oil rate) that guide or trigger specific actions in the operation mode of the valves. For instance, monitoring water cut (Wcut) allows for the well to be scheduled for shutdown when it reaches a predefined threshold.

These rules can be classified as reactive, proactive, or hybrid. The reactive approach (Barreto et al., 2010; Van Essen et al., 2009; Van Essen et al., 2011) consists of responding to undesired events, such as partially closing a valve after a specific volume of water is produced by the well. Notably, this type of control typically demands less computational effort compared to other methods, as it involves assessing fewer variables in most cases (Pinto et al., 2012).

In proactive control strategies (Bellout et al., 2012; Isebor, 2009; Isebor et al., 2014), preventive measures are taken to mitigate undesirable phenomena. For instance, one might reduce the injection rate of a well to impede the influx of water. The proactive approach demands increased computational effort when implemented over the field's lifespan. This is because, depending on the optimization procedure, a new variable is introduced for each valve and time period. Consequently, the optimization task becomes a product of the number of valves and its intervention periods.

Applying both reactive and proactive control extends to the hybrid approach. In this scenario, the optimization parameters comprise variables from both proactive and reactive methods. Consequently, this procedure requires a higher computational cost (Pinto, 2013), but it also encompasses more solutions that can yield better results if properly optimized. To alleviate computational challenges, one can optimize in stages, with proactive control being initially implemented, followed by reactive control, or vice versa (Santos, 2017).

Ebadi and Davies (2006) presented a study that underscores all three mentioned control types. In the reactive strategy, they partially close the valve upon reaching a specific Wcut threshold. The proactive approach involves preemptively reducing the valve opening before water breakthrough. Finally, they implement the hybrid approach, using proactive control initially, and if there's an excess of Wcut, applying the reactive method to further diminish the valve orifice.

The G2L process is generally quite costly, and measures to expedite it can contribute to the production strategy selection stage and the overall CLFDM. This computational effort is attributed to various factors, such as the multitude of variables involved, numerous uncertainties (e.g., geological, operational, and economic), the use of numerical reservoir simulation, and the complex nature of oil fluids as a mixture of reservoir hydrocarbons.

Uncertainties present additional challenge in the selection of the production strategy because they demand the evaluation of numerous scenarios. Nevertheless, it is imperative to consider these uncertainties to minimize risk (Schiozer et al., 2004). Failing to address them adequately during the determination of the optimal strategy can result in a production plan that may appear optimal in simulation models but fail to deliver the expected results when applied in real-world field operations (Botechia et al., 2018).

In general, the incorporation of uncertainties follows a systematic process. Initially, relevant uncertainties (e.g., geological, fluid properties and relative permeability) are identified and assessed during the reservoir characterization. These uncertainties are then combined using sampling techniques to generate a suitable number of scenarios for numerical reservoir simulation. Subsequently, data assimilation methods are employed to reduce uncertainties, resulting in a set of simulation models for production strategies optimization. The reservoir characterization, data assimilation, and production strategy optimization process can be revisited using a closed-loop approach as additional data, such as dynamic and seismic data, becomes available over time.

Two most common approaches are employed in the pursuit of developing optimal production strategies under uncertainties. The first approach, known as ensemble nominal optimization, involves optimizing deterministic objective function(s) for each individual model derived from the data assimilation process. This generates multiple production strategies, each one corresponding to a specific model. The second approach, referred to as robust optimization, focuses on optimizing probabilistic objective function(s) across the entire set of data assimilated models. This method aims to create a single production strategy that exhibits the best average performance considering the entire ensemble of models.

Optimizing the production strategy through ensemble nominal optimization or robust optimization approaches can be challenging and resource-intensive when dealing with a large number of scenarios derived from the data assimilation process and optimizing a large set of variables. To address this issue, Meira et al. (2017) introduced the RMFinder 2.0 tool that utilizes a mathematical function to identify representative models (RMs) within the entire ensemble. These RMs capture the probability distribution of input variables, such as reservoir and operational uncertainties, as well as the variability of important output variables like production and injection forecasts. By working with RMs, both RM robust optimization and RM nominal optimization can be performed, eliminating the need to optimize the ensemble of models selected during the data assimilation process. Subsequently, the optimal strategies derived from RMs are often assessed within the ensemble of models before being implemented in the real field to ensure a more accurate evaluation of the risk curve.

As mentioned earlier, another significant factor contributing to the high computational effort in the strategy selection process is the numerical reservoir simulation. This process involves solving several finite difference equations in the reservoir model grid cell at each time step, adding to the resource demands. However, it is a necessary tool as it allows for a more precise estimation of the behavior of pressures, saturations, compositions, and productions in a hydrocarbon reservoir over time (Mello, 2015). In order to mitigate the computational burden, the commonly employed approach is to utilize the Black-Oil model, which simplifies the composition of reservoir fluids into three primary phases: water, oil, and gas. Nonetheless, advanced compositional modeling is necessary to accurately represent the fluid behavior in complex scenarios characterized by light oil, high gas content, water alternating gas (WAG) injection as recovery mechanism and fully gas reinjection, as observed in certain Brazilian pre-salt fields. This enhanced accuracy comes at the expense of increased computational demands in reservoir simulations, as a higher number of pseudo-components are required to precisely

describe the fluids present in the field (Schlijper, 1986). These demanding reservoir simulations are performed thousands of times during the optimization process due to the large number of variables to be optimized, uncertainties to be incorporated into the reservoir models, and the iterative nature of the closed-loop process.

Machine learning (ML), which has been widely applied in the oil industry and other engineering fields over the last few decades (Tariq et al., 2021), can be an ally in easing computational efforts. Overall, machine learning revolves around the use of algorithms and statistical models to enable computers to learn from data and make predictions or decisions without the need for explicit programming for each specific task. One advantage of ML lies in its ability to swiftly process vast amounts of data, identifying trends and patterns with good precision at a much faster pace than traditional physics-based simulation models.

Strategy optimization problems present other challenges beyond computational effort, especially related to G2L parameterization, which is the focus of this thesis. One primary issue arises from the vast number of optimization variables, leading the algorithm to prematurely converge towards suboptimal solutions. An example of such a parameterization is the stepwise approach, commonly found in literature (Bellout et al., 2012; Forouzanfar et al., 2015; Humphries et al., 2014; Wang et al., 2019). In this approach, the optimization variables consist of the bottom-hole pressure (BHP) or flow rate values at each time interval and well throughout the field's lifespan. While this method works well for simple models with a limited number of wells and intervals, it faces significant challenges when applied to real-world projects as the number of variables grow significantly.

Another concern with an inappropriate G2L parameterization is the potential generation of strategies with adverse effects from a production-engineering perspective. For instance, strategies involving repeated large changes in BHP can lead to equipment and formation damage (Sorek et al., 2017). Moreover, the parameterization approach may produce highly complex solutions, making their implementation in actual field operations significantly challenging.

In this study we investigate and propose methodologies to expedite the optimization process of G2L variables under different recovery methods, aiming to generate feasible solutions for real-world applications while maintaining or enhancing the economic return compared to traditional approaches.

1.1 Motivation

The computational burden associated with each step of the oil field development and management plan can hinder the decision-making process, leading to delays and potentially requiring simplifications or even omitting certain steps. These constraints can ultimately have negative impacts on production levels and economic returns. Therefore, methodologies that effectively reduce the execution time for critical steps, including reservoir characterization, data assimilation modeling, and production strategy selection, are highly valued.

There are several research endeavors in making simplifications that ultimately help to expedite the production strategy selection process. One such approach involves replacing high-fidelity models with low-fidelity models that feature a reduced number of grid blocks during the reservoir characterization step. Another approach involves simplifying fluid characterization using a Black-Oil model, which requires solving fewer equations per block and per time step. Also, many studies developed techniques to select representative models specifically for the optimization of production strategies, rather than employing the entire set of data assimilated models.

Despite the use of various techniques, production strategy optimization can still be costly, particularly when multiple scenarios are required to accurately represent the uncertainties and when numerous variables need to be optimized. Thus, there is room for the development of procedures that can complement the aforementioned techniques and further expedite the production strategy selection process and subsequent decision-making. One opportunity for improvement lies in enhancing the parameterization of optimization variables to reduce their number. Another approach involves leveraging machine learning techniques to predict production behavior, which can serve as a valuable tool for reducing the reliance on extensive simulations.

Aside from reducing computational time, it is essential that the method used to select the production strategy yield results of equivalent or superior quality compared to traditional approaches. Furthermore, a common limitation observed in many optimization approaches discussed in the literature is their tendency to generate strategies, particularly for life-cycle control rules (G2L), that prove excessively complex to implement or unfavorable from a production-engineering standpoint.

In essence, finding the ideal balance between developing a well-designed G2L strategy and ensuring manageable computational effort remains a compelling topic.

1.2 Objectives

The primary goal of this work is to propose procedures that expedite the optimization process of wells' life cycle control variables (G2L) in studying cases involving water alternating gas injection and water flooding recovery methods. These procedures have been meticulously designed to address two key challenges that hinder the practical application of methodologies in real-world scenarios. Firstly, they aim to significantly reduce computational time, thereby increasing the likelihood of successful project implementation within the specified timeline. Secondly, the procedures aim to generate well-behaved and practical-to-implement G2L strategies, overcoming the complex strategies derived from traditional approaches in the literature.

The reduction of computational effort while providing practical solutions is achieved through two approaches:

1. Reducing the number of variables to be optimized and consequently narrowing down the search space by utilizing parametric equations designed to produce solutions that closely align with the practical operational behavior of the wells.
2. Replacing a portion of time-consuming full physical simulations by incorporating low-cost machine learning models into the optimization process.

Most of the proposed procedures are validated under geological uncertainties, acknowledging the inherent uncertainty in real-world scenarios. Validating the procedures under geological uncertainties ensures their applicability and robustness, enabling more accurate and trustworthy results.

1.3 Description of thesis

The thesis comprises four scientific articles, each one contributing to the overall objectives of this thesis. In this subsection, we provide a summary of these studies and their relevance. The subsequent chapters delve into the detailed presentation of them.

1.3.1 Investigation of Well Control Parameterization with Reduced Number of Variables Under Reservoir Uncertainties (published work)

Santos, D. R., Fioravanti, A. R., Santos, A. A. S., Schiozer, D. J. Investigation of Well Control Parameterization with Reduced Number of Variables Under Reservoir Uncertainties.

Presented at the SPE Europec featured at 82nd EAGE Conference and Exhibition held in Amsterdam, The Netherlands, October. 2021. https://doi.org/10.2118/205207-MS.

This study aims to overcome the limitations of traditional approaches in optimizing well control life-cycle variables (G2L) in oil reservoirs under waterflooding recovery. The conventional approach, which entails determining optimal values of bottom-hole pressure (BHP) or flow rates at each time step, encounters challenges in real-world scenarios. These challenges include impractical changes in pressure and production curves that can lead to equipment and formation damage. Additionally, the high number of variables in non-convex optimization problems can cause algorithms convergence issues.

We explored four G2L parameterizations with a reduced number of decision variables to maximize the net present value (NPV) of the field. The first parameterization specifically focuses on determining the optimal well rates and shut-in time. The remaining three parameterizations involve defining the bottom-hole pressure (BHP) curve over time using first-order and second-order polynomials, as well as a logistic equation. We then optimize the coefficients of these equations, resulting in a significant reduction in the number of decision variables from thousands to fewer than a hundred. Next, we compared each parameterization with a well control short-term strategy that prioritizes production in wells with a higher oil-water ratio and aims to replicate the general industry practice. The subsequent analysis involved applying the best parameterization to select the G2L strategy under reservoir uncertainties and on a reference model.

The results revealed that the logistic equation surpassed other parameterizations in generating higher NPV and ensuring a smooth well production curve. Furthermore, this parameterization significantly outperformed the well control short-term strategy when subjected to uncertainties and applied to the reference model.

The results are consistent with the objective of the thesis of proposing methods that reduce computational effort and are applicable in practical scenarios. The proposed solution, which reduces the number of variables and computational burden, generates smooth BHP curves and achieves promising results under uncertainties and in the reference case.

Moreover, this study serves as the foundation for the subsequent articles of this thesis, which either utilize the proposed G2L parameterization or incorporate modified parametric equations.

1.3.2 A Machine Learning Approach to Reduce the Number of Simulations for Long-term Well Control Optimization (published work)

Santos, D. R., Fioravanti, A. R., Santos, A. A. S., Schiozer, D. J. A Machine Learning Approach to Reduce the Number of Simulations for Long-Term Well Control Optimization. Presented at the SPE Annual Technical Conference and Exhibition, Virtual, October. 2020. https://doi.org/10.2118/201379-MS.

This article aims to improve upon the work presented in the previous paper by incorporating machine learning algorithms to reduce the number of simulations needed to achieve optimal solutions. The proposed methodology, known as IDLHC-ML, combines several machine learning techniques with iterative discrete Latin hypercube (IDLHC) to optimize the coefficients of the logistic equation that governs the well's bottom-hole pressure over time in a deterministic study case.

The IDLHC is an iterative process that involves discretizing the optimization variables into levels, each associated with a corresponding probability mass function (PMF). The Discrete Latin Hypercube (Maschio and Schiozer, 2016) sample technique is then utilized to generate a set of strategies by combining the discretized variables based on the prior PMF. These strategies are subsequently simulated, and the best strategies are selected based on the objective function to update the PMFs of each variable. This iterative process continues until convergence is attained.

The IDLHC-ML utilizes the strategies generated in each iteration of IDLHC to train machine learning models, which are then employed to predict the best strategies for the next iteration. As a result, instead of simulating all the strategies generated by the DLHC, only a fraction that is predicted by the ML model to have the highest economic return is simulated.

This approach efficiently reduces the total number of expensive full-physics reservoir simulations, owing to the utilization of fast and cost-effective ML models. In addition to reducing computational effort, this paper also assesses the performance of different machine learning algorithms, explores various input variables for training these algorithms, and investigates different configurations of IDLHC parameters.

1.3.3 Optimizing Well Control Strategies with IDLHC-MLR: A Machine Learning Approach to Address Geological Uncertainties and Reduce Simulations (published work)

Santos, D. R., Fioravanti, A. R., Botechia, V. E., Schiozer, D. J. Optimizing Well Control Strategies with IDLHC-MLR: A Machine Learning Approach to Address Geological Uncertainties and Reduce Simulations. Presented at the Offshore Technology Conference Brasil, October. 2023.

The IDLHC-ML was originally developed and tested without considering uncertainties. However, it remains essential to evaluate its effectiveness when uncertainties are considered.

In the third article, we present the IDLHC-MLR as an enhanced version of IDLHC-ML that addresses this limitation by integrating uncertainties through the utilization of RMs. To accomplish this, we modified the objective function and target variables used in training the machine learning models. The IDLHC-MLR robustly optimizes the coefficients of the logistic equation to accurately determine the BHP over the field's lifespan.

The results demonstrate that the IDLHC-MLR exhibits efficiency in handling uncertainties, yielding a reduction of 4050 simulations (equivalent to a 45% less) compared to the traditional IDLHC model, while maintaining a similar expected monetary value.

In conclusion, recognizing and accounting for uncertainties in real-world problems play a critical role in mitigating costly errors and accurately evaluating project economic returns. In this regard, the IDLHC-MLR assumes significant importance for this thesis as it enhances the practicality of the IDLHC-ML algorithm by incorporating uncertainties.

1.3.4 Accelerated optimization of CO2-miscible water-alternating-gas injection in carbonate reservoirs using production data-based parameterization (published work)

Santos, D. R., Fioravanti, A. R., Botechia, V. E., Schiozer, D. J. Accelerated optimization of CO2-miscible water-alternating-gas injection in carbonate reservoirs using production data-based parameterization. Journal of Petroleum Exploration and Production Technology, 13, 1833–1846. 2023. https://doi.org/10.1007/s13202-023-01643-0.

The last study expands on the concept of parametric equations introduced in the first paper, with a specific focus on optimizing miscible water alternating gas (WAG) injection in

reservoirs containing light oil and high gas content. Traditionally, this type of recovery method requires computationally intensive compositional simulation.

The main objective of this work is to enhance the convergence speed of the optimization algorithm by introducing a novel parametric equation for WAG injection. The equation defines a dynamic priority rank for each well over time, determining the gas or water injection schedule throughout the entire life cycle of the field. The calculation of the priority rank incorporates reservoir production data, including water cut (W_{CUT}), gas-oil ratio (GOR), and cumulative gas production from the wells. The equation terms have been meticulously designed to induce desirable effects on production and WAG profiles, ensuring the creation of well-behaved WAG strategies suitable for real-world applications. Additionally, the approach emphasizes flexibility by accommodating individual WAG profiles for each injector, regardless of cyclic patterns. This flexibility enables the exploration of diverse solutions that can yield improved NPV.

To assess the effectiveness of the proposed procedure, two alternative approaches were employed for comparison: a benchmark solution that optimizes the injected fluid for each well over time, and a traditional baseline strategy with fixed six-month WAG cycles. The proposed method effectively reduces computational requirements while ensuring consistent patterns across injectors. This feature is crucial when designing practical WAG strategies.

2 Investigation of Well Control Parameterization with Reduced Number of Variables Under Reservoir Uncertainties

Authors: Santos D. R., Fioravanti, A. R., Santos, A. A. S., Schiozer, D. J.
Presented at the SPE Europec featured at 82nd EAGE Conference and Exhibition held in Amsterdam, The Netherlands, October, 2023.
(https://doi.org/10.2118/205207-MS)

Abstract

Although several studies have shown that life-cycle well control strategies can significantly improve a field's economic return, the industry often relies on short-term strategies. One drawback of traditional parameterization, adopted for well control life-cycle numerical optimization, is that it often generates control strategies that yield impractical abrupt changes in production curves. Another issue, especially in cases with a large number of decision variables, is the local optima convergence related to the non-convex optimization problems. In this context, we proposed and compared four life-cycle well control parameterizations to maximize the net present value (NPV) of the field under uncertainties, which are able to mitigate both the above-mentioned problems.

The first parameterization is applied to optimize the apportionment of well rates at the beginning of the field management and well shut-in time. The other three are based on optimizing the coefficients of parametric equations (first-and second-order polynomials, and logistic equation) to guide the bottom-hole pressure (BHP) over time. We executed each parameterization five times in a deterministic reservoir scenario and compared them with well control short-term strategy that prioritizes production in wells with higher oil-water ratio and that aimed to replicate the general industry practice. In this strategy, the wells' priority rank was updated at every 30-simulation days. Subsequently, the best parameterization was used to select the well control life-cycle strategy under reservoir uncertainties and this strategy was applied to the reference model representing a real reservoir.

The results showed that all the proposed parametrizations significantly improved the NPV in comparison to the well control short-term strategy, while simultaneously ensuring a

smooth well production curve. The logistic equation presented the best result among all parameterizations, as it delivered both the highest average of NPV and the smallest dispersion over the five experiment repetitions. This parameterization also produced similar results when applied under uncertainties and for the reference model. These results endorse the importance of not only relying on a short-term strategy, but also planning it for the life-cycle.

2.1 Introduction

A broad study is necessary to mitigate the risks in high investment oil projects. This type of study involves reservoir characterization, production data assimilation, and selecting the production strategy under uncertainty (Schiozer et. al, 2015). In this context, a high-quality production strategy plays a key role as it directly affects the oil recovery factor and economic return of the project. Such a strategy is considerably influenced by uncertainties (e.g.: reservoir, economical and operational), which should be considered when selecting a production strategy.

Specifically, there are several reservoir uncertainties to represent a real field; thus, a large number of reservoir scenarios should be analyzed to define a suitable production strategy. Since this analysis demands high computational effort, many authors (e.g.: Meira et al., 2017; Meira et al., 2020; Sarma et al., 2013; Shirangi and Durlofsky, 2016) suggest developing the production strategy considering representative models (RM), which aims to typify the reservoir scenarios that honored the field's observed data. Each step of this process – representing the real field through production data assimilated reservoir models and then, characterizing these models with RM – may carry over bias. Hence, it is particularly important to evaluate the strategy in the set of history-matched reservoir models and, for the purpose of this study, test within the reservoir model, which represents the real reservoir.

Apart from the reservoir uncertainties, the complexity of defining the appropriate production strategy also incorporates the high number of optimization variables. To facilitate this process, we can divide these variables into three groups – according to their investments and the period of the field when they are implemented – and optimize them separately. These groups are denoted here as design variables (G1), control variables (G2), and revitalization variables (G3) – see Gaspar et al., 2016b for further details. The study of G1 and G3 are out of the scope of this work and we only focus on the G2 strategy selection.

The G2 represent the equipment operations (e.g.: control valves choke in platform, well or region level) over time using field's life-cycle control rules (G2L) and short-term controls

(G2S) where the operations can be changed any time during production at no significant costs. A suitable G2L strategy may have a significant influence in the life-cycle economic return of the project. However, according to van Essen et al. (2011), the industry often relies only on G2S strategies. These strategies usually prioritize production in wells according to a short-term control rule (e.g.: wells with higher oil production or with higher oil-water ratio). One possible explanation for industries to prioritize G2S strategies application in real field could be the difficulty in performing G2L strategies generated by the traditional approach of controlling the opening fraction of well valve chokes indirectly by BHP or rates – also called stepwise method – at each time interval (see examples in Bellout et al., 2012; Forouzanfar et al., 2015; Humphries et al., 2014; Wang et al., 2019). The difficulty arises from the fact that this approach often provides G2L strategies with repeated large changes in BHP or rates and, in general, it is unfavorable from a production-engineering perspective as this might cause damage to the equipment and formation (Sorek et al., 2017).

Generally, adopting gradient-free algorithms to optimize G2L (broadly applied in literature – see Echeverría et al., 2011; Han et al., 2021; Isebor et al., 2014) using the stepwise approach provides good results for the OF in projects with few variables (as shown in Awotunde, 2014). Conversely, controlling the BHP along time using the stepwise method in real projects often leads to more than a thousand decision variables. Thus, another drawback from the traditional approaches evaluated in reservoir simulators is related to the high-dimensional search space – given by the number of wells multiplied by the number of time intervals – in non-convex optimization problems, which may produce premature convergence of gradient-free optimization algorithms towards local minima (as verified in Pinto et al., 2019a, and Sorek et al., 2017; for some related problems).

The aim of this work is to propose and compare different G2L parameterizations able to overcome the problems just mentioned – high variation in well rates and BHP along time, and gradient-free algorithm optimization convergence to a poor solution – while generating a strategy that is considerably superior to the industry regular G2S strategy when performed in the reference model.

In this work, we investigated four G2L parameterizations with a reduced number of decision variables to maximize the economic return in a field waterflooding project. The first parameterization aimed to find the best well rates apportionment at the start of the field management timeline and well shut-in time, simultaneously. This approach is very similar to

the one applied by Santos and Schiozer (2017), but with two main differences: (1) the authors first optimized the well rates and then the well shut-in time in a sequential process and (2) a different reservoir benchmark case was adopted in that work. We chose to replicate this parameterization because it provided good results in their experiment. Therefore, we can compare this parameterization to the others proposed in this work. The three remaining parameterizations consisted in representing the BHP trajectories over time using: (1) first-and (2) second-order polynomials, and (3) logistic equation. To define these trajectories, the goal was to optimize the coefficients of those three equations instead of directly optimizing the BHP at each time interval, as is done in the traditional methods. Through the mathematical models, we substantially reduced the number of decision variables of our case from thousands to less than a hundred. It is particularly noteworthy that in these three parameterizations, the wells were controlled in two stages, which prioritize G2L and adopted the G2S control whenever the problems constraints were reached (e.g.: wells BHP limits, maximum rates for wells and platform).

Finally, we applied the best parameterization investigated to perform a nominal optimization for each RM of the field, i.e., we optimize each RM individually yielding one optimum strategy (or specialized strategy) per RM. The specialized strategies were then evaluated in a large set of scenarios that honored the field production data history and the one that provided the highest economic return was applied and compared to the G2S industry strategy applied in the reference model. It is noteworthy that evaluating the proposed method in the reference and production data assimilated models provides more reliability that the G2L parameterization proposed would be suitable when applied to a real field.

We have organized the rest of this paper in the following manner. In the next subsections, we review the literature of the G2L optimization based on parametric equations, define the G2S employed in this work to replicate industry practices, briefly describe the first parameterization, which was applied in a similar way to Santos and Schiozer (2017), and summarize how the gradient-free algorithm and the method to select the RMs selected here works. The subsequent section of this paper describes the methodology of this work, explaining in detail the three remaining parameterizations proposed for G2L. After, we briefly introduce the case study of this work and present the economical parameters, operational constraints, and important dates to represent the field timeline and cash flow update. Next, we compare the results of the G2L parameterization and the G2S strategies and present the efficiency of our best method under uncertainties and for the reference model. In the succeeding section, we

discuss our findings and compare them with existing literature. Finally, we conclude the paper and discuss future research directions.

2.1.1 Life-cycle well control rule (G2L) optimization based on parametric equations

Few researchers (e.g.: Awotunde, 2014; Pinto et al., 2019a; Sorek et al., 2017) have addressed the G2L optimization issue by using parametric equations to define the well BHP or rate trajectories over time. Sorek et al. (2017) investigated the BHP control over time using first, second, third and quartic polynomial trajectories and compared them to the stepwise approach that optimizes the BHP at each time interval. They observed that the higher the polynomial order, the higher the value of the objective function, but also with an increment of the number of simulations up to convergence. In addition, as the problem becomes more complex, the stepwise formulation containing 1500 decision variables did not converge within their stop criteria. Awotunde (2014) controlled the well rate trajectories using first, cubic and fifth order polynomial equation, apart from the cosine function, and compared these to the stepwise method. For smaller problems, the stepwise turned out to be the best approach; however, as the problem increased, the first and cosine formulation outperformed the stepwise method. The author verified that, contrary to Sorek et al. (2017), higher objective function values were obtained using lower polynomial degrees. Both aforementioned studies do not consider reservoir uncertainties in their analysis, despite the importance of including it to generate high-quality production strategies for oil projects (Meira et al., 2020, and Rahim and Li, 2015). Thus, there is no assurance that the parametric equations proposed by Sorek et al. (2017) and Awotunde (2014) would provide good solutions when considering distinct scenarios of the same field. In this regard, it is imperative to test the G2L parameterization in such a situation.

Pinto et al. (2019a) performed the robust optimization of G2L based on polynomial equation to guide BHP trajectories over time considering uncertainties. They first optimized the G2L into a group of five RM selected among 50 scenarios and, then, evaluated the best strategy among the larger group. The G2L strategy applied to the 50 scenarios reasonably improved the expected monetary value (EMV) compared to a reactive strategy (roughly 4%), which shut producers when water cut (W_{CUT}) for that well becomes uneconomic. The authors did not implement the final G2L strategy in any reference model.

2.1.2 Well control in the short-term (G2S)

The well short-term control (G2S) adopted in this work, which aims to represent industry practices, consists of a reactive control rule to define the wells rate apportionment to meet the target group rate accordingly to a general priority rank described by **Equation 2.1**. This rule is set using the INGUIDE keyword (see Pinto et al., 2019b for more details) from the commercial simulator software IMEX 2016.10.

$$Priority(we) = \frac{A_0(ig) + \sum_{i=1}^{nph} A_i(ig) \times q_i(we)}{B_0(ig) + \sum_{i=1}^{nph} B_i(ig) \times q_i(we)} \qquad \text{Equation 2.1}$$

where we is the index for each well contributing to a target group ig, i is the index of a specific stream, nph is the number of fluids to be included in the formula, q_i is the stream rate under the most restrictive reservoir and operational constraint or in the user predefined bottom-hole pressure (BHP), A_0 and B_0 are, respectively, the numerator and denominator constants for group ig, and A_i and B_i are the numerator and denominator of the stream i in the order given.

In this equation the higher the $Priority(we)$, the proportionally higher the rate of a specific well will be in relation to other wells from the same group. The coefficients A_0, A_i, B_0, and B_i are set by the simulator default or can be defined by the user.

It is important to mention that, if any constraint is violated, the well or group will be left out of the apportionment with the rate being reduce to the maximum rate under the violated constraint. Moreover, INGUIDE is activated only when a group of wells or platform operates under its maximum rate capacity.

2.1.3 PGUIDE parameterization

PGUIDE G2L parameterization combines the proactive control rule to determine the producers' and injectors' apportionment liquid rates in the beginning of the field management with the reactive rule to define well optimum shut-in time (check Santos and Schiozer, 2017 for further information). These rules were optimized simultaneously in this work. The rate distribution among wells is done through the GUIDE keyword from IMEX 2016.10 commercial software. This key word specifies user-supplied weights for each well (p_{i_we}) that is included to stipulate the fraction of liquid in relation to a target group ig according to **Equation 2.2**. Thus, changing the weights associated to each well also changes their rates for the stream i (q_{i_we}).

$$q_{i\_we} = \frac{p_{i\_we}}{\sum_{we=1}^{n_{we}} p_{i\_we}} \times q_{i\_ig\_target} \qquad \text{Equation 2.2}$$

where n_{we} denotes the total number of wells in the group ig, and $q_{i\_ig\_target}$ is the rate restriction for the target group (ig) considering the stream i. In this work the stream considered for the producers and injectors were the produced liquid rate ($q_o + q_w$) and the water injection rate (q_{winj}).

Similar to INGUIDE, GUIDE obeys the reservoir, wells and platform constraints and it is activated when a group of wells or platform operates under its maximum rates capacity.

To define optimum shut-in time for wells, we applied a "IF ... ELSE" condition. Producers are shut whenever the W_{CUT_WE} is greater than $W_{CUT\_LIMIT\_WE}$, where W_{CUT_WE} is the water cut for the producer we at a given moment, and $W_{CUT\_LIMIT\_WE}$ is the limit value to shut the producer we. Injectors are shut whenever the cumulative water production of the field (W_p) is greater than $W_{p\_LIMIT\_WE}$, which represents the cumulative water produced by the field to shut the injector we. Note that each producer and injector have its own value of $W_{CUT\_LIMIT\_WE}$ and $W_{p\_LIMIT\_WE}$ limit to be shut.

2.1.4 Designed Exploration and Controlled Evolution (DECE) algorithm

In this paper, we optimized the decision control variables of each G2L parameterization with the Designed Exploration and Controlled Evolution (DECE) gradient-free algorithm. This algorithm iteratively performs two stages in series: Designed Exploration and Controlled Evolution. The Designed Exploration is first applied to random explore the search space in order to achieved as much information about the solution space. To this end, it is employed a Tabu search (Glover, 1990) and experimental design (Yang et al., 2007) techniques. Subsequently, the Controlled Evolution stage is applied, which consists of statistically analyzing the data from the simulations performed in the previous stage in order to verify whether, by rejecting a candidate value of a given parameter is there a better chance of improving the solution. To reduce the chance of being trapped in local minima, the rejected candidate values are check from time to time to make sure if their rejection remains valid or if they should be reused during optimization.

It is important to mention that DECE is a commercial optimization algorithm and its mathematical formulation is not publicly available. Apart from that, we choose this algorithm as it has provide good results in similar optimization problems (check: Gaspar et al., 2016b;

Pinto et al., 2018; Santos and Schiozer, 2017; Schiozer et al., 2019; Sidahmed et al., 2019; Yang et al., 2007).

2.1.5 Representative model selection general overview

To select the representative models (RMs) of the data assimilated scenarios, we applied the RMFinder 2.0 method proposed by Meira et al. (2017). The RMFinder 2.0 consists of defining and describing mathematically the representativeness concept of a given subset of scenarios. Subsequently, this mathematical function is optimized in order to find the RMs. The representativeness concept is based on three criteria. The first and the second consist of keeping the main input variables well distributed in their cross plots and in the associated risk curves, respectively. These main input variables (such as NPV, cumulative oil and water production) are defined and weighted by the user. The third criterion tries to preserve the proportion of the levels of uncertain attributes in the RMs as closest as possible to the proportion observed in the data assimilation models. Additionally, the mathematical formulation of RMFinder 2.0 penalizes sets of RMs that lack at least one instance for each level of every attribute. Finally, the method considers a different probability for each RMs delivering more accurate risk curve for the RMs.

We chose this approach because it yielded appropriate set of RMs for different cases of studies (including a high fractures pre-salt carbonate reservoir and a highly heterogeneous heavy oil offshore field). Note that the RMFinder 2.0 mathematical equation description is beyond the purpose of this work, thus the keen reader can check it directly at Meira et al. (2017).

2.2 Methodology

This section of the paper describes the methodology of this work, which can be divided in two parts: validation and application. In validation, we compared four proposed parameterizations for life-cycle well control rule (G2L) optimization and a regular short-term well control (G2S) that is normally applied in the industry for the nominal reservoir case. In application, we selected the parametrization that yielded the highest outcome in the validation part and investigated its capability to be applied under reservoir uncertainties. Moreover, we evaluated its performance in a benchmark for the real reservoir.

2.2.1 Validation part

As PGUIDE was presented before, we focus here on explaining the three remaining parameterizations proposed in this work, which consist of a proactive control rule for the field's life-cycle formulated through an equation that determines the well BHP limit along the field management. The goal was to define the coefficients of each equation that maximizes the net present value (NPV). The equations allow us to reduce significantly the number of parameters, while generating a smooth and executable BHP control for real cases.

In this paper, each parameterization was executed five times to obtain statistically meaningful interpretations regarding their robustness because different results for the objective function may be achieved every time DECE gradient-free algorithm is performed.

2.2.1.1 PLOG parameterization

PLOG parameterization resides in optimizing the coefficients a_{we}, b_{we} and c_{we} of logistic equation (**Equation 2.3**).

$$BHP_{we,t} = BHP_{Min} + \frac{(BHP_{Max} - BHP_{Min})}{1+e^{\left(a_{we} + b_{we} \times \left(\frac{t}{t_{final}}\right) + c_{we} \times \left(\frac{t}{t_{final}}\right)^2\right)}} \qquad \text{Equation 2.3}$$

where $BHP_{we,t}$ describes the BHP in the well we for time t, BHP_{Max} and BHP_{Min} represent the maximum and minimum BHP that each well can operate, respectively, and t_{final} denotes the whole management period.

The limits of the coefficients of the equations (a_{we_limits}, b_{we_limits}, c_{we_limits}) were chosen first by setting the a_{we_limits} to guarantee by itself that the $BHP_{we,t}$ can vary within the well BHP constraints over time. The b_{we_limits} and c_{we_limits} were defined to have a similar influence in the BHP curve than a_{we_limits}. The approach to do this was to ensure that all coefficients in their limits have the same integral along the normalized time ($t_{norm} = t/t_{final}$), which varies between [0, 1], as following: $b_{we\_limit} = \frac{a_{we\_limit}}{\int_0^1 t_{norm}\, dt_{norm}}$, and $c_{we\_limit} = \frac{a_{we\_limit}}{\int_0^1 {t_{norm}}^2\, dt_{norm}}$.

An interesting propriety of logistic equation is that there is no need to define hard bounds for the well BHP limit as – due to its formulation – the BHP is automatically kept in between the maximum and minimum reservoir pressure. In this sense, the logistic equation

avoids subtle changes when BHP reaches reservoir pressure constraint. Additional reasons for applying the logistic equation in this work are presented in **Appendix A.1**.

2.2.1.2 P_{LIN} parameterization

P_{LIN} has two coefficients to be defined, a_{we} and b_{we}. **Equation 2.4** describes BHP over time for producers and injectors, respectively. Differently from P_{LOG}, it is necessary to set hard bounds defined by Min and Max terms in the linear equation to avoid the violation of reservoir pressure constraints.

$$BHP_{we,t} = Min\left(BHP_{Max}, Max\left(BHP_{Min}, a_{we} + b_{we} \times \frac{t}{t_{final}}\right)\right) \qquad \text{Equation 2.4}$$

2.2.1.3 P_{SEC} parameterization

P_{SEC} is also a polynomial equation parameterization. However, P_{SEC} correspond to the second order equation and encompass the linear equation while add the non-linear term represented by the c_{we} coefficient. Similar to P_{LIN}, it is required to set hard bounds for the **Equation 2.5**.

$$\mathrm{BHP}_{we,t} = \mathrm{Min}\left(\mathrm{BHP}_{\mathrm{Max}}, \mathrm{Max}\left(\mathrm{BHP}_{\mathrm{Min}}, \mathrm{a_{we}} + \mathrm{b_{we}} \times \frac{\mathrm{t}}{\mathrm{t_{final}}} + \mathrm{c_{we}} \times \left(\frac{\mathrm{t}}{\mathrm{t_{final}}}\right)^2\right)\right)$$

Equation 2.5

One may inquire about the rationale behind not testing polynomials of higher orders. This matter is elucidated in the thesis **Appendix A.2**.

2.3 Application part

The parameterization, which provided the best NPV in the previous section, was addressed under uncertainties in the application. A large computational effort would be required to create one strategy for each scenario that honored the dynamic data from the field's historical period (filtered scenarios). Therefore, we used representative models (RM) to typify the group of filtered scenarios and performed a nominal optimization for each RM. After each strategy was applied to all filtered scenarios and the more robust specialized strategy in terms of expected monetary value (EMV) was selected, we investigated its performance in the reference model.

2.3.1 Reservoir simulation case

In this section, we briefly describe the UNISIM-I-R model and UNISIM-I-M benchmark case (see Avansi and Schiozer, 2015, and Gaspar et al., 2016a, respectively) employed to evaluate our methodology. Also, we present important dates, economical parameters, operational constraints, and the parameters of the control rules.

The UNISIM-I-R is a high-resolution simulation model – approximately 3.5 million active cells in a grid of 326 × 234 × 157 blocks – built considering petrophysical, facies and structural public data from the Namorado Field located in the Campos Basin, Brazil. The UNISIM-I-R serves as a benchmark to test the applicability of different methodologies before applying it to a real reservoir.

The UNISIM-I-M is a simulation model based on information from UNISIM-I-R and consists in a synthetic, three-dimensional, Black-oil reservoir, including 36,739 active cells designed to optimize control (G2) and revitalization (G3) variables. The case has a history period of 7 years from 14 producer and 11 injector wells, and a management period of almost 23 years. Moreover, the case considers an exploitation strategy (G1), already implanted by Avansi and Schiozer (2015) during the development phase, which must not be altered. **Figure 2.1** presents the wells placement for this strategy.

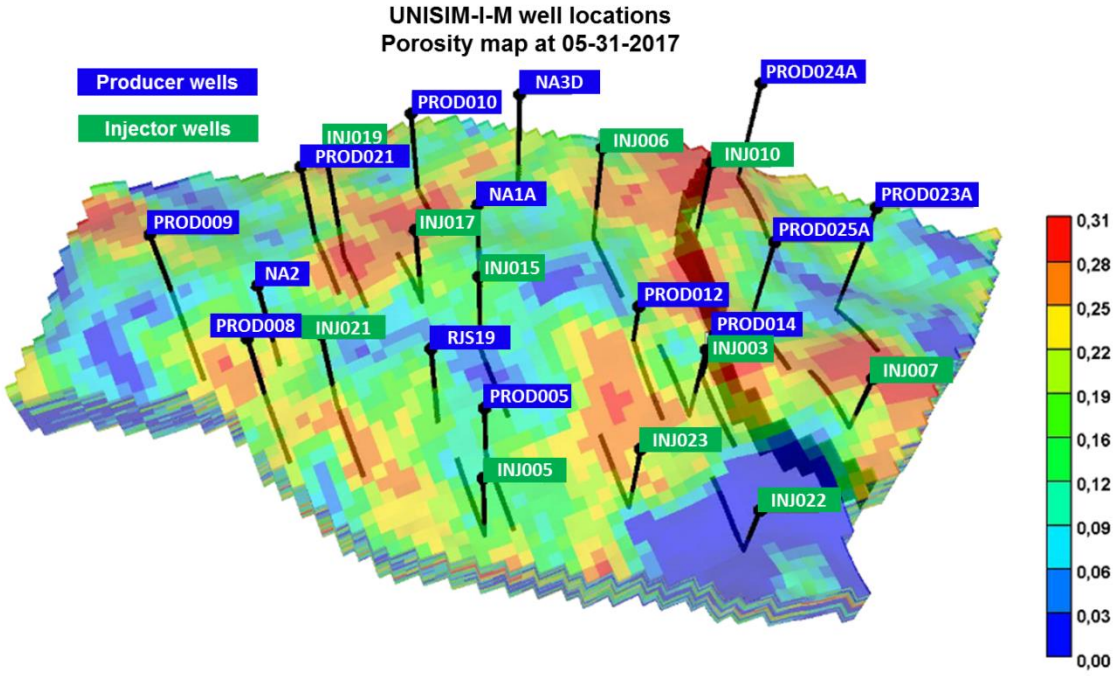

Figure 2.1–Map porosity at 05-31-2017 with the well placement strategy defined in the development phase for UNISIM-I-M.

To account for the reservoir uncertainties of the reference model (UNISIM-I-R), we evaluated our methodology in 48 scenarios selected by Gaspar et al. (2016a) obtained through the data assimilation process. From these scenarios, nine representative models (RM) were chosen using the RMFinder 2.0 previously presented.

These models were used for the purpose of production strategy selection, as the whole set of scenarios (48) may be computationally expensive. To more accurately estimate the risk curves, each RM has the individual probability of occurrence shown in **Table 2.1**.

Table 2.1–Probabilities of each representative model (RM) selected using RMFinder 2.0 to represent the 48 scenarios that typify the field uncertainties.

Representative model	Probability
RM1	0.167
RM2	0.104
RM3	0.146
RM4	0.125
RM5	0.042
RM6	0.063
RM7	0.063
RM8	0.167
RM9	0.125

In the validation section of the methodology, each parameterization was conducted nominally in RM4, chosen for its proximity to the P50 percentile in NPV.

To calculate the objective functions, the economic parameters from **Table 2.2** were applied to the same equations of the NPV and EMV presented by Gaspar et al. (2015). We reinforce that costs related to design variables (G1) – e.g.: platform, well drilling and completion – are not considered because the G1 strategy was already implemented when we started updating the cash flow. **Table 2.3** and **Table 2.4** summarize important dates and the operational constraints used in this paper, respectively.

Table 2.2–Economic parameters to calculate the NPV.

Parameter	Description	value
Market Values	Oil price (USD/m^3)	314.5
	Discount rate (%)	9
	Royalties (%)	10
Taxes	Special taxes on gross revenue (%)	9.25
	Corporate taxes (%)	34
Costs	Oil production (USD/m^3)	62.9
	Water production (USD/m^3)	6.29
	Water injection (USD/m^3)	6.29

Table 2.3–Dates to represent simulation field timeline and cash flow update.

Time (year - days)	Description
05/31/2013 - 0	Simulation and production starting time
06/30/2020 - 2587	Opening the last well from the development phase
	End of history production
07/31/2020 - 2618	Start of management period
	Time for updating cash flow
05/31/2043 - 10957	Maximum simulation period
	Field abandonment

Table 2.4–Wells and platform operational constraints.

Type	Producer	Injector	Platform production	Platform injection
[1]Max water rate (m^3/d)	-	5,000	13,950	21,700
[2]Min oil rate (m^3/d)	20	-	13,950	-
[1]Max liquid rate (m^3/d)	2,000	5,000	15,500	-
BHP (kgf/cm^2)	[2]Min 190	[1]Max 350	-	-
[1]Max gas oil ratio	200	-	-	-
Shut-in W_{CUT}	0.9	-	-	-

[1] Maximum. [2] Minimum.

In this paper, the strategy originated from INGUIDE control rule is denominated S0. This control rule was defined differently for the group of producers and for the group of injector wells. The $Priority(we)$ for each producer is calculating using **Equation 2.6**. This equation came from **Equation 2.1** by adjusting its parameters in order to prioritize the production in wells with higher oil-water ratio. Note that the constant value in the denominator was set to a relatively small value (equals 10) in order to avoid division by zero.

$$Priority\ (we) = \frac{q_{o\_we}}{10 + q_{w\_we}} \qquad \text{Equation 2.6}$$

where q_{o_we} and q_{w_we} are the oil and water rate of the producer we. Regarding injectors, the higher the injection potential of the well at the maximum BHP constraint (350 kgf/cm^2), the higher the injector $Priority(we)$. The injection potential consists in the capacity of each well to inject water if there were no restriction for rates. The priority rank of both injection and production wells was updated every 30-simulation days.

To optimize $P_{GUIDE,}$ the GUIDE weights (p_{i_we}) can vary between 0 and 5 units in 0.5 intervals and the GUIDE function is declared when production starts. These are the same weights applied by Santos and Schiozer (2017). Apart from that, the $W_{CUT\_LIMIT\_WE}$ values range from 0.68 to 0.98 every 0.03 units and the $W_{\text{p\_LIMIT\_WE}}$ values were selected aiming to

test if a particular injector should be shut every year from the half-life until the field abandonment. In this parameterization, we have two variables to be optimized per well and a total of 50 variables for the 25 wells from UNISIM-I-M.

Concerning the parametric equations to control BHP over time, we used nine interval values between the ranges presented in **Table 2.5** for each coefficient. The limit values for the coefficients in the logistic equation were determined based on the explanation in Section 2.2.1.1, while the details for the linear and second-order equations are available in **Appendix A.3.** In the BHP parametric equations, the number of decision variables are given by the product of the number of wells times the number of the equation coefficients. Hence, P_{LOG}, P_{LIN} and P_{SEC} have 75, 50 and 75 decision variables each.

Table 2.5–Range of values to optimize the BHP parametric equations.

Parametric equation	Coefficients	Range [Minimum, Maximum]
P_{LOG}	a_{we}	[-5, 5]
	b_{we}	[-10, 10]
	c_{we}	[-15, 15]
P_{LIN}	a_{we} for producers	[190, 300]
	a_{we} for injectors	[190, 350]
	b_{we}	[-40, 40]
P_{SEC}	a_{we} for producers	[190, 300]
	a_{we} for injectors	[190, 350]
	b_{we}	[-40, 40]
	c_{we}	[-60, 60]

In this work, we assume that the well BHP limit can be changed in a minimal interval of 3 months (92 control intervals) and that one thousand simulations are performed in each optimization process. It is interesting to note that, for this number of intervals, the traditional BHP control approach (stepwise method) would require 2,300 decision variables for the 25 wells of UNISIM-I-M.

2.4 RESULTS

The results are first discussed by comparing the NPV of each parameterization for life-cycle well control strategy with the well control short-term strategy (S0) in a nominal approach (validation part). After, the parameterization that provided the best outcome was applied under uncertainties and for the reference model (application part).

It is worth to remind that the life-cycle parametric equations are applied on the top of the G2S control using INGUIDE where the $Priority(we)$ of each producer is calculated with

Equation 2.6 and the $Priority(we)$ of each injector is defined using the injection potential of the well at the maximum well BHP constraint.

2.4.1 Comparison of well control parameterizations for the nominal approach (validation part)

First, we applied the four parameterizations to the representative model RM4 of UNISIM-I-M. To obtain statistically meaningful interpretations, we executed the optimization of each parameterization five times and compared the NPV statistic samples to the S0 strategy. The S0 provided a NPV of 2.55 billion USD when applied to the RM4 and it is clear that all proposed parameterizations outperformed this strategy (**Figure 2.2**).

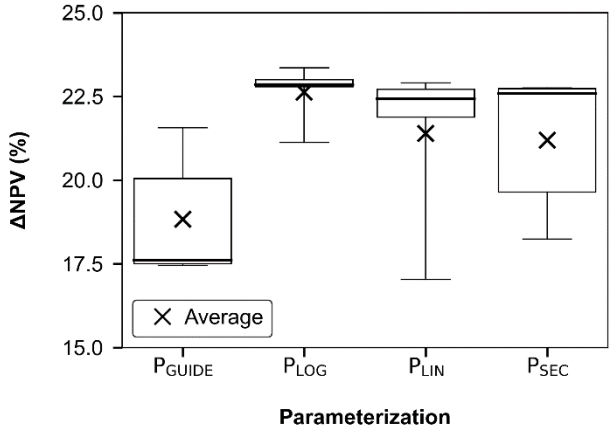

Figure 2.2–Boxplot of the percentage variation of Net Present Value (ΔNPV) of the best strategies of each parameterization compared to the well control short-term strategy (S0) applied to the representative model RM4 of UNISIM-I-M.

It is possible to notice that the parameterizations that define the BHP along time provided better results than P_{GUIDE}. P_{LOG} yields a slightly better $\overline{NPV}$ (roughly 1%) than P_{LIN} and P_{SEC} (**Table 2.6**). Therefore, we decided to carry on the P_{LOG} to be applied under uncertainties because it not only yielded the higher $\overline{NPV}$, but also turned out to be the most robust approach (smaller standard deviation).

Table 2.6–Comparison between the best strategies yielded for each parametrization performed five times in the representative model RM4 of UNISIM-I-M with well control short-term strategy (S0).

Parameterization	$\overline{NPV}$ (10^9 USD)	σ (10^6 USD)	$\Delta\overline{NPV}$ (%)
S0	2.55	-	-
P_{LOG}	3.13	20	22.6%
P_{LIN}	3.09	56	21.4%
P_{SEC}	3.09	48	21.2%
P_{GUIDE}	3.03	42	18.8%

It is interesting to observe that the proposed parameterization to control the BHP over time produced a smooth BHP profile for all wells. **Figure 2.3** includes examples of BHP curves imposed by the parametric equations and the actual BHP that the well is operating while it is open. Those curves are not necessarily equal, as seen for P_{SEC} (Injector INJ017) until about 6,000 days, because there are other constraints in the problem (e.g.: wells BHP restrictions and rate limits for wells and platform). Note that we did not show necessarily the same wells for all parameterization because depending on the best strategy for each parameterization, a specific well may not have been following the G2L rule. For those situations, the BHP imposed is different from the actual BHP.

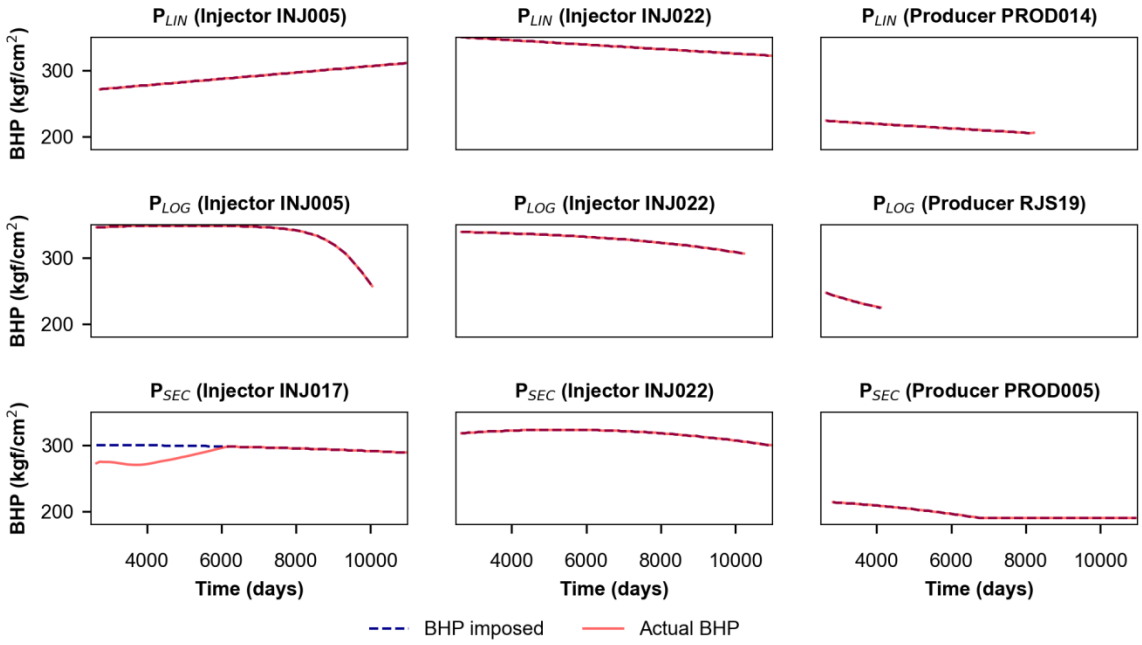

Figure 2.3–Examples of actual and imposed BHP curves over time for each parametric equation applied to the representative model RM4 of UNISIM-I-M.

The NPV improvement for the best parameterization proposed here (P_{LOG}) is mainly related to rising oil production while reducing water production relative to the S0 strategy (**Figure 2.4**).

In **Figure 2.5**, it is possible to observe that, for most of the P_{LOG} execution, the DECE algorithm converge to the 1% best solution with roughly 500 simulations (except for execution 1 and 4), therefore, one could perform P_{LOG} with half of the simulations if the project has a short time constraint. We also verify that, for the initial simulations, there was a great variation between the curve of execution 5 and the other curves. This is explained by the intrinsic characteristic of the optimization algorithm that starts by generating random strategies in order

to obtain the maximum information about the solution space. The quantity of random strategies created depends on the number of optimization variables and for P_{LOG} – which contains 75 variables – a total of 201 arbitrary simulation strategies were generated.

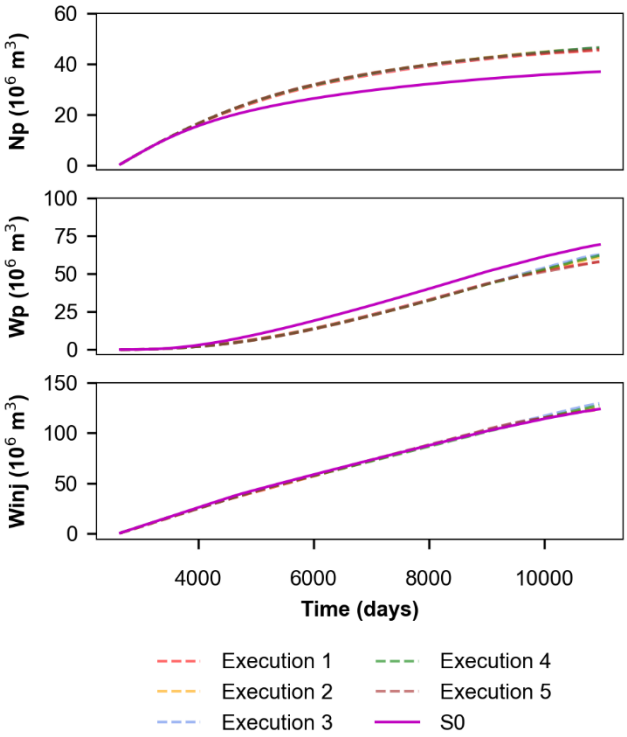

Figure 2.4–Comparison of the field's cumulative oil (Np), cumulative water produced (Wp) and cumulative water injected (Winj) between the best P_{LOG} strategies for each five executions applied in the representative model RM4 of UNISIM-I-M and the well control short-term strategy (S0).

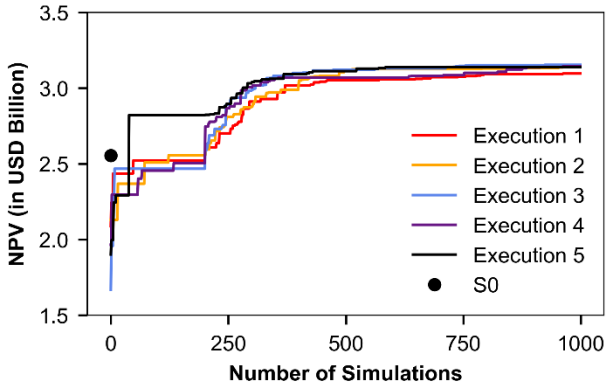

Figure 2.5–Maximum NPV over the DECE algorithm simulation runs for each execution using P_{LOG} parameterization applied to the representative model RM4 from UNISIM-I-M. In this graph, the NPV found in the next simulation is plotted only if it is larger than the previous one, otherwise the previous value is repeated.

2.4.2 Evaluation of the best parameterization under uncertainties (application)

After finding the P_{LOG} as the best parameterization nominally, we applied it under uncertainties. To do so, first we performed the G2L nominal optimization using P_{LOG} to select one strategy for each RM, leading us to nine different strategies. After the G2L optimization, the specialized strategy from each RM is denoted as S_x, where x corresponds to the RM that the strategy was designed for (e.g.: after optimizing RM1, the specialized strategy is named S1). We then applied each of those strategies to the 48 filtered scenarios and selected the one that provided the best EMV to be employed in the UNISIM-I-R. The goal was to evaluate the P_{LOG} robustness in relation to EMV and its applicability for real reservoirs.

All strategies significantly increased the EMV (over 16%) compared with the S0 strategy (**Table 2.7**). In addition, the specialized strategies considerably increased the Np for each of the 48 scenarios and most of the strategies decreased the Wp for the majority of the scenarios (**Figure 2.6**). Besides, depending on the strategies, there was an increment (e.g.: S1, S3 and S9) or a reduction (e.g.: S6, S7, and S8) in relation to Winj for most of the scenarios. These results show that, due the high nonlinearity of the reservoir model, different strategies could provide a similar economic return.

Table 2.7–Comparison of expected monetary value (EMV) and production data between each specialized strategy and the well control short-term strategy (S0) applied to the 48 filtered scenarios of UNISIM-I-M. The strategies are shown in the EMV descending order.

G2 Strategy	**S5**	**S2**	**S8**	**S9**	**S3**	**S4**	**S7**	**S1**	**S6**	**S0**
EMV (10^9 USD)	3.14	3.14	3.13	3.13	3.12	3.12	3.12	3.12	3.02	2.59
ΔEMV (%)	21.4	21.3	21.0	20.9	20.6	20.6	20.5	20.4	16.7	0
EMV σ (10^8 USD)	1.1	1.0	1.0	1.1	1.1	0.9	0.9	1.0	0.9	1.1
$\overline{Np}$ (10^7 m^3)	5.6	5.6	5.5	5.6	5.5	5.6	5.5	5.6	5.4	4.8
$\overline{\Delta Np}$ (%)	17.3	17.1	15.5	17.4	16.7	16.9	16.3	17.0	14.5	0
NP σ (10^6 m^3)	1.5	1.4	1.3	1.5	1.6	1.2	1.3	1.5	1.3	1.6
$\overline{Wp}$ (10^7 m^3)	6.5	6.3	5.4	6.7	6.5	5.9	5.5	6.5	5.9	7
$\overline{\Delta Wp}$ (%)	-7.4	-10.3	-22.6	-4.4	-6.2	-15.2	-20.8	-6.4	-15.6	0
Wp σ (10^6 m^3)	4.0	3.9	4.0	4.2	3.9	5.0	3.9	3.6	4.8	3.9
$\overline{Winj}$ (10^8 m^3)	1.4	1.4	1.3	1.4	1.4	1.3	1.3	1.4	1.3	1.4
$\overline{\Delta Winj}$ (%)	2.8	1.1	-6.6	4.3	3.7	-1.6	-4.1	4.1	-3.3	0
Winj σ (10^6 m^3)	4.6	4.4	4.4	4.9	4.8	5.6	4.2	4.6	5	4.7

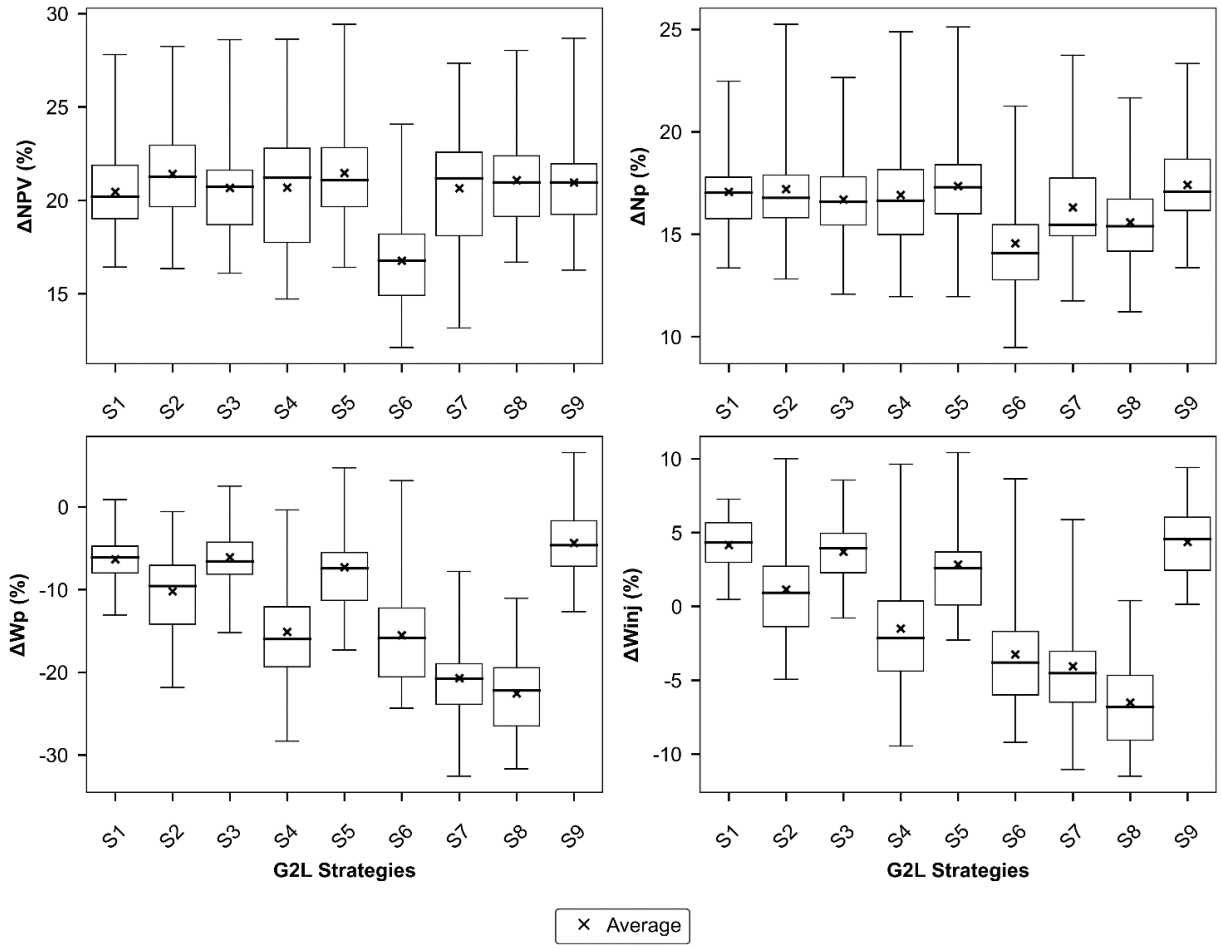

Figure 2.6–Boxplot of the percentage variation of Net Present Value (△NPV) and production data (△Np, △Wp and △Winj) of each specialized strategy and the well control short-term strategy (S0) applied to the 48 filtered scenarios from UNISIM-I-M.

It should be noted that in terms of economic return, many strategies provided equivalent results (e.g.: S2, S5 and S8) but, as the S5 specialized strategy delivered the highest EMV considering the 48 filtered scenarios – 21.4% or over 600 million USD higher than S0 – we chose this strategy to perform in UNISIM-I-R.

Figure 2.7 and **Figure 2.8** present, respectively, the cross plots and risk curves of economic and production data for the 9 RMs and the 48 scenarios under S5 and S0 strategies. We can observe that the 9 RMs are still representative of the 48 scenarios even after the optimization of the production strategy. This is a good indicator that the RMs were well selected by RMFinder 2.0 method.

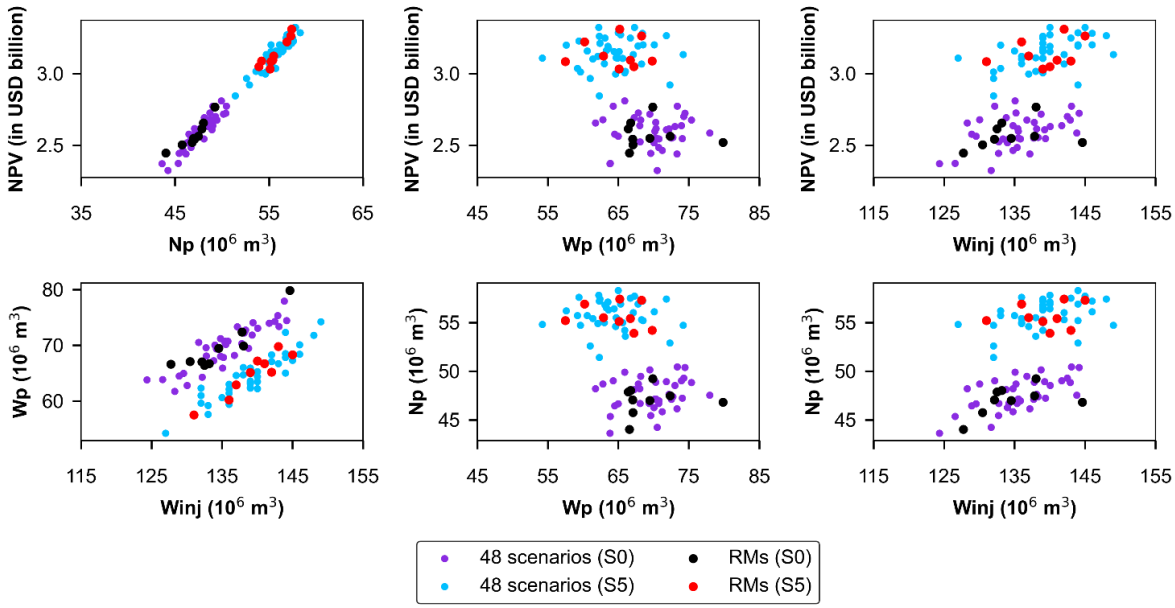

Figure 2.7–Cross plots of economic and production data for the 9 RMs and the 48 filtered scenarios under S5 and S0 strategies.

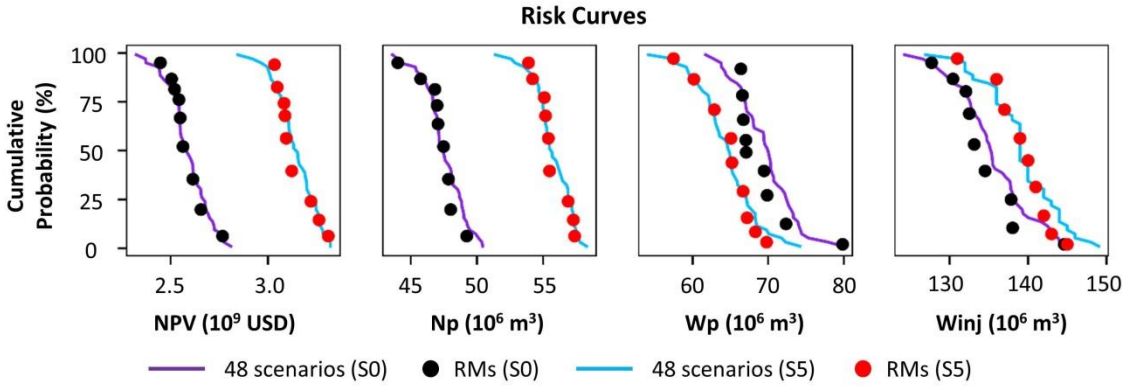

Figure 2.8–Risk curves of economic and production data for the 9 RMs and the 48 filtered scenarios under S5 and S0 strategies.

Also note from **Figure 2.9** that the S5 strategy is fairly robust as the NPV of applying S5 for each RMs were very close to the NPV of the specialized strategies applied to their respective RM (e.g.: S1 employed to RM1, S2 employed to RM2 etc.). Kindly navigate to the zoomed-in version of **Figure 2.9** in **Appendix A.4** for a closer view that highlights the disparity in NPV between strategy S5 and the RM-specialized strategies.

The S5 satisfactorily improved the NPV of UNISIM-I-R compared to the S0 strategy (roughly 260 million USD), as seen in **Figure 2.10**. It is interesting to note that the NPV for the reference model were well represented by the 48 filtered scenarios under the S0 strategy as the NPV risk curve includes the NPV for UNISIM-I-R. In contrast, the 48 scenarios were unable to represent the NPV uncertainties from the reference model when S5 was applied. This

fact indicates that the production strategy directly affects the models that honor the field's observed data. One possible solution to solve the lack of representativeness of the data assimilated scenarios would be to perform the closed-loop reservoir management method (Schiozer et. al, 2015). However, it is beyond the scope of this work.

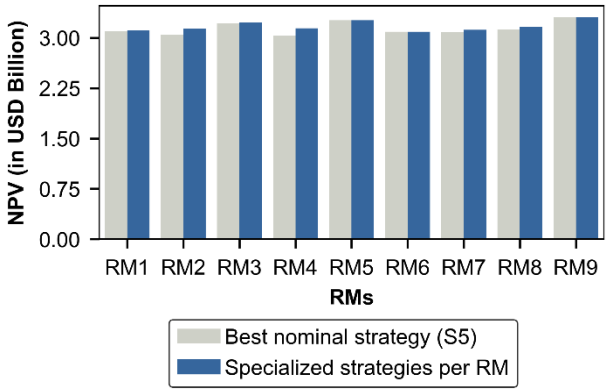

Figure 2.9–Comparison between the most robust specialized strategy (S5) with each specialized strategy applied in the RM that they were optimized for.

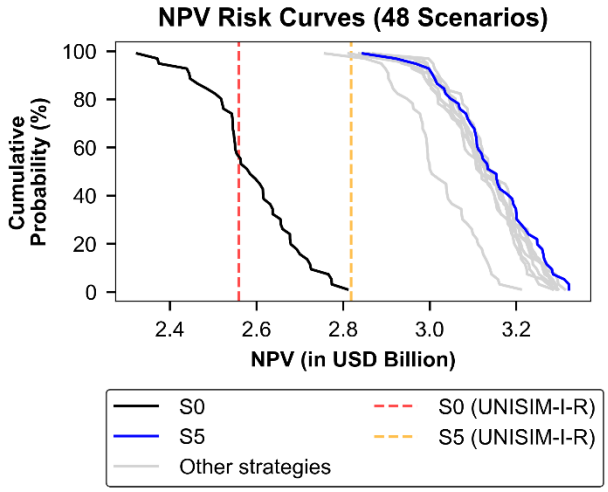

Figure 2.10–Comparison of NPV risk curves between the specific strategies and the well control short-term strategy (S0) applied to the 48 filtered scenarios of UNISIM-I-M. The blue curve is the best specialized strategy (S5) and the grey curves are the other specialized strategies of each RM. The red and yellow vertical dashed lines represent, respectively, the NPV of S0 and S5 when applied to the reference model (UNISIM-I-R).

Apart from the meaningful NPV addition (10%) in relation to the S0 strategy (**Table 2.8**), S5 apparently swept the reservoir more efficiently as the Np satisfactorily increased (11%) and Wp and Winj reasonably decreased (11% and 4%, respectively).

Table 2.8–Comparison of net present value (NPV) and production data between the best specialized strategy (S5) and the well control short-term strategy (S0) applied to the reference model (UNISIM-I-R).

Parameters	G2 strategy		Difference	Percentage Difference (%)
	S5	S0		
NPV (10^9 USD)	2.82	2.56	0.26	+10
Np (10^6 m^3)	52.8	47.4	5.4	+11
Wp (10^6 m^3)	65.9	74.2	-8.3	-11
Winj (10^6 m^3)	134.4	139.6	-5.2	-4

Particularly, we observed that four injection wells that were open using S0 strategy, ended up being shut when the S5 strategy was applied (**Figure 2.11**). This fact indicates that, with the information available, the costs of drilling and perforating those wells could have possibly been avoided. In **Appendix A.5**, we demonstrate that the wells also shut-in for the 48 filtered scenarios and provide explanations for the reasons behind these shut-ins.

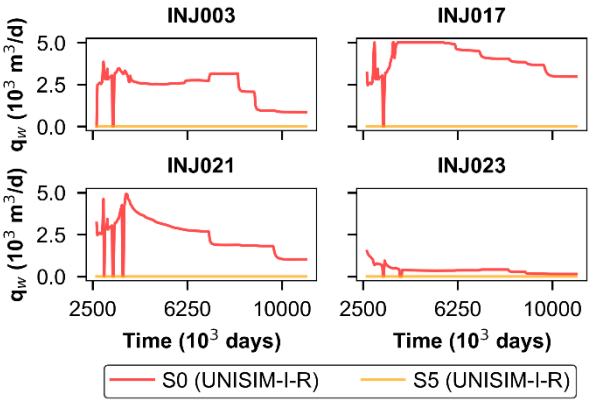

Figure 2.11–Injections water rate for well control short-term strategy (S0) and the best specialized strategy (S5) performed in UNISIM-I-R. It is verified that those wells ended up being shut when S5 was applied.

In addition, S5 also produced a smooth solution for the BHP over time (**Figure 2.12**). In **Figure 2.12a**) and **Figure 2.12b**), it is possible to observe two peaks and valleys for the actual BHP, respectively. This happened first because the wells detected with low flow rate were shut for a period and then because the platform stopped for maintenance.

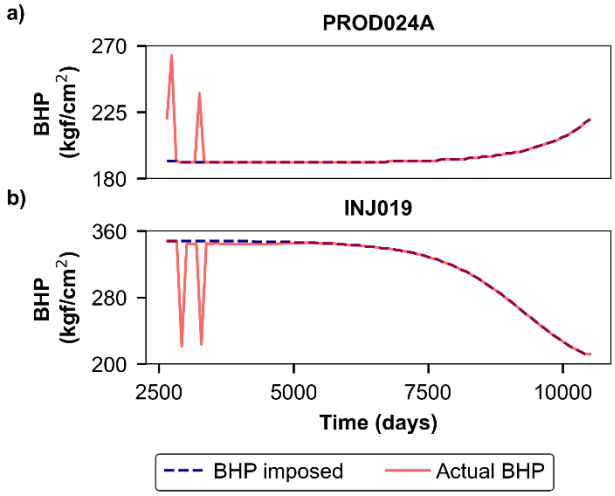

Figure 2.12–Example of actual and imposed BHP curves over time for the best specialized strategy (S5) applied to the reference model (UNISIM-I-R). The peaks (a) and valleys (b) in the actual BHP represent wells shut-in due to low flow rate detection and platform downtime, respectively.

2.5 Discussion

This paper aimed to propose and compare different practical parameterizations to manage well control operations that are able to overcome the issue of premature convergence in economic maximization studies for petroleum reservoir. Also, we investigated the ability, in economic terms, of the best parameterization of this work to be applied under reservoir uncertainties and in the reference model. The latter approach aimed to confirm the reliability of the best parameterization when performed in an existing field.

All proposed parameterizations, solved by DECE, successfully avoided convergence to a poor solution in a high dimensional problem (relative to literature studies) while it considerably increased NPV compared to a G2S strategy representing industry applications (S0). This finding agrees with those obtained by Sorek et al. (2017), who compared polynomial equations with the traditional approach of controlling BHP at each time interval.

Contrary to Sorek et al. (2017), our results support those of Awotunde (2014), who obtained better NPV using lower polynomial degrees than higher ones. Although this result may seem counterintuitive, it can be explained by the reduction in the solution search space that enables the optimization algorithm to find good solutions. Thus, different polynomial order may perform better depending on the case study and the gradient-free algorithm chosen.

The parameterization, based on logistic equation (P_{LOG}), proved to slightly outperform all parameterizations in terms of NPV and robustness, including the polynomial equations,

which have already been explored by other authors, as mentioned in this work. These results were valid for a nominal case and have to be extrapolated with caution when considering uncertainties. Another advantage of the logistic over polynomial equations is that there is no need to set hard bounds in its formulation, which enables smooth solutions even on BHP boundaries. For the other equations, this may not always be true.

P_{LOG} also expressively increased the expected monetary value (EMV) compared to the S0 strategy when performed under uncertainties. By controlling BHP over time, it was possible to produce more oil (Np) and reduce water production (Wp) in average for the scenarios honoring field history data, thus, improving the sweep efficiency, which is one of the main purposes of waterflooding.

With respect to the reference model, P_{LOG} also provided a superior NPV and Np while substantially reducing Wp in relation to the S0 strategy. It can therefore serve as a justification for companies to ally the operator experience with a planned G2L strategy. This analysis in the field, which represents the real reservoir, is particularly important because it ensures that P_{LOG} parameterization would work effectively in realistic cases. This statement could not be true by just analyzing the results in a coarse model, which generally oversimplifies the reservoir's geology and carries a bias from the real reservoir.

It should be noted that this study has only examined well control parameterizations considering waterflooding recovery method and for constrained platform. Specifically, the latter situation is usually favorable for G2 optimization because we can redistribute rates among producer and injector wells without delaying or reducing oil production while reducing water production (Santos and Schiozer, 2017). Further studies would be necessary to evaluate the applicability of the proposed parameterizations in a non-restricted platform and under other recovery methods (e.g.: polymer and CO_2 injection).

2.6 Conclusion

The well control life-cycle rule (G2L) parameterization, based on logistic equation (P_{LOG}), yielded an achievable strategy for a benchmark case emulating the real reservoir as it provided amenable well production trajectories over time, which may be required from an engineering perspective, while leading to a relevant gain in the economic return. P_{LOG} also generated robust strategies when applied under uncertainties. Also, by proposing G2L parameterizations with a reduced number of variables in relation to traditional approach, we

were able to circumvent the problem of premature convergence towards a poor solution of gradient-free optimization in the oil field with a relatively large number of decision variables.

The P_{LOG} outperformed parameterizations based on polynomial equations in terms of economic return for a nominal experiment of this paper. As opposed to polynomial equation, positive features regarding P_{LOG} do not require setting hard bounds for this function, which prevents abrupt changes in BHP values, even within the reservoir's operational constraints. Additionally, one of the objectives of this study was to confirm an end-to-end application for life-cycle well control management (nominal, probabilistic, and reference reservoir analysis) and, since the G2L optimization provided good results for the model representing the real reservoir, it may be interesting for companies not only to rely on operator experience, but also plan the G2L strategy.

Finally, since consistent well control parameterizations were compared and applied to a field subject to waterflooding and to constrained platform, the next steps should focus on well control optimization for a platform operating below its production and injection rate limits, and other types of recovery methods (e.g.: polymer, carbon dioxide or water alternating gas injection). Another direction for future work would be to try to decrease the number of simulations in the G2L optimization process as this could hinder its applicability depending on the complexity of the case (compositional simulator and fine scale models) and the project deadline.

2.7 Nomenclature

A_0, B_0	Numerator and denominator constants for a target group ig
A_i, B_i	Numerator and denominator of the stream i
a_{we}, b_{we}, c_{we}	Coefficients of parametric equations
$a_{we\_limits}, b_{we\_limits}, c_{we\_limits}$	Range limits for the coefficients of parametric equations
BHP	Bottom-hole pressure
BHP_{Max}	Maximum BHP limit of the reservoir
BHP_{Min}	Minimum BHP limit of the reservoir
$BHP_{we,t}$	BHP in well we for time t
DECE	Designed Exploration and Controlled Evolution
EMV	Expected monetary value
G1	Design variables

G2	Well control variables
G2L	Life-cycle control rules
G2S	Short-term control
G3	Revitalization variables
INGUIDE	Internally generated guide rates
Max	Maximum value of determined indicator
Min	Minimum value of determined indicator
Np	Field's cumulative oil production
nph	Number of fluids to be included in the INGUIDE formula
NPV	Net present value
P50	Percentile 50
p_{i_we}	GUIDE weight for the stream i and for the well we
q_i	Stream rate i under the most restrictive constraint or a predefined BHP
$q_{i\_ig\_target}$	Stream rate i for the target group ig
q_{i_we}	Stream rate i for the well we
q_o, q_w, q_{winj}	Oil rate, production water rate and injection water rate
q_{o_we}, q_{w_we}	Oil and water rate for well we
RM	Representative model
S0	Base well control short-term strategy
σ	Standard deviation
S_x	Best strategy optimized for the representative model
t_{final}	Management period interval
USD	United States Dollar
we	Wells index
W_{CUT}	Water cut
$W_{CUT\_LIMIT\_WE}$	Water cut limit for well we
$W_{\mathrm{CUT\_WE}}$	Monitor water cut for well we
$W_{\mathrm{p\_LIMIT\_WE}}$	Field's cumulative water production limit to shut well we
Winj	Field's cumulative water injection
Wp	Field's cumulative water production

2.8 Acknowledgments

We conducted this work with the support of Libra Consortium (Petrobras, Shell Brasil, Total, CNOOC, CNPC, and PPSA) within the ANP R&D levy as “commitment to research and development investments” and Energi Simulation. The authors are grateful for the support of the Center of Petroleum Studies (CEPETRO-UNICAMP/Brazil), the Department of Energy (DE-FEM-UNICAMP/Brazil), and Research Group in Reservoir Simulation and Management (UNISIM-UNICAMP/Brazil). Besides, special thanks to CMG for software licenses. Finally, thank you to the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)* for the financial support.

3 A Machine Learning Approach to Reduce the Number of Simulations for Long-term Well Control Optimization

Authors: Santos, D. R., Fioravanti, A. R., Santos, A. A. S., Schiozer, D. J.

Presented at the SPE Annual Technical Conference and Exhibition, Virtual, October, 2020.

(https://doi.org/10.2118/201379-MS)

Abstract

A long-term well control strategy is frequently selected using optimization methods applied to reservoir simulations. However, this approach usually requires a large number of simulations that can be computationally demanding. In this paper, we evaluated several machine learning (ML) techniques to reduce the number of simulations for optimizing long-term well control strategy while preserving the quality of the solution.

We proposed a methodology, denoted as IDLHC–ML, which combines many ML techniques with iterative discrete Latin hypercube (IDLHC) – a gradient-free optimization algorithm that was successfully applied in previous work – to optimize the coefficients of the logistic equation that guides the well's bottom-hole pressure along the time horizon. In IDLHC-ML, we used a set of simulation runs from the first iteration to train the initial ML models. From the second iteration onwards, we employed the trained ML models to predict the net present value (NPV) and only a percentage of the scenarios, which were expected to have the best NPV, were then simulated. As we simulated new scenarios, we updated our ML models to further improve predictions. For a fair comparison, we set the same values for the optimization parameters of IDLHC to the IDLHC–ML and, then, we compared the NPV and the number of simulation runs considering different configurations of IDLHC parameters. In this paper, we evaluated a total of twelve ML regression techniques, such as Bayesian Ridge, Random Forest, and stacked ensemble learning, which consists in using the predictions from multiple ML algorithms as input to a second-level learning model.

To minimize random effects, we repeatedly applied IDLHC and IDLHC–ML five times in a single reservoir model (nominal optimization). The results showed that, depending on the IDLHC optimization parameters, IDLHC-ML reduced at least 27% of simulations while

keeping the equivalent NPV statistical metrics calculated in all five repetitions, when compared to IDLHC. Moreover, the best ML technique for IDLHC–ML varied with the IDLHC set of optimization parameters. To conclude, the method proposed here was able to reduce a significant amount of computational time by curtailing the total number of full-physics expensive reservoir simulations, with the help of fast and low-cost ML models.

There are many published studies in well control optimization, but these generally involve high computational demand. In this sense, ML methods revealed to be an adequate and inexpensive alternative in reducing the number of simulation runs in well control optimization. The methodology is generic and it can be applied under uncertainties, and for more complex cases.

Keywords: machine learning, simulation reduction, long-term well control optimization, waterflooding, field management, UNISIM-I-M benchmark case.

3.1 Introduction

Before implementing a production strategy in the oil field, a comprehensive study has to be performed in order to lessen risks associated with high project investments. Roughly, this kind of study comprehends reservoir characterization, production data assimilation, and selecting a production strategy under uncertainties (Schiozer et. al, 2015).

Ideally, to adequately define the final strategy of the field, a large number of scenarios honoring the field's observed data should be analyzed and a high number of variables optimized. This large number of variables can be divided into three groups: design variables (G1) representing the infrastructure to be implemented during field development (e.g.: well types and placements, platform capacity, well opening schedule, etc.); well control variables (G2) related to operational specifications to guide production along the management period (e.g.: control valves choke on the platform, the well or level region) in long (G2L) and short-term (G2S); and revitalization variables (G3) that consist of the field's infrastructure modifications employed during the management phase (e.g.: re-perforation, infill drilling, well conversion) – see Gaspar et al., 2016b for further details.

Particularly, the use of large number of scenarios and variables mentioned can render the production strategy selection very time consuming. Thus, it is important to employ techniques to accelerate this process in order to respect the oil project schedule. One technique to this end is to represent the larger group of scenarios that honored the field's observed data

thorough a smaller set of reservoir models, denominated representative model (RM), which serves to generate production strategies under uncertainties (Meira et al., 2017; Meira et al., 2020; Sarma et al., 2013; Shirangi and Durlofsky, 2015). Another method consists of variable reduction in numerical optimization, such as well long-term control rule (G2L), based on parametric equations to guide bottom-hole pressure (Pinto et al., 2019a; Sorek et al., 2017) or well rates (Awotunde, 2014) over time. In this approach, the coefficients of the equations are optimized, and these substantially reduce the number of decision variables (from thousands to less than a hundred, for some related problems) compared with the traditional method (also called step-wise) of optimizing the BHP or well rates at each time interval.

Despite the mentioned techniques, the production strategy selection process may still demand a high number of simulations runs depending on the type and number of variables to be optimized, the optimizer, the number of RM adopted to generate the strategies and the amount of history-matched reservoir scenarios where the strategies are evaluated before they are implemented to the real field. For example, Silva (2018) needed a total of 36.000 simulations runs to execute a robust optimization for G1 variables using nine RMs. In relation to G2L parameters optimization, Silva et al. (2019) performed between 7.000 and 28.000 simulations runs, depending on the number of models to describe the geological uncertainty (5, 10, 15, 20, 25, 50). Thus, the optimization process may take several months to be completed depending on the amount of computational resource available, the capability of parallelism from the optimization algorithm, and the complexity of the case studied (e.g.: Black-oil or compositional reservoir simulation and the scale of the simulation model). In this context, some of the optimization steps can be simplified or even skipped if the evaluation time of the production strategy does not fit in the project schedule. In **Appendix B.1**, we elucidate the concept of a computationally expensive simulation model.

Therefore, methods to speed-up the variable optimization process, without compromising the quality of the results, may help the production strategy to be implemented on time in the real field. Based on that, the objective of this work is to reduce the number of simulation runs to optimize the parameters of a long-term well control rule (G2L) by combining a gradient-free algorithm with ML techniques, without affecting the economic return.

The results are evaluated by comparing the net present value (NPV) and the number of simulation runs required to optimize the coefficients of the logistic equation to guide the BHP curve over time, both using Iterative discrete Latin hypercube (IDLHC) optimization algorithm

and the method proposed in this paper, which combines the IDLHC with ML techniques (IDLHC–ML). It is important to mention that these methods were applied to a single representative model of a reservoir benchmark case and the work considering uncertainties is currently under way.

This paper is structured as follows: in the next subsections, we review the literature that employed ML techniques to optimize production strategies, we briefly describe the IDLHC algorithm (von Hohendorff et al., 2016), explain the well control simulator default, which serves as a reference to check if the IDLHC and the G2L parametrization proposed here were working correctly, and we then explain the G2L parameterization adopted in this work. We also present some basic ML concepts. The second section introduces the general methodology of this work and provides details on how IDLHC–ML was designed. Section 3 concisely presents the case study of this work and shows the parameters adopted for IDLHC and IDLHC–ML. In Section 4, we compare the results of IDLHC and IDLHC–ML. Section 5 concludes and discusses future research ventures.

3.1.1 Optimization based on machine learning approaches

Some researchers applied machine learning in order to speed up the process of optimizing the production strategy variables (Bruyelle and Guérillot, 2019; Busby et al., 2017; Doraisamy et al., 1998; Jang et al., 2018; Min et al., 2011; Nwachukwu, 2018; Sarma et al., 2018; Teixeira and Secchi, 2019). For example, Teixeira and Secchi (2019) applied multilayer perceptron (MLP) neural network to represent the nonlinear dynamic behavior of the UNISIM-I-M benchmark reservoir, which is a synthetic model containing roughly 37,000 active grid cells and 11 injector and 14 producer wells. The dataset of their analysis was the injector water rates (q_w), producer oil rates (q_o), and BHP acquired monthly during the field's lifetime and the data was split into training, validation, and test sets in the respective proportions of 70%, 15%, and 15%. Next, the MLP model was built considering the q_w, q_o and BHP of 15-time delays as input and the q_o of the next time step as output. Although the q_o predicted by the MLP presented some high amplitude spikes, the neural network was able to follow the behavior of the q_o curves for all wells along the UNISIM-I-M lifetime compared to a commercial reservoir simulator. The authors state that their methodology is suitable to determine the optimum control in reservoir management; nevertheless, they did not perform the optimization process.

Jang et al. (2018) employed MLP to optimize the position of a single horizontal infill well to maximize the cumulative gas production in a coalbed methane reservoir containing 6,771 grid cells and six vertical producers already operating for three years. The authors alternate reservoir simulations and MLP model application to reduce, iteratively, the search space, initially composed of 3,550 combinations, of the infill well position in order to find the global optimal solution. The MLP model built in each iteration was applied to predict all possible solutions and the best ones, which were above a predefined cutoff value, were select to create the search space of the next iteration. The MLP approach required a maximum of 70 simulation runs to find the global optimal solution compared to at least 124 simulations of a regular particle swarm optimization algorithm.

Although some works employed machine learning for production strategy, it is still necessary to perform it for different study cases and adopt other approaches for a better understanding of the ML potential in the production strategy optimization area.

3.1.2 Iterative discrete Latin hypercube (IDLHC) optimization algorithm

In this work, we used an iterative sampling method denoted as IDLHC (von Hohendorff et al., 2016) to optimize the coefficients of the logistic equation that guides the well's BHP over time. In each IDLHC iteration (I), the optimization variables are discretized into values (or levels) and the variable values are combined using discrete Latin hypercube sampling method (DLHC, see Maschio and Schiozer, 2016), which considers a prior probability density function (PDF), to randomly generate a set of N samples. Then, the objective function (OF) of the samples are evaluated and a threshold cut percentage (F) is used to select the best samples (of size $F \times N$) to update each variable frequency level that will be employed to then create the next iteration samples. The optimization search space is gradually reduced along the iterations and the process ends when a convergence criterion is reached (e.g.: maximum number of iterations – I_{FINAL}). Generally, in the first iteration, the levels of each optimization variable have equal probability of occurrence. The IDLHC framework is presented in **Figure 3.1**.

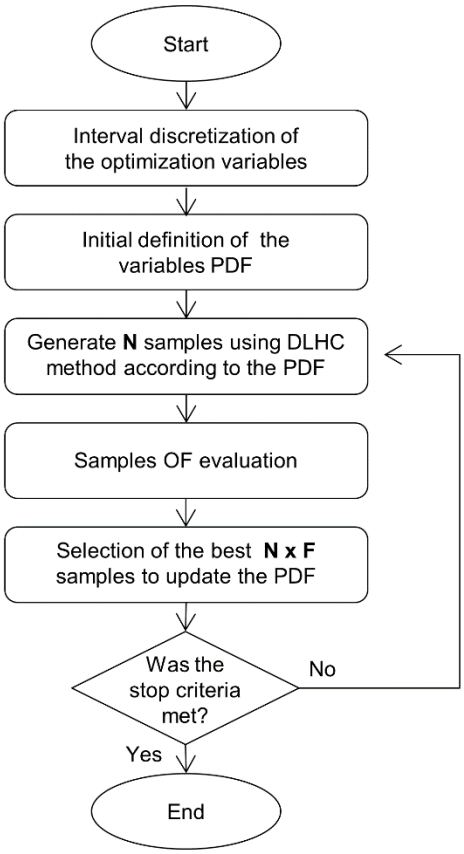

Figure 3.1–Iterative discrete Latin hypercube (IDLHC) optimization algorithm flowchart.

According to von Hohendorff et al. (2016), the choice of N and F parameters depends on the case studied and has a great influence in the optimization performance. The authors tested different parameters for three different cases (**Table** 3.1) and all of them satisfactorily converged towards the maximum value. Thus, these values can be a guide for choosing IDLHC parameters in a production strategy optimization problem.

Table 3.1–Iterative discrete Latin hypercube parameters tested in Von Hohendorff et al. (2016).

Case tested	Parameters		
	N	F (%)	I_{FINAL}
I	20	20	5
II	2500	20	20
III	100	10	9

It is important to note that we selected the IDLHC optimization algorithm due its code availability, which facilitates our methodology implementation and because it provided a good convergence rate to maximize the OF in von Hohendorff et al. (2016) for the production optimization problem, compared to a well-established optimization algorithm (designed exploration stage and controlled evolution).

3.1.3 Well control reference strategy (S0)

In this work, we adopted a reference strategy (named here as S0) to check if the G2L parametrization proposed and the IDLHC algorithm were performing well, that is, if the IDLHC optimization is delivering a considerably superior NPV than the reference strategy.

The S0 strategy is based in a well short-term rule (G2S) defined internally by the commercial simulator software IMEX 2016.10 (see Pinto et al., 2019b for more details). This control rule, named INGUIDE, determines the rate apportionment among wells to meet the target group (e.g.: platform production or injection capacity), according to a priority rank. In this paper, the producer wells' priority rank was defined in order to prioritize production in wells with higher oil-water ratio (**Equation 3.1**).

$$Priority\ (we) = \frac{q_o}{10 + q_w} \qquad \text{Equation 3.1}$$

where $Priority(we)$, q_o and q_w are the priority rank, the oil rate, and the water rate of the producer we, respectively. In this equation, each producer well receives a fraction of the platform production capacity in proportion to its value of $Priority(we)$. Note that, the constant value in the denominator was set to a relatively low value (equals 10) in order to avoid division by zero whenever $q_w = 0$.

Regarding injectors, the priority rank was built considering each well injection potential at the maximum well BHP constraint, that is, the capacity of each well to inject water if there were no restriction for rates. Therefore, the contribution of each well is proportional to its well injection potential value. The priority rank for both producer and injector wells was set to be reevaluated at every 30 simulation days.

It is important to mention that, if the rate assigned by the command INGUIDE to specific well violates the most restrictive constraint, the well will be left out of the apportionment with the rate being reduced to its maximum rate and the remainder of the platform capacity rate is still split to the remaining wells proportionally to their priority rank value.

3.1.4 Parametrization for well long-term control rule (G2L)

The well long-term control rule (G2L) applied in this work consists of optimizing the logistic equation coefficients (a_{we}, b_{we}, and c_{we}) to guide the BHP limit of each well over time (**Equation 3.2**).

$$BHP_{we,t} = BHP_{MIN} + \frac{(BHP_{MAX} - BHP_{MIN})}{1 + e^{\left(a_{we} + b_{we} \times \left(\frac{t}{t_{FINAL}}\right) + c_{we} \times \left(\frac{t}{t_{FINAL}}\right)^2\right)}} \qquad \text{Equation 3.2}$$

where $BHP_{we,t}$ is the BHP for the well we at time t, BHP_{MAX} and BHP_{MIN} are the maximum and minimum well BHP constraints, in the order given, and t_{FINAL} represents the entire management period.

The range values for each coefficient are defined as follows: The limit values of a_{we} (a_{we_lim}) are defined in order to guarantee singly that the $BHP_{we,t}$ can be carried out through the well BHP constraints over time. Note that t/t_{FINAL} varies from [0, 1], thus, the b_{we} and c_{we} limits (b_{we_lim} and c_{we_lim}, respectively) must be higher than the a_{we_lim} to have a similar influence in the BHP curve. One way to do this is to set their limits in order to ensure that all coefficients have the same integral in the interval of [0, 1] as follows: $b_{we\_lim} = \frac{a_{we\_lim}}{\int_0^1 t * dt}$, and $c_{we\_lim} = \frac{a_{we\_lim}}{\int_0^1 t^2 * dt}$.

It is important to mention that whenever the G2L cannot be applied to a specific well in a certain period of the simulation due to any constraints of the problem (e.g.: wells BHP limits, maximum rate for wells and platform), the INGUIDE rule described in the previous section is applied in that period for that well.

3.1.5 Machine learning general framework and basic concepts

In this work, we applied a supervised machine learning task (Goodfellow et al., 2016) meaning that, from input data (denominated features, e.g. $x_1, x_2, \dots, x_n$), there is a corresponding output (named label, e.g.: y). In this task, the objective is to create a function $h(x)$, also called ML model, capable of mapping all n features to their labels based on each training example i. For this purpose, a set of examples consisting of a pair of features and labels (x^i and y^i) is gathered to create the ML model, which will be used to predict or classify labels from a new set of examples (**Figure 3.2**).

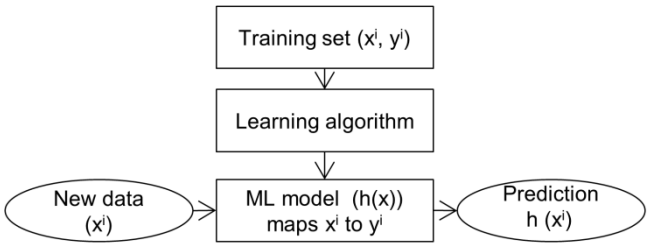

Figure 3.2–Supervised machine learning intuition.

Figure 3.3 shows the general steps of the supervised machine learning framework adopted in this work. We first defined the set of features and labels related with the task to be studied (problem definition). These features and labels must be available for collection (data collection). Next, the data was pre-processed, a step which comprises any transformations applied to the data before feeding it to the algorithm. The dataset chosen in this work was reasonably clean as it did not have missing values, the data was balanced, and there was no categorical data. Thus, data normalization was the only transformation implemented in this work since it helps speed up the convergence of some ML algorithms (Lecun et al., 2012).

Subsequently, the ML algorithm was selected and trained. Ideally, a good ML model should have low bias (small error on training data) and low variance (generalize well to new unseen data). A common problem when training a ML algorithm is the overfitting phenomenon (Duda et al., 2001), which happens when a ML model perfectly fits the training data (low bias), but its performance considerably decreases into new data (high variance). There are several techniques to avoid overfitting (Ying, 2019), for example, increasing the number of training examples, reducing the number of features, adjusting hyperparameters, which are parameters that refer to the model selection task and are not learning in a regular training process, and separating the available data into training, validation, and testing sets (cross-validation).

In relation to cross-validation techniques, there are several approaches (e.g.: hold-out, k-fold and leave-one-out, see Bishop, 2006). The k-fold cross-validation, for instance, consists of dividing the available data in train and test set and, then, split the training data into k groups. For each group, the data is trained in $k - 1$ folds and the accuracy is measured in the data left (validation set). This process is performed k times and an average of the performance measure (e.g. mean absolute error) reported in each validation set is used to select the set of hyperparameters or the ML algorithm. After tuning the hyperparameters or choosing the ML algorithm, all the training and validation data can be used to create the new ML model. This model is then applied in the test set in order to evaluate how it would perform when new data

is collected. Although, k-fold cross-validation can be computationally expensive because the ML algorithm is trained *k* times, this technique is indicated when there is little data (Hastie et al., 2009).

Specifically, in this work, we applied the k-fold cross-validation slightly modified as we did not split the data into train and test, but directly fed it to perform the cross-validation process and, then, the samples from succeed iteration of IDLHC were used as test set. We decided to apply the k-fold cross-validation in this manner to mitigate adverse effects due to the small number of training examples.

After cross-validating, the next step consisted of monitoring the ML model performance and, as new data is collected, the entire process should be repeated.

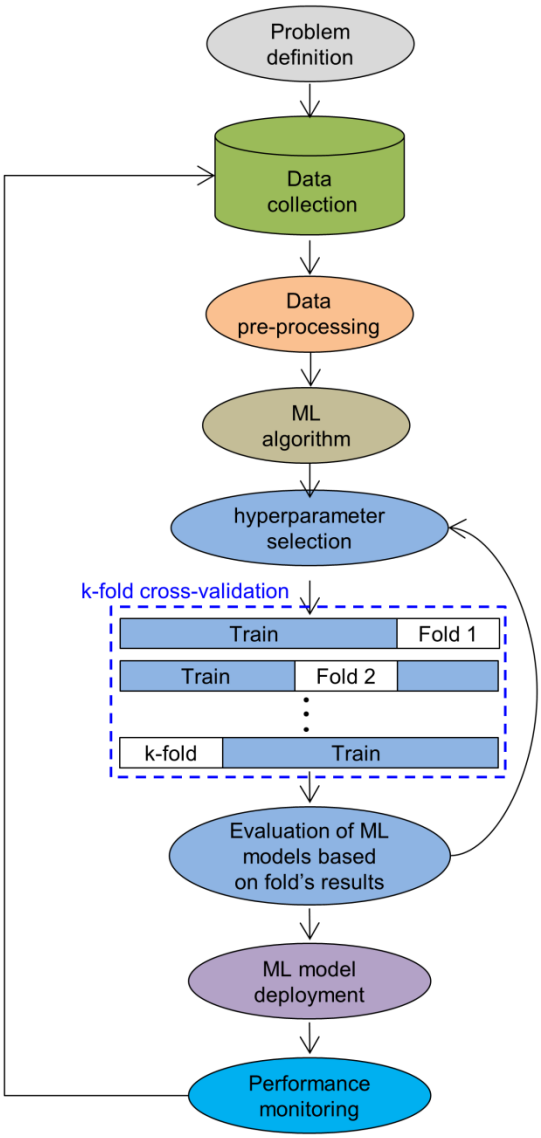

Figure 3.3–General machine learning framework based.

3.1.6 Ensemble learning for regression

Instead of using a single individual machine learning model, it is possible to associate multiple models to reduce bias and variance. In this process, the final algorithm (meta-learner) uses the prediction of multiple ML base models to make the final prediction. This technique is denominated ensemble learning and may be executed in three main ways: boosting, bagging, and stacking (Hastie et al., 2009).

Bagging (Breiman, 1996) consists of randomly splitting the training data into several groups of the same size using sampling with or without replacement, training the same ML base algorithm for each subgroup, and creating a meta-model that averages the outputs of each ML model to produce the final prediction. Bagging is designated to reduce variance; thus, it is often applied using base learner algorithms that yield high variance and low bias (such as tree-based algorithms).

In contrast, the boosting method aims to build different base models in series so that the bias of the meta-model is minimized (Freund and Schapire, 1997). Specifically, the same base learner is applied to the whole training set and then another model is built considering a higher weight to the data that was poorly predicted by the prior model. Finally, the meta-model weight-averages the output of each model previously created. Higher weights are given to the models with better performance in the training data. In general, this method is applied to algorithms that generate results with low variance and relatively high bias.

In the stacking learning (Wolpert, 1992), several ML base algorithms are applied in the training data and the prediction of each model is used as input to train the final meta-learner. Like boosting, the stacking technique is designed to reduce the bias of the base learners.

3.2 Methodology

This section outlines the specific method used within this research to achieve our main objective of reducing the number of simulations for a derivative-free optimization algorithm by combining it with ML techniques. The general methodology is described below (**Figure 3.4**).

a) First, the IDLHC parameters (N, F and I_{FINAL}) were defined.
b) Next, we performed the IDLHC optimization for a single reservoir model (nominal optimization) five times. The goal of repeating it many times is to lessen random effects associated with stochastic derivative-free algorithms. The results of this item served as

reference to be compared with the optimization algorithm associated with the ML techniques (IDLHC–ML) proposed in this work.

c) After, we checked if the IDLHC algorithm, using the parameters N, F and I_{FINAL} defined in **a)**, was suitable to optimize the G2L parameters of our problem by comparing its NPV with the one delivered by the G2S simulator default (S0 strategy) described in introduction. If the parameters selected did not deliver good results, we either changed them or ended the process.

d) Subsequently, we performed a sensitivity analysis to pre-evaluate several ML techniques using the strategies simulated in the Item **b)**. Concisely, the ML technique predicted the economic return among the N samples (or strategies) randomly generated by IDLHC at a current iteration, and the subset m of N strategies expected to have the best outcome was picked. Subsequently, we measured the percentage (or accuracy) of the $F \times N$ best samples of IDLHC that were actually among the best m samples predicted by the ML. It is important to mention that, if the pre-defined value of m leads to a poor accuracy, one could redefine this value. The full explanation of this process is in the next subsection.

e) Next, the ML technique that delivered the highest accuracy on average in the sensitivity analysis was selected to perform the IDLHC–ML method. Here, the process was very similar to item **d)** and the details of this process are also elucidated in the following subsection. As in Item **b)**, we performed the IDLHC–ML nominal optimization five times for the same reservoir model adopted before.

f) In each iteration, the strategies left out in Item **e)** by IDLHC–ML were simulated in order to assess its performance, that is, how many correct scenarios to update the PDF were in fact simulated in step **e)**. This step would not be performed in a real project and it was conducted only to evaluate the reliability of IDLHC–ML.

g) If the NPV statistical samples (mean, median, dispersion, maximum, and minimum) for IDLHC and IDLHC–ML were similar, we repeated item a) with alternative values for the N, F, and I_{FINAL} parameters of IDLHC. This was done to evaluate if IDLHC, with different configurations, yields comparable results to IDLHC–ML using the same number of simulations, as this would not justify the application of the proposed method. Otherwise, the methodology ends.

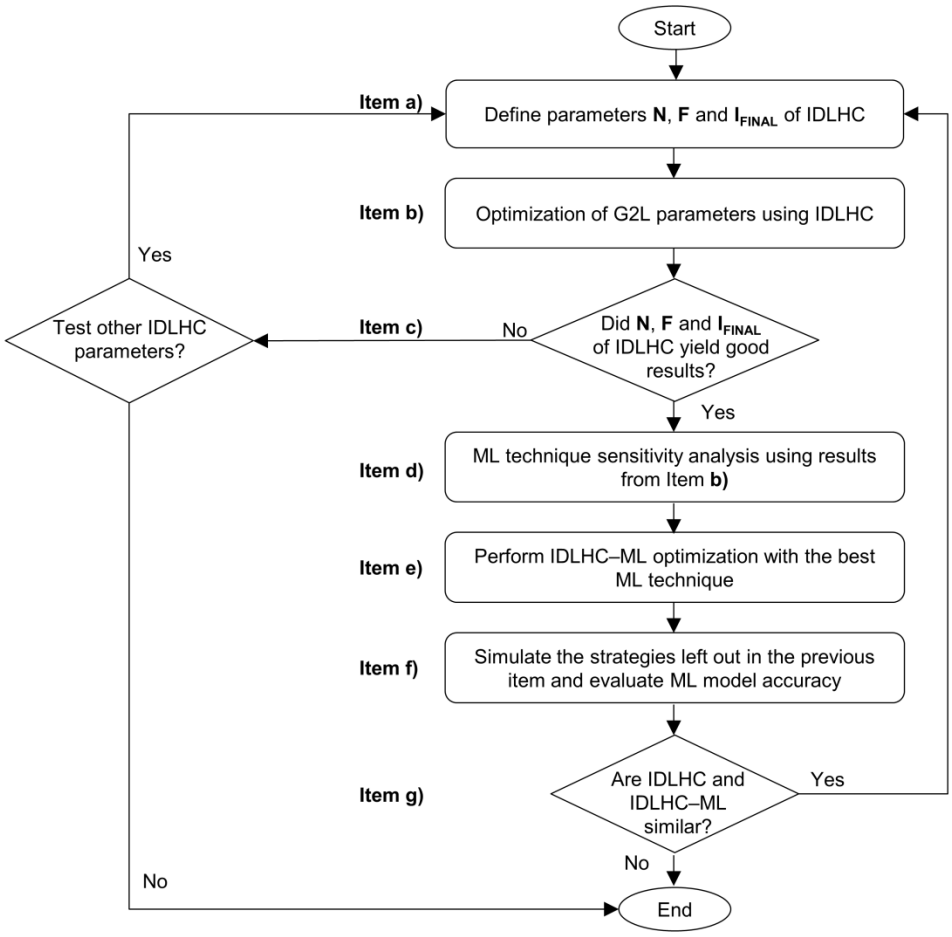

Figure 3.4–Global methodology to perform and evaluate the IDLHC–ML method proposed that consists of combining the Iterative discrete Latin hypercube (IDLHC) optimization algorithm with machine learning (ML).

3.2.1 Machine learning sensitivity analysis and IDLHC–ML evaluation details

Figure 3.5 presents the diagram used in the ML technique sensitivity analysis and IDLHC–ML evaluation steps, which is described as follows: first, we defined the features and labels to be used as dataset. We then chose the ML technique to be tested. After, the features and labels of all the strategies simulated (N) in the first iteration ($I = 1$) of IDLHC were scaled and used to train the ML technique through k-fold cross validation process. In this part, we selected the hyperparameters that minimize the root mean squared error of the ML algorithm. We subsequently tested the ML trained model (h) in the N strategies created in the next iteration of IDLHC in order to predict and select the m strategies with higher OF values. Then, we measured the accuracy from iteration $I + 1$, which is the percentage of $F \times N$ real best strategies that are within m strategies predicted by the ML model. Finally, the data from the m strategies become part of the dataset to be used to retrain the ML technique. This process was repeated until $I_{FINAL} - 1$ iterations for each ML techniques and feature tested.

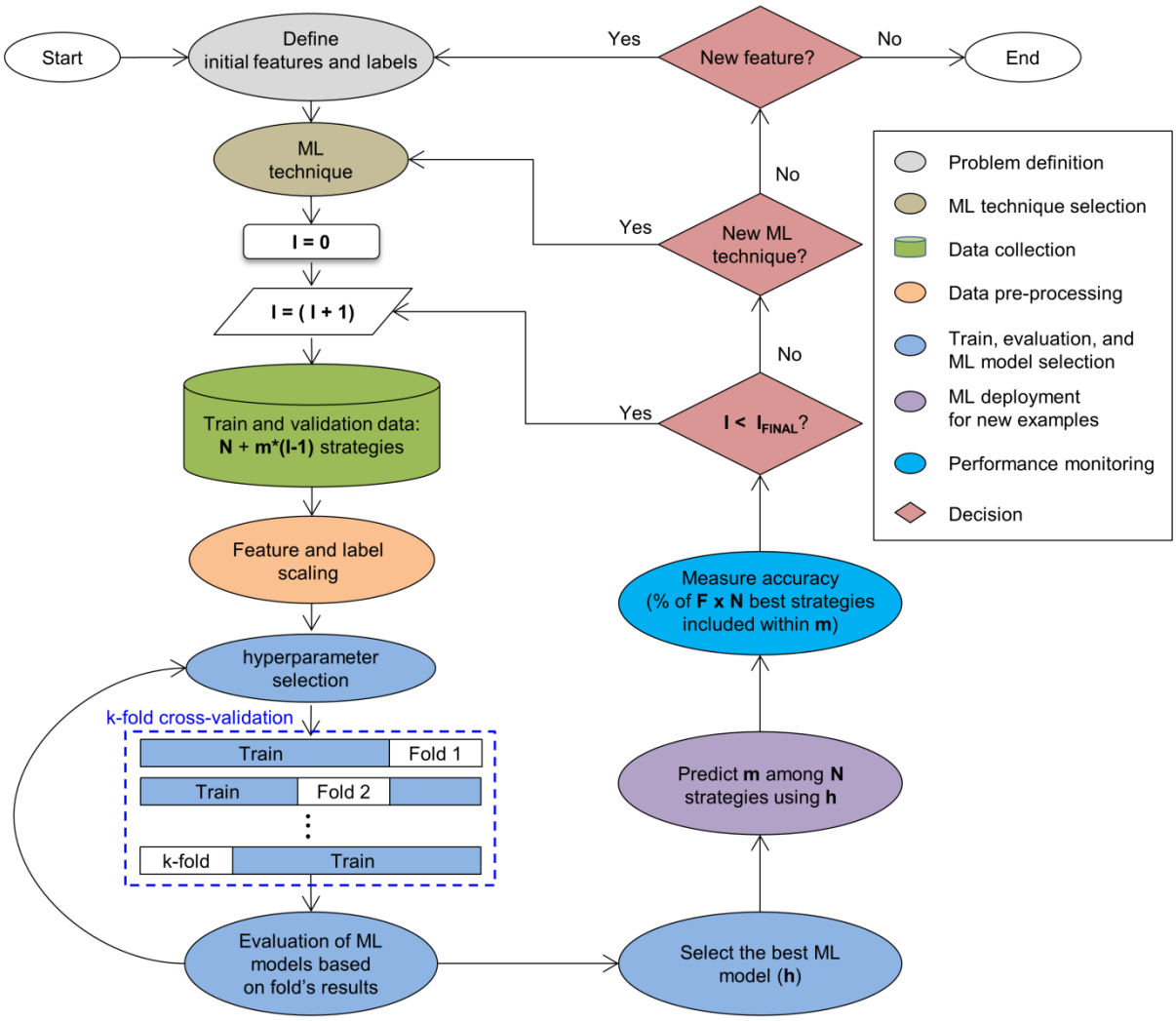

Figure 3.5–Procedure to evaluate the ability of the IDLHC–ML of predicting the correct strategies used by the Iterative discrete Latin hypercube (IDLHC) to update the probability density function (PDF) of the optimization parameters.

The sensitivity analysis was carried out for every five IDLHC executions. Then, the ML technique, which provided the highest mean for the OF, was selected to perform the IDLHC–ML evaluation step.

The single difference in the evaluation step from the sensitivity analysis is that the simulations were actually executed and, thus, from the second iteration onwards, the best strategies were not necessarily selected to update the PDF of the next iteration. It is important to mention that the approach set for IDLHC–ML evaluation requires $N + m \times (I_{FINAL} - 1)$ simulations instead of $N \times I_{FINAL}$ simulations required for IDLHC.

Also, we emphasize that the discarded part of the strategies from each iteration $(N - m)$ were simulated in the evaluation step only for the purpose of measuring the accuracy of the IDLHC–ML method, and it would otherwise not be performed.

3.3 Application

In this topic, we briefly describe the characteristics of the UNISIM-I-M benchmark case in which our methodology was tested (see Gaspar et al., 2016a for full model description). In addition, we present the information to calculate the objective function (NPV), such as dates and economical parameters, G2L parameters values, operational constraints, and the configuration for the optimization algorithm (IDLHC) and for the ML techniques.

3.3.1 UNISIM-I-M benchmark case

The UNISIM-I-M is a 3D Black-oil reservoir simulation model, which comprises a 100 × 100 × 8 grid cell resolution, 20 layers, and 36,739 active grid blocks. The model contains 14 producer and 11 injector wells and a history production data of 7 years. This case was designed to study well control (G2) and revitalization (G3) variables strategy selection during the field management phase. In this work, we focused on the long-term well control rule (G2L) optimization. It should be mentioned that the G1 exploitation strategy – well type and placement, platform limit, etc. – was already defined by Avansi and Schiozer (2015) in the development stage of the field and this must not be modified.

The original optimization algorithm (IDLHC) and the proposed method (IDLHC–ML) were performed nominally in a representative model of UNISIM-I-M, denoted as RM4. The RM4 was chosen because it is the closet model to the percentile P50 in NPV.

3.3.2 Information to calculate the objective function and operational data

The NPV was calculated using the same equation, economic parameters (**Table 3.2**), and dates (**Table 3.3**) presented in Gaspar et al. (2015). We stress that there is no cost associated with design variables (G1) because the development strategy was implemented before the date that the cash flow was updated.

Table 3.2–Economic data to calculate the net present value.

Parameter	Description	value
Market Values	Oil price (USD/m^3)	314.5
	Discount rate (%)	9
Taxes	Royalties (%)	10
	Special taxes on gross revenue (%)	9.25
	Corporate taxes (%)	34
Costs	Oil production (USD/ m^3)	62.9
	Water production (USD/ m^3)	6.29
	Water injection (USD/ m^3)	6.29

Table 3.3–Simulation timeline and date to update the cash flow.

Time (year - days)	Description
05/31/2013 - 0	Simulation and production starting time
06/30/2020 - 2587	Opening the last well from development phase
	End of production history
07/31/2020 - 2618	Start of management period
	Time for updating cash flow
05/31/2043 - 10957	Maximum simulation period
	Field abandonment

The coefficients (a_{we}, b_{we}, c_{we}) of the logistic equation to optimize the BHP limits over time were linearly discretized into 15 values between the ranges exposed in **Table 3.4**.

Table 3.4–Range and number of interval values to optimize the coefficients of the logistic equation.

Parametric equation	Coefficients	Range [Minimum, Maximum]
P_{LOG}	a_{we}	[-5, 5]
	b_{we}	[-10, 10]
	c_{we}	[-15, 15]

We emphasize that the logistic equation BHP limit must respect all constraints presented in **Table 3.5**. Therefore, the limit of BHP set by the logistic equation will always be in between the range of 190 and 350 (kgf/cm²).

Table 3.5–Operational constraints for wells and platform.

Type	Producer	Injector	Platform production	Platform injection
[1]Max water rate (m^3/d)	-	5,000	13,950	21,700
[2]Min oil rate (m^3/d)	20	-	13,950	-
[1]Max liquid rate (m^3/d)	2,000	5,000	15,500	-
BHP (kgf/cm²)	[2]Min 190	[1]Max 350	-	-
[1]Max gas oil ratio	200	-	-	-
Shut-in W_{CUT}	0.9	-	-	-

[1] Maximum. [2] Minimum.

3.3.3 Configuration for IDLHC and for IDLHC–ML

Initially, we set $N = 100$, $F = 20\%$ and, $I_{FINAL} = 10$ for IDLHC and IDLHC–ML, and the IDLHC–ML extra parameter m was first defined as equal to 50. We highlight that if the results were not satisfactory in relation to G2S simulator default, the parameter values of the optimizer can be modified.

Whenever the NPV statistical samples of IDLHC and IDLHC–ML were comparable to a specific N, F, and I_{FINAL}, we performed a new IDLHC test maintaining the premise that F x N should be constant and the new N value should be equal to the IDLHC–ML m value. The I_{FINAL} was then chosen as the minimal number that would guarantee that the number of simulations performed by IDLHC was at least equal to the ones performed by IDLHC–ML.

3.3.4 ML techniques and features

The sensitivity analysis was performed considering two groups of features: 1) the coefficients of the logistic equation for each well, and 2) the BHP over the time for each well originated from the logistic equation calculations. These features were normalized before the training process using min-max scaling method, presented in **Equation 3.3**.

$$x_{norm}^{i} = \frac{x^{i} - x_{min}}{x_{max} - x_{min}} \qquad \text{Equation 3.3}$$

where x_{norm}^{i} and x^{i} are the scaled and non-scaled values for a single feature and for the training example i, and x_{max} and x_{min} are the maximum and minimum values of a specific feature. This equation is applied to all n features. **Table 3.6** shows the ML techniques tested in this work. Specifically, the multilayer perceptron stacking learner (MLP-S) was trained on all predictions of the ML techniques from **Table 3.6**, excluding the support vector machine-stacking learner (SVM-S). Similarly, the SVM-S used all but the output of MLP-S. The hyperparameters and the number of folds used in k-fold cross-validation for each ML technique are presented in **Table 3.8** from Appendix A. The approach chosen for determining the combination of hyperparameter values tested was the randomized search (refer to Bergstra and Bengio, 2012).

Table 3.6–Machine learning techniques applied in this work.

Machine Learning technique	Abbreviation
Elastic-Net regression	ENET
Lasso regression	LSR
Bayesian ridge regression	BRR
Gradient boosting regression	GBR
Random forest regression	RF
Bagging regression with gradient boosting as base learner	BGR-GBR
Bagging regression with Bayesian ridge as base learner	BGR-BRR
Bagging regression with random forest as base learner	BGR-RF
Adaboost regression with gradient boosting as base learner	ABR-GBR

Adaboost regression with random forest as base learner	ABR-RF
Support vector machine stacking	SVM-S
Multilayer perceptron stacking	MLP-S

We emphasize that all ML algorithms of this work were applied using scikit-learn version 0.22 (Pedregosa et al., 2011), which is a Python open-source library incorporating several state-of-the-art ML algorithms. Therefore, this work had no objective to improve the existing ML techniques or scrutinize their best configuration or architecture. The intention was to try some well-known ML techniques and evaluate their performance on the case studied. For the explanation of how each ML algorithm works, please refer to Pedregosa et al. (2011).

3.4 Results

Henceforth, the term IDLHCX will be used here to refer to the IDLHC method, where X is the N parameter of IDLHC. Apart from this, the term IDLHCX–MLY_z represents the IDLHC–ML method where X, Y, and z represent N, m, and the machine learning technique used for the IDLHC–ML, respectively.

In order to evaluate the quality of the solutions yielded by the IDLHC algorithm for the first configuration tested ($N = 100$, $F = 20\%$, and $I_{FINAL} = 10$), we compared the NPV of the best strategies for every five executions with the one provided by the well control short-term strategy (S0). As presented in **Figure 3.6**, the IDLHC100 significantly outperformed (22.3% in average) the NPV of 2.55 USD billion from S0 strategy.

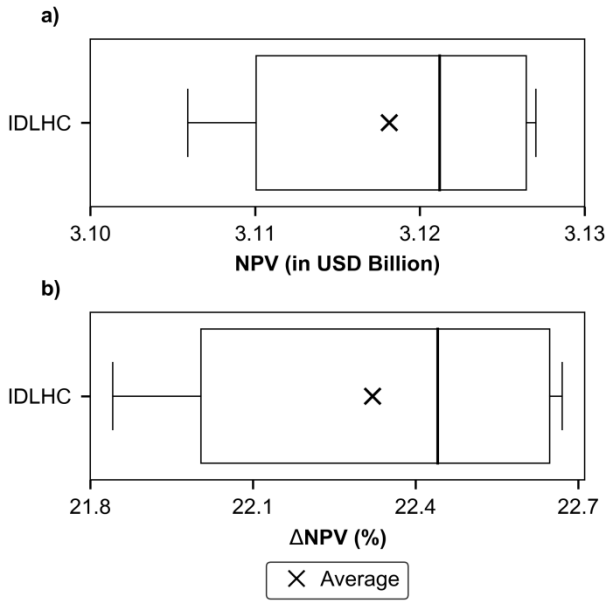

Figure 3.6–Boxplot of the net present value (NPV) from the best strategy of each IDLHC100 execution considering (a) the absolute values of NPV and (b) the NPV percentage variation (ΔNPV) in relation to

the one yielded by the well control short-term strategy (S0).

Subsequently, we reused the strategies generated during IDLHC100 nominal optimization process for the five executions to perform a sensitivity analysis of the ability of each ML technique to include the best $F\ x\ N$ strategies among the top fifty ($m = 50$) strategies predicted by the ML model from the second iteration onwards. Also, we assessed which features would be more suitable for the method proposed in this paper.

It is possible to observe that the ML techniques that employed the BHP features clearly yielded a better accuracy, calculated by equally averaging their value in all iterations and for each execution, than the features formed by the logistic equation coefficients (**Figure 3.7**). In addition, MLP-S presented the highest minimum, maximum, median and average accuracy. Thus, this ML technique was chosen to perform the IDLHC–ML method considering $N = 100$, $F = 20\%$, and $I_{FINAL} = 10$. We can note, however, that a few ML techniques provided very close accuracy (especially SVM-S) and they may provide similar results if applied to the IDLHC–ML method.

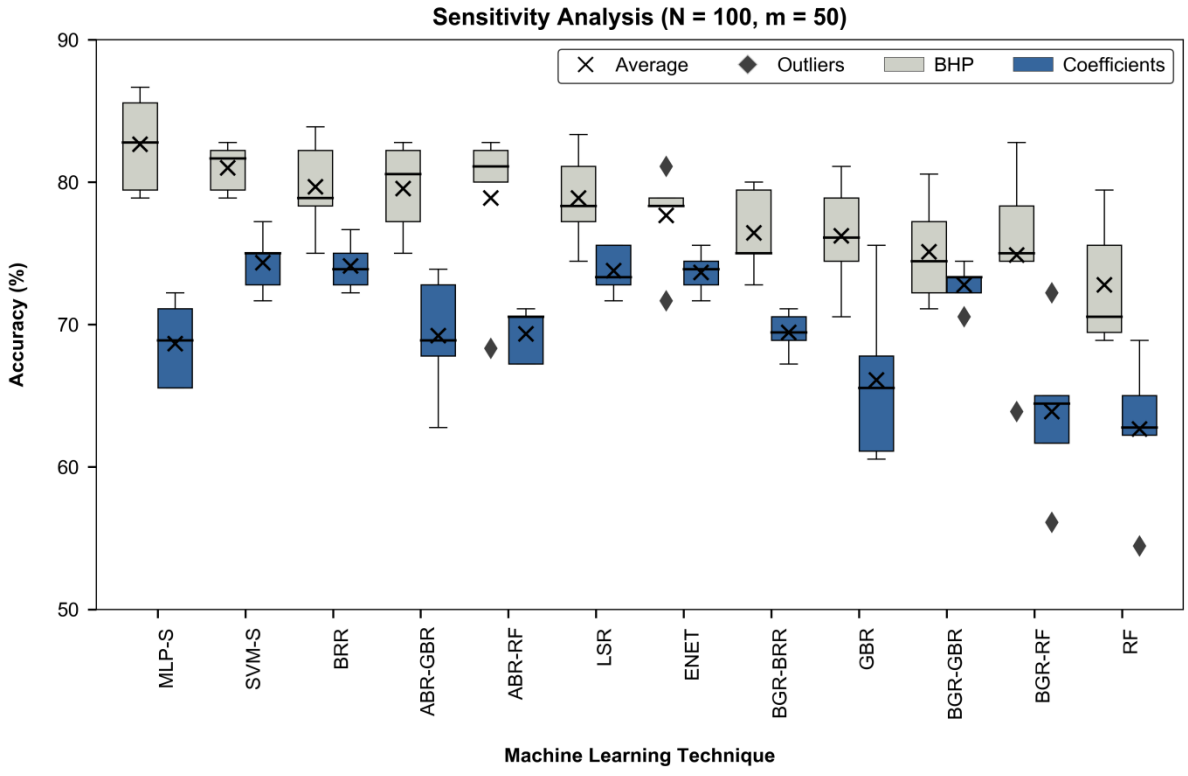

Figure 3.7–Boxplot of machine learning (ML) technique accuracy averaged over all iterations and for each of five IDLHC100 executions. The analysis was carried out considering two sets of features: the BHP and the logistic equation coefficients.

After the sensitivity analysis step, we executed the IDLHC100–ML50$_{\text{MLP-S}}$ five times. **Figure 3.8** demonstrates that the IDLHC100–ML50$_{\text{MLP-S}}$ provided a comparable $\overline{\text{NPV}}$ (-0.01%) and statistic samples in the five executions, in relation to the original IDLHC100, requiring 45% less simulations.

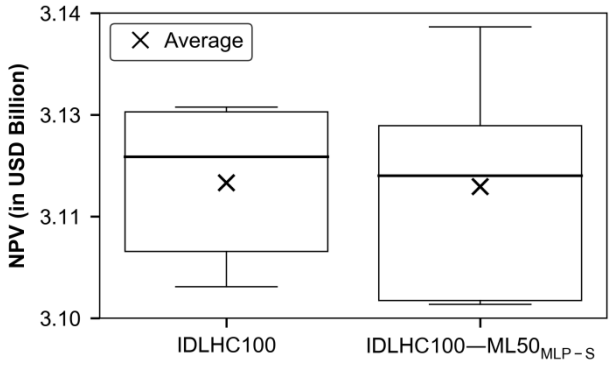

Figure 3.8–Boxplot of net present value (NPV) IDLHC100 and IDLHC100–ML50$_{\text{MLP-S}}$.

As the IDLHC100 and IDLHC100–ML50$_{\text{MLP-S}}$ had an equivalent performance, we carried out a second test for IDLHC, considering N, F, and I_{FINAL} equals 50, 40%, and 11, respectively. As seen in **Figure 3.9**, this new configuration delivered similar results in terms of NPV, requiring the same 550 simulations of IDLHC100–ML50$_{\text{MLP–S}}$.

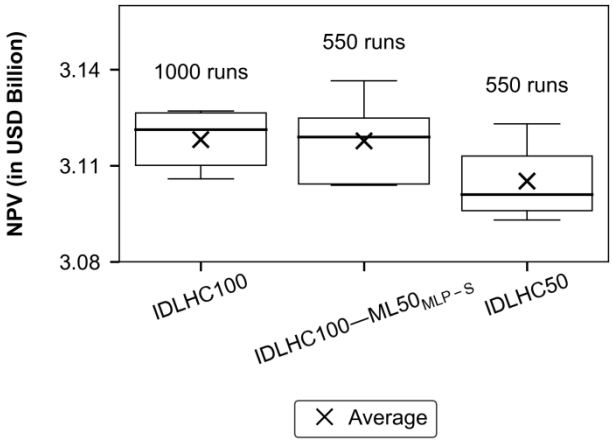

Figure 3.9–Boxplot of the net present value (NPV) for IDLHC100, IDLHC100–ML50$_{\text{MLP-S}}$ and IDLHC50.

Therefore, we executed the IDLHC–ML also using N, F, and I_{FINAL} equals 50, 40% and 11, respectively. For choosing the m value and the ML algorithm for IDLHC–ML, we once again executed the sensitivity analysis using the BHP as features and the simulations already performed by IDLHC50. This was done because the new optimizer parameters can influence the ML algorithm accuracies and, thus, the m value of the first test may not be suitable for the second one.

The values analyzed for m was 25 and 30. It can be seen from **Figure 3.10** that the SVM-S provided the highest average accuracy among all ML techniques for both m equals 25 and 30. Also, the SVM-S average accuracy using $m = 25$ (75.6%) was considerably lower compared to the first sensitivity analysis test (82.6%), but SVM-S guaranteed a similar average accuracy with $m = 30$ (84.3%). Therefore, we opted to perform the IDLHC–ML using the SVM-S and $m = 30$.

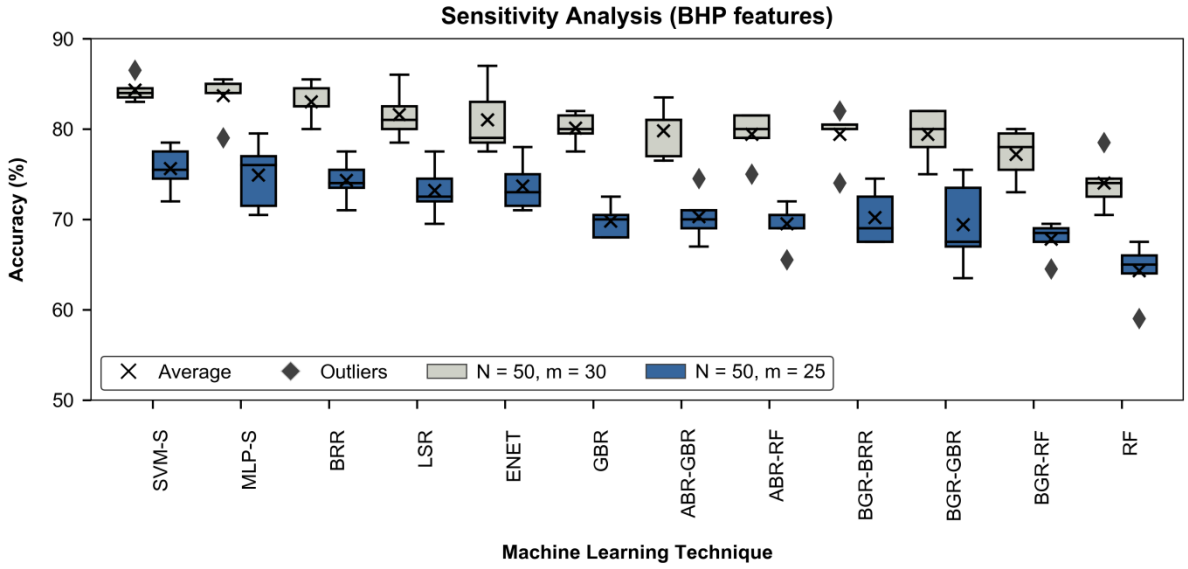

Figure 3.10–Boxplot of machine learning (ML) technique accuracy averaged over all iterations and for each of the five IDLHC50 executions considering m = 25 and m = 30. The analysis only considered the BHP as features.

It is possible to note from **Figure 3.11** that the IDLHC50–ML30$_{\text{SVM-S}}$ delivered a worse result compared to IDLHC100 for all metrics analyzed (mean, median, minimum, maximum and dispersion). Therefore, we carried out the IDLHC50–ML35$_{\text{SVM-S}}$ followed by the IDLHC35. The mean in the five IDLHC50–ML35$_{\text{SVM-S}}$ executions was slightly higher than IDLHC50 (+0.2%) and required roughly 27% less simulations runs. On the other hand, the IDLHC35 delivered the worst result among all other IDLHC and IDLHC–ML configurations, thus, the methodology process was finished.

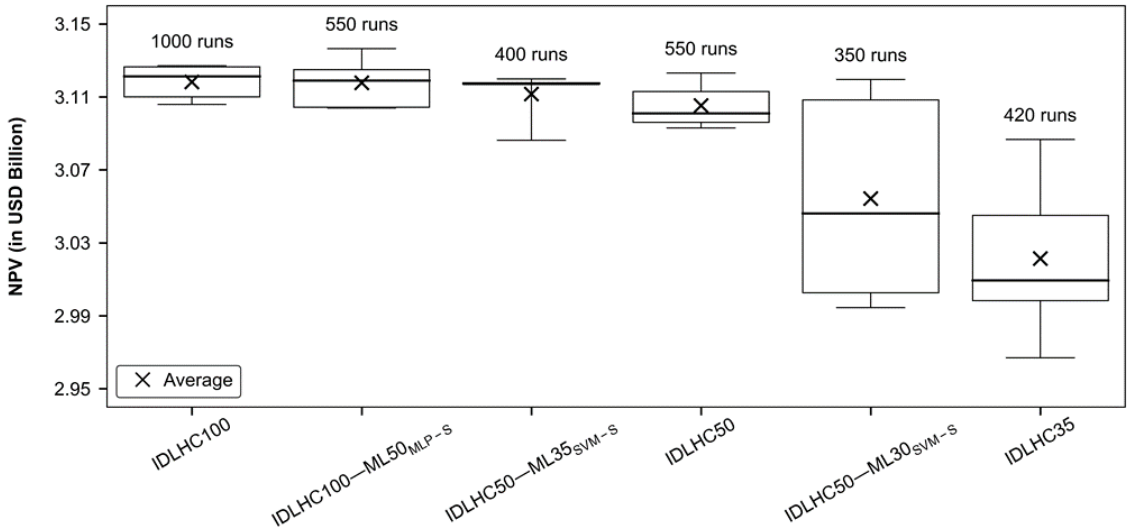

Figure 3.11–Boxplot of net present value (NPV) for all IDLHC and IDLHC–ML configurations tested in this work.

Table 3.7 summarizes the results for all IDLHC and IDLHC–ML configurations tested in this work and compares them with the values achieved by the IDLHC100. We observe that the NPV results worsened as the number of simulations for the IDLHC decreased. However, regardless of the IDLHC configuration, the IDLHC–ML proposed in this work reduced the number of simulations and kept the NPV statistical samples results very close.

Table 3.7–Summary of the IDLHC and IDLHC–ML parameter configurations and of the net present value (NPV) percentage variation in relation to IDLHC100.

Configuration and method	NPV percentage variation (%)				Configuration				
	[1]Max	[2]Min	[3]Avg	Median	N	F (%)	I_{FINAL}	m	No simulations
IDLHC100	–	–	–	–	100	20	10	–	1000
IDLHC100–$ML50_{MLP-S}$	0.30	-0.07	-0.01	-0.07	100	20	10	50	550
IDLHC50	-0.13	-0.42	-0.42	-0.65	50	40	11	–	550
IDLHC50–$ML35_{SVM-S}$	-0.23	-0.63	-0.21	-0.12	50	40	11	35	400
IDLHC50–$ML30_{SVM-S}$	-0.24	-3.59	-2.05	-2.41	50	40	11	30	350
IDLHC35	-1.29	-4.47	-3.11	-3.58	35	57	12	–	420

[1] Maximum. [2] Minimum. [3]Average.

Figure 3.12 demonstrates that the IDLHC100–$ML50_{MLP–S}$ and the IDLHC50–$ML35_{SVM–S}$ had a better convergence rate than all IDLHC configurations while maximized the NPV to a close-to-optimum value obtained by IDLHC100.

Considering the results obtained, the IDLHC–ML techniques are a good alternative to the original IDLHC, mainly when there is limited time to define G2L strategies, or even to previously assess which G2L control rule would perform better in the field studied.

We simulated the IDLHC and IDLHC–ML optimization method considering 16 simulations in parallel and six processors for each simulation. For such configuration, the average time to simulate and calculate the NPV of a single strategy was roughly 127 seconds. In contrast, the time to perform MLP-S and SVM-S flowchart per simulation was significantly lower (roughly 10 seconds). Therefore, IDLHC100–ML50$_{MLP\text{-}S}$ saved roughly 40% in computational time compared to IDLHC100 and the IDLHC50–ML35$_{SVM\text{-}S}$ reduced the time in approximately 22% compared to IDLHC50. In **Appendix B.1**, we provide the time for training the models, along with IDLHC elapsed time.

Additionally, it is important to mention that the reservoir simulation computational cost should be higher as the complexity of the reservoir model increases. Therefore, as it does not affect the ML efficiency, the percentage of time saved using IDLHC–ML would get closer to the percentage reduction of simulation runs (45% and 27% for the two IDLHC parameter configurations tested).

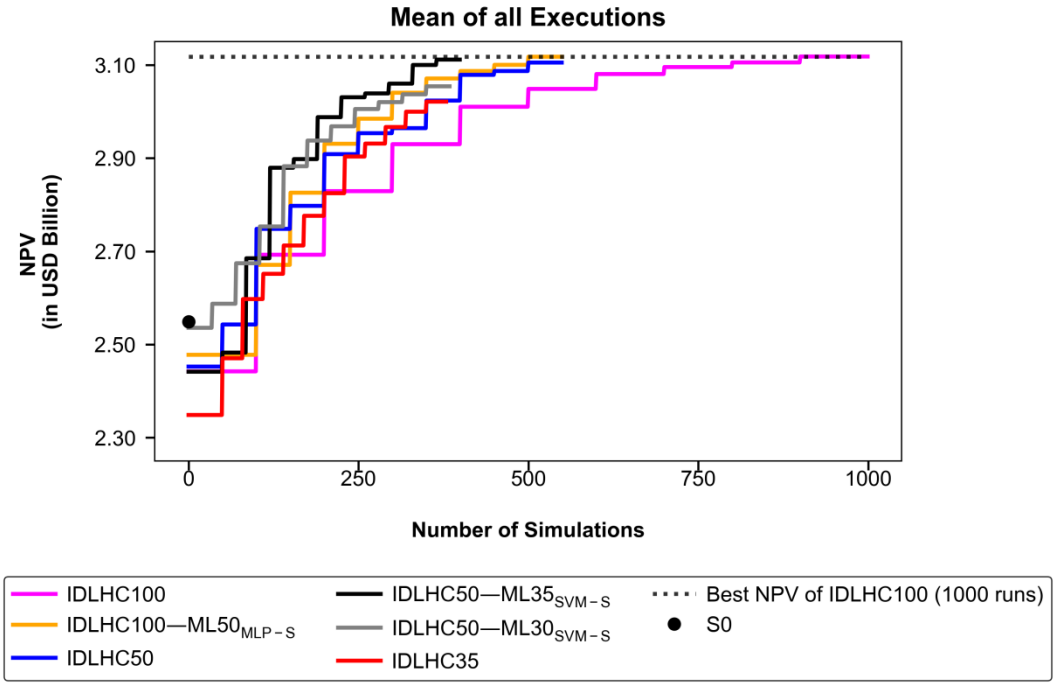

Figure 3.12–Comparison of the average of the maximum net present value (NPV) per iteration calculated over the five executions performed by each IDLHC and IDLHC–ML parameter configurations.

3.5 Conclusions

By combining the IDLHC optimization algorithm with machine learning techniques (IDLHC–ML) to perform a nominal optimization of long-term well control rules, it was

possible to decrease the number of reservoir model simulation runs while maintaining an equivalent net present value (NPV) compared to the original IDLHC method.

We tested three IDLHC sets of parameter configurations, which are denoted here as IDLHC100, IDLHC50, and IDLHC35 and these required 1000, 550, and 420 simulation runs, respectively. The IDLHC configurations with a larger number of simulations provided better quality solutions in terms of NPV. However, the results from IDLHC100 and IDLHC50 were very close (within 0.4%). The IDLHC–ML versions for the former and for the latter (namely here as IDLHC100–ML50$_{\text{MLP-S}}$ and IDLHC50–ML35$_{\text{SVM-S}}$, respectively) were able to reduce the number of simulations in 45% and 27 %, respectively, while maintaining an equivalent NPV. Considering the time to perform the ML technique, the computational time saved for IDLHC100–ML50$_{\text{MLP-S}}$ and IDLHC50–ML35$_{\text{SVM-S}}$ was roughly 40% and 22% in the order given. For a more complex simulation model, the reduction in time tends to get closer to the percentage reduction of simulation runs as the time to train the ML model is not influenced by the model's complexity.

We also showed that a different configuration of the IDLHC optimizer parameters affects the selection of the ML technique and the minimum number of best strategies predicted by the ML technique to be simulated (m) to render the IDLHC–ML method effective.

Finally, once we have tested the ML for well control optimization in a nominal case, the next step should focus on decision-making strategy selection, which includes uncertainties and fine-scale reservoir analysis. Another line for future work would be trying to perform IDLHC–ML considering other types of variables (e.g.: well type and placement) and another configuration for IDLHC parameters (N and F). These investigations would confirm the ability of the method to be generalized or to define in which cases the IDLHC–ML is more suitable to be applied.

For future study cases, there is no need to test other machine learning algorithms again. We recommend directly applying the Multi-Layer Perceptron Stacking (MLP-S), which yielded the best results in this paper. The rationale for this choice is thoroughly explained in **Appendix B.2**.

3.6 Nomenclature

ABR-GBR	Adaboost regression using gradient boosting as base-learner
ABR-RF	Adaboost regression using random forest as base-learner

a_{we}, b_{we}, c_{we}	Coefficients of parametric equations
a_{we_lim}, b_{we_lim}, c_{we_lim}	Parametric equation coefficient limits
BGR-BRR	Bagging regression using Bayesian ridge as base-learner
BGR-GBR	Bagging regression using gradient boosting as base-learner
BGR-RF	Bagging regression using random forest as base-learner
BHP	Bottom-hole pressure
BHP_{Max}	Maximum BHP limit of the reservoir
BHP_{Min}	Minimum BHP limit of the reservoir
$BHP_{we,t}$	BHP in well we for time t
BRR	Bayesian ridge regression
DLHC	Discrete Latin hypercube sampling
ENET	Elastic-Net regression
F	Threshold cut percentage to select the best samples
G1	Design variables
G2	Well control variables
G2L	Long-term control rules
G2S	Short-term effective control
G3	Revitalization variables
GBR	Gradient boosting regression
h	Generic representation for base-learner
I	Iteration index from IDLHC
i	Index representing a single training example
IDLHC	Iterative discrete Latin hypercube
IDLHC–ML	Iterative discrete Latin hypercube combined with Machine Learning
I_{FINAL}	Last iteration from IDLHC
LSR	Lasso regression
Max	Maximum value of determined indicator
m	Number of best strategies predicted by ML model
Min	Minimum value of determined indicator
ML	Machine Learning
MLP	Multilayer perceptron

MLP-S	Stacking using multilayer perceptron as the final meta-learner
N	Sample size generated in each IDLHC iteration
NPV	Net present value
OF	Objective Function
P50	Percentile 50
PDF	Probability density function
RF	Random forest regression
RM	Representative models
S0	Reference well control short-term strategy
SVM-S	Stacking using support vector machine as the final meta-learner
t	Time passed from initial management date
t_{FINAL}	Management period interval
USD	United States Dollar
x	Generic representation of a feature
x_{norm}^{i}	Generic representation of a normalized feature for example i
x_{max}	Maximum value for a specific feature
x_{min}	Minimum value for a specific feature
y	Generic representation of a label

3.7 Acknowledgements

This work was conducted with the support of *Energi Simulation* and *Petrobras* within the ANP R&D tax as “commitment to research and development investments”. The authors are thankful for the support of the Center of Petroleum Studies (CEPETRO-UNICAMP/Brazil), the Department of Energy (DE-FEM-UNICAMP/Brazil) and Research Group in Reservoir Simulation and Management (UNISIM-UNICAMP/Brazil). Also, a special thanks to *CMG* for software licenses.

3.8 Appendix A

Table 3.8 presents the hyperparameters and the k-fold cross-validation specifications to train each ML technique employed in this work. A randomized search (Bergstra and Bengio, 2012) was applied to define the combination of hyperparameter values to be tested. The hyperparameters from the ML algorithms that are not displayed in the in **Table 3.8** were set with their default value, and these can be confirmed in Pedregosa et al. (2011).

Table 3.8–Machine learning (ML) algorithms, hyperparameters, and k-fold cross-validation specifications for training the data in the sensitivity analysis and evaluation steps.

ML algorithm		[1]No of folds	No of combinations of the hyperparameters per fold	Hyperparameters
Individual learner	Elastic-Net	10	10	'alpha': (10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 100); 'l1_ratio': [0.0, 1.0] equally spaced in 10 values
	Lasso	10	10	'alpha': [-6, -1] logarithmic spaced in 40 values
	Bayesian ridge	10	10	'alpha_1': (10^{-6}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 3); 'alpha_2': (10^{-6}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 3); 'lambda_1': (10^{-6}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 5, 10); 'lambda_2': (10^{-6}, 10^{-3}, 10^{-2}, 10^{-1})
Ensemble learner	Gradient boosting	5	10	'learning rate': (0.1, 0.5, 0.7); 'subsample': (0.5, 0.7, 1.0); '[2]max depth': (2, 10, 20, 50); '[3]min samples leaf': (5, 20, 45, 75); 'No estimators': (250, 500, 750); 'min samples split': (5, 20, 45, 75), 'loss': ([4]'ls', [5]'lad', 'huber', 'quantile'); 'max features': (0.4, 0.5, 0.8, 1.0); 'max leaf nodes': (10, 50, 100)
	Random forest	5	10	'bootstrap': (True, False); 'max depth': (5, 10, 50, 100); 'max features': (0.3, 0.7, 5, 20); 'min samples leaf': (5, 20, 45, 75); 'min samples split': (5, 20, 45, 75); 'No estimators': (75, 100, 125); 'max leaf nodes': (10, 50, 100)
	Bagging regression with gradient boosting as base learner	-	-	'No estimators': (10). Hyperparameter tuning not performed
	Bagging regression with bayesian ridge as base learner	-	-	'No estimators': (10). Hyperparameter tuning not performed
	Bagging regression with random forest as base learner	-	-	'No estimators': (10). Hyperparameter tuning not performed
	Adaboost regression with gradient boosting as base learner	-	-	'No estimators': (10). Hyperparameter tuning not performed

	Adaboost regression with random forest as base learner	-	-	'No estimators': (10). Hyperparameter tuning not performed
	Support vector machine stacking	10	20	'gamma': (1e-6, 1e-5, 1e-4, 1e-3, 0.005); '[6]C': (1e-3, 0.05, 0.01, 5, 10, 50, 100, 250); 'kernel': ('linear','[7]rbf','sigmoid'); 'shrinking': (True, False)
	Multilayer perceptron stacking	10	20	'Initial learning rate': [-4, 0] logarithmic spaced in 8 intervals ''hidden layer sizes': ((10, 10, 10, 10), (10, 10, 10, 10, 10, 10), (100, 100, 100, 100)); 'activation': ('relu'), 'solver':('adam'), 'alpha': (10^{-5}, 10^{-4}, 10^{-3}, $5*10^{-3}$, 10^{-2}, 10^{-1}, 1); 'learning rate':('adaptive'), 'max iter':(1000); 'early stopping': (True)

[1]Number. [2]Maximum. [3]Minimum. [4]Least squares regression. [5]Least absolute deviation. [6]Regularization parameter. [7]Radius basis function.

4 Optimizing Well Control Strategies with IDLHC-MLR: A Machine Learning Approach to Address Geological Uncertainties and Reduce Simulations

Authors: Santos, D. R., Fioravanti, A. R., Botechia, V. E., Schiozer, D. J.

Presented at the Offshore Technology Conference Brasil held in Rio de Janeiro, Brazil, October, 2023.

Abstract

This paper presents an advanced version of the previous IDLHC-ML approach, designed to enhance life-cycle well control optimization by reducing simulations. Unlike its predecessor, this updated method, called IDLHC-MLR, uses representative models (RMs) to address the effect of geological uncertainties on production strategies. Despite presenting additional computational challenges, considering uncertainties in determining effective strategies is crucial, making the new IDLHC-MLR approach a valuable solution.

The IDLHC-MLR combines the iterative discrete Latin hypercube optimization algorithm (IDLHC) with machine learning (ML) to robustly optimize the well's bottom-hole pressure (BHP) throughout the field management period. The method is applied to the UNISIM-I-M benchmark of Namorado Field, located in the Campos Basin, Brazil. The IDLHC-MLR method trains the initial ML model with well BHP strategies robustly applied to all RMs in the first iteration of IDLHC. In subsequent iterations, the trained ML model is used to predict the expected monetary value of the RMs, and only a subset of new strategies with the highest expected outcome is selected for simulation. In addition, the ML algorithms are retrained with newly generated strategies over the iterations to improve the model's accuracy.

The IDLHC-MLR incorporates stacked ensemble learning, which leverages predictions from various base machine learning models to train a secondary algorithm. In this approach, the IDLHC-MLR employs multiple base learners such as Lasso, Gradient Boosting, and Random Forest to make predictions, which are then inputted into a multi-layer perceptron for training purposes. This integration of multiple base models results in a more robust and accurate prediction and provides a 45% reduction in the number of simulations required compared to the traditional IDLHC method while maintaining similar expected monetary

value. To conclude, utilizing inexpensive ML models effectively reduces computational time by substituting costly full-physics reservoir simulations.

The significant computational time required for full-physics simulations, particularly when considering multiple scenarios to account for uncertainties, can pose a challenge to meeting project deadlines. The IDLHC-MLR methodology, incorporating low-cost ML models, offers a practical solution to reduce computational time, increasing the likelihood of successful project implementation within the given timeline.

4.1 Introduction

Defining a suitable production strategy in a petroleum field is a complex task due to the many uncertain factors involved (e.g., reservoir characteristics, market conditions, and technical constraints) and the numerous variables that must be considered.

Failing to properly account for significant uncertainties when making decisions can lead to biased simulation models. This can result in a production strategy that may seem optimal in the simulation models but is far from being the most effective solution when implemented in the actual field. Botechia et al. (2018), for example, conducted a study that identified petrophysical uncertainties as the cause of discrepancies observed in a case where reservoir simulation models were used. Although the simulation models showed an increase of 29% in the expected monetary value (EMV), the actual net present value (NPV) decreased by 2% in the reference case. Therefore, accurately quantifying and reducing uncertainties is critical to mitigate the risk of making costly errors and to optimize the economic outcome of oil projects (Schiozer et al., 2004; Santos et al., 2017a).

In general, many uncertainties (e.g., geological, fluid properties) are identified and mapped during reservoir characterization. These uncertainties may be combined using sampling techniques (Schiozer et al., 2017) to generate a feasible number of scenarios that can be used for numerical reservoir simulation. Data assimilation methods (Naevdal et al., 2005; Gu and Oliver, 2007; Maschio and Schiozer, 2016) are then applied to reduce uncertainties, resulting in a set of simulation models for optimizing production strategies. As new data, such as dynamic and seismic data, becomes available over time, the reservoir characterization, data assimilation, and production strategy optimization process can be revisited through a closed-loop approach (Lorentzen et al., 2009; Schiozer et al., 2019; Hidalgo et al., 2017; Shirangi, 2019).

Two approaches, ensemble nominal optimization and robust optimization, can be used to develop optimal production strategies that account for uncertainties (Paiaman et al., 2021). Nominal ensemble optimization involves optimizing a deterministic objective function(s) for each individual model resulting from the data assimilation process, leading to one production strategy per model. Robust optimization (Hanea et al., 2019; Jahandideh and Jafarpour, 2020), on the other hand, involves optimizing a probabilistic objective function(s) over the set of data assimilated models, creating a single production strategy with the best average performance for the entire ensemble of models.

Optimizing the production strategy through ensemble nominal optimization or robust optimization approaches can be costly when multiple scenarios need to be considered to accurately represent the real field and when numerous variables need to be optimized. This computational burden can slow down the decision-making process, leading to delays in production optimization and reducing economic performance. To mitigate these issues, various methods have been studied to reduce the computational effort required while maintaining the accuracy of the results. These methods include simplifying the model by using low-fidelity-models (Kou et al., 2022; Ramos et al., 2012), or proxy models (Bahrami and James, 2023; He et al., 2018; Chen et al., 2017) to represent high-fidelity models, dividing the reservoir model into sectors and using them as isolated models (Pires et al., 2020; Dzyuba et al., 2012), simulating only a partial lifetime of the field (Loomba et al., 2022), grouping and optimizing variables hierarchically (Gaspar et al., 2016b; Humphries et al., 2014), reducing dimensionality by parametrizing optimization variables (Pinto et al., 2019; Santos et al., 2021; Santos et al., 2023; Sorek et al., 2017), representing the entire ensemble of models with a subgroup called representative models (Meira et al., 2017; Meira et al., 2020; Sarma et al., 2013), and applying machine learning to reduce the number of simulations (Jang et al., 2018; Santos et al., 2020).

For instance, Meira et al. (2017) proposed the RMFinder 2.0 tool, which employs a mathematical function to identify representative models (RMs) by considering both the probability distribution of input variables (e.g., reservoir and operational uncertainties) and the variability of important output variables (e.g., production and injection forecasts). The utilization of RMs enables the implementation of RM robust optimization (Silva et al., 2016; Loomba and Schiozer, 2022; Peters et al., 2010), as well as RM nominal optimization (Morosov and Schiozer, 2017; Perrone et al., 2015; Santos et al., 2021; Santos and Schiozer, 2017; Silva et al., 2019). These approaches involve performing optimization in the RMs, rather than in the ensemble of models selected during the data assimilation process. To obtain a more

precise evaluation of the risk curve, the RM optimal strategies are generally assessed in the models' ensemble before their implementation in the actual field.

Santos et al. (2020) addressed the issue of computational time by reducing the number of simulations in the iterative discrete Latin hypercube (IDLHC) optimization algorithm. They accomplished this by combining the algorithm with machine learning techniques, which enabled them to optimize the well bottom-hole pressure (BHP) over time with fewer simulations. They tested different machine learning algorithms and IDLHC parameters to determine the most efficient configuration. They were able to achieve significant computational savings without sacrificing performance. However, it is important to note that the optimization was performed nominally, which means that the method may not be suitable when uncertainties are considered.

In this context, our study introduces the IDLHC-MLR method, designed to enhance the IDLHC-ML algorithm by incorporating uncertainties through the use of RMs. To achieve this, we modified the IDLHC-ML to perform RM robust optimization and assessed whether the results were still effective. These modifications required adjusting the objective function and target variables used to train the models. It is essential to validate this method under uncertainties since ignoring uncertainties can result in solutions that may appear optimal but are considerably suboptimal when implemented in the actual field.

Once the IDLHC-MLR approach for optimizing the best strategy for RMs has been performed, we evaluate the gains obtained with this strategy in a large set of data assimilated models and for the reference case, which aims to represent the real field. This validation is crucial to ensure that the selected RMs are appropriate and that the provided production strategy is robust enough to yield positive results in the real case, where information is limited. It is important to note that when dealing with uncertainties, more models are considered, leading to an increase in the number of simulations required.

4.2 Methodology

In the methodology section, we introduce the two optimization methods to be compared: the IDLHC-MLR and the IDLHC. Both methods are employed to optimize the coefficients of the logistic equation, which defines the BHP profile over time for each individual well. These methods are compared with a base strategy (S0) that serves as the benchmark for EMV. This allows us to evaluate whether the solutions obtained by the

optimization algorithms are satisfactory. The S0 strategy is a default control rule internally defined by the commercial simulator IMEX 2016. It determines the allocation of production rates among wells based on a priority ranking system (for further details, refer to Santos et al., 2021).

To account for uncertainties, we perform a RM robust optimization, and apply the optimal robust strategies to a large set of scenarios adhering to the field's production history data and to the reference model, which represents the actual field (more details about the reference model can be consulted in the Study Case Section). Evaluating the proposed method in both the reference and production data assimilated models enhances the reliability of the IDLHC-MLR framework for real-world field applications.

The following subsections provide a comprehensive description of the logistic equation, the machine learning framework utilized for training, and the functioning of both the IDLHC and IDLHC-MLR methods.

4.2.1 Logistic equation parameterization

The IDLHC-MLR and IDLHC methods are employed to optimize the coefficients a_{we}, b_{we}, c_{we} of the logistic equation, **Equation 4.1**, which defines the BHP of the wells over time (Santos et al., 2021).

$$BHP_{we,t} = BHP_{Min} + \frac{(BHP_{Max} - BHP_{Min})}{1+e^{\left(a_{we} + b_{we} * \left(\frac{t}{t_{final}}\right) + c_{we} * \left(\frac{t}{t_{final}}\right)^2\right)}} \qquad \text{Equation 4.1}$$

where $BHP_{we,t}$ refers to the BHP in well we at time t. BHP_{Max} and BHP_{Min} represent the maximum and minimum constraints of BHP for the reservoir, respectively. Finally, t_{final} denotes the entire management period under consideration.

We have opted to employ this parameterization method because its demonstrated superiority over other parametric equations and established techniques, such as shutting down producers based on specific criteria. Additionally, the logistic equation is advantageous as it avoids subtle changes in BHP over time, making it a practical approach.

To optimize the BHP limits over time, we discretized the coefficients a_{we}, b_{we}, c_{we} of the logistic equation into 15 linearly spaced values within the ranges specified in **Table 4.1**. The selection of these specific values is well explained in Santos et al. (2021).

Table 4.1–Range of values to optimize the coefficients of the logistic equation.

Coefficients	Range [Minimum, Maximum]
a_{we}	[-5, 5]
b_{we}	[-10, 10]
c_{we}	[-15, 15]

4.2.2 Machine learning training framework

In **Figure 4.1** we described the machine learning framework to train the algorithms used in this paper.

1. The first step involves defining the input features and target variable to be used in the machine learning algorithm.
2. Next, we normalize the input features and target variable to ensure they are on the same scale. Scaling helps to improve the performance and stability of machine learning algorithms by preventing features with larger values or different ranges from dominating the training process. In this study, we used the min-max normalization method.
3. The third step is to select the machine learning algorithm to be trained using the input features and target variable.
4. Afterwards, we specify and discretize the hyperparameters to be tested during the training of each machine learning algorithm.
5. We then perform k-fold cross-validation to evaluate the algorithm's performance and ensure it can generalize well to new data. In k-fold cross-validation the dataset is split into k equal-sized subsets, and the model is trained on $k - 1$ subsets and evaluated on the remaining subset in each iteration. This process is repeated k times, with each subset used for validation exactly once. The results obtained from k-fold cross-validation enable us to estimate the model's performance on independent validation data and help us determine the optimal hyperparameters for the model.
6. Once the machine learning algorithm has been trained using cross-validation, we evaluate its performance by calculating the average root mean squared error (RMSE) across all folds.
7. If additional hyperparameters need to be tested, the process returns to step 5 to train and evaluate the algorithm with the new set of hyperparameters. If the performance is

satisfactory, and no additional hyperparameters need to be tested, we move on to testing other algorithms as needed.

8. To enhance the model's performance, we adopt a stacking ensemble learning approach. This involves training multiple base-learners on the same dataset to predict the target variable. Their predictions are then used as input features to train a meta-learner, which makes the final prediction. In this work, we use the base-learners shown in **Figure 4.2a** and the multilayer perceptron (MLP) algorithm as meta-learner (**Figure 4.2b**). The base-learners use the injectors' and producers' BHP over time as input features, and the EMV as the target variable. Finally, the MLP is trained using the predicted EMV values obtained from each base-learner as input features, with the true EMV values serving as the target variable.

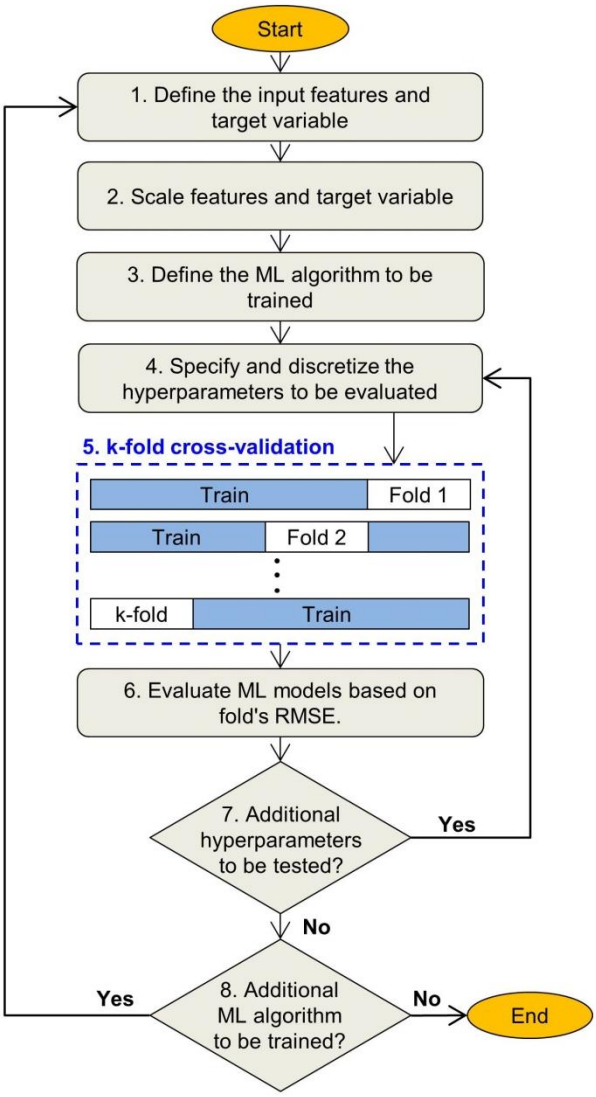

Figure 4.1–Machine learning framework for training the algorithms in this work.

The stacking ensemble learning approach, utilizing the specified base learners and meta-learner algorithms, was chosen due to its ability to deliver superior results, as demonstrated in Santos et al. (2020). We employed Scikit-learn, an open source data analysis library for Python, to implement machine learning algorithms. For a detailed description of the Scikit-learn tool and of the functioning of these algorithms, please refer to Pedregosa et al. (2011).

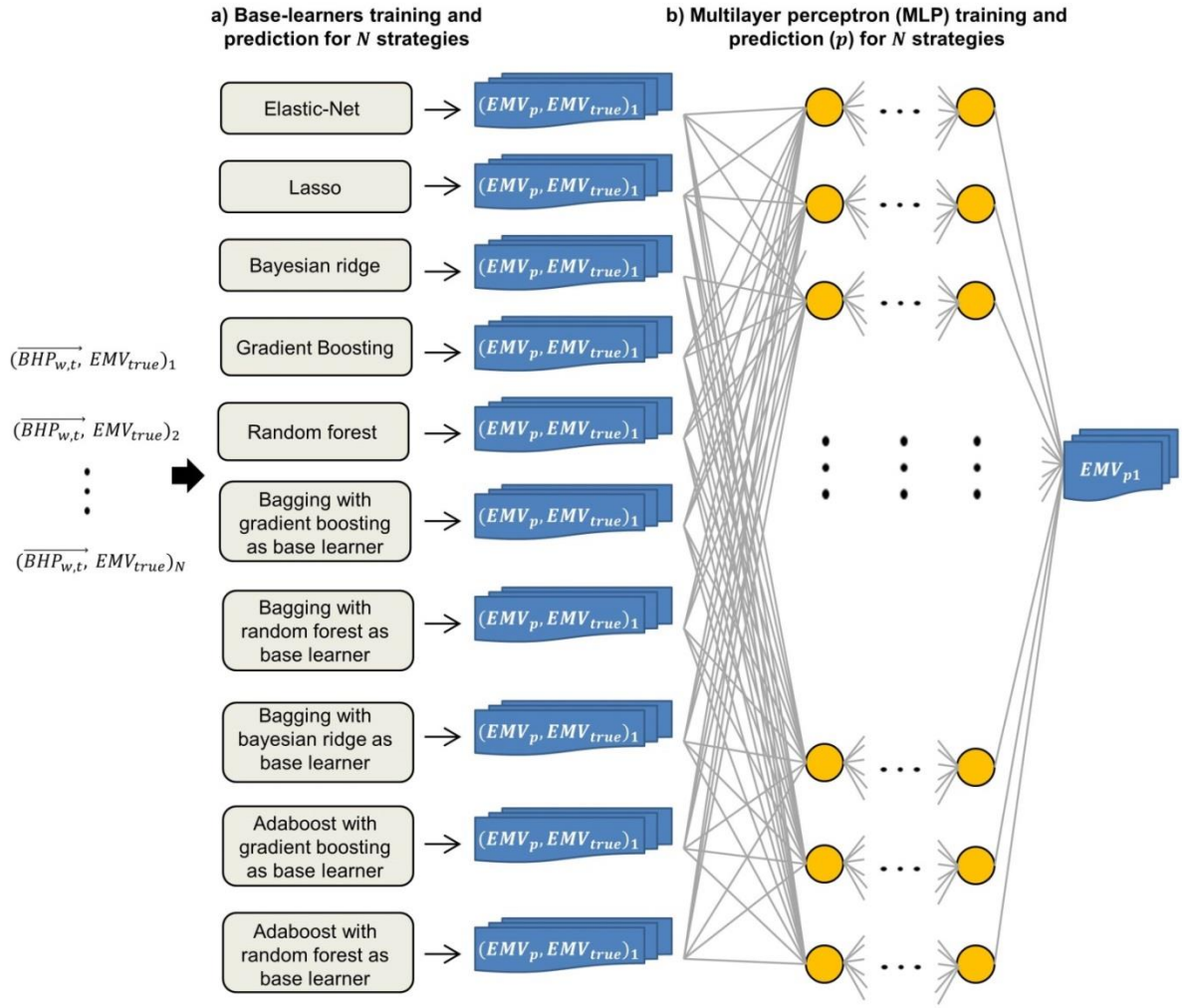

Figure 4.2–Diagram of the stacking learning approach, depicting the utilized base learners, the chosen meta learner, and the pair of features and target variables at each stage of the stacking learning training process.

4.2.3 Iterative Discrete Latin Hypercube (IDLHC)

In **Figure 4.3**, the regular IDLHC is depicted by the black arrows, which includes the following steps:

- Discretize each optimization variable into levels and define the initial probability mass function (PMF). Typically, in the initial iteration of the optimization process, the probability of occurrence for each level of an optimization variable is assumed to be equal (uniform distribution).
- Generate N samples (or strategies) using the Discrete Latin Hypercube (DLHC) sampling method, as described by Maschio and Schiozer (2016), based on the prior PMF of each variable. Each strategy is then combined with each RM. As a result, the number of models to be evaluated is the product of the number of RMs (N_{RM}) and N strategies.
- Simulate the $N_{RM} \times N$ models and calculate the expected monetary value (EMV) for each strategy. The EMV is determined by taking a weighted average of the net present value (NPV) in each RM, with weights derived from the probability of each RM occurring.
- Update the variables' PMF using the F percent best strategies based on the EMV. The updated PMF is then used as the input for the next iteration (I) of the IDLHC method.

This process is repeated until a stopping criterion is met. In this study, we used the number of iterations as the stopping criterion.

4.2.4 Robust Iterative Discrete Latin Hypercube combined with machine learning (IDLHC-MLR)

The IDLHC-MLR method integrates the steps indicated by the green arrows in **Figure 4.3** (consult **Appendix C** for the revised figure) into the IDLHC. The initial iteration of this process mirrors the traditional IDLHC approach, but with the addition of utilizing a set of simulation runs from the first iteration to train the initial machine learning (ML) models. These steps happen in parallel with the PMF update and sample generation steps. From the second iteration onwards, the following iterative steps are performed:

- Generate N strategies using the DLHC method based on the prior PMF of each variable.
- Predict the EMV for the N strategies using the ML model trained in the previous iteration.
- Rank the top m strategies among N strategies.
- Simulate the $N_{RM} \times m$ strategies and calculate their EMV.

- Train the base learners using BHP strategies-EMV pairs. The number of samples used in the training process is given by $N + m \times (I - 1)$. Each base learner delivers an EMV vector that serves as input to train the multi-layer perceptron, which is the meta-learning algorithm adopted in this work.
- Update the variables' PMF using the F percent best strategies based on the EMV.

These steps are repeated until the same number of iterations as in the traditional IDLHC framework is reached. At each iteration, the ML algorithms are retrained with the newly simulated strategies to further improve prediction.

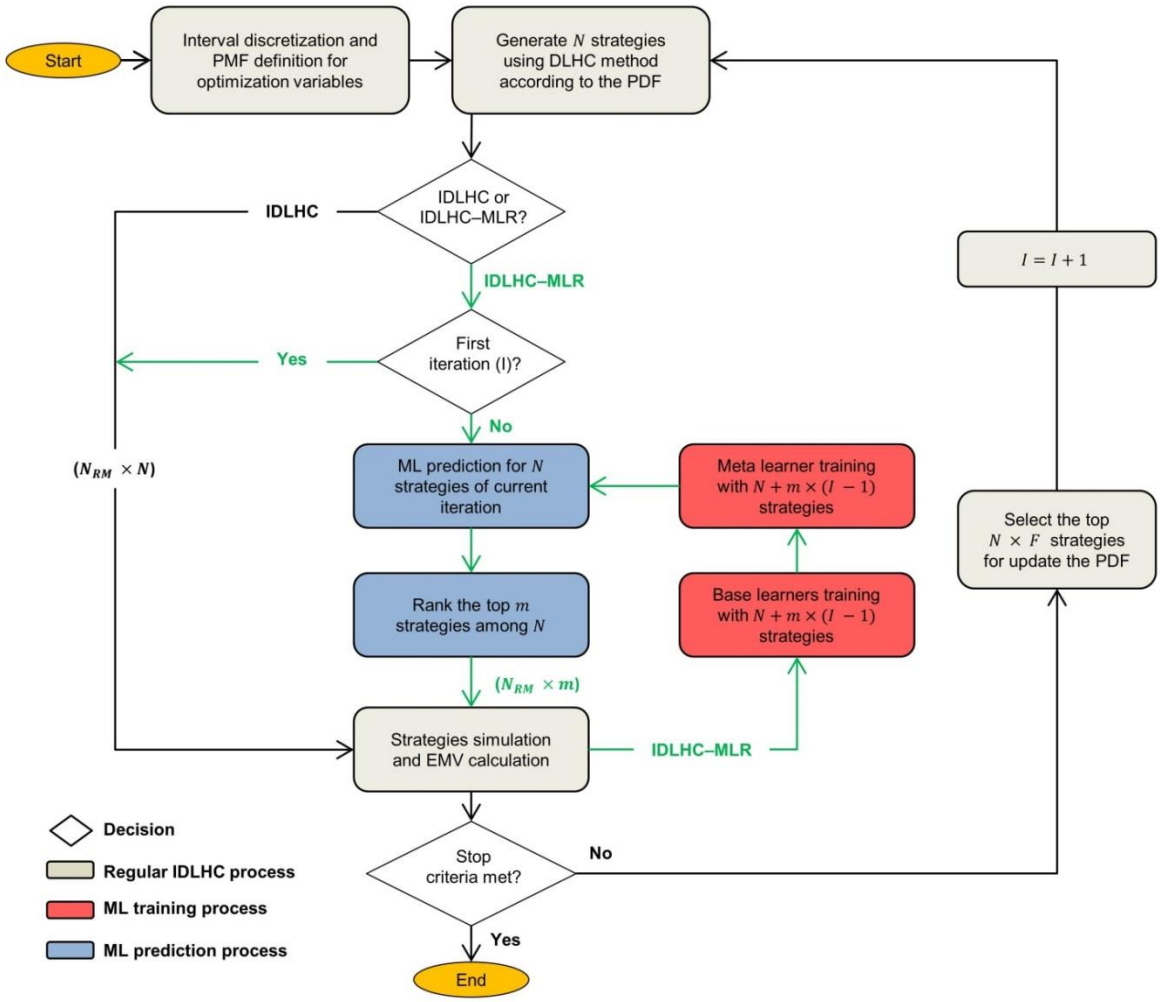

Figure 4.3–Flowchart of IDLHC and IDLHC-MLR methods of optimization.

4.3 Study Case

This section provides a technical overview of the UNISIM-I-R (Avansi and Schiozer, 2015) and the UNISIM-I-M benchmark case (Gaspar et al., 2016a), which were used to

evaluate the efficacy of our methodology. The UNISIM-I-R is a high-fidelity simulation model with roughly 3.5 million active cells organized in a grid of 326 × 234 × 157 blocks. It was constructed utilizing public petrophysical, facies, and structural data extracted from the Namorado Field situated in the Campos Basin, Brazil. This serves as a standard for assessing the feasibility of various methodologies before they are practically implemented in real reservoirs. The UNISIM-I-R, which emulates a real reservoir, is employed as the reference model for this study.

The UNISIM-I-M benchmark is based on information obtained from the UNISIM-I-R. The case presents a challenging scenario for reservoir simulation, as it involves a synthetic three-dimensional Black-oil reservoir with multiple sources of uncertainty. These uncertainties include PVT properties, net-to-gross ratio, porosity, horizontal and vertical permeability, oil-water contact, water relative permeability, rock compressibility, structural models, vertical permeability multiplier, and 500 images with characteristics related to facies and systems availability.

The UNISIM-I-M is designed to optimize control and revitalization variables. It has a history period of 7 years, during which 14 producer and 11 injector wells were utilized, and a management period of nearly 23 years. The exploitation strategy for this case was established by Avansi and Schiozer (2015) during the development phase and must remain the same.

To account for uncertainties, we consider the 48 simulation models selected by Gaspar et al. (2016a) through a data assimilation process. From this set of 48 scenarios, 9 representative models were selected using the RMFinder 2.0 developed by Meira et al. (2017) to optimize the production strategy. To calculate the NPV, we utilize the parameters listed in **Table 4.2** and **Equation 4.2**, retrieved and slightly modified from Gaspar et al. (2015). Since this study specifically focuses on the management period, the costs associated with investments in equipment facilities are not considered and therefore can be excluded from the calculations.

$$NPV = \sum_{j=1}^{N_t} \frac{[(R - Roy - ST - OC) \times (1 - T)]_j}{(1+i)^{t_j}} \qquad \text{Equation 4.2}$$

where R represents the gross revenues obtained from selling oil, Roy refers to the royalties expressed as a percentage of gross revenue, ST denotes the amount paid in social taxes also measured as a percentage of gross revenue, and OC is the operational production costs associated with oil and water production as well as water injection. Additionally, N_t

represents the total number of periods and j the specific time period, i is the discount rate, while t_j the average time of period j relative to the date of analysis.

Table 4.2–Parameters to calculate the NPV.

Parameter	Description	value
Market Values	Oil price (USD/m^3)	314.5
	Discount rate (%)	9
Taxes	Royalties (%)	10
	Special taxes on gross revenue (%)	9.25
	Corporate taxes (%)	34
Costs	Oil production (USD/m^3)	62.9
	Water production (USD/m^3)	6.29
	Water injection (USD/m^3)	6.29
Dates	Start of management period and time for update the cash flow	07/31/2020
	Maximum simulation period and field abandonment	05/31/2043

The EMV, shown in **Equation 4.3**, is calculated using the weighted average of NPV by the probabilities of occurrence of each RM shown in **Table 4.3**.

$$EMV = \sum_{i=1}^{n} p_i \times NPV_i \qquad \text{Equation 4.3}$$

where p_i is the probability of occurrence of scenario i.

Table 4.3–Probabilities for representative models (RMs) selected using RMFinder 2.0 (adapted from Santos et al., 2021).

Representative model	Probability
RM1	0.167
RM2	0.104
RM3	0.146
RM4	0.125
RM5	0.042
RM6	0.063
RM7	0.063
RM8	0.167
RM9	0.125

In this study, we adopted the configuration that yielded the highest economic return in Santos et al. (2020). Specifically, for both the IDLHC and IDLHC-MLR methods, the parameters F and N were set to 100 and 20%, respectively. Additionally, the extra parameter m from IDLHC-MLR was set to 50.

4.4 Results

The optimization process was conducted across the 9 RMs, and the best strategy derived from these optimizations was subsequently replicated in the 48 data assimilated models. Each run from **Figure 4.4** represents the simulation of the 9 RMs with the same strategy. Comparing the evolution of the EMV throughout the runs, it becomes evident that the IDLHC-MLR approach exhibited a superior convergence rate compared to IDLHC. Furthermore, it effectively maximized the EMV, achieving a value close to the optimum obtained by IDLHC. Specifically, the IDLHC-MLR yielded an EMV of 3.14 billion USD after 4,950 simulations, whereas the IDLHC approach yielded a slightly higher EMV of 3.15 billion USD, but required 9,000 simulations. Consequently, the IDLHC-MLR approach necessitated 45% fewer simulations while delivering a marginally lower EMV of only 0.3%. Furthermore, both the IDLHC and IDLHC-MLR strategies exhibited significant improvements in EMV compared to the base strategy (S0), with an increase of over 21% or 550 million USD. This observation underscores the effectiveness of the selected parameterization for controlling the BHP over time.

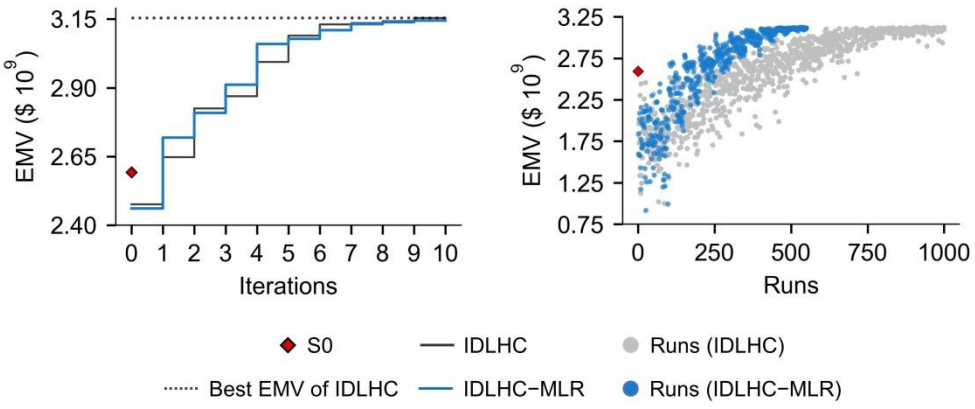

Figure 4.4–Evolution of EMV per iteration and per run for the IDLHC and IDLHC-MLR methods, along with a comparison to the base strategy (S0). Each run represents the simulation of the 9 representative models with the same strategy.

Table 4.4 depicts the accuracy of the IDLHC-MLR method from the second iteration onwards, which represents its capability to incorporate the F strategy among the m predicted strategies by the multi-layer perceptron algorithm. On average, the IDLHC-MLR approach demonstrated a good accuracy of approximately 88%. It is worth noting that although the accuracy dipped to around 80% during certain iterations, these deviations did not adversely impact the overall results.

Table 4.4–IDLHC-MLR accuracy over the iterations.

Iteration	Accuracy (%)
2	95
3	95
4	80
5	85
6	85
7	100
8	85
9	80
10	85
Average	87.8

The findings presented in **Figure 4.5** reveal that both the IDLHC and IDLHC-MLR strategies demonstrate notable improvements in NPV throughout all RMs when compared to the base strategy. Furthermore, it is noteworthy that the EMV obtained by both strategies remains stable across the 48 scenarios. This is also reflected in the distribution of Net Present Value (NPV), oil production (Np), water production (Wp), and water injection (Winj) as depicted in **Figure 4.6**. We observe a slight average increase of 0.8% in water production and 0.6% in water injection, along with a 0.7% decrease in average oil production for the IDLHC-MLR strategy compared to IDLHC.

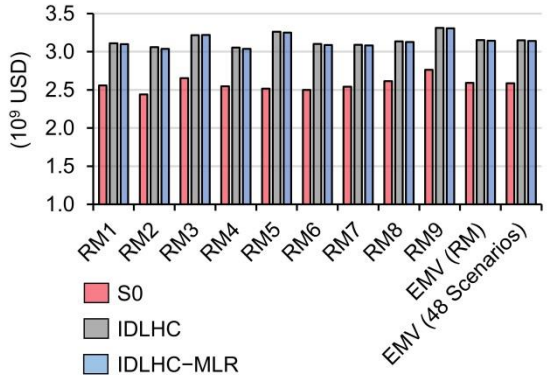

Figure 4.5–NPV for each RMs and EMV for both the RMs and the 48 data assimilated scenarios. The NPV and EMV are compared among the base strategy (S0), the best strategy of IDLHC, and the best strategy of IDLHC-MLR.

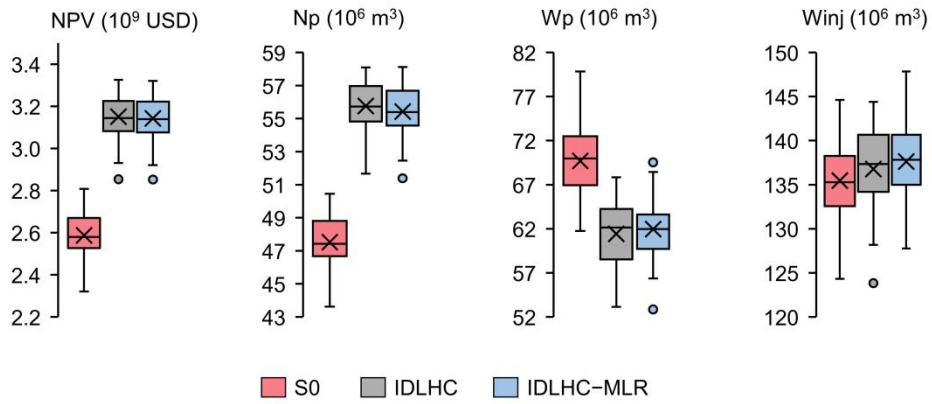

Figure 4.6–NPV, Np, Wp and Winj boxplots for the 48 data assimilated scenarios. The boxplots are presented for the base strategy (S0) and for the best strategies from IDLHC and IDLHC-MLR.

Figure 4.7 provides a visualization of the risk curves for economic and production data in both the 9 RMs and the 48 scenarios, comparing the S0 and IDLHC-MLR strategies. Remarkably, the 9 RMs continue to accurately represent the overall behavior observed in the 48 scenarios even after implementing the optimized production strategy. This outcome serves as a positive indication that the RMs were effectively selected using the RMFinder 2.0 method, ensuring their representativeness and suitability for the analysis.

Additionally, the implementation of the IDLHC-MLR strategy yielded an average increase in oil production (17%) and water injection (approximately 2%), while simultaneously achieving a reduction in water production (11%) when compared to the base strategy. These outcomes collectively indicate a highly effective reservoir sweep, leading to improved oil recovery and decreased water production.

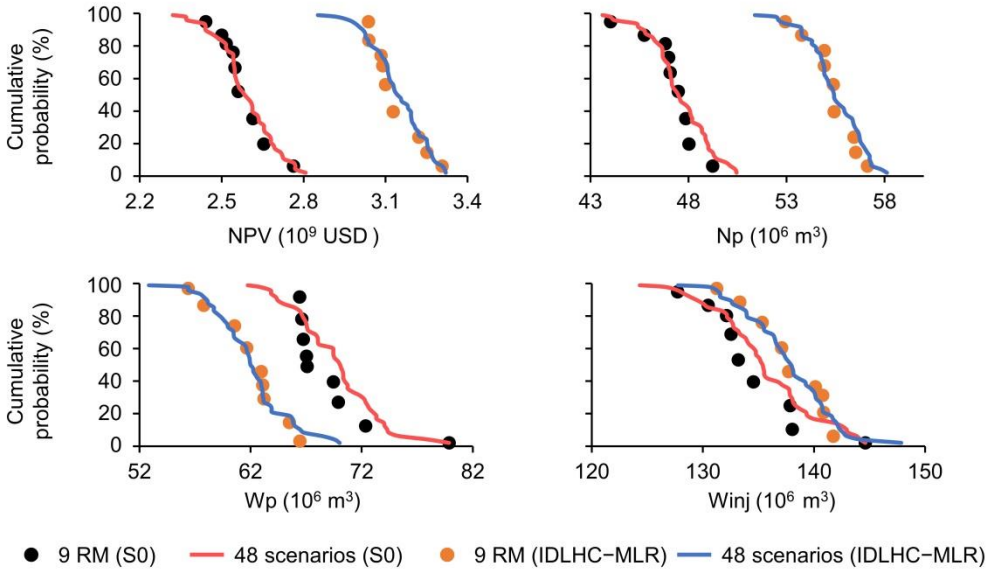

Figure 4.7–Risk curves for NPV, Np, Wp, and Winj in the 48 data assimilated scenarios and the RMs. The risk curves are presented for both the base strategy (S0) and the best strategy of IDLHC-MLR.

Afterwards, we executed the three strategies in the reference case, which emulates the actual field. The objective was to ascertain the suitability of the IDLHC-MLR strategy in such a situation. The IDLHC and IDLHC-MLR strategies demonstrated significant improvements in the NPV of UNISIM-I-R when compared to the S0 strategy, with an increase of 216 (8.4%) and 228 (8.9%) million USD, respectively. The IDLHC-MLR strategy slightly outperformed the IDLHC strategy, highlighting that the modeling error in accurately representing the field is greater than the marginal 0.3% EMV advantage achieved by IDLHC during the optimization of the 9 RMs.

Interestingly, it is worth noting from **Figure 4.8** that the uncertainties in the NPV for the reference model were well captured by the 48 filtered scenarios under the S0 strategy, as the NPV risk curve encompasses the NPV values for UNISIM-I-R. On the other hand, when the IDLHC and IDLHC-MLR strategies were applied, the 48 scenarios failed to adequately represent the NPV observed in the reference model. This indicates that the production strategy directly influences the behavior of the models that incorporate the field's observed data. To address the problem of insufficient representation in the assimilated scenarios, it would be necessary to revisit and refine these scenarios using a closed-loop approach. However, conducting such an analysis is outside the scope of the current study.

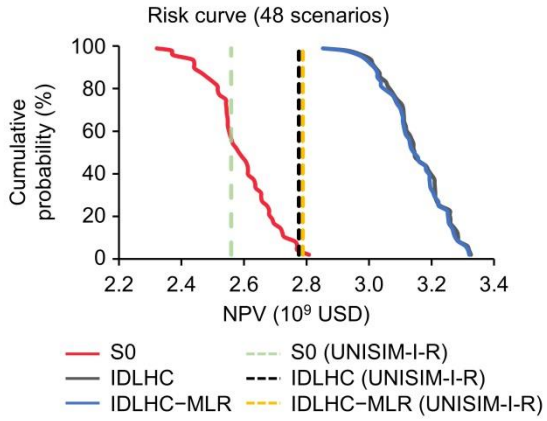

Figure 4.8–Comparison of NPV risk curves in 48 scenarios for the base strategy (S0) and the best strategies from IDLHC and IDLHC-MLR. The vertical dashed lines represent NPV values obtained by each strategy on the reference model (UNISIM-I-R).

Based on **Figure 4.9**, the main drivers behind the higher NPV achieved by the IDLHC-MLR strategy compared to the base strategy for the UNISIM-I-R are the increase in oil production (11.4%) and the decrease in water production (-8.7%). The volume of injection remained relatively stable with a slight decrease of -0.7%.

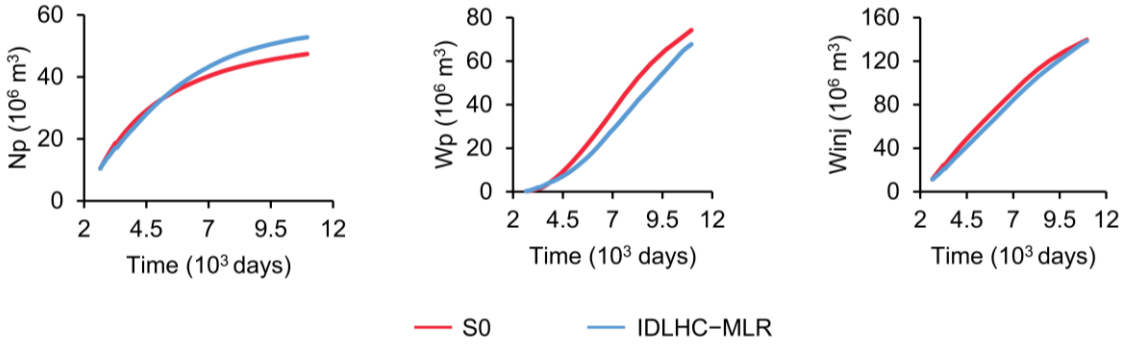

Figure 4.9–Comparison of Np, Wp and Winj between the base strategy (S0) and the best IDLHC-MLR strategy applied to the reference model (UNISIM-I-R).

4.5 Conclusions

In this study, our primary objective was to modify the Iterative Discrete Latin Hypercube combined with Machine Learning (IDLHC-ML) by incorporating uncertainties through robust optimization (IDLHC-MLR) and validating its performance under uncertainties. We successfully accomplished this objective, as the IDLHC-MLR method significantly reduced the required number of simulations by 45% (from 9,000 to 4,950 simulations) without compromising the expected monetary value for long-term well control optimization compared to the traditional IDLHC approach.

Furthermore, our findings demonstrate the robustness of the production strategy provided by the IDLHC-MLR method. It yielded positive outcomes in scenarios representing uncertainties, resulting in a considerable improvement of 21.5% (equivalent to 555 million USD) compared to the base strategy. Additionally, in the real field, the IDLHC-MLR method exhibited an improvement of 8.9% (228 million USD).

Lastly, the successful application of machine learning (ML) for well control optimization opens avenues for future research. Further investigations involving the utilization of the IDLHC-MLR method in different study cases would contribute to validating its generalizability and determining the specific scenarios where the IDLHC-MLR approach is most effective. Additionally, the machine learning approach has demonstrated its benefits and could be applied considering other variables, such as well type and locations.

4.6 Acknowledgements

We conducted this work with the support of Libra Consortium (Petrobras, Shell Brasil, TotalEnergies, CNOOC, CNPC, and PPSA) within the ANP R&D levy as “commitment to research and development investments” and Energi Simulation. The authors are grateful for the support of the Center for Petroleum Studies (CEPETRO-UNICAMP/Brazil), the Department of Energy (DE-FEM-UNICAMP/Brazil), and the Research Group in Reservoir Simulation and Management (UNISIM-UNICAMP/Brazil). A special thanks also goes to CMG for software licenses.

5 Accelerated optimization of CO_2-miscible water-alternating-gas injection in carbonate reservoirs using production data-based parameterization

Authors: Santos, D. R., Fioravanti, A. R., Botechia, V. E., Schiozer, D. J.

Published on Journal of Petroleum Exploration and Production Technology, 2023
Volume 13, pages 1833–1846.
(https://doi.org/10.1007/s13202-023-01643-0)

Abstract

Enhancing oil recovery in reservoirs with light oil and high gas content relies on optimizing the miscible water alternating gas (WAG) injection profile. However, this can be costly and time-consuming due to computationally demanding compositional simulation models and numerous other well control variables. This study introduces WAG_{eq}, a novel approach that expedites the convergence of the optimization algorithm for miscible water alternating gas (WAG) injection in carbonate reservoirs. The WAG_{eq} leverages production data to create flexible solutions that maximize the net present value (NPV) of the field, while providing practical implementation of individual WAG profiles for each injector.

The WAG_{eq} utilizes an injection priority index to rank the wells and determine which should inject water or gas at each time interval. The index is built using a parametric equation that considers factors such as producer and injector relationship, water cut (W_{CUT}), gas-oil ratio (GOR), and wells cumulative gas production, to induce desirable effects on production and WAG profile.

To evaluate WAG_{eq}'s effectiveness, two other approaches were compared: a benchmark solution named WAG_{bm}, in which the injected fluid is optimized for each well over time, and a traditional baseline strategy with fixed six-month WAG cycles. The procedures were applied to a synthetic simulation case (SEC1_2022) with characteristics of a Brazilian pre-salt carbonate field with karstic formations and high CO_2 content. The WAG_{eq} outperformed the baseline procedure, improving the NPV by 6.7% or 511 USD million. Moreover, WAG_{eq} required fewer simulations (less than 350) than WAG_{bm} (up to 2,000), while delivering a

slightly higher NPV. The terms of the equation were also found to be essential for producing a WAG profile with regular patterns on each injector, resulting in a more practical solution.

In conclusion, WAG_{eq} significantly reduces computational requirements while creating consistent patterns across injectors, which are crucial factors to consider when planning a practical WAG strategy.

Keywords: simulation reduction, practical solution, flexible solution, water alternating gas, production data, parameterized equation, field management.

5.1 Introduction

The Brazilian pre-salt is one of the world's largest polygons of oil and gas discovered in recent decades (Godoi and dos Santos, 2021). The oil from many pre-salt fields has high levels of gas-oil ratio (GOR) and CO_2 component (Pasqualette et al., 2017). This high amount of gas offers the opportunity to supply the gas market but also poses challenges related to gas handling, storage, and transportation (Ligero and Schiozer, 2014). Some of the pre-salt reservoirs also present gas contaminant contents that make their commercialization difficult. To avoid greenhouse gas emissions in the cases where the gas is not profitable, there is a special interest in reinjecting the gas produced. Wang et al. (2023) proposed a solution for CO_2 storage safety by investigating the feasibility of water alternating gas (WAG) injection and brine extraction in a deep saline aquifer. Their study found that WAG injection and brine extraction can enhance CO_2 injectivity and storage safety, with WAG injection reducing structural trapping contribution and brine extraction decreasing the maximum averaged reservoir pressure.

The WAG injection method involves alternating the gas with water, which has a synergistic effect that arises from the properties of both fluids. The water's primary role are pressure maintenance, macroscopic displacement, and improving gas sweep efficiency by controlling gas mobility and stabilizing the gas front, while gas decreases oil viscosity and residual oil saturation, thereby increasing the efficiency of microscopic displacement (Christensen et al., 2001; Kulkarni and Rao et al., 2005; Arogundade et al., 2013; Ramachandran et al., 2010; Afzali et al., 2018; Janssen et al., 2020).

Several studies in the literature have reported the effectiveness of the WAG method in increasing oil recovery and the economic return of oil fields (Chen et al., 2010; Duchenne et al., 2014; Hu et al., 2020; Kong et al., 2021; Mousavi et al., 2011; Sampaio et al., 2020;

Schaefer et al., 2017; Teklu et al., 2016). Schaefer et al. (2017) and Hu et al. (2020) conducted a comparison between CO_2-WAG and continuous CO_2 injection using reservoir simulation. In both studies, the WAG method resulted in increased oil production in comparison to continuous CO_2 injection, with approximately 30% and 14% higher oil production reported, respectively. Duchenne et al. (2014) performed a laboratory study to investigate the microscopy efficiency of CO_2-WAG injection on horizontal carbonate cores with light oil under reservoir conditions, and they showed that CO_2-WAG injections provide a faster and better oil recovery than pure CO_2 injection.

To improve production management strategies, it is crucial to optimize the WAG profile, which involves the alternating injection of water and gas into wells. A range of methods have been identified in the literature to optimize the WAG profile, with the most prevalent being the optimization of WAG cycles (Bahagio, 2013; Esmaiel et al., 2005; Nait et al., 2018; Pal et al., 2018; Pereira et al., 2022), i.e., determining the optimal duration of water injection and gas injection before switching back to gas from water, as well as the WAG ratio, which is the ratio between the volume of water and gas injected (Chen et al., 2010; Chen and Reynolds, 2016; Kazakov and Bravichev, 2015; Panjalizadeh et al., 2015).

Chen and Reynolds (2016) implicitly optimized the WAG ratio for each cycle by optimizing both the target for gas and water injection rates and the target for the producers' bottom-hole pressure (BHP) over time. They considered different numbers of WAG cycles (4, 8, 16, and 32 cycles) and aimed to maximize the net present value (NPV) of the field. The authors obtained better results by increasing the number of cycles, and they suggested that fixing a WAG ratio for the entire field life cycle is inappropriate. Pereira et al. (2022) tested five approaches to optimize the WAG cycle's duration to maximize the NPV of a synthetic reservoir simulation model with pre-salt characteristics. The two approaches that delivered better results were: 1) the simultaneous optimization of WAG cycles (same cycle size for all injectors) and the gas-oil ratio (GOR) limit to shut producers; 2) using the best solution from the previous approach, the authors optimized the cycle individually for each injector, followed by a re-optimization of the GOR limit to shut-in producer wells. An important conclusion from their work was that the optimum cycle size could be changed meaningfully according to the operation of other control variables (such as the shut-in of producers).

Supporting the findings of Pereira et al. (2022), Chen et al. (2010) stated that inappropriate selection of the control of injectors and producers may result in unstable pressure

distribution, early gas breakthrough and, consequently, a lower oil recovery factor. Bahagio (2013) showed the importance of controlling the WAG parameters (e.g., WAG cycle duration) in a synthetic reservoir of $7 \times 7 \times 3$ grid blocks with a 5-spot pattern containing four injectors and one producer in the middle. The author observed a significant increase in the NPV by optimizing both the duration of WAG cycles and the BHP injection, compared to optimizing only the BHP and maintaining a fixed WAG half-cycle duration of six months for both water and gas. The number of water cycles alternated with gas was kept fixed at 30, while allowing the duration of gas and water cycles to vary during each fluid exchange. The author used an equal WAG profile for all injectors, which is appropriate for the homogeneous case studied, but may lead to suboptimal solutions when accounting for reservoir heterogeneities.

Kazakov and Bravichev (2015) argue that describing WAG injection solely by time-dependent periods of gas and water injection is not informative, as injectivity may not remain constant due to varying saturation and reservoir pressure during field production. Similarly, when accounting for geological uncertainties, a time-dependent WAG injection alone may be less practical as the injection and production capacities may vary due to reservoir uncertainties. To address these limitations, a WAG injection control that takes into account production monitoring variables like water cut, GOR, and fluid rates could provide a more flexible and adaptable solution for the real reservoir's changing behavior over time.

Compositional simulation is generally considered a more reliable approach for miscible enhanced oil recovery (e.g., miscible WAG injection) in reservoirs with light oil and high gas content, such as those found in Brazilian pre-salts. However, the reliability comes at the cost of augmented computational effort as the number of pseudo components required to accurately describe the field's fluids increases (Schlijper, 1986). The computational cost is amplified by the optimization of several well control variables, including WAG control, as well as the presence of uncertainties and high heterogeneity of the fields, which demand a large number of blocks for accurate modeling.

Therefore, efforts to reduce computational costs while maintaining high oil production and economic return are crucial for the optimization of miscible enhanced oil recovery. This can be achieved using various techniques, such as representing a larger group of models that honor observed data through a small set of scenarios that still accurately capturing field uncertainties (Meira et al., 2017; Meira et al., 2020; Sarma et al., 2013; Shirangi and Durlofsky, 2015), numerically tuning high-complexity reservoir models (Mello et al., 2022), and using

coarse models to represent high fidelity ones. For example, Kou et al. (2022) proposed upscaling methods for CO_2 migration in 3D heterogeneous geological systems that reduce computational costs while preserving fine-scale flow mechanisms. The authors tested two percolation-based methods demonstrating the robustness of the upscaling methods with less than 10% computational errors between fine- and coarse-scale models.

A supplementary strategy to alleviate the computational load involves minimizing the number of simulations required for optimization by constraining the search space. By doing so, the optimization algorithm can converge more rapidly towards satisfactory solutions. Expanding on these considerations, the present study introduces a novel WAG parameterization rule that utilizes reservoir production data to expedite the optimization process, while simultaneously maximizing the economic return over the life cycle of the WAG strategy. The proposed rule offers a high degree of flexibility by accommodating individual WAG profiles for each injector, regardless of whether they follow a cyclic pattern or not, while still being easy to implement in practical cases.

As previously mentioned, optimizing the WAG injection profile also requires consideration of the producer wells' operation. Thus, we simultaneously optimize the GOR limit (GOR_{limit}) to shut each producer and the WAG injection under the total gas reinjection constraint to maximize the field's life cycle NPV for two different procedures. The first consists of a WAG parameterization that defines the set of possible solution space for the fluid to be injected by each well at each time step. The solution achieved by this procedure is only time-dependent, and serves as a benchmark in terms of number of simulations and NPV. Our second and primary procedure involves utilizing a parametric equation to determine the WAG injection profile. This equation considers the gas injection volume during each cycle and the most influential producers for each injector. Each term in the equation has been carefully designed to produce a specific effect on the WAG profile, which will be explained in detail in the methodology section of this paper.

The optimization process involves tuning the coefficients of the parametric equation, which significantly reduces the number of optimization variables compared to the first procedure. The results obtained from both procedures are then compared to the baseline strategy, where WAG cycles are set to six months and only the GOR limit for shut-in producers is optimized. This comparison allows the evaluation of the NPV gain achieved by optimizing the WAG profile.

The paper's organization is as follows: first, a brief explanation of the optimization algorithm used in this study is provided. Next, the methodology section presents comprehensive details on the adopted procedures. The case study, including production constraints and underlying assumptions, is then presented succinctly. We then compare the results obtained from the simulations and the net present value (NPV) achieved for both the procedures and the baseline approach. The following section investigates and discusses whether the terms included in the proposed equation are generating the expected effect on the WAG profile. The paper concludes by summarizing the advantages of the proposed method. It also highlights the limitations of the study and potential directions for future research.

5.1.1 IDLHC optimization algorithm

The optimization algorithm applied in this work is based on the discrete Latin hypercube sampling method (DLHC, see Maschio and Schiozer, 2016), a technique that recreates the entire space distribution by randomly sampling the discrete candidate values of each optimization variable. The probability mass function (PMF) of each optimization variable determines the frequency with which each discrete value appears in the sample. For instance, if variable Y has a sample size of 10, and its PMF is given by $P_Y(x) = \{0.6, 0.2, 0.2\}$, where the discrete values of Y are represented by $x = \{0, 2, 4\}$, then the number of samples with the values 0, 2, and 4 would be 3, 1, and 1, respectively (as depicted in **Figure 5.1**).

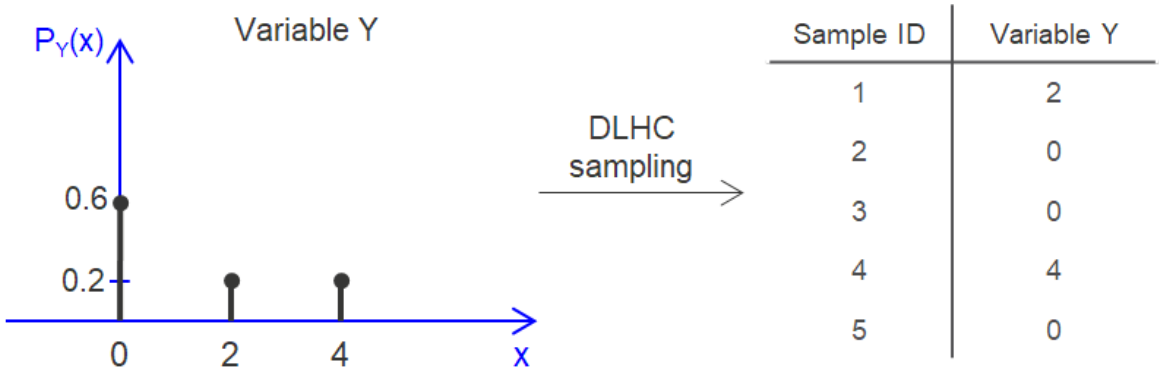

Figure 5.1–Discrete Latin hypercube sampling method example for the variable Y. The samples of the variable Y generated with the value x depend on the probability $P_Y(x)$.

The Iterative discrete Latin hypercube (IDLHC, see von Hohendorff et al., 2016), depicted in **Figure 5.2**, is a widely-used optimization technique that performs the following steps: (1) defining the optimization variables and discretizing them into values, (2) generating N samples using the discrete Latin hypercube method based on a prior probability mass

function (PMF) defined by the user for each variable, (3) evaluating the samples using an objective function, and (4) selecting the *F* percent best samples to update the PMF of each variable independently. The updated PMF is then used to generate the samples in the next iteration. The loop continues until a specific condition is reached, such as a maximum number of iterations or a threshold of minimum difference between the maximum and minimum values obtained in the iteration. It is important to note that in the first iteration, the probability mass function of all variables is generally set as equiprobable, and the value of *N* must be greater than the number of candidate values for each variable.

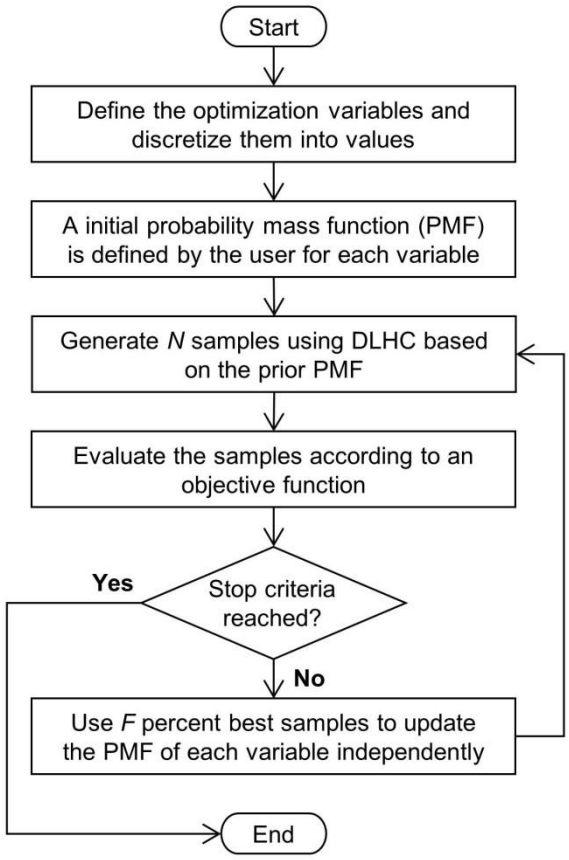

Figure 5.2–IDLHC algorithm workflow, which iteratively generates samples using discrete Latin hypercube (DLHC) method, evaluates them using an objective function, and updates the probability mass function (PMF) of each variable based on the *F* percent best samples. The loop continues until specific criteria are reached.

We selected the IDLHC algorithm for optimization because it has been demonstrated to generate high-quality solutions in numerous published works (Botechia et al., 2021; Loomba et al., 2022; Pereira et al., 2022; Santos et al., 2020; von Hohendorff et al., 2016; von Hohendorff and Schiozer, 2018). Some of these works have compared the IDLHC algorithm with other well-established optimization techniques, and the IDLHC has been shown to

generate better results. For instance, von Hohendorff et al. (2016) reported that the IDLHC algorithm outperformed a well-established optimization technique in a production optimization problem, demonstrating a faster convergence rate and resulting in a slightly better objective function. Another study by von Hohendorff and Schiozer (2018) showed that the IDLHC algorithm has several advantages, including its simplicity and rapid convergence near the global optimum, compared to a genetic algorithm method, when applied to optimize the well placement production strategy.

5.2 Methodology

The method presented herein aims to maximize the net present value (NPV) of a reservoir through an optimization problem, as formulated in **Equation 5.1**. This involves the optimization of parameters related to the baseline procedure ($Well_{gor\_limit}$), the benchmark strategy (WAG_{bm}), and the main procedure which uses a parametric equation to define the WAG injection profile (WAG_{eq}). The WAG_{bm} and WAG_{eq} parameters are jointly optimized with the well limit of gas-oil ratio (GOR_{limit}), which is used to shut-in each producer individually. Furthermore, all wells from WAG_{bm} and WAG_{eq} are capable of injecting alternating water and gas.

$$\begin{aligned} &Maximize\ NPV(x) \\ &Subject\ x_{lb} \leq x \leq x_{ub}, \\ &C(x) \leq 0, \\ &Eq(x) = 0. \end{aligned} \qquad \text{Equation 5.1}$$

where x is the decision variable vector subject to lower (x_{lb}) and upper bounds (x_{ub}). The objective function is $NPV(x)$, and the inequality and equality constraints are represented by $C(x)$ and $Eq(x)$, respectively. In the reservoir simulation problem, there are multiple constraints to consider, including well bottom-hole pressures, well and platform rates (inequality constraint), full gas re-injection, and reservoir pressure maintenance (equality constraints).

The methodology's results are first discussed by comparing the best NPV, the optimization algorithm convergence rate, WAG profile, and field production for the $Well_{gor\_limit}$, WAG_{bm} and WAG_{eq}. Subsequently, an analysis of the parametric equation's terms is conducted to determine if each of them are generating the anticipated effect on the WAG

profile and whether they should be retained or eliminated. The aim of this examination is to offer significant insights into the effectiveness of the WAG_{eq} and the validity of the modeled behavior.

5.2.1 Baseline procedure ($Well_{gor\_limit}$)

This procedure utilizes the measured gas-oil ratio of each producer p (GOR^p) as a monitoring variable. The control rule, $f_p\left(GOR^p_{limit}\right)$, operates by shutting down producer p once it reaches the GOR_{limit}, as shown by **Equation 5.2**. As a result, the number of optimization variables for this procedure is equal to the number of producer wells, and the value of GOR_{limit} can be different for each producer. Additionally, the WAG cycles are fixed in this procedure, providing a baseline NPV that enables the evaluation of the economic increment capacity of optimizing the WAG profile.

$$f_p\left(GOR^p_{limit}\right) = \begin{cases} shut\ producer,\ GOR^p_{limit} \geq GOR^p \\ keep\ producer\ open,\ GOR^p_{limit} < GOR^p \end{cases} \quad \text{Equation 5.2}$$

This procedure was conducted by Botechia et al. (2021), and the candidate values for the GOR_{limit} and the NPV obtained from the optimized strategy are presented in the application section.

5.2.2 WAG benchmark procedure (WAG_{bm})

Here, we present a parameterization that aims to find the best combination of water injectors and gas injectors over time, while adhering to the constraint that all produced gas must be reinjected. Additionally, each injector well can switch between water and gas injection, with a minimum interval of t_{min} between switching. The solution space in this approach consists of the combination C^n_m of all injectors (n) where at least a pre-defined number of gas injectors (m) are operating to ensure complete reinjection (**Equation 5.3**).

$$C^n_m = \left(\frac{n!}{m! \times (n-m)!}\right)^{nt} \quad \text{Equation 5.3}$$

where nt represents the number of minimal time intervals. This procedure is solely dependent on time and is performed to establish a benchmark value for the number of simulations and NPV that can be expected by optimizing the WAG injection profile. Our primary approach,

WAG_{eq}, is expected to deliver an NPV performance that is at least equivalent to this benchmark, while requiring fewer simulations.

5.2.3 WAG parametric equation rule (WAG_{eq})

This procedure aims to optimize the WAG profile over time in a more efficient way while maintaining or even increasing the NPV compared to the previous procedure (WAG_{bm}). It involves the development of a parametric equation rule using the injection priority index (**Equation 5.4**), which determines the order of gas and water injection at each time interval. The index value is used to rank the injectors, with those above a predefined threshold rank injecting gas and the others injecting water. The index of each injector (I^i) is determined by several variables, including the field's lifetime (t), the volume of gas injected (V_g^i) since the well last water injection, the WCUT of the n_p^i producers influenced by the water injector i, and GOR of the m_p^i producers influenced by the gas injector i. The influence is determined by analyzing the streamlines from the injectors to the producers, without utilizing any data beyond that which was available during the WAGeq procedure. The fluid of each injector can be altered within the t_{min} interval, similar to the WAGbm. Alternative methods, such as trace or Euclidean distance, could be used instead of streamlines to determine the influence of each injector.

$$I^i = \frac{1}{n_p^i} \times \sum_{p=1}^{n_p^i} \left(\alpha_p^i \times W_{CUT}^p\right) - \frac{1}{m_p^i} \times \sum_{p=1}^{m_p^i} \left(\frac{\beta_p^i \times GOR^p}{GOR_{max}}\right) + \gamma^i \times \left(\frac{t - t_{init}^i}{t_{end} - t_{init}^i}\right) - \delta^i \times \left(\frac{V_g^i}{V_{norm}^i}\right)$$

Equation 5.4

The optimization variables are the coefficients of the producer p that are influenced by the injector i (α_p^i, β_p^i), the coefficients γ^i and δ^i, and the gas volume normalizer for the injector i (V_{norm}^i).

To ensure consistent influence on the priority index calculation, all terms in the equation are normalized. The time term is normalized by the difference between the whole simulation period (t_{end}) and the moment that the injector i starts to operate (t_{init}^i). The maximum gas-oil ratio (GOR_{max}) is used as the normalizer for GOR^p. The value for GOR_{max} should be set equal to the maximum candidate value of GOR_{limit} because it represents the highest value a producer can attain before being shut-in. Finally, the W_{CUT} is already a normalized term between 0 and 1.

Each term in the equation used for calculating the injection priority index has a specific role in shaping the WAG profile, as explained below:

- The $\frac{-\delta^i \times V_g^i}{V_{norm}^i}$ represents the fraction of gas injected by the well in relation to the gas volume normalizer. The gas volume normalizer (V_{norm}^i) is chosen such that the ratio of gas volume injected by the well $\frac{V_g^i}{V_{norm}^i}$ falls within the range of approximately 0.1 to 1. The negative sign (-) incorporated in the term decreases the value of I^i for wells that have injected gas since their last water injection (**Figure 5.3a**). Conversely, water-injecting wells are exempt from this term, resulting in a higher value of I^i and thereby promoting their gas injection during the next fluid exchange period. Therefore, this term is conducive to promoting water alternating with gas injection in the well.

- The expression $\gamma^i \times \left(\frac{t - t_{init}^i}{t_{end} - t_{init}^i}\right)$ exhibits a constant value of $\left(\frac{t - t_{init}^i}{t_{end} - t_{init}^i}\right)$ among all wells with the same t_{init}^i. Therefore, the contribution of this term to I^i increases with positive values of γ^i, which correspond to a greater tendency towards gas injection throughout the well's lifetime, on the other hand, negative values of γ^i lead to a preference for water injection over the entire life cycle (**Figure 5.3b**). The objective of this term is to establish a consistent injection pattern for each well, so that if a well initiates a water-gas cycle, this term fosters its continuation until the end of the field's lifetime or the opening of new wells.

- The $\alpha_p^i \times W_{CUT}^p$ is always positive and contributes to an increase in the value of I^i. As such, the higher the W_{CUT} of the producers influenced by injector i the higher the value of I^i for that injector (**Figure 5.3c**). As a result the injector tends to transition from water to gas injection during the next t_{min} interval. This term promotes the uniform growth of W_{CUT} among the field's producers over time, potentially improving reservoir sweep efficiency. In contrast, the $\frac{-\beta_p^i \times GOR^p}{GOR_{max}}$ works in a similar manner, but the negative sign results in a decrease of I^i value (**Figure 5.3d**), and the injectors tend to shift from gas to water injection leading to homogeneous GOR growth among producers.

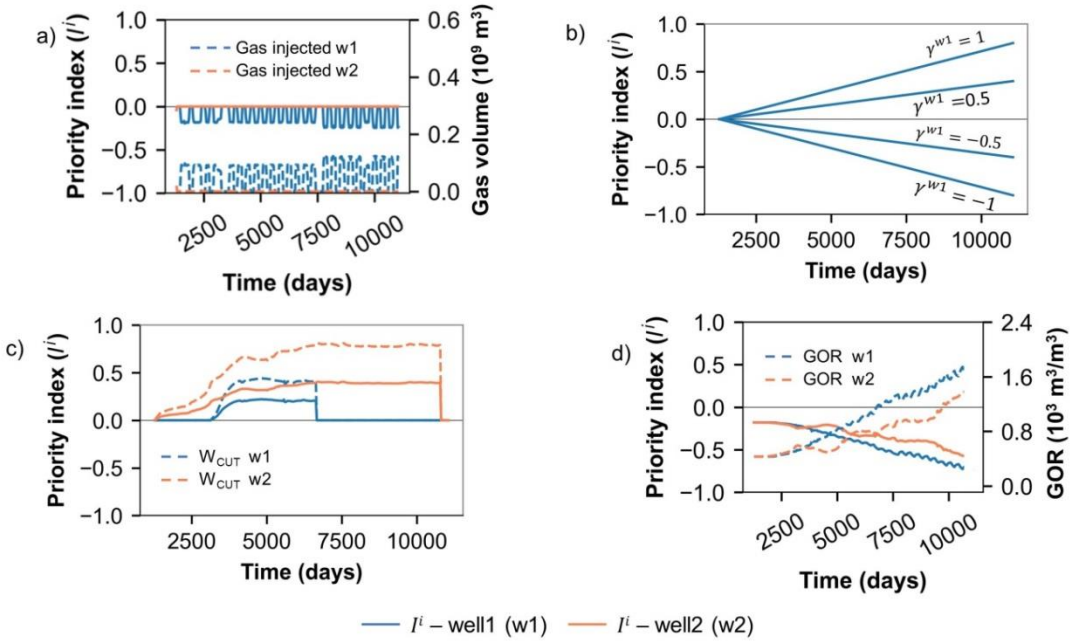

Figure 5.3–Impact of each term in the WAG_{eq} procedure. (a) Well 1 (w1) gas injection lowers I^i and promotes water injection in the next cycle. Well 2 (w2) with zero gas injection does not affect I^i. (b) I^i linearly varies over time and its slope depends on gamma. (c) A higher W_{CUT} increases I^i and gas injection tendency. (d) A higher GOR reduces I^i, favoring water injection.

5.3 Application

We begin by describing the benchmark and the premises adopted for the study. Subsequently, we present important dates that are necessary to compute the NPV and also show the production operational constraints. Furthermore, we provide information about the distribution and range of values for the optimization variables, as well as the IDLHC configuration parameters adopted in the procedures.

5.3.1 Study case

Our methodology was applied to the SEC1_2022 benchmark, which is a reservoir simulation model that has been designed to replicate the geological and fluid characteristics of Brazilian pre-salt fields (Chaves, 2018; Correia et al., 2020). This model represents a carbonate reservoir with light oil containing high levels of CO_2, and comprises 63 × 120 × 309 grid cells and 77,071 active blocks. Each block corresponds to an average volume of 200 × 200 × 5 cubic meters. Due to the complex characteristics of the fluid, the GEM compositional simulator (version 2020.10) developed by the Computer Modelling Group (CMG) was used to simulate the SEC1_2022 model.

Figure 5.4 displays the opening of 17 wells in two phases. During the first phase, six producers and seven injectors were opened between May and December/2021. In the second phase, which began in February 2027, the remaining two producers (P17 and P18) and two injectors (I18 and I19) were put into operation.

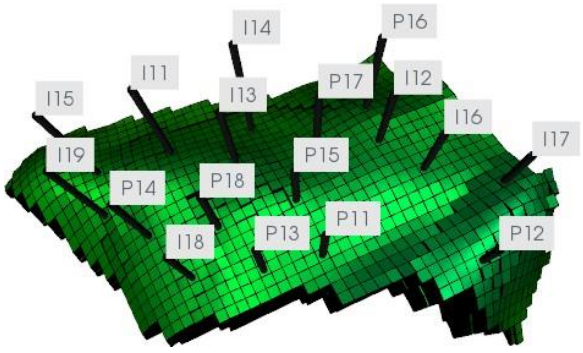

Figure 5.4–Reservoir wells placement.

The case includes historical data from October/2018 to February/2022, and the simulation continues until December/2048. The reference date for updating the NPV is February/2022. The costs associated with the wells and platform installed in first phase are not considered in the NPV calculation because they were installed before the reference date. However, costs associated with the wells drilled and perforated in the second period are considered. The operational constraints for both wells and the platform are provided in **Table 5.1** and **Table 5.2**.

Table 5.1–Wells' operational constraint.

Type of well	Constraint	Value
Producers	Minimum bottom-hole pressure (kPa)	50,000
	Maximum liquid production (m^3/day)	8,000
Injectors	Maximum bottom-hole pressure (kPa)	75,000
	Maximum water injection (m^3/day)	10,000
	Maximum gas injection (m^3/day)	4,000,000

Table 5.2–Platform's operational constraints.

Constraint	Value (m^3/day)
Maximum oil rate	28,617
Maximum liquid rate	28,617
Maximum water production rate	23,848
Maximum water injection rate	35,771
Maximum gas production rate	12,000,000

5.3.2 Assumptions and optimization parameters

The assumptions and optimization parameters for the baseline strategy and the two procedures are outlined below:

- The average reservoir pressure of 61,000 kPa is maintained through water injection rate control.
- All produced gas must be re-injected into the reservoir.
- Exactly four wells must inject gas at the same time throughout the entire field management.
- All injectors can inject both water and gas.
- The minimum period (t_{min}) to switch from one fluid to another is six months.

The baseline strategy includes a fixed six-month period to switch the injected fluid from each injector. One well (I16) is designated to inject only gas during the entire period to ensure full gas re-injection. In WAG_{bm} and WAG_{eq}, all injectors can switch between water and gas, meaning that I16 does not have to inject only gas like in the baseline strategy. The $Well_{gor\_limit}$ strategy was optimized by Botechia et al. (2021), who used candidate values for GOR_{limit} ranging from 600 to 2400 sm^3/sm^3 at 200 sm^3/sm^3 intervals. The optimized strategy delivered an NPV of 7.68 USD billion requiring 150 simulations.

For both the WAG_{bm} and WAG_{eq} procedures, the optimization is conducted simultaneously with WAG parameters and GOR_{limit}, utilizing the same candidate values for GOR_{limit} as Botechia et al. (2021). In the case of WAG_{bm}, the parameters cover all possible combinations of open injectors, with four gas injectors operating throughout the entire management period. When considering GOR_{limit} variables, the search space includes approximately 5.71×10^{113} potential solutions for this procedure.

In the WAG_{eq} procedure, the coefficients and V_{norm}^{i} are linearly spaced with 11 values between the ranges displayed in **Table 5.3**. Furthermore, GOR_{max} is defined as the maximum candidate value for GOR_{limit} (2400 sm^3/sm^3).

Table 5.3–Range of the optimization variables for WAG_{eq} procedure.

Variable	Range [min, max]
α^i_p	[0, 1]
β^i_p	[0, 1]
γ^i	[-1, 1]
δ^i	[0, 1]
V^i_{norm} (m^3)	$[5 \times 10^8, 5 \times 10^9]$

The minimum V^i_{norm} value was set slightly below the maximum volume of gas that can be injected by one well in the t_{min} interval (approximately 7.2×10^8 m^3). The maximum value was defined as ten times greater than the minimum value to ensure that the gas volume term can assume magnitudes similar to the other terms of the equation (between 0 and 1).

We have set the number of samples generated for each iteration based on the IDLHC requirement that N be greater than or equal to the number of candidate values for each variable. For WAG_{bm}, given that the combination of C_4^9 is 126, we set the number of samples to 150. For the WAG_{eq} simulation, we set N to 50. Both simulations had an F value of 30% and were run for 15 iterations.

5.4 Results

First, we identified the producers affected by each injector and the respective fluids (water or gas) in the non-optimized study case, which enabled us to determine the WAG_{eq} terms related to GOR^p and W^p_{CUT} monitoring variables (**Table 5.4**). By influence, we mean any fluid streamline from the injector arriving to the producer at any time during the management period. It is important to note that we assumed that the same producers affected by gas injection from I16 would also be affected by the water injected from this well. This is due to the fact that I16 solely injects gas in the non-optimized scenario, and therefore, it would not have any water streamlines originating from it to impact any of the producers.

Table 5.4–Producer wells influenced by gas and water streamlines from each injector.

Injectors	Producers (gas)	Producers (water)
I11	P14, P18	P18
I12	P11, P17	P13, P15, P17
I13	P13, P15, P17, P18	P15
I14	P16	P16
I15	P14, P18	P14
I16	P11, P13	---
I17	P12	P12
I18	P13, P18	P13, P14, P18
I19	P14	P14

The WAG_{bm} and WAG_{eq} approaches significantly enhanced the NPV, resulting in approximately 8.11 and 8.18 USD billion, respectively. In comparison to the NPV of $Well_{gor\_limit}$ procedure (7.68 USD billion), the WAG_{eq} yielded a similar value with 50 simulations completed, while the WAG_{bm} approach required 150 simulations to achieve a slightly higher NPV relative to the baseline strategy. The WAG_{bm} and WAG_{eq} improved the NPV by 436 USD million (5.7%) and 511 USD million (6.7%) compared to $Well_{gor\_limit}$, respectively. This finding highlights the importance of optimizing the WAG profile alongside well producer control variables. Although the maximum NPV achieved by WAG_{eq} (after 720 simulations) was only slightly higher than that of WAG_{bm} (0.9% or 75 USD million), WAG_{eq} required significantly fewer simulations and iterations to converge. At iteration 7 and with fewer than 350 simulations, the NPV was comparable, with the maximum NPV of WAG_{bm} achieved after 14 iterations and with 2051 simulations, as depicted in **Figure 5.5**.

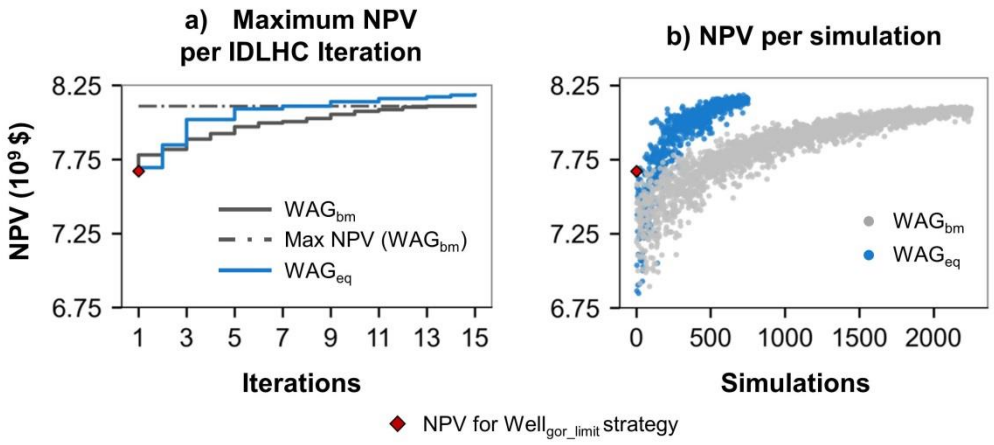

Figure 5.5–Comparison of NPV between the proposed WAG procedures (WAG_{bm} and WAG_{eq}) and the baseline strategy ($Well_{gor\_limit}$). The figure shows a) the maximum NPV achieved by each approach throughout the iterations of IDLHC and b) the NPV over the simulations.

The optimal WAG solutions did not exhibit a cyclical pattern, as shown in **Figure 5.6**. This result underscores the potential limitations of defining the search space based solely on regular cycles, as it may result in suboptimal outcomes. Notably, most of the wells in WAG_{eq} predominantly injected one fluid all over the management period (for considerations on WAG injection well preparation in cases where it operates as a single-fluid injection, please consult **Appendix D**). In particular, the I16 and I15 wells had the highest rank of injector priority index (I^i) for the majority of the period, while I11, I12, and I13 exhibited the lower values. The I^i was observed to alternate between I12, I13, I14, and I17 before the second wave and between I17 and I19 after it.

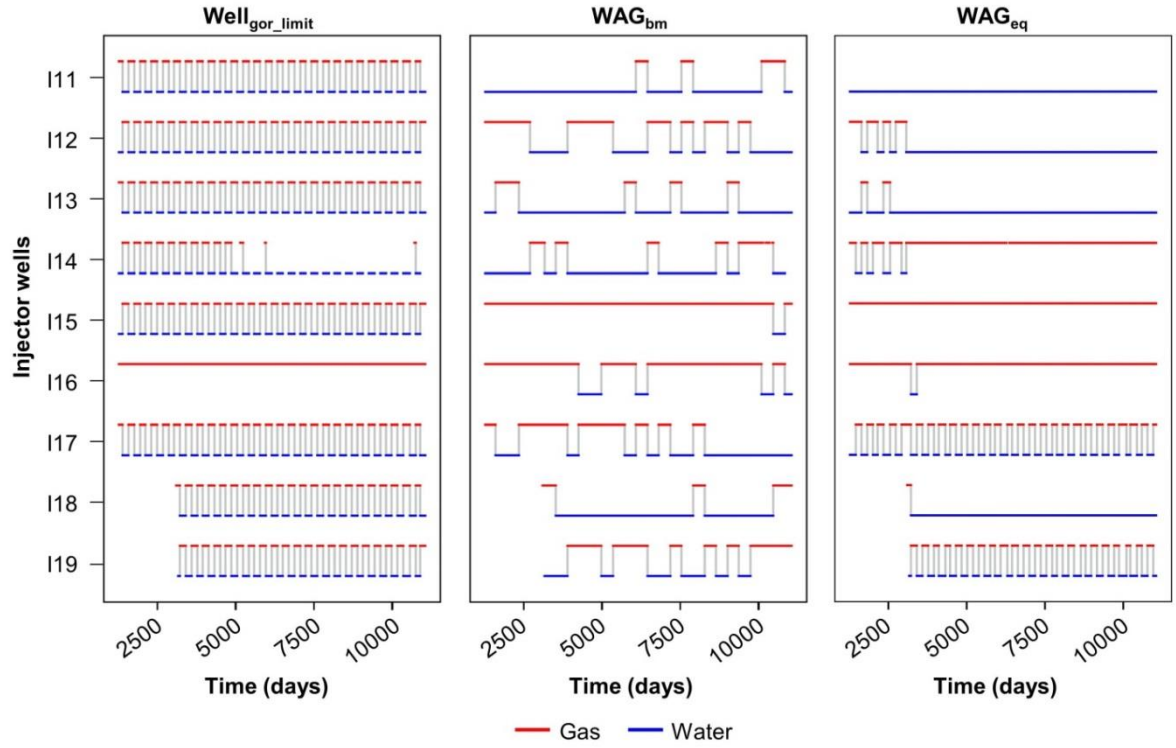

Figure 5.6–Comparison of WAG profile between the baseline strategy (Well$_{gor_limit}$) and the best strategy obtained using WAG$_{bm}$ and WAG$_{eq}$ procedures.

The WAG$_{eq}$ and WAG$_{bm}$ approaches significantly increased oil and water production compared to the baseline strategy. Precisely, WAG$_{eq}$ increased oil production by 11.4% and water production by 172%, while WAG$_{bm}$ resulted in a 9.4% increase in oil production and a 160% increase in water production. Moreover, both approaches injected considerably more water than the Well$_{gor_limit}$ strategy, with WAG$_{eq}$ and WAG$_{bm}$ injecting 43% and 38% additional water, respectively. This maintained average reservoir pressure and rise water and oil production (**Figure 5.7**) indicating the successful enhancement of the reservoir's sweep efficiency.

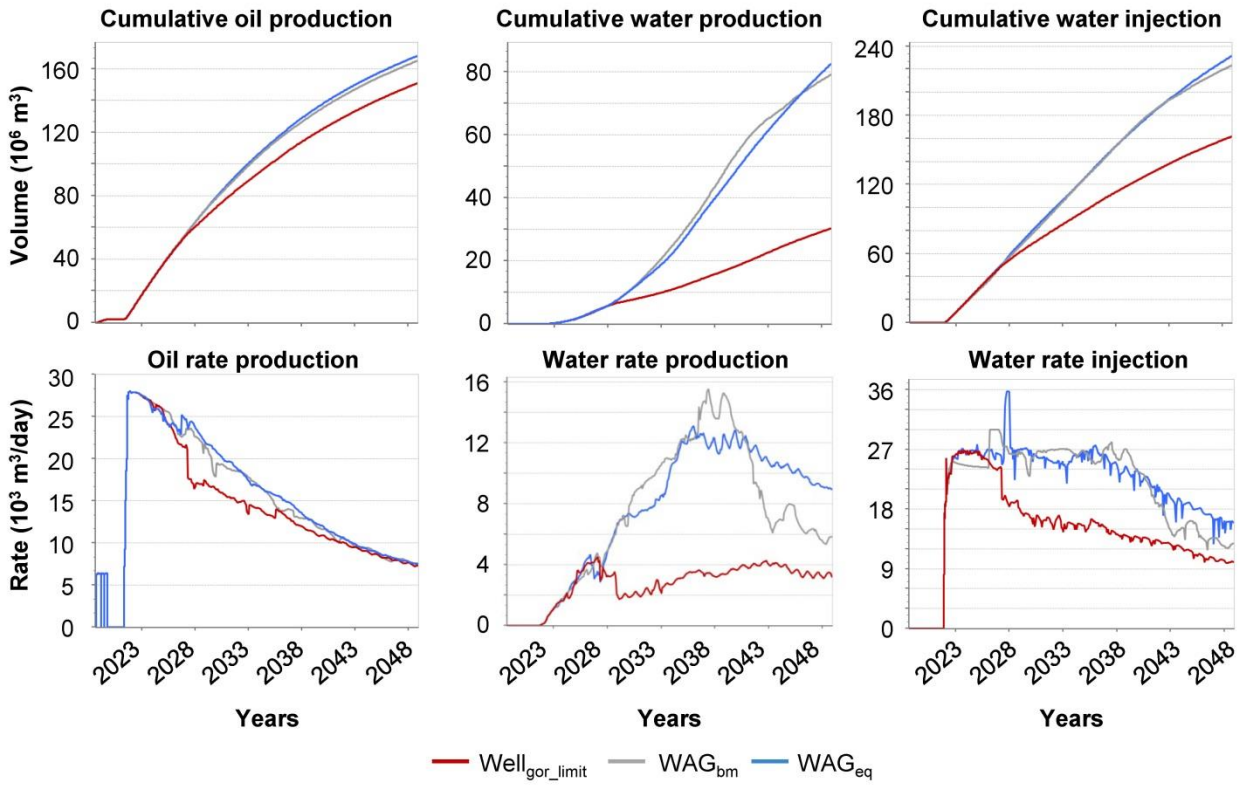

Figure 5.7–Volume and rates of production and injection for the baseline strategy ($Well_{gor\_limit}$) and for the best strategy of WAG_{bm} and WAG_{eq} procedures.

The production of gas for all three strategies was limited by platform capacity and the gas full re-injection constraint was also observed, as shown in **Figure 5.8**. This demonstrates that the optimal strategies prioritize maximizing gas injection, even if it results in higher gas production costs.

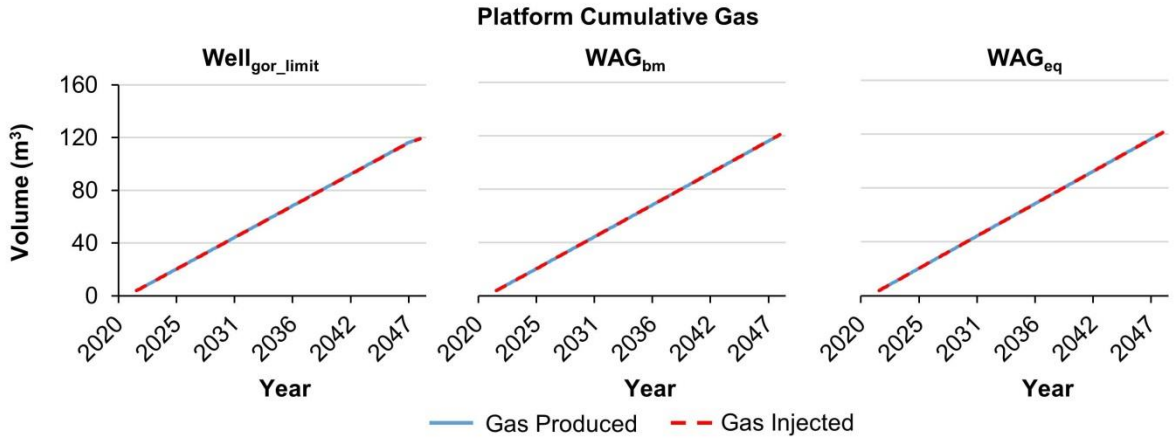

Figure 5.8–Cumulative gas production and injection for the baseline strategy ($Well_{gor\_limit}$) and for the best strategy of WAG_{bm} and WAG_{eq} procedures.

5.4.1 Investigation of the terms of the equation for WAG_{eq} procedure

We conducted further analysis to examine the impact of each term in the parametric equation on the WAG profile and NPV. This was done to determine the necessity of all terms in the equation. To accomplish this, we optimized the parametric equation with all terms except one to observe if the resulting WAG profile lost the specific behavior associated with that term. The optimization parameters and assumptions used were the same as those employed in the WAG_{eq} procedure.

First, we removed $\frac{-\delta^i \times V_g^i}{V_{norm}^i}$ from the equation, which promotes water and gas exchange between injectors as described in the methodology. **Figure 5.9b** shows that when this term is removed, the wells indeed tend to inject the same fluid over time. Wells that alternated between water and gas injection in WAG_{eq} (such as I19, I17, and I14) injected almost exclusively one fluid when the gas volume term was removed. The I11, I13, I15, I16, and I18 wells also injected almost solely one fluid over the field management in both WAG solutions.

The WAG profile solution obtained after removing the term $\gamma^i \times \left(\frac{t - t_{init}^i}{t_{end} - t_{init}^i}\right)$ appears less regular for injectors like I12, I14, and I17, which do not primarily inject a single type of fluid for most of the time (**Figure 5.9c**). This indicates that the temporal term is necessary to maintain a regular injection pattern from each injector throughout the field life cycle or until a significant event occurs, such as the opening of new wells in the second wave. The temporal term generates more stable strategies that are easier to implement in practical scenarios.

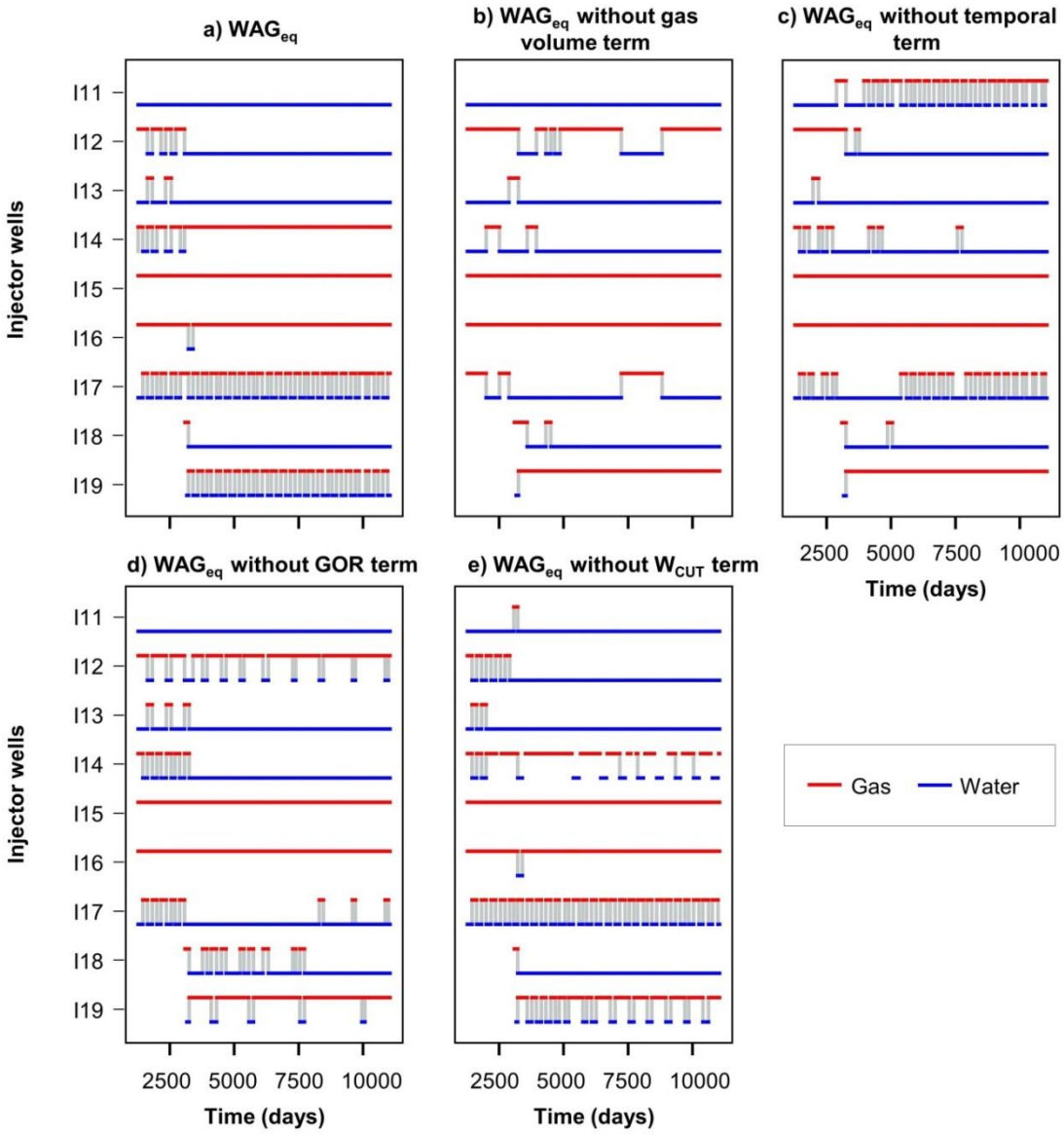

Figure 5.9–Comparison between WAG profile for the best strategy of a) WAG$_{eq}$ and the WAG$_{eq}$ without one of the following terms: b) gas volume, c) temporal term, d) GOR, and e) W$_{CUT}$.

We found that removing the term $\frac{-\beta_p^i \times GOR^p}{GOR_{max}}$ from the equation leads to more dispersed GOR curves over time compared to the original WAG$_{eq}$ strategy, as shown in **Figure 5.10**. This indicates that the GOR term does indeed induce the desired behavior of making GOR more even among the producers, which could improve sweep efficiency.

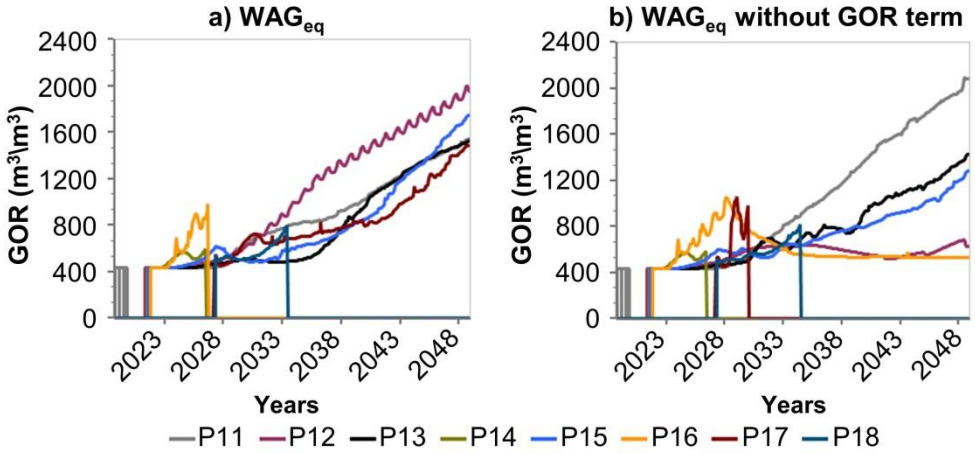

Figure 5.10–Gas oil ratio (GOR) evolution over time for producers P11 to P18 analyzed in two scenarios: a) WAG_{eq} and of b) WAG_{eq} without GOR term.

After removing the W_{CUT} term from the equation, we did not observe significant changes in the injection profiles of the wells (**Figure 5.9e**). This indicates that the W_{CUT} term had a minor effect on the WAG_{eq} strategy, and the top four injectors' rank, which determines the gas injection wells, remained largely unchanged. The optimal WAG_{eq} strategy may have prioritized gas production control over water production control, as the former was deemed more pressing. According to **Figure 5.11a** and **Figure 5.11b**, most wells exhibited W_{CUT} values below 0.6 for both the WAG_{eq} original strategy and the strategy without the W_{CUT} term, indicating that water production was not a significant concern in this study case. Nonetheless, the inclusion of the W_{CUT} term may prove beneficial in other cases where water production is a more critical issue.

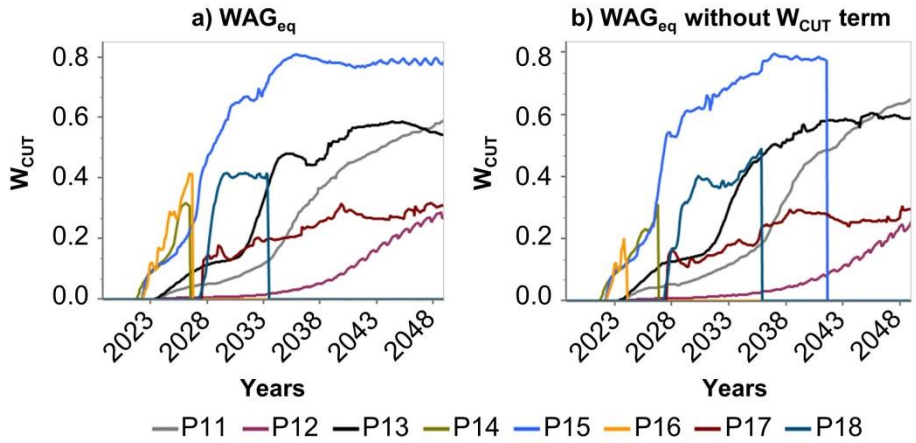

Figure 5.11–Water cut evolution over time for producers P11 to P18 analyzed in two scenarios: a) using the WAG_{eq} method, and b) using the WAG_{eq} method without the W_{CUT} term.

The results indicate that removing one term of the equation had a minimal impact on both the optimization convergence of the algorithm and the NPV of the best solution, as seen in **Figure 5.12**. The maximum decrease in NPV was 1.2% (95 million dollars) for the WAG_{eq} solution without the gas volume term. Despite the slightly higher NPV of the original WAG_{eq}, it is advisable to retain all the equation terms as it offers a more practical solution from an engineering perspective. It also widens the search space without affecting the convergence rate of the algorithm, which can be important for other study cases.

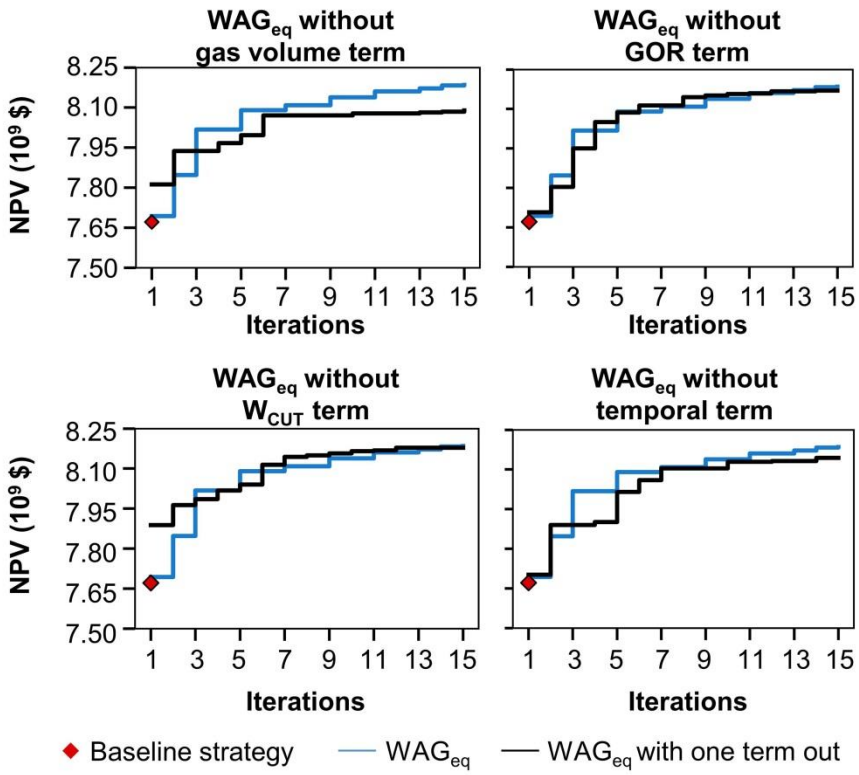

Figure 5.12–Comparison of maximum NPV until each iteration from the best strategy of WAG_{eq} and each of the WAG_{eq} best strategies with one term excluded.

5.5 Conclusions

Our proposed method, WAG_{eq}, aims to optimize the miscible water alternating gas (WAG) process by using a tailored parametric equation based on reservoir production data. This method offers significant advantages over other methods, and the major conclusions drawn from this study are:

- The WAG_{eq} method successfully achieved its main goal of accelerating the optimization process compared to the benchmark WAG_{bm} procedure, which determines the optimal type of fluid injection (gas or water) for each well over time. The WAG_{eq} reduced the

simulations by almost sixfold, while also delivered a slightly higher net present value (NPV) compared to WAG_{bm}.

- The WAG_{eq} substantially increased the NPV by 511 USD million (6.7%) compared with the baseline strategy ($Well_{gor\_limit}$), which has fixed WAG cycles of six months. This result highlights the significance of optimizing the WAG profile over time for improved economic return.
- The WAG_{eq} enables well-behaved WAG profiles for each injector, facilitating practical implementation.
- The non-cyclical pattern observed in the optimal WAG solution for most injectors underscores the importance of WAG_{eq} being able to include flexible solutions in the optimization search space.
- The WAG_{eq} equation terms, designed to favor alternating water and gas injection, regular injection patterns, and uniform gas-oil ratio production, were successful in generating these desired behaviors in the optimal WAG solution profile. Although the inclusion of a term aiming for homogeneous water cut (W_{CUT}) growth did not demonstrate any significant impact on the optimal WAG solution in this study, it may be relevant in cases where water production is a concern.

Overall, this paper presents a significant contribution in the development of a practical procedure that addresses crucial factors in real-world scenarios. Specifically, it minimizes computational effort, ensures well-behaved WAG profiles across all injectors and delivers a high NPV, which is helped by the procedure's ability to produce flexible solutions. Future research opportunities include testing its validity for studies cases considering uncertainties and where water production is a more significant issue, as well as combining this approach with machine learning techniques to further reduce computational effort.

5.6 Acknowledgments

We conducted this work with the support of Libra Consortium (Petrobras, Shell Brasil, TotalEnergies, CNOOC, CNPC, and PPSA) within the ANP R&D levy as “commitment to research and development investments” and Energi Simulation. The authors are grateful for the support of the Center for Petroleum Studies (CEPETRO-UNICAMP/Brazil), the Department of Energy (DE-FEM-UNICAMP/Brazil), and the Research Group in Reservoir

Simulation and Management (UNISIM-UNICAMP/Brazil). A special thanks also goes to CMG for software licenses.

5.7 Statements and Declarations

5.7.1 Ethical statements

The authors confirm that this manuscript has not been previously published and is not currently under consideration by any other journal. Each named author has substantially contributed to conducting the underlying research and drafting this manuscript. Additionally, all of the authors have approved the contents of this paper. They have also agreed to the submission policies of Journal of Petroleum Exploration and Production Technology.

5.7.2 Funding and conflicts of interests/competing interests

On behalf of all authors, the corresponding author states that there is no conflict of interest. Besides that, the authors have no relevant financial or non-financial interests to disclose.

5.8 List of symbols

C_m^n	Combination of n elements taken m at a time
F	Threshold cut percentage to select the best samples
$\mathrm{f_p}(\mathrm{GOR}^{\mathrm{p}}_{\mathrm{limit}})$	Control rule function to shut the producer p
$\mathrm{GOR}^{\mathrm{p}}_{\mathrm{limit}}$	Gas-oil ratio limit value to shut the producer p
$\mathrm{GOR}_{\mathrm{max}}$	Gas-oil ratio maximum value normalizer
$\mathrm{GOR}^{\mathrm{p}}$	Gas-oil ratio of producer p
i	Index representing the injector
I^i	Priority index of injector i
N	Sample size generated in each Iterative discrete Latin hypercube iteration
$NPV(x)$	Net present value objective function depending on the decision variable vector x
$C(x)$	Inequality constraint of the maximization problem
$Eq(x)$	Equality constraint of the maximization problem
m	Number of gas injectors
m_p^i	Number of producers influenced by the gas injector i

n	Number of injectors
n_p^i	Number of producers influenced by the water injector i
nt	Number of minimal time intervals
p	Index representing the producer
t	Time passed from initial management date
t_{end}	Total simulation period
t_{init}^i	Initial time the injector i starts to operate
t_{min}	Minimal time interval
V_g^i	Volume of gas injected by injector i since it has injected water
V_{norm}^i	Gas volume normalizer for injector i
$\text{Well}_{\text{gor\_limit}}$	Baseline strategy
W_{CUT}^p	Water cut of producer p
x	Generic decision variables
x_{lb}	Lower bound of the generic decision variable x
x_{min}	Minimum value for a specific feature
x_{up}	Upper bound of the generic decision variable x

5.8.1 Greek letters

α_p^i	Coefficient of the producer p that are influenced by the water injector i
β_p^i	Coefficient of the producer p that are influenced by the gas injector i
γ^i	Coefficient associate with time term of the parametric equation
δ^i	Coefficient associate with the volume of gas produced term of the parametric equation

5.9 Abbreviations

DLHC	Discrete Latin hypercube sampling
GOR	Gas-oil ratio
$\text{GOR}_{\text{limit}}$	Gas-oil ratio limit
IDLHC	Iterative discrete Latin hypercube
NPV	Net present value
PMF	Probability mass function
USD	United states dollar

WAG	Water alternating gas
WAG_{bm}	Water alternating gas benchmark procedure
WAG_{eq}	Water alternating gas parametric equation procedure
W_{CUT}	Water cut

6 Conclusions

This work addressed the challenges encountered in the literature concerning the optimization methods for life-cycle control (G2L) that hinder their real-world applicability. It effectively tackled two key issues: high computational efforts and methods that produce excessively complex or non-applicable strategies from an engineering standpoint.

To achieve our objectives, we employed innovative approaches to parameterize the G2L variables for reducing the search space. This reduction led to the algorithm requiring fewer simulations to converge to good solutions, while effectively mitigating the risk of being trapped in local minimum solutions. Moreover, we ensured that the resulting solutions remained practical and applicable in real-world scenarios. The parameterizations were designed to explore a wide range of solutions, ultimately contributing to the attainment of favorable economic outcomes.

To enhance efficiency, we integrated machine learning (ML) techniques into the optimization algorithm for some parameterizations. This integration enabled us to replace computationally expensive physics-based simulations with more efficient machine learning-driven approaches. This hybrid approach not only reduced computational efforts but also enhanced the optimization process outcome compared to traditional approaches and with well short-term control strategy (G2S) broadly applied in industry.

The first study case for G2L involved a Black-oil model under uncertainties, where water injection was considered as the recovery method. We began by investigating four parameterizations for G2L and compared them with a G2S strategy that prioritizes production in wells with higher oil-water ratio and aims to replicate industry practice. The first parameterization aimed to determine both the producers' and injectors' apportionment of liquid rates at the beginning of field management, as well as to define the optimal shut-in time for the wells. In contrast, the other three parameterizations focused on optimizing the coefficients within first and second-order polynomial equations, as well as the logistic equation. The purpose of these parametric equations was to govern the bottom-hole pressure (BHP) over time. We summarize the key contributions of this part below:

- All the proposed parameterizations not only effectively facilitate convergence to good solutions in a high-dimensional problem (compared to existing literature studies) but also

demonstrated a significant increase in net present value (NPV) when compared to the G2S strategy representing industry applications.

- By employing equations for G2L control, a considerable reduction in the number of variables was achieved, decreasing from 2300 to fewer than 100, as compared to the commonly studied stepwise method in the literature, which involves defining the well's BHP at each time step.
- The logistic parametric equation (P_{LOG}) demonstrated superior economic performance compared to all other parameterizations in a nominal experiment of this studied case.
- The P_{LOG} parameterization exhibited a significant enhancement in economic returns when implemented under uncertainties and in the reference case, exceeding the performance of the G2S strategy by 21% and 10%, respectively.
- The stepwise approach in literature results may result in strategies with frequently abrupt BHP changes between time steps, which may poses risks to equipment and formation integrity. Conversely, the intrinsic nature of the logistic equation effectively addressed this issue by prioritizing smoother BHP curves across all cases (nominal, probabilistic, and reference case).
- The successful outcomes obtained through G2L optimization, which demonstrated favorable results in representing the real reservoir model, suggest that companies can benefit from not solely relying on operator experience but also strategically planning the G2L approach for enhanced well control management.
- Furthermore, the implementation of the optimal P_{LOG} strategy led to certain injectors remaining inactive during management. This underscores the importance of G2L strategy planning even during field development phase, as it can help avoid unnecessary costs associated with drilling and perforating of these wells.
- We demonstrated the advantages of adopting equation-based parameterization, which not only significantly reduces the number of variables to optimize but also allows us to obtain solutions that closely align with specific operational requirements by accurately defining the right parameters.

Having established the optimal parameterization for G2L, we integrated the iterative discrete Latin Hypercube optimization algorithm (IDLHC) with machine learning techniques (IDLHC-ML) to streamline the P_{LOG} optimization process for nominal and probabilistic cases.

The approach involved simulating a subset of the best solutions predicted by the machine learning model in each IDLHC iteration. The key findings from applying machine learning to expedite the optimization process are as follows:

- The nominal case was used to validate the machine learning algorithms, their training features, and IDLHC optimization parameters.
- The coefficients of the logistic equation and the BHP over time, derived from the equation, were evaluated as features. Incorporating BHP as features yielded enhanced accuracy in correctly classifying optimal strategies for simulation in each IDLHC iteration, surpassing the utilization of coefficients alone. This improvement held true regardless of the IDLHC configuration or the specific machine learning algorithm employed.
- A total of twelve machine learning regression techniques were tested. Among them, the stacked learning algorithms, which leverage predictions from multiple ML algorithms as input to a meta-learning model, demonstrated superior performance. Specifically, the multi-layer perceptron meta-learner (MLP-S) achieved the best NPV outcome, while the Support Vector Machine meta-learner (SVM-S) provided a similar NPV with fewer simulations. The choice between the two models depends on the project's time constraints and the desired level of precision.
- Both MLP-S and SVM-S versions achieved significant simulation reductions of 45% and 27%, respectively, compared to the standard IDLHC algorithm, without compromising NPV.
- We enhanced the IDLHC-ML algorithm by integrating uncertainties through representative models (RMs) for robust optimization, creating the IDLHC-MLR method. The IDLHC-MLR achieved a remarkable 45% reduction in required simulations (from 9,000 to 4,950 simulations), while preserving the expected monetary value for logistic equation optimization, compared to the traditional IDLHC approach.
- Our findings confirm the robustness of the IDLHC-MLR method, showing a improvement in the Expected Monetary Value (EMV) for both data assimilation models and the reference case, compared to the traditional IDLHC approach.

- The machine learning techniques offer a compelling alternative to the original IDLHC method, especially when time constraints are present in defining G2L strategies or to make a prior assessment of the G2L control rule for the case studied.

Given the importance of Water-Alternating-Gas (WAG) injection, particularly in fields with light oil and high CO_2 content requiring full reinjection, we developed a technique for controlling WAG injection with reduced computational effort. The technique was applied in a second study case using a reservoir simulation model designed to replicate the geological and fluid characteristics of Brazilian pre-salt fields. Our approach involved formulating a WAG-based equation to calculate an injection priority index for each well. This index dictates the priority of each well for injecting water or gas at each time interval. The WAG equation approach (WAG_{eq}) resulted in the major following findings:

- We presented the benchmark method (WAG_{bm}) for comparison in terms of the number of simulations and NPV. The WAG_{bm} method involves determining the optimal type of fluid injection (gas or water) for each well over time.
- The WAG_{eq} method remarkably reduced the number of required simulations by nearly sixfold, and yielded a slightly higher NPV compared to WAG_{bm}.
- The equation is built upon four terms, each parametrically designed to induce anticipated behaviors in the WAG profile from an engineering perspective. The expected effects of inducing alternating water and gas injection, promoting regular injection patterns, and achieving a uniform gas-oil ratio production among the wells were successfully achieved. While the incorporation of a term aiming for homogeneous water-cut (W_{CUT}) growth in the producers did not exhibit any significant impact on the optimal WAG profile, it may hold relevance in scenarios where water production is a concern.
- The WAG_{eq} significantly improved the NPV compared to a traditional method that utilizes fixed WAG cycles of six months. While fixed cycle patterns for all wells are commonly applied in literature, adopting more flexible profiles can potentially lead to better economic returns.
- The optimal strategy derived from WAG_{eq} is found to be straightforward to implement in practical cases.

- Most injectors exhibited a non-cyclical behavior in both the optimal solutions of WAG_{eq} and WAG_{bm}, underscoring the significance of WAG_{eq}'s flexibility in exploring non-cyclical and individual patterns for each well during the optimization process.

Overall, this work provided valuable procedures for optimizing well life cycle control variables that allow a reduction in computational effort while enabling the exploration of a wide range of solutions. The inherent ability of these procedures to generate flexible solutions enhanced the potential for achieving high economic returns. Furthermore, the approaches have proven to be particularly advantageous in producing practical and well-behaved strategies.

7 Suggestions for Future Studies

This research has addressed crucial questions by developing and presenting computationally efficient techniques for optimizing well life-cycle control (G2L). These techniques guarantee the generation of practical solutions while also yielding a favorable economic return. However, there is still untapped potential to further expand in the following areas:

- To validate and strengthen the parameterization of the logistic equation for wells bottom-hole pressure proposed in Chapter 2, it would be beneficial to test it in other study cases. Additionally, since the assimilated models did not fully capture the uncertainties from the reference model (UNISIM-I-R), an opportunity arises to implement well control in closed-loop cycles.
- The method that combines the iterative discrete Latin-hypercube optimization algorithm with machine learning has exhibited favorable performance for the nominal and probabilistic approaches. Nevertheless, it is crucial to extend and explore its effectiveness in other studied scenarios, particularly those with higher computational demands. These explorations would contribute significantly to validating the method's generalizability and identifying specific scenarios where it excels. Moreover, given the successful application of the machine learning approach, it would be advantageous to adapt and employ it for analyzing other types of variables, such as for the water alternating gas injection discussed in Chapter 5.
- An alternative approach involves optimizing the variables in the management phase, leveraging and adapting the machine learning and parameterization techniques developed in this work. In this case, the variables extend beyond G2L to include revitalization variables (G3) such as wells conversion, infill drilling, and recompletion.
- Another avenue for future research involves the application of machine learning to reduce simulation time. This can be accomplished by partially running the simulations and employing a machine learning for time series forecasting to identify strategies that yield the best economic return. By doing so, we can harness the benefits of the method proposed in this work, which effectively reduces the number of simulations, along with methods that cut down simulation time.

- In Chapter 5, we implemented the WAG parameterization using an equation to determine the priority of each well for gas or water injection in a nominal case. To ensure the method's reliability, it is imperative to test its validity in scenarios under uncertainties. Additionally, it is essential to investigate cases where water production becomes a more significant concern, as the term in the priority equation specifically intended for water control did not demonstrate any effect.

8 REFERENCES

AFZALI, S., REZAEI, N., ZENDEHBOUDI, S. A Comprehensive Review on Enhanced Oil Recovery by Water Alternating Gas (WAG) Injection. **Fuel**, v. 227, pp. 218-246. 2018. https://doi.org/10.1016/j.fuel.2018.04.015.

AROGUNDADE, O. A., SHAHVERDI, H. R., SOHRABI, M. **A Study of Three Phase Relative Permeability and Hysteresis in Water Alternating Gas (WAG) Injection.** SPE Enhanced Oil Recovery Conference**.** 2013**.** https://doi.org/10.2118/165218-MS.

AVANSI, G. D., SCHIOZER, D. J. UNISIM-I: Synthetic Model for Reservoir Development and Management Applications. **International Journal of Modeling and Simulation for the Petroleum Industry**, v. 9(1), pp. 21–30. 2015.

AWOTUNDE, A. A. **On The Joint Optimization of Well Placement and Control**. SPE Saudi Arabia Section Technical Symposium and Exhibition, 21–24 April, Al-Khobar, Saudi Arabia. 2014. https://doi.org/10.2118/172206-MS.

BAHAGIO, D. N. T. **Ensemble Optimization of CO2-WAG EOR**. 2013. Master's thesis - Delft University of Technology, Delft, Netherlands.

BAHRAMI, P., JAMES, L. A. Screening of Waterflooding Using Smart Proxy Model Coupled with Deep Convolutional Neural Network. **Geoenergy Science and Engineering**, v. 221. 2023. https://doi.org/10.1016/j.petrol.2022.111300.

BARRETO, C. E. A. G., MAZO, E. O. M., SCHIOZER, D. J. **Análise da Determinação do Corte de Água Ótimo para o Gerenciamento de Águas sob Restrições Operacionais. Rio Oil & Gas Expo and Conference**, Rio de Janeiro. 2010.

BELLOUT, M. C., CIAURRI, D. E., DURLOFSKY, L. J., FOSS, B., KLEPPE, J. Joint Optimization of Oil Well Placement and Controls. **Computational Geosciences**, v. 16, pp. 1061–1079. 2012. https://doi.org/10.1007/s10596-012-9303-5.

BERGSTRA, J., BENGIO, Y. Random Search for Hyper-Parameter Optimization. **Journal of Machine Learning Research**, v. 13(10), pp. 281–305. 2012.

BISHOP, C. M. **Pattern Recognition and Machine Learning (Information Science and Statistics)**. New York: Publisher Springer, 2006.

BOTECHIA, V. E., GASPAR, A. T. F., AVANSI, G. D., DAVOLIO, A., SCHIOZER, D. J. Investigation of Production Forecast Biases of Simulation Models in A Benchmark Case. **Oil & Gas Science and Technology - Revue d'IFP Energies nouvelles**, v. 73(23). 2018. https://doi.org/10.2516/ogst/2018014.

BOTECHIA, V. E., LEMOS, R. A., VON HOHENDORFF FILHO, J. C., SCHIOZER, D. J. Well and ICV Management in a Carbonate Reservoir with High Gas Content. **Journal of Petroleum Science and Engineering**, v. 200. 2021. https://doi.org/10.1016/j.petrol.2021.108345.

BOTECHIA, V. E., SCHIOZER, D. J. Model-Based Life Cycle Control of ICVs in Injectors in a Benchmark Analogous to a Pre-Salt Field. **Journal of Petroleum Science and Engineering**, v. 215(B). 2022. https://doi.org/10.1016/j.petrol.2022.110707.

BREIMAN, L. Bagging Predictors. **Machine Learning**, v. 24(2), pp. 123–140. 1996. https://doi.org/10.1023/A:1018054314350.

BRUYELLE, J., GUÉRILLOT, D. **Well Placement Optimization with an Artificial Intelligence Method Applied to Brugge Field**. SPE Gas & Oil Technology Showcase and Conference, Dubai, UAE, 21–23 October. 2019. https://doi.org/10.2118/198656-MS.

BUSBY, D., PIVOT, F., TADJER, A. **Use of Data Analytics to Improve Well Placement Optimization Under Uncertainty**. Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 13–16 November. 2017. https://doi.org/10.2118/188265-MS.

CHAVES, J. M. P. **Multiscale approach to construct a carbonate reservoir model with karstic features and Brazilian pre-salt trends using numerical simulation**. 2018. 103 p. Master's thesis – Mechanical Engineering Faculty and Geosciences Institute, University of Campinas, Campinas.

CHEN, B., REYNOLDS, A. C. Ensemble-Based Optimization of the Water-Alternating-Gas-Injection Process. **SPE Journal**, v. 21(03), pp. 0786–0798. 2016. https://doi.org/10.2118/173217-PA.

CHEN, B., HE, J., WEN, X. H., CHEN, W., REYNOLDS, A. C. Uncertainty Quantification and Value of Information Assessment Using Proxies and Markov Chain Monte Carlo Method for a Pilot Project. **Journal of Petroleum Science and Engineering**, v. 157, pp. 328–339. 2017. https://doi.org/10.1016/j.petrol.2017.07.039.

CHEN, S., LI, H., YANG, D., TONTIWACHWUTHIKUL, P. Optimal Parametric Design for Water-Alternating-Gas (WAG) Process in a CO_2-Miscible Flooding Reservoir. **Journal of Canadian Petroleum Technology**, v. 49(10), pp. 75–82. 2010. https://doi.org/10.2118/141650-PA.

CHRISTENSEN, J. R., STENBY, E. H., SKAUGE, A. Review of WAG Field Experience. **SPE Reservoir Evaluation & Engineering**, v. 4(02), pp. 97–106. 2001. https://doi.org/10.2118/71203-PA.

CORREIA, M. G., BOTECHIA, V. E., PIRES, L. C. O., RIOS, V. S., SANTOS, S. M. G., RIOS, V. S., VON HOHENDORFF FILHO, J. C., CHAVES, J. M. P., SCHIOZER, D. J. **UNISIM-III: Benchmark Case Proposal Based on a Fractured Karst Reservoir**. Proceedings of the European Association of Geoscientists & Engineers. 2020. https://doi.org/10.3997/2214-4609.202035018.

DORAISAMY, H., ERTEKIN, T., GRADER, A. S. **Key Parameters Controlling the Performance of Neuro-Simulation Applications in Field Development**. SPE Eastern Regional Meeting, Pittsburgh, USA, 9–11 November. 1998. https://doi.org/10.2118/51079-MS.

DUCHENNE, S., PUYOU, G., CORDELIER, P., BOURGEOIS, M., HAMON, G. **Laboratory Investigation of Miscible CO_2 WAG Injection Efficiency in Carbonate**. SPE EOR Conference at Oil and Gas West Asia. 2014. https://doi.org/10.2118/169658-MS.

DUDA, R. O., HART, P. E., STORK, D. G. **Pattern Classification.** 2. ed. New York: Wiley, 2001.

DZYUBA, V. I., LITVINENKO, Y. V., BOGACHEV, K. Y., MIGRASIMOV, A. R., SEMENKO, A. E., KHACHATUROVA, E. A., EYDINOV, D. A. **Application of Sector Modeling Technology for Giant Reservoir Simulations**. SPE Russian Oil & Gas Exploration & Production Technical Conference and Exhibition held in Moscow, Russia. 2012. https://doi.org/10.2118/162090-MS.

EBADI, F., DAVIES, D. R. **Should "Proactive" or "Reactive" Control be Chosen for Intelligent Well Management?** Intelligent Energy Conference and Exhibition, Amsterdam. 2006.

ECHEVERRÍA, C. D., ISEBOR, O. J., DURLOFSKY, L. J. Application of Derivative-Free Methodologies to Generally Constrained Oil Production Optimisation Problems. **International Journal of Mathematical Modelling and Numerical Optimisation**, v. 2(2), pp. 134–161. 2011. https://doi.org/10.1504/IJMMNO.2011.039425.

ESMAIEL, T. E., FALLAH, S. B., VAN KRUISDIJK, C. **Determination of WAG Ratios and Slug Sizes Under Uncertainty in a Smart Wells Environment**. SPE Middle East Oil and Gas Show and Conference, Kingdom of Bahrain. 2005. https://doi.org/10.2118/93569-MS.

FREUND, Y., SCHAPIRE, R. E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. **Journal of Computer and System Sciences**, v. 55(1), pp. 119–139. 1997. https://doi.org/10.1006/jcss.1997.1504.

FOROUZANFAR, F., POQUIOMA, W. E., REYNOLDS, A. C. Simultaneous and Sequential Estimation of Optimal Placement and Controls of Wells With a Covariance Matrix Adaptation Algorithm. **SPE Journal**, v. 21(2), pp. 501–521. 2016. https://doi.org/10.2118/173256-PA.

GASPAR, A. T. F. S., AVANSI, G. D., SANTOS, A. A., VON HOHENDORFF FILHO, J. C., SCHIOZER, D. J. UNISIM-I-D: Benchmark Studies for Oil Field Development and Production Strategy Selection. **International Journal of Modeling and Simulation for the Petroleum Industry**, v. 9(1), pp. 47–55. 2015.

GASPAR, A. T. F. S., AVANSI, G. D., SCHIOZER, D. J. **UNISIM-I-M: Benchmark Case Proposal for Oil Reservoir Management Decision-Making**. SPE Trinidad and Tobago Section Energy Resources Conference, 13–15 June, Port of Spain, Trinidad and Tobago. 2016a. https://doi.org/10.2118/180848-MS.

GASPAR, A. T. F., BARRETO, C. E. A., SCHIOZER, D. J. Assisted Process for Design Optimization of Oil Exploitation Strategy. **Journal of Petroleum Science and Engineering**. v. 146, pp. 473–488. 2016b. https://doi.org/10.1016/j.petrol.2016.05.042.

GLOVER, F. Tabu Search–Part II. **ORSA Journal on Computing**, v. 2(1), pp. 4–32. 1990. https://doi.org/10.1287/ijoc.2.1.4.

GODOI, J. M. A., DOS SANTOS, M. P. H. L. Enhanced Oil Recovery with Carbon Dioxide Geosequestration: First Steps at Pre-Salt in Brazil. **Journal of Petroleum Exploration and Production Technology**, v. 11(03), pp. 1429–1441. 2021. https://doi.org/10.1007/s13202-021-01102-8.

GOODFELLOW, I., BENGIO, Y., COURVILLE, A. **Deep Learning**. Publisher: The MIT Press, 2016.

GU, Y., OLIVER, D. An Iterative Ensemble Kalman Filter for Multiphase Fluid Flow Data Assimilation. **SPE Journal**, v. 12(4), pp. 438–446. 2007. https://doi.org/10.2118/108438-PA.

HANEA, R., CASANOVA, P., HUSTOFT, L., BRATVOLD, R., NAIR, R., HEWSON, C. W., LEEUWENBURGH, O., FONSECA, R. M. Drill and Learn: A Decision-Making Work Flow to Quantify Value of Learning. **SPE Reservoir Evaluation and Engineering**, v. 22(03), pp. 1131–1143. 2019. https://doi.org/10.2118/182719-PA.

HAN, X., ZHONG, L., WANG, X., LIU, Y., WANG, H. Well Placement and Control Optimization of Horizontal Steamflooding Wells Using Derivative-Free Algorithms. **SPE Reservoir Evaluation & Engineering,** v. 24(01), pp. 174–193. 2021. https://doi.org/10.2118/203821-PA.

HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. H. **The elements of statistical learning: data mining, inference, and prediction**. 2. ed. New York: Publisher Springer, 2009.

HE, J., SARMA, P., BHARK, E., TANAKA, S., CHEN, B., WEN, X. H., AND JAIRAM, K. Quantifying Expected Uncertainty Reduction and Value of Information Using Ensemble-Variance Analysis. **SPE Journal**, v. 23, pp. 428–448. 2018. https://doi.org/10.2118/182609-PA.

HIDALGO, D. M., EMERICK, A. A., COUTO, P., ALVES, J. L. D. **Closed-loop Field Development under Geological Uncertainties: Application in a Brazilian Benchmark Case**. Offshore Technology Conference, Rio de Janeiro, Brasil, pp. 24–26. 2017. https://doi.org/10.4043/28089-MS.

HU, G., LI, P., YI, L., ZHAO, Z., TIAN, X., LIANG, X. Simulation of Immiscible Water-Alternating-CO_2 Flooding in the Lihue Oilfield Offshore Guangdong, China. **Energies**, v. 13(9). 2020. https://doi.org/10.3390/en13092130.

HUMPHRIES, T. D., HAYNES, R. D., JAMES, L. A. Simultaneous and sequential approaches to joint optimization of well placement and control. **Computational Geosciences**, v. 18, pp. 433–448. 2014. https://doi.org/10.1007/s10596-013-9375-x.

ISEBOR, O. J. **Constrained Production Optimization With an Emphasis on Derivative-Free Methods**. 2009. 92 p. Master's thesis – Department of Energy Resources Engineering, Stanford University, California.

ISEBOR, O. J., DURLOFSKY, L. J., ECHEVERRÍA, C. D. A Derivative-Free Methodology With Local and Global Search for The Constrained Joint Optimization of Well Locations and Controls. **Computational Geosciences**, v. 18, pp. 463–482. 2014. https://doi.org/10.1007/s10596-013-9383-x.

JAHANDIDEH, A., JAFARPOUR, B. Closed-Loop Stochastic Oilfield Optimization for Hedging Against Geologic, Development, and Operation Uncertainty. **Computational Geosciences**, v. 24, pp. 129–148. 2020. https://doi.org/10.1007/s10596-019-09902-y.

JANG, I., OH, S., KIM., Y., PARK, C, KANG, H. Well-Placement Optimisation Using Sequential Artificial Neural Networks. **Energy Exploration & Exploitation**, v. 36(3), pp. 433–449. 2018. https://doi.org/10.1177/0144598717729490.

JANSSEN, M. T., TORRES, M. F. A., ZITHA, P. L. Mechanistic Modeling of Water-Alternating-Gas Injection and Foam-Assisted Chemical Flooding for Enhanced Oil Recovery. **Industrial & Engineering Chemistry Research**, v. 59(8), pp. 3606-3616. 2020.

KAZAKOV, K. V., BRAVICHEV, K. A. **WAG Optimization Algorithms Based on the Reaction of Producing Wells**. SPE Russian Petroleum Technology Conf. 2015. https://doi.org/10.2118/176576-MS.

KONG, D., GAO, Y., SARMA, H., LI, Y., GUO, H., ZHU, W. Experimental Investigation of Immiscible Water-Alternating-Gas Injection in Ultra-High Water-Cut Stage Reservoir. **Advances in Geo-Energy Research**, v. 5(2), pp. 139-152. 2021. https://doi.org/10.46690/ager.2021.02.04.

KOU, Z., WANG, H., ALVARADO, V., MCLAUGHLIN, J. F., QUILLINAN S. A. Method for Upscaling of CO_2 Migration in 3d Heterogeneous Geological Models. **Journal of Hydrology**, v. 613(A). 2022. https://doi.org/10.1016/j.jhydrol.2022.128361.

KULKARNI, M. M., RAO, D. N. Experimental Investigation of Miscible and Immiscible Water-Alternating-Gas (WAG) Process Performance. **Journal of Petroleum Science and Engineering**, v. 48(1), pp. 1–20. 2005. https://doi.org/10.1016/j.petrol.2005.05.001.

LECUN, Y. A., BOTTOU, L., ORR, G. B., MÜLLER, K. R. **Efficient BackProp**. Montavon, G., Orr, G.B., Müller, KR. (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, v. 7700. Publisher Springer, Berlin, Heidelberg. 2012. https://doi.org/10.1007/978-3-642-35289-8_3.

LI, L., JAFARPOUR, B. A Variable-Control Well Placement Optimization for Improved Reservoir Development. **Computational Geosciences**, v. 16(4), pp. 871–889, 2012.

LIGERO, E. L., SCHIOZER, D. J. **Miscible WAG-CO_2 Light Oil Recovery from Low Temperature and High Pressure Heterogeneous Reservoir**. SPE Latin America and Caribbean Petroleum Engineering Conf. 2014. https://doi.org/10.2118/169296-MS.

LOOMBA, A. K., BOTECHIA, V. E., SCHIOZER, D. J. A Comparative study to accelerate field development plan optimization. **Journal of Petroleum Science and Engineering**, v. 208(D). 2022. https://doi.org/10.1016/j.petrol.2021.109708.

LORENTZEN, R. J., SHAFIEIRAD, A., NÆVDAL, G. **Closed-loop Reservoir Management Using the Ensemble Kalman Filter and Sequential Quadratic Programming**. SPE Reservoir Simulation Symposium, The Woodlands, Texas, USA, 2–4 February. 2009. https://doi.org/10.2118/119101-MS.

MASCHIO, C., SCHIOZER, D. J. Probabilistic History Matchingusing Discrete Latin Hypercube Sampling and Nonparametric Density Estimation. **Journal of Petroleum Science and Engineering**, v. 147, pp. 98–115. 2016. https://doi.org/10.1016/j.petrol.2016.05.011.

MEIRA, L. A., COELHO, G. P., SILVA, C. G., ABREU, J. L. A., SANTOS, A. A. S., SCHIOZER, D. J. Improving Representativeness in a Scenario Reduction Process to Aid Decision Making in Petroleum Fields. **Journal of Petroleum Science and Engineering**, v. 184. 2020. https://doi.org/10.1016/j.petrol.2019.106398.

MEIRA, L. A., COELHO, G. P., SILVA, C. G., SCHIOZER, D. J., SANTOS, A. A. S. **RMFinder 2.0: An Improved Interactive Multicriteria Scenario Reduction Methodology**. SPE Latin America and Caribbean Petroleum Engineering Conference,17–19 May, Buenos Aires, Argentina. 2017. https://doi.org/10.2118/185502-MS.

MELLO, S. F. **Caracterização de Fluido e Simulação Composicional de Injeção Alternada de Água e CO_2 para Reservatórios Carbonáticos Molháveis à Água**. 2015. 298 p. Ph.D. thesis - University of Campinas, Campinas.

MELLO, S. F., AVANSI, G. D., RIOS, V. S., SCHIOZER, D. J. Computational Time Reduction of Compositional Reservoir Simulation Model with WAG Injection and Gas Recycle Scheme Through Numerical Tuning of Submodels. **Brazilian Journal of Petroleum and Gas**, v. 16(1). 2022. http://dx.doi.org/10.5419/bjpg.

MIN, B. H., PARK, C., KANG, J. M., PARK, H. J., JANG, I. S. Optimal Well Placement Based on Artificial Neural Network Incorporating the Productivity Potential. **Energy Sources, Part A: Recovery, Utilization, and Environmental Effects**, v. 33(18), pp. 1726–1738. 2011. https://doi.org/10.1080/15567030903468569.

MOROSOV, A. L., SCHIOZER, D. J. Field-Development Process Revealing Uncertainty-Assessment Pitfalls. **SPE Reservoir Evaluation and Engineering**, v. 20, pp. 765–778. 2017. https://doi.org/10.2118/180094-PA.

MOUSAVI, M. S. M., HOSSEINI, S. J., MASOUDI, R., ATAEI, A., DEMIRAL, B. M., KARKOOTI, H. **Investigation of Different I-WAG Schemes Toward Optimization of Displacement Efficiency**. SPE Asia Pacific Enhanced Oil Recovery Conference. 2011. https://doi.org/10.2118/144891-MS.

NAEVDAL, G., JOHNSEN, L. M., AANONSEN, S. I., VEFRING, E. H. Reservoir Monitoring and Continuous Model Updating Using Ensemble Kalman Filter. **SPE Journal**, v. 10(1), pp. 66–74. 2005. https://doi.org/10.2118/84372-PA.

NAIT, A. M., ZERAIBI, N., REDOUANE, K. Optimization of WAG Process Using Dynamic Proxy, Genetic Algorithm and Ant Colony Optimization. **Arabian Journal for Science and Engineering**, v. 43(11), pp. 6399–6412. 2018. https://doi.org/10.1007/s13369-018-3173-7.

NWACHUKWU, A., JEONG, H., PYRCZ, M., LAKE, L. W. Fast Evaluation of Well Placements in Heterogeneous Reservoir Models Using Machine Learning. **Journal of Petroleum Science and Engineering**, v. 163, pp. 463–475. 2018. https://doi.org/10.1016/j.petrol.2018.01.019.

PAL, M., PEDERSEN, R. B., GILANI, S. F., TARSAULIYA, G. **Challenges and Learnings from Operating the Largest Off-Shore WAG in the Giant Al-Shaheen Field and Ways to Optimize Future WAG Developments**. SPE EOR Conference at Oil and Gas West Asia. 2018. https://doi.org/10.2118/190343-MS.

PAIAMAN, A. M., SANTOS, S. M. G., SCHIOZER, D. J. A Review on Closed-Loop Field Development and Management. **Journal of Petroleum Science and Engineering**, v. 201, pp. 1–23. 2021. https://doi.org/10.1016/j.petrol.2021.108457.

PANJALIZADEH, H., ALIZADEH, A., GHAZANFARI, M., ALIZADEH, N. Optimization of the WAG Injection Process. **Petroleum Science and Technology**, v. 33(3), pp. 294–301. 2015. https://doi.org/10.1080/10916466.2014.956897.

PASQUALETTE, M. A., REMPTO, M. J., CARNEIRO, J. N. E., FONSECA R., CIAMBELLI, J. R. P., JOHANSEN, S. T., LØVFALL, B. T. **Parametric Study of the Influence of GOR and CO_2 Content on the Simulation of a Pre-Salt Field Configuration**. Offshore Technology Conference. 2017. https://doi.org/10.4043/28093-MS.

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M., DUCHESNAY, E. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, pp. 2825–2830. 2011.

PEREIRA, F. G. A., BOTECHIA, V. E., SCHIOZER, D. J. Model-Based Optimization of Cycles of CO_2 Water-Alternating-Gas (CO_2-WAG) Injection in Carbonate Reservoir. **Brazilian Journal of Petroleum and Gas**, v. 15(4), pp. 139–149. 2022. https://doi.org/10.5419/bjpg2021-0012.

PERRONE, A., ROSSA, E. D. **Optimizing Reservoir Life-Cycle Production Under Uncertainty: A Robust Ensemble-Based Methodology**. SPE Reservoir Characterisation and Simulation Conference and Exhibition, Abu Dhabi, 14-16 September. 2022. http://dx.doi.org/10.2118/175570-MS.

PETERS, L., ARTS, R., BROUWER, G., GEEL, C., CULLICK, S., LORENTZEN, R., CHEN, Y., DUNLOP, N., VOSSEPOEL, F., XU, R., SARMA, P., ALHUTALI, A. H., REYNOLDS, A. C. Results of The Brugge Benchmark Study for Flooding Optimization and History Matching. **SPE Reservoir Evaluation and Engineering**, v. 13, pp. 391–405. 2010. https://doi.org/10.2118/119094-PA.

PINTO, J. W. O., TUEROS, J. A. R., HOROWITZ, B., SILVA, S. M. B. A., WILLMERSDORF, R. B., OLIVEIRA, D. F. B. Gradient-Free Strategies to Robust Well Control Optimization. **Computational Geosciences**, v. 24, pp. 1959–1978. 2019a. https://doi.org/10.1007/s10596-019-09888-7.

PINTO, M. A. S., GASPAR, A. T. F. S., SCHIOZER, D. J. Optimisation of Well Rates Under Production Constraints. **International Journal of Oil, Gas and Coal Technology**, v. 21(2), pp. 131–148. 2019b. https://doi.org/10.1504/IJOGCT.2019.099586.

PINTO, M. S., HERRERA, D. M., ANGARITA, J. C. G. Production Optimization for a Conceptual Model Through Combined use of Polymer Flooding and Intelligent Well Technology Under Uncertainties. **Revista fuentes el reventón enrgético**, v. 16(1), pp. 37–45. 2018. https://doi.org/10.18273/revfue.v16n1-2018003.

PIRES, L. O., BOTECHIA, V. E., SCHIOZER, D. J. **Application of Sector Modeling Approach in a Probabilistic Study of a Giant Reservoir**. ECMOR XVII. 2020. https://doi.org/10.3997/2214-4609.202035102.

RAMACHANDRAN, K. P., GYANI, O. N., SUR, S. **Immiscible Hydrocarbon WAG: Laboratory to Field**. SPE Oil and Gas India Conference and Exhibition. 2010. https://doi.org/10.2118/128848-MS.

RAMOS, J. A. F, FILHO, D. S. M, LYRA, P. R. M, WILLMERSDORF, R. B., SILVA, R. S., CARVALHO, D. K. E. Automatic Upscaling in Petroleum Reservoir Simulation Using Classical Techniques. **International Journal of Simulation Modelling**, v. 6(2), pp. 43–50. 2012.

RAHIM, S., LI, Z. Well Placement Optimization with Geological Uncertainty Reduction. **IFAC-PapersOnLine**, v. 48(8), pp. 57–62. 2015. https://doi.org/10.1016/j.ifacol.2015.08.157.

RISSO, F., RISSO, V., SCHIOZER, D. J. Risk Analysis of Petroleum Fields Using Latin Hypercube, Monte Carol [sic] and Derivative Tree Techniques. **Journal of Petroleum and Gas Exploration Research**, v. 1(1), pp. 014-021. 2011.

SAMPAIO, M. A., DE MELLO, S. F., SCHIOZER, D. J. Impact of Physical Phenomena and Cyclical Reinjection in Miscible CO_2-WAG Recovery in Carbonate Reservoirs. **Journal of Petroleum Exploration and Production Technology**, v. 10(8), pp. 3865–3881. 2020. https://doi.org/10.1007/s13202-020-00925-1.

SANTOS, S. M. G., BOTECHIA, V. E., SCHIOZER, D. J., GASPAR, A. T. F. S. Expected Value, Downside Risk and Upside Potential as Decision Criteria in Production Strategy Selection for Petroleum Field Development. **Journal of Petroleum Science and Engineering**, v. 157, pp. 81–93. 2017a. https://doi.org/10.1016/j.petrol.2017.07.002.

SANTOS, S. M. G., GASPAR, A. T. F. S., SCHIOZER, D. J. Risk Management in Petroleum Development Projects: Technical and Economic Indicators to Define a Robust Production Strategy. **Journal of Petroleum Science and Engineering**, v. 151, pp. 116-127. 2017b. https://doi.org/10.1016/j.petrol.2017.01.035.

SANTOS, S. M. G., GASPAR, A. T. F. S., SCHIOZER, D. J. Comparison of Risk Analysis Methodologies in a Geostatistical Context: Monte Carlo with Joint Proxy Models and Discretized Latin Hypercube. **International Journal for Uncertainty Quantification**, v. 8(1), pp. 23-41. 2018a. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2018019782.

SANTOS, S. M. G., GASPAR, A. T. F. S., SCHIOZER, D. J. Managing Reservoir Uncertainty in Petroleum Field Development: Defining a Flexible Production Strategy from a Set of Rigid Candidate Strategies. **Journal of Petroleum Science and Engineering**, v. 171, pp. 516-528. 2018b. https://doi.org/10.1016/j.petrol.2018.07.048.

SANTOS, D. R. **Influência das Variáveis de Controle de Poços Durante o Desenvolvimento de Campos de Petróleo sob Incertezas**. 2017. 154 p. MSc thesis – University of Campinas, Campinas.

SANTOS, D. R., FIORAVANTI, A. R., BOTECHIA, V. E., SCHIOZER, D. J. Accelerated Optimization of CO_2-Miscible Water-Alternating-Gas Injection in Carbonate Reservoirs Using Production Data-Based Parameterization. **Journal of Petroleum Exploration and Production Technology**, v. 13, pp. 1833–1846. 2023. https://doi.org/10.1007/s13202-023-01643-0.

SANTOS, D. R., FIORAVANTI, A. R., SANTOS, A. A. S., SCHIOZER, D. J. **A Machine Learning Approach to Reduce the Number of Simulations for Long-Term Well Control Optimization**. SPE Annual Technical Conference and Exhibition. 2020. https://doi.org/10.2118/201379-MS.

SANTOS, D. R., FIORAVANTI, A. R., SANTOS, A. A. S., SCHIOZER, D. J. **Investigation of Well Control Parameterization with Reduced Number of Variables Under Reservoir Uncertainties**. SPE EUROPEC, 18-21 October, Amsterdam, Holland. 2021. https://doi.org/10.2118/205207-MS.

SANTOS, D. R., SCHIOZER, D. J. **Influence of Well Control Parameters in the Development of Petroleum Fields Under Uncertainties**. SPE Latin America and Caribbean Mature Fields Symposium, 15–16 March, Salvador, Bahia, Brazil. 2017. https://doi.org/10.2118/184933-MS.

SARMA, P., CHEN, W. H., XIE, J. **Selecting Representative Models from a Large Set of Models**. SPE Reservoir Simulation Symposium, 18–20 February, The Woodlands, Texas, USA. 2013. https://doi.org/10.2118/163671-MS.

SARMA, P., LAWRENCE, K., ZHAO, Y., KYRIACOU, S., SAKS, D. **Implementation and Assessment of Production Optimization in a Steamflood Using Machine-Learning Assisted Modeling**. SPE International Heavy Oil Conference and Exhibition, Kuwait City, Kuwait, 10–12 December. 2018. https://doi.org/10.2118/193680-MS.

SCHAEFER, B. C., REIS, M. H., SCHAEFER, M. F. L., ZULIANI, P., PINTO, M. A. S. **Technical-Economic Evaluation of Continuous CO_2 Reinjection, Continuous Water Injection and Water Alternating Gas (WAG) Injection in Reservoirs Containing CO_2**. XXXVIII Iberian-Latin American Congress on Computational Methods in Engineering. 2017.

SCHIOZER, D. J. Estimation of time for the closed-loop implementation in complex reservoirs. UNISIM ON-LINE 150th edition, 2022. Disponível em: https://www.unisim.cepetro.unicamp.br/online/UNISIM_ON_LINE_N150.PDF. Acesso em: 06 dez 2023.

SCHIOZER, D. J., AVANSI, G. D., SANTOS, A. A. S. Risk Quantification Combining Geostatistical Realizations and Discretized Latin Hypercube. **Journal of the Brazilian Society of Mechanical Sciences and Engineering**, v. 39(2), pp. 575–587. 2017. https://doi.org/10.1007/s40430-016-0576-9.

SCHIOZER, D. J., LIGERO, E., SUSLICK, S., COSTA, A., SANTOS, J. Use of Representative Models in the Integration of Risk Analysis and Production Strategy Definition. **Journal of Petroleum Science and Engineering**, v. 44(1–2), pp. 131–141. 2004. https://doi.org/10.1016/j.petrol.2004.02.010.

SCHIOZER, D. J., SANTOS, A. A. S., DRUMOND, P. S. **Integrated Model Based Decision Analysis in Twelve Steps Applied to Petroleum Fields Development and Management**. SPE EUROPEC, 1–4 June, Madrid, Spain. 2015. https://doi.org/10.2118/174370-MS.

SCHIOZER, D. J., SANTOS, A. A. S., SANTOS, S. M. G., FILHO J. C. H. Model-Based Decision Analysis Applied to Petroleum Field Development and Management. **Oil & Gas Science and Technology - Rev. IFP Energies nouvelles**, v. 74(46). 2019. https://doi.org/10.2516/ogst/2019019.

SCHIOZER, D. J., SANTOS, S. M. G., SANTOS, A. A. S., FILHO, J. C. H. **Model-Based Life-Cycle Optimization for Field Development and Management Integrated with Production Facilities**. SPE EUROPEC, June, Madrid, Spain. 2022. https://doi.org/10.2118/209681-MS.

SCHLIJPER, A. G. Simulation of Compositional Processes: The Use of Pseudo Components in Equation-Of-State Calculations. **SPE Reservoir Engineering**, v. 1(5), pp. 441–452. 1986. https://doi.org/10.2118/12633-PA.

SHIRANGI, M. G. Closed-Loop Field Development with Multipoint Geostatistics and Statistical Performance Assessment. **Journal of Computational Physics**, v. 390, pp. 249–264. 2019. https://doi.org/10.1016/j.jcp.2019.04.003.

SHIRANGI, M. G., DURLOFSKY, L. J. Closed-Loop Field Development Under Uncertainty by Use of Optimization with Sample Validation. **SPE Journal**, v. 20(5), pp. 908–922. 2015. https://doi.org/10.2118/173219-PA.

SIDAHMED, A., NEJADI, S., NOURI, A. A Workflow for Optimization of Flow Control Devices in SAGD. **Energies**, v. 12. 2019. https://doi.org/10.3390/en12173237.

SILVA, L. O. M. **Seleção de Estratégia de Produção Robusta com o Uso de Modelos Representativos para Campos de Petróleo na Fase de Desenvolvimento**. 2018. 108 p. MSc thesis - University of Campinas, Campinas.

SILVA, L. O. M., SANTOS, A. A. S., SCHIOZER, D. J. **Otimização da Estratégia de Produção sob Incertezas Geológicas e Econômicas**. Rio Oil & Gas Expo and Conference, 24–27 October, Rio de Janeiro, Brazil. 2016.

SILVA, V. L. S., CARDOSO, M. A., OLIVEIRA, D. F. B., MORAES, R. J. Stochastic Optimization Strategies Applied to the OLYMPUS Benchmark. **Computational Geosciences**, v. 24, pp. 1943–1958. 2019. https://doi.org/10.1007/s10596-019-09854-3.

SOREK, N., GILDIN, E., BOUKOUVALA, F., BEYKAL, B., FLOUDAS, C. A. Dimensionality Reduction for Production Optimization Using Polynomial Approximations. **Computational Geosciences**, v. 21, pp. 247–266. 2017. https://doi.org/10.1007/s10596-016-9610-3.

TARIQ, Z., ALJAWAD, M. S., HASAN, A., MURTAZA, M., MOHAMMED, E., EL-HUSSEINY, A., ALARIFI, S. A., MAHMOUD, M., ABDULRAHEEM, A. A Systematic

Review of Data Science and Machine Learning Applications to the Oil and Gas Industry. **Journal of Petroleum Exploration and Production Technology**, v. 11, 4339–4374. 2021. https://doi.org/10.1007/s13202-021-01302-2.

TEIXEIRA, A. F., SECCHI, A. R. Machine Learning Models to Support Reservoir Production Optimization. **IFAC-PapersOnLine**, v. 52(1), pp. 498–501. 2019. https://doi.org/10.1016/j.ifacol.2019.06.111.

TEKLU, T. W., ALAMERI, W., GRAVES, R. M., KAZEMI, H., ALSUMAITI, A. M. Low-Salinity Water-Alternating-CO_2 EOR. **Journal of Petroleum Science and Engineering**, v. 142, pp. 101–118. 2016. https://doi.org/10.1016/j.petrol.2016.01.031.

VAN ESSEN, G., VAN DEN HOF, P., JANSEN, J. D. Hierarchical Long-Term and Short-Term Production Optimization. **SPE Journal**, v. 16(1), pp. 191–199. 2011. https://doi.org/10.2118/124332-PA.

VON HOHENDORFF FILHO, J. C., MASCHIO, C., SCHIOZER, D. J. Production Strategy Optimization Based on Iterative Discrete Latin Hypercube. **Journal of the Brazilian Society of Mechanical Sciences and Engineering**, v. 38, pp. 2473–2480. 2016. https://doi.org/10.1007/s40430-016-0511-0.

VON HOHENDORFF FILHO, J. C., SCHIOZER, D. J. **Integrated Production Strategy Optimization Based on Iterative Discrete Latin Hypercube**. ECMOR XVI - 16th European Conference on the Mathematics of Oil Recover. 2018. https://doi.org/10.3997/2214-4609.201802213.

WANG, H., KOU, Z., JI, Z., WANG, S., LI, Y., JIAO, Z., JOHNSON, M., MCLAUGHLIN, J. F. Investigation of Enhanced CO_2 Storage in Deep Saline Aquifers by WAG and Brine Extraction in the Minnelusa Sandstone, Wyoming. **Energy**, v. 265. 2023. https://doi.org/10.1016/j.energy.2022.126379.

WANG, X., HAYNES, R. D., HE, Y., FENG, Q. Well Control Optimization Using Derivative-Free Algorithms and a Multiscale Approach. **Computers & Chemical Engineering**, v. 123, pp. 12–33. 2019. https://doi.org/10.1016/j.compchemeng.2018.12.004.

WOLPERT, D. H. Stacked Generalization. **Neural Networks**, v. 5(2), pp. 241–259. 1992. https://doi.org/10.1016/S0893-6080(05)80023-1.

YANG, C., NGHIEM, L., CARD, C., BREMEIER, M. **Reservoir Model Uncertainty Quantification Through Computer-Assisted History Matching**. SPE Annual Technical Conference and Exhibition, 11–14 November, Anaheim, California. 2007. https://doi.org/10.2118/109825-MS.

YANTING, Z., LIYUN, X. Research on Risk Management of Petroleum Operations. **Energy Procedia**, v. 5, pp. 2330-2334. 2011. https://doi.org/10.1016/j.egypro.2011.03.400.

YING, X. An Overview of Overfitting and its Solutions. **Journal of Physics: Conference Series**, v. 1168(2). 2019. https://doi.org/10.1088/1742-6596/1168/2/022022.

ZANDVLIET, M. J., HANDELS, M., VAN ESSEN G. M., BROUWER D. R., JANSEN, J. D. Adjoint-Based Well-Placement Optimization Under Production Constraints. **SPE Journal**, v. 13(4), pp. 392–399, 2008. https://doi.org/10.2118/105797-PA.

Appendix A: Explanation and corrections for the paper titled “Investigation of Well Control Parameterization with Reduced Number of Variables Under Reservoir Uncertainties”

Appendix A.1: Criteria for choosing the logistic parametric equation

The logistic equation was chosen for our study based on three key factors:

1. Its intrinsic ability to confine the solution within the BHP limits of the problem.
2. Its capability to produce smooth curves, which is desirable in practical cases.
3. Its ability to fit with the BHP curves of the base case (**Figure A. 1**).

In **Figure A. 1** and **Figure A. 2**, we illustrate the fits for the logistic equation and second-order polynomial equation only for a subset of injection wells (INJ007, INJ010, INJ022, INJ023) and production wells (NA3D, PROD010, PROD023A, PROD025A, RJS19) from the base case. Nevertheless, a similar fit is observed for the remaining wells in the case.

From the figures, it is clear that the logistic equation provides a better fit to the BHP curves than the second-order polynomial, as evidenced by lower mean absolute error (MAE) and higher R-squared values for almost all wells. The NA3D well is the only exception, but the fit is still close compared to the second-order equation. The linear equation fit is not depicted as the second-order equation already encompasses its potential solutions.

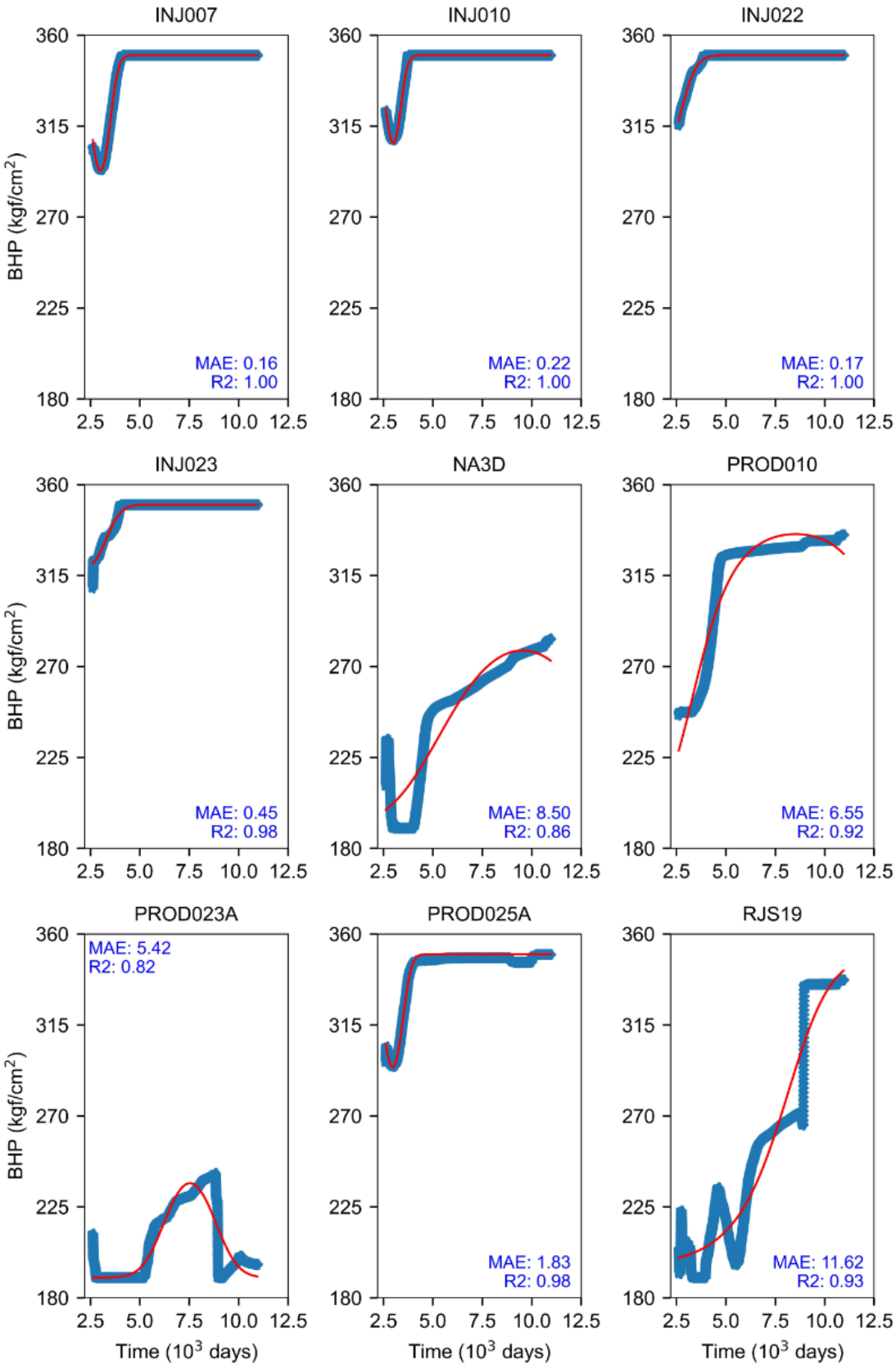

Figure A. 1–Logistic equation fit applied to the BHP curves of the wells in the UNISIM-I-M base case.

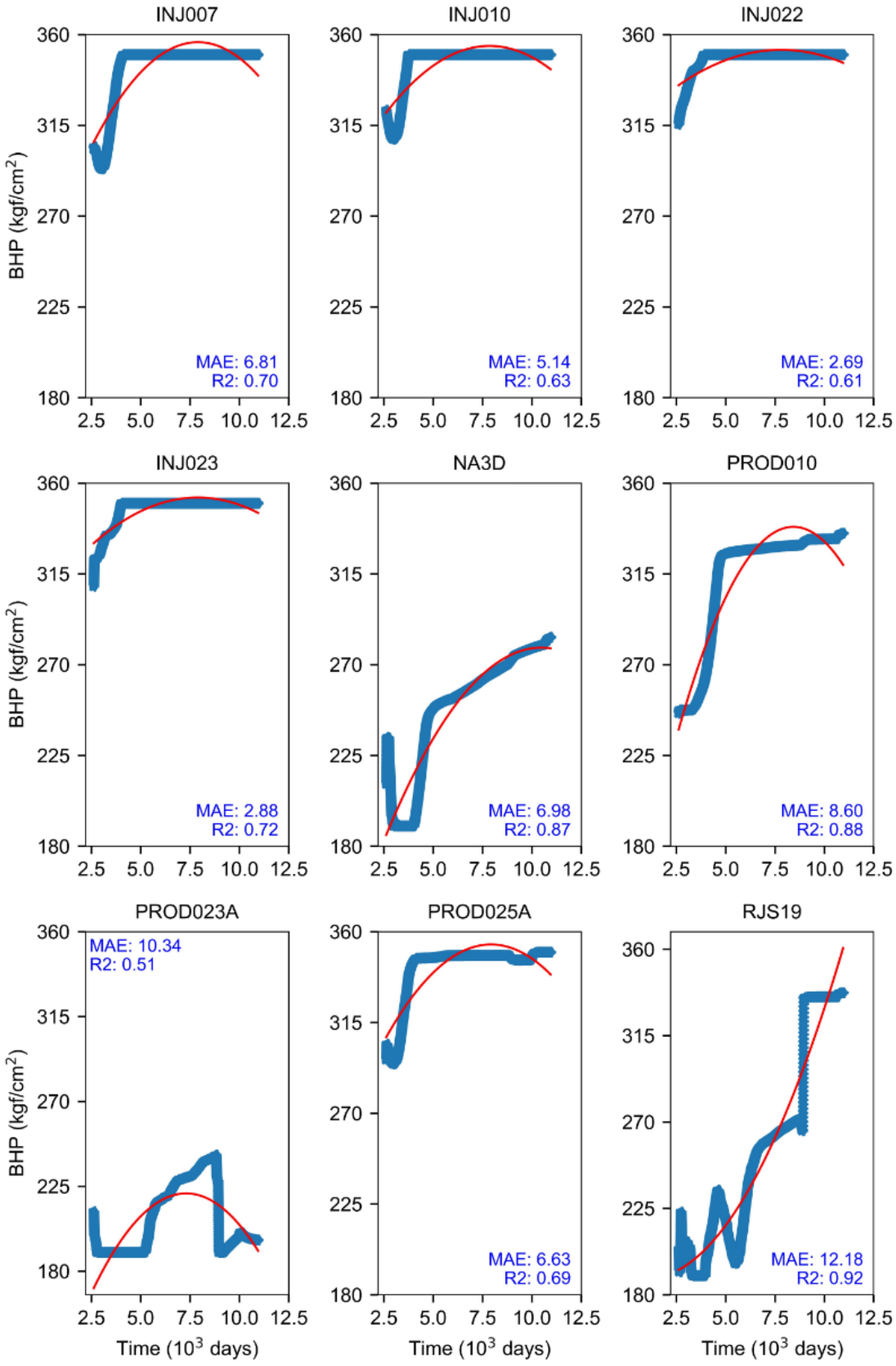

Figure A. 2–Second-order polynomial equation fit applied to the BHP curves of the wells in the UNISIM-I-M base case.

Appendix A.2: Definitions of parametric equation orders to be tested

In Section 2 of our study, we explored first and second-order polynomial equations, alongside the logistic equation with the exponential term raised to a second-order polynomial. We refrained from testing higher-order polynomial equations for two primary reasons: 1) our aim was to reduce the number of optimization variables, mitigating the need for an extensive amount of simulations, and 2) the main contribution of our research centered on evaluating the logistic equation — that was not yet explored in the literature concerning the optimization of well control variables. As we conducted tests where the logistic equation's exponential term was of the second order, we also chose parametric polynomial equations up to the second order to ensure a comparable number of optimization variables.

It is important to mention that we would test higher polynomial orders if the second-order yields a greater Net Present Value (NPV) compared to the first-order. However, the second-order equation resulted in a slightly lower average NPV than its first-order counterpart. While this result may appear counterintuitive, it can be explained by a reduction in the solution search space enabling the optimization algorithm to find better results and avoid local maxima (Awotunde, 2014). This result highlights the optimization algorithm's challenge in converging to optimal solutions in the well control optimization problem, with higher polynomials potentially worsening convergence towards suboptimal solutions.

Appendix A.3: Boundaries for polynomial equations coefficients

In this segment of this Appendix, we provide clarity on how we defined the limits for the coefficients (**Table 2.5)** for both first (**Equation 2.4)** and second-order (**Equation 2.5**) polynomial equations. We set the limits for the a_{we} coefficient in injectors to match the pre-established operating pressure range for the study case (190 and 350kgf/cm^2).

We define the upper limit of a_{we} for producers at 300 kgf/cm^2 because some producers shut down at this value in the base case (**Figure A. 3**). Therefore, our goal was to narrow the search space, avoiding large value increments within the specified limits that may make it difficult to evaluate the intermediate values. Although we opted for this approach, we analyzed whether the majority of well coefficients concentrated at the upper bound after the optimization process. We would reassess the values in case this trend was observed. However, this did not occur, as evidenced in the **Figure A. 4a** and **Figure A. 5a**.

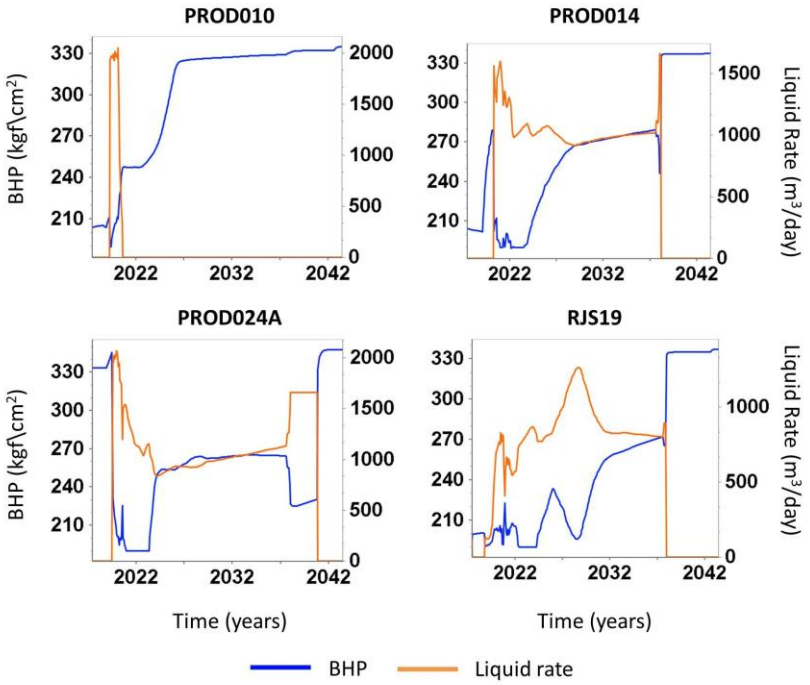

Figure A. 3–Example of producer wells that shut in with BHP below 300 kgf/cm² in the base case.

We set the limits of -40 and 40 kgf/cm² for the b_{we} coefficient in both linear and quadratic equation to reduce significant variations over time in a specific well. The choice of limits for the c_{we} in the second-order equation follows the same rationale as in the logistic equation. The aim is to ensure comparability between the terms within the equation. To achieve this, we match the area under the BHP curve over time for both terms associated with b_{we} and c_{we} coefficients. Therefore, c_{we} is calculated as shown in **Equation A. 1**.

$$c_{we\_limit} = \frac{\int_0^1 (b_{we\_limit} \times t_{norm}) dt_{norm}}{\int_0^1 {t_{norm}}^2 \, dt_{norm}} \qquad \text{Equation A.1}$$

where t_{norm} is the time normalized ($t_{norm} = t/t_{final}$). Similar to the a_{we} coefficient for producers, the limits for b_{we} and c_{we} would be reconsidered if their values reached lower or upper bounds across most of wells. However, this was not observed, as shown in **Figure A. 4b** and **Figure A. 5b** for b_{we}, and in **Figure A. 5c** for c_{we}.

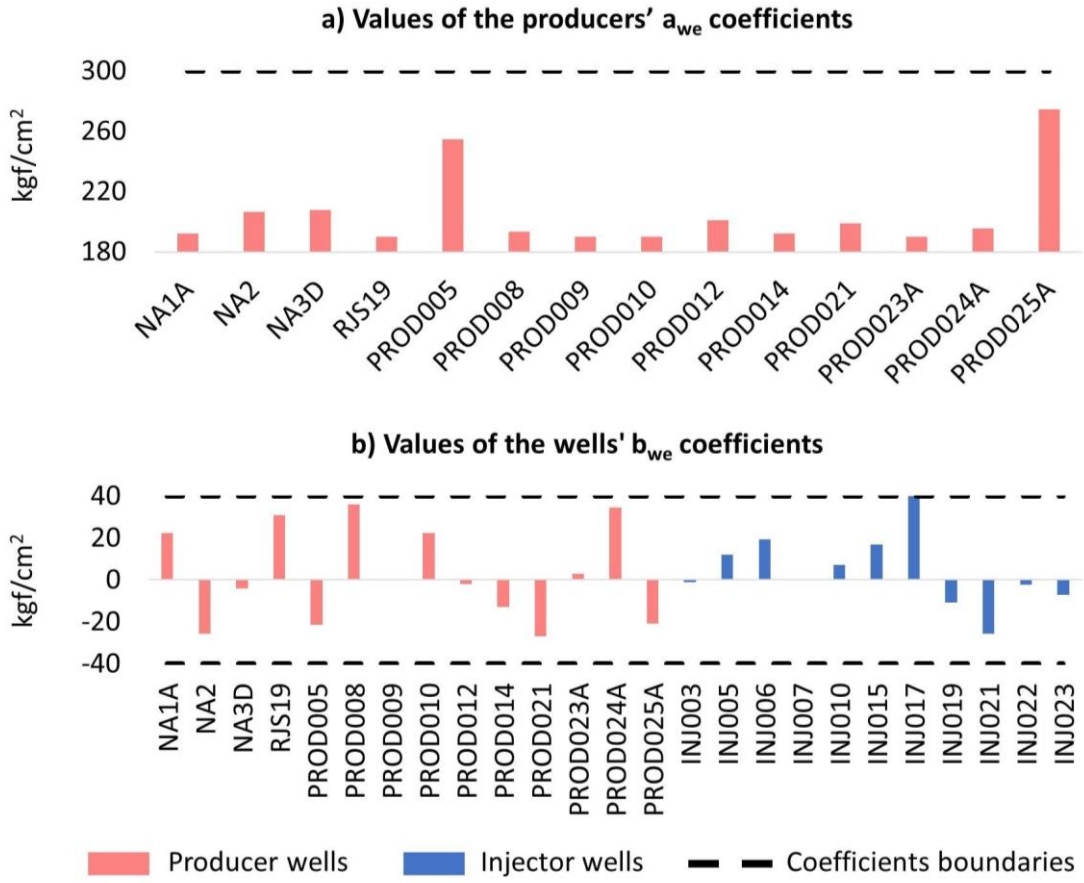

Figure A. 4–Coefficient values for the optimal strategy of the linear equation in the first execution out of the five performed. In (a), optimal a_{we} coefficients for producers are highlighted, while b) shows the optimal b_{we} coefficients for both producer and injector wells.

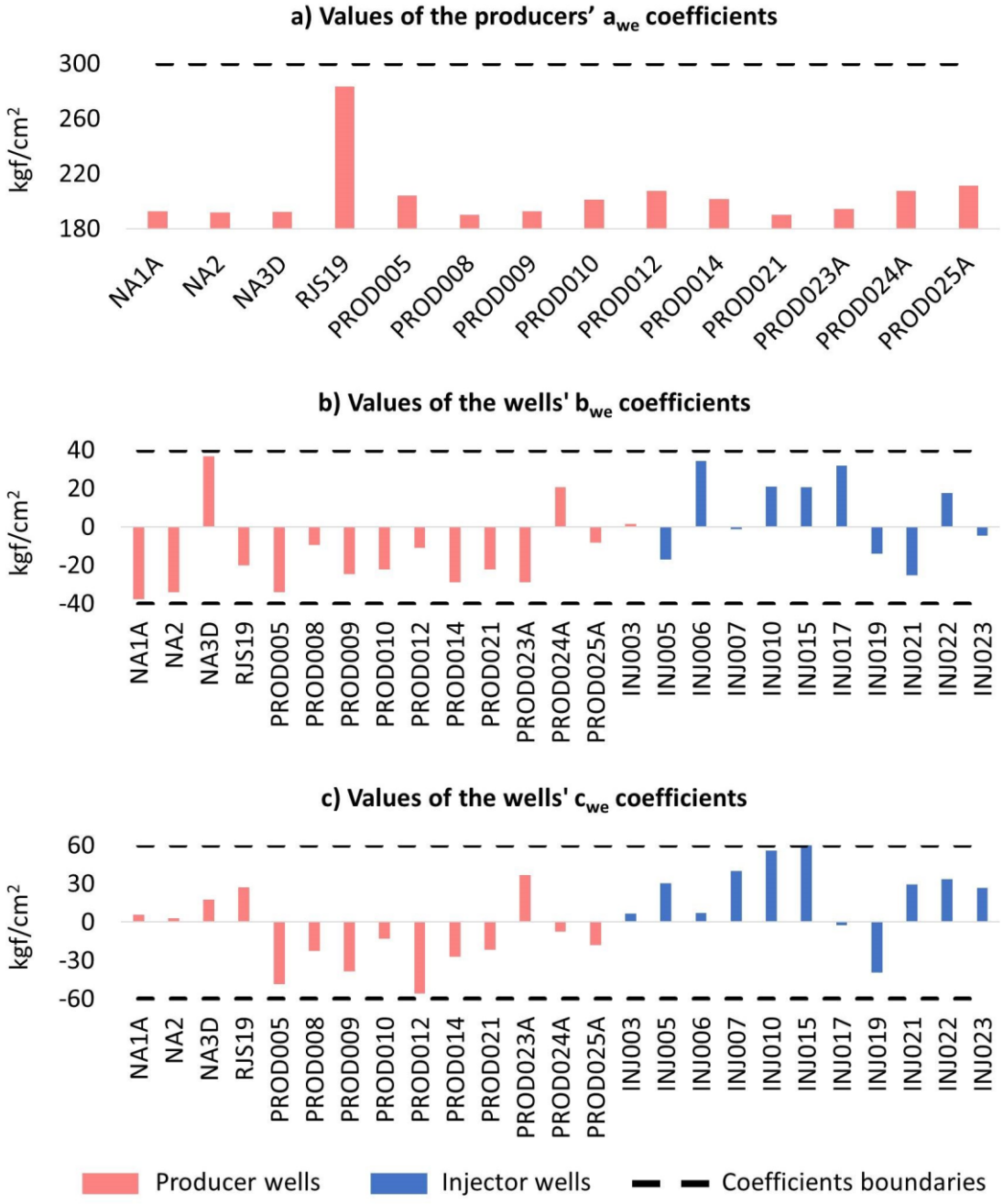

Figure A. 5–Coefficient values for the optimal strategy of the second-order polynomial equation for the first execution out of the five performed. In (a), optimal a_{we} coefficients for producers are highlighted, while the optimal b_{we} and c_{we} coefficients for both producer and injector wells are shown in (b) and (c), respectively.

Appendix A.4: Adjustment in Figure 2.9 to enhance visibility

In **Figure A. 6**, we present a zoomed-in image of **Figure 2.9** to clearly highlight the difference between strategy S5 and the other specialized strategies. The intention is to demonstrate that strategy S5 is relatively robust since it generates similar NPV results across

each RM compared to strategies developed specifically for each of them (e.g., S1 employed in RM1, S2 employed in RM2, etc.). Exceptions occurred in the cases of RM2 and RM4, where S5 yielded NPV values 2.9% and 3.6% lower than those of S2 and S4, respectively.

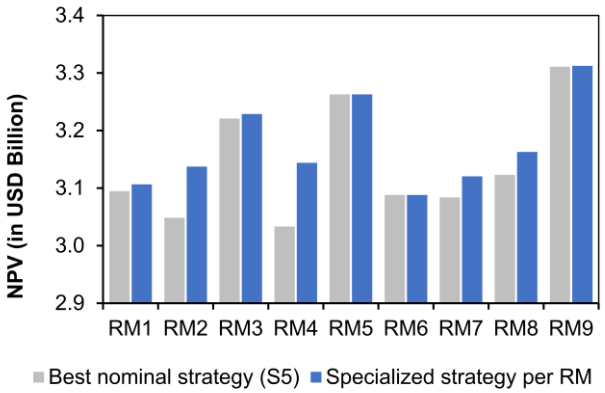

Figure A. 6–Zoomed-in image from Figure 2.9 that allows for a comparison between the most robust specialized strategy (S5) and the specialized strategies applied in the RM for which they were optimized.

Appendix A.5: Correction in cost savings statement for well drilling and perforation in UNISIM-I-R

We previously pointed out that four injectors from **Figure 2.11** did not open in the UNISIM-I-R when we employed the best nominal strategy (S5). Initially, we stated that is unnecessary to consider drilling these wells based on this observation. However, the decision regarding these wells should be based on the results from the 48 filtered scenarios rather than on UNISIM-I-R. This is because UNISIM-I-R represents the actual field, and we are considering that we are planning the strategy at the beginning of the management phase. Therefore, we lack information about whether these wells will be shut in the real case since the strategy has not been applied to it yet.

When employing S5 to the filtered scenarios, it was indeed verified that none of these wells became operational across all 48 models. This outcome can be attributed to the consistently low maximum BHP values imposed by the logistic equation over time, as demonstrated in **Figure A. 7**. These values significantly fall below the maximum BHP constraints set for the case study, which are at 350 kgf/cm^2. As a result, this emphasizes the importance of optimizing the G2L even during the development phase to mitigate unnecessary costs associated with drilling and well completion. It is crucial to note that, operating under the assumption that we are in the management phase, these costs are irreversible.

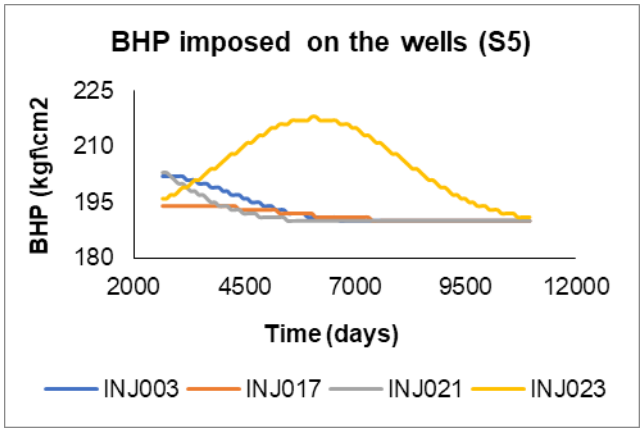

Figure A. 7–Maximum BHP imposed by the logistic equation on injector wells (INJ003, INJ017, INJ021, INJ023). These wells must operate within this BHP limit.

Appendix B: Extra details for the paper titled "A Machine Learning Approach to Reduce the Number of Simulations for Long-term Well Control Optimization"

Appendix B.1: Time for training the multi-layer perceptron stacking technique and example of an expensive study case

In this segment of the Appendix, we offer supplementary information concerning the training duration needed for each of the models adopted in multi-layer perceptron stacking (MLP-S) technique, which yielded the best results among all. We also provide clarification on what constitutes a computationally expensive simulation in production strategy selection problem

The MLP-S is trained using the output of all algorithms from **Table B. 1**, excluding itself. **Table B. 1** shows the average training time across the five executions for IDLHC100–ML50$_{\text{MLP-S}}$. In the final row are both the training time per iteration and the total cumulative time for all iterations.

Table B. 1–Average training time across five executions for IDLHC100–ML50$_{\text{MLP-S}}$.

Algorithms	Time per iteration (minutes)									Total
	1	2	3	4	5	6	7	8	9	
ABC-GBR	1.40	2.19	2.79	3.31	3.99	4.60	5.14	5.75	5.84	35.00
ABC-RF	0.41	0.67	0.90	1.19	1.51	1.84	2.03	2.20	2.40	13.14
BGR-BRR	0.02	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.06	0.32
BGR-GBR	0.14	0.15	0.14	0.16	0.34	0.36	0.21	0.30	0.22	2.02
BGR-RF	0.23	0.34	0.49	0.67	0.81	0.99	1.25	1.29	1.77	7.85
BRR	0.20	0.24	0.31	0.34	1.75	0.91	0.50	0.60	0.65	5.51
ENET	1.05	1.15	1.20	1.57	2.12	1.81	1.58	2.06	2.02	14.56
GBR	0.38	0.58	0.70	1.51	2.13	2.45	2.55	3.62	3.31	17.23
LSR	0.81	0.92	0.87	1.02	1.18	1.27	1.12	1.38	1.35	9.93
RF	0.36	0.57	1.04	1.29	1.90	1.43	2.89	2.40	2.21	14.10
MLP-S	0.08	0.10	0.12	0.12	0.19	0.24	0.22	0.19	0.22	1.47
Total	**5.09**	**6.93**	**8.59**	**11.19**	**15.94**	**15.95**	**17.54**	**19.84**	**20.04**	**121.12**

Table B. 2 illustrates both the total time and the time required for each iteration in the five executions of IDLHC. Additionally, the average time across the five runs is presented. As clarified in the article, we implemented the IDLHC optimization method with 16 simulations running concurrently, each utilizing six processors.

Table B. 2–Total time, time per iteration, and average time for five executions of IDLHC.

IDLHC-100 iteration	Time of each executions (minutes)					Average
	1	2	3	4	5	
1	189	258	243	192	207	218
2	209	266	267	222	182	229
3	206	273	255	168	182	217
4	190	300	199	164	209	212
5	190	330	164	168	297	230
6	215	263	182	167	188	203
7	195	267	177	179	181	200
8	191	232	185	179	187	195
9	200	220	197	212	187	203
10	217	249	196	218	200	216
Total time	**2,001**	**2,659**	**2,064**	**1,869**	**2,019**	**2,123**

As observed, the total training time for the entire MLP-S process is less than 6% when compared to the total simulation time. The simulation time for the UNSIM-I-M model is relatively fast, averaging around 2 minutes. Consequently, the impact on time savings for UNISIM-I-M may not be particularly significant. However, the implementation of Machine Learning would bring substantial benefits in complex reservoirs, reminiscent of pre-salt fields with WAG-CO2 injection, where a considerable computational effort is required.

As illustrated by Schiozer (2022), an example of complex reservoirs requires a runtime of 3 hours. If we consider the same 1000 simulations conducted in Section 3 and utilize 9 representative models to address uncertainties, concurrently running 16 simulations, the well control life cycle (G2L) optimization alone would demand approximately 190 days (slightly over 6 months). In this context, the implementation of machine learning would notably streamline the process, saving approximately 3 months in simulation time. Moreover, the training of ML models, which is independent of the study case, would remain at 2 hours.

Furthermore, Schiozer et al. (2019) recommend incorporating both nominal and robust optimization processes into their 12-step closed-loop field development and management (CLFDM) methodology when sufficient time is available. The nominally and robustly optimized strategies play a crucial role in the step of identifying potential changes in the production strategy to manage uncertainty and improve the chance of success, based on analyses of the value of information and the value of flexibility.

In cases where both robust and nominal optimization are required, the simulation time practically doubles, further emphasizing the significance of our IDLHC-ML methodology.

Appendix B.2: Criteria for choosing the machine learning algorithms in future study cases

This appendix aims to clarify whether retesting all machine learning algorithms from Section 3 is necessary when applying the IDLHC-ML methodology to other case studies. Our recommendation is to directly apply IDLHC-ML using the best-found algorithm, which is the multi-layer perceptron stacking (MLP-S). Machine learning algorithms are primarily influenced by features and targets, showing less dependence on the specific case study. Consequently, implementing the IDLHC-ML methodology with MLP-S in the well control optimization problem using the logistic equation should yield similar results in other scenarios, as long as the same features (bottom-hole pressure over time) and target (economic return) are employed.

Moreover, the IDLHC-ML does not aim to pinpoint the exact economic return value but rather focuses on eliminating suboptimal solutions generated from the logistic equation. Consequently, the machine learning algorithm does not require an exceptionally high level of precision, enhancing the likelihood of MLP-S effectively generalizing to different case studies.

Appendix C: Explanation and corrections for the paper titled “Optimizing Well Control Strategies with IDLHC-MLR: A Machine Learning Approach to Address Geological Uncertainties and Reduce Simulations”

This note serves to rectify an error in **Figure 4.3**. In the initially published version, there was a loop without a decision, making it impossible to exit the loop. We have now addressed this issue by adjusting the position of the arrows within the red box and introducing the parallel process in green (**Figure C. 1**). It's important to note that the parallel process exclusively applies to the IDLHCR-MLR method; the IDLHC method follows the black arrows.

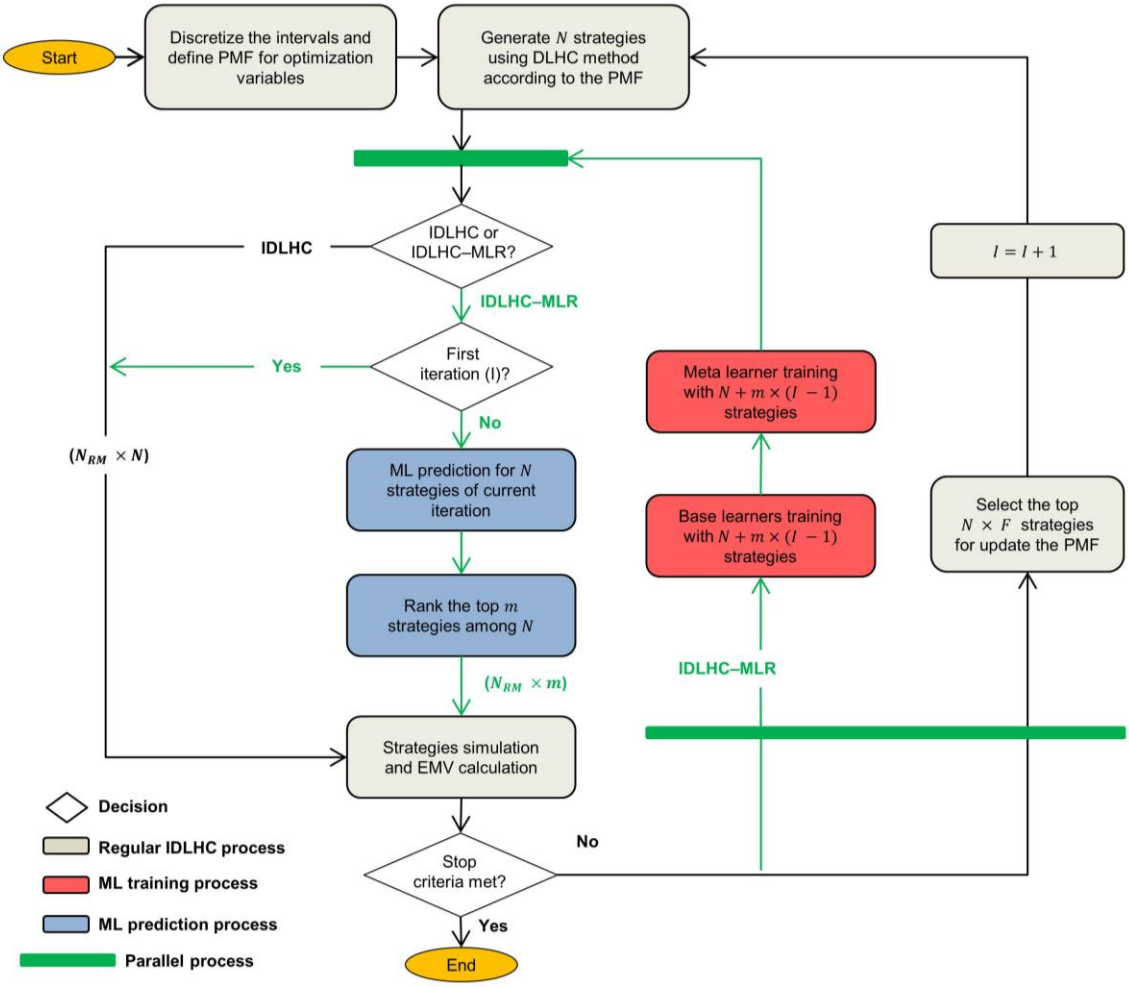

Figure C. 1–Flowchart of IDLHC and IDLHC-MLR methods of optimization (figure adjustment).

Appendix D: Explanation for the paper titled "Accelerated optimization of CO_2-miscible water-alternating-gas injection in carbonate reservoirs using production data-based parameterization"

In this appendix, we provide an additional note to the WAG_{eq} profile showed in **Figure 5.6** of Section 5.4. In the aforementioned section, we noted that some wells in the WAG_{eq} strategy injected a single type of fluid throughout the entire management period. With this context in mind, we can make the following observation: during the development phase, it may not be economically advantageous to retrofit these wells for dual-fluid injection. However, it's essential to emphasize that throughout this thesis, we have been operating under the assumption that we are in the management phase. As a result, the costs associated with equipment installation during the development phase cannot be recovered or reversed.

We also noticed wells consistently injecting a specific fluid for most of the field's lifespan, with sporadic fluid switches occurring within a single time period. Changing the injected fluid type for these wells in a single instance may be operationally inefficient. Within the context of a reservoir management decision analysis framework, the automated WAG strategy is further analyzed to address the sporadic fluid changes issue mentioned and anticipate other problems that may arise with the strategy. For instance, in Step 11 of the oil field development and management framework proposed by Schiozer et al. (2019), the objective is to explore potential adaptations to the production strategy to enhance the chances of success. It's important to note, however, that this level of refinement falls outside the scope of this thesis.

Appendix E: Comparison between the WAG_{eq} and $Well_{gor\_limit}$ strategies in the reference case (SEC1-R)

In this appendix, we show the results obtained by applying the optimal strategy derived from WAG_{eq} to the reference case and compare it with the baseline strategy ($Well_{gor\_limit}$). The aim is to assess the real-world implications of implementing the WAG_{eq} control rules.

The reference case (SEC1-R) serves as our "true-to-life representation" to simulate real-field challenges and difficulties in a controlled environment, allowing us to mimic real-world scenarios. As described by Botechia and Schiozer (2022), this benchmark was developed using a high-resolution geocellular model, leveraging publicly accessible data from a pre-salt carbonate field.

SEC1-R utilizes a corner-point grid comprising 250 × 478 × 640 blocks, totaling 2,728,823 active blocks, each with dimensions measuring 50 × 50 × 2 m. The entire simulation process takes approximately four days on a 16-processor cluster (Botechia and Schiozer, 2022). In **Figure E. 1a** it is possible to observe that the SEC1-R employs a significantly finer grid compared to the SEC1_2022 simulation model (**Figure E. 1b**).

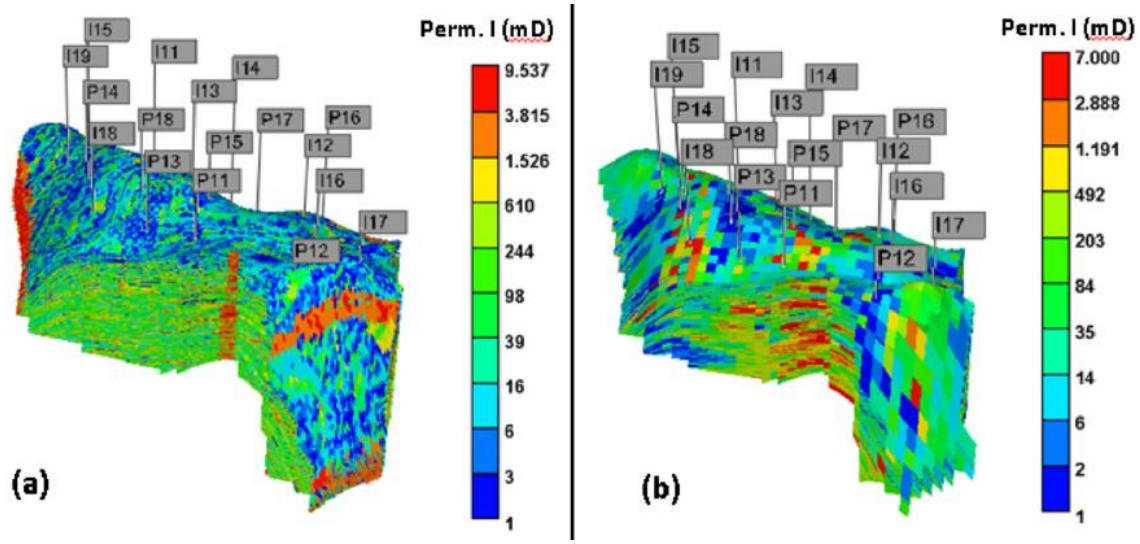

Figure E. 1–Permeability map in 3D view and well distribution. (a) Depicts the configuration for the reference case, and (b) illustrates the coarser simulation model (Botechia and Schiozer 2022).

Table E. 1 provides an overview of the differences in performance indicators for the $Well_{gor\_limit}$ and WAG_{eq} strategies. These evaluations encompass both the reference case and the simulation model. Notably, it becomes evident that the disparities in outcomes between these two strategies vary depending on the model being considered. The percentage increase

in the WAG_{eq} strategy in relation to the $Well_{gor\_limit}$ strategy is lower when both are applied in the reference case than when both are applied in the simulation model for all indicators except for the cumulative gas production (Gp). In terms of Gp, the WAG_{eq} strategy generates the maximum allowable gas quantity for the platform in the simulation model, while it operates with some margin in the reference case.

The smaller increase in NPV in the reference case was somewhat expected for two reasons: first, we conducted optimization in a nominal scenario, which may not adequately capture the uncertainties of the field. Utilizing a wide range of uncertain scenarios would help identify a more robust strategy with a higher likelihood of outperforming in the reference case. The second reason is that when we optimize for the coarse model, we tend to develop a strategy that is optimistically biased towards it. Consequently, this strategy does not perform as well in the reference case, which significantly differs from the simulation model and features a much finer grid. However, it's important to highlight that the direction of the gain is consistent in both scenarios, meaning that an improvement in the simulation model also leads to an improvement in the field's NPV, albeit to a lesser degree. These results emphasize that we can achieve better field management using an optimal WAG strategy than the traditional approach of a fixed six-month WAG cycle, as commonly observed in the industry.

Table E. 1– Comparative analysis of performance indicators between $Well_{gor\_limit}$ and WAG_{eq} strategies.

Case	Strategy	NPV (10^9 USD)	Np (10^6 m^3)	Gp (10^9 m^3)	Wp (10^6 m^3)	Winj (10^6 m^3)
Reference case (SEC1-R)	WAG_{eq}	8.06	164	113	63	210
	$Well_{gor\_limit}$	7.77	156	120	43	182
	Relative difference (%)	3.7	5.2	-6.2	47	15
Simulation model (SEC1_2022)	WAG_{eq}	8.18	166	120	83	232
	$Well_{gor\_limit}$	7.68	149	120	30	162
	Relative difference (%)	6.7	11.4	0	172	43