

Universidade Estadual de Campinas
Instituto de Computação

Jeanfranco David Farfan Escobedo

Model Uncertainty Quantification to Improve Oil Reservoir Simulations

Quantificação da Incerteza do Modelo para Melhorar Simulações de Reservatórios de Petróleo

CAMPINAS
2025

Jeanfranco David Farfan Escobedo

Model Uncertainty Quantification to Improve Oil Reservoir Simulations

Quantificação da Incerteza do Modelo para Melhorar Simulações de Reservatórios de Petróleo

Tese apresentada ao Instituto de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Ciência da Computação.

Thesis presented to the Institute of Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Computer Science.

Supervisor/Orientador: Prof. Dr. Anderson de Rezende Rocha
Co-supervisor/Coorientador: Prof. Dr. Gabriel Cirac Mendes Souza

Este exemplar corresponde à versão final da Tese defendida pelo aluno Jeanfranco David Farfan Escobedo e orientada pelo Prof. Dr. Anderson de Rezende Rocha.

CAMPINAS
2025

Ficha catalográfica
Universidade Estadual de Campinas (UNICAMP)
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Silvania Renata de Jesus Ribeiro - CRB 8/6592

Farfan Escobedo, Jeanfranco David, 1994-
F224m Model uncertainty quantification to improve oil reservoir simulations / Jeanfranco David Farfan Escobedo. – Campinas, SP : [s.n.], 2025.

Orientador: Anderson de Rezende Rocha.
Coorientador: Gabriel Cirac Mendes Souza.
Tese (doutorado) – Universidade Estadual de Campinas (UNICAMP), Instituto de Computação.

1. Aprendizado de máquina. 2. Aprendizado profundo. 3. Reservatórios de petróleo. 4. Reservatórios (Simulação). I. Rocha, Anderson de Rezende, 1980-. II. Souza, Gabriel Cirac Mendes. III. Universidade Estadual de Campinas (UNICAMP). Instituto de Computação. IV. Título.

Informações complementares

Título em outro idioma: Quantificação da incerteza do modelo para melhorar simulações de reservatórios de petróleo
Palavras-chave em inglês:
Machine learning
Deep learning
Petroleum reservoirs
Reservoirs (Simulation)
Área de concentração:
Titulação: Doutor em Ciência da Computação
Banca examinadora:
Anderson de Rezende Rocha [Orientador]
Karla Tereza Figueiredo Leite
George Luiz Medeiros Teodoro
Marcelo da Silva Reis
Celmar Guimaraes da Silva
Data de defesa: 24-11-2025
Programa de Pós-Graduação: Ciência da Computação

Objetivos de Desenvolvimento Sustentável (ODS)
Não se aplica

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0001-6733-6615
- Currículo Lattes do autor: http://lattes.cnpq.br/6242935985641657

Anderson de Rezende Rocha

Karla Tereza Figueiredo Leite

George Luiz Medeiros Teodoro

Marcelo da Silva Reis

Celmar Guimaraes da Silva

A ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Acknowledgments

I extend my deepest gratitude to my parents, Senovia and David, my eternal heroes, whose unwavering support, love, and lessons in perseverance made it possible for me to pursue my studies in Brazil. Their belief in me laid the foundation for every dream I've dared to chase. I am deeply grateful to my brother, Werner, for taking on the responsibility of caring for our parents during my absence.

This dissertation is dedicated to my beloved wife, Samara, and my dear son, Mateo. You became my constant inspiration and the steadfast reason to persist through every challenge and moment of doubt during the last part of my doctoral journey. Your encouragement was invaluable.

My heartfelt thanks go to my extended family—Delfina, Doris, Yuliana, Flor, Maira, Zoe, Marcelo, Aidan, and Carla—for their unwavering moral and emotional support and to all other relatives for their good wishes. This work also serves as a dedication in eternal memory of my beloved cousin, Rodrigo; my dear aunt, Georgina; my nephew, Liam; and my cherished grandparents, Carlota, Victor, Basilio, and Ricardina, as well as all other relatives and acquaintances who have passed on to a better life.

I am especially grateful to my Peruvian friends, including Rodolfo, John, José Luis, Jesús, Elian, Viviane, Manuel, Yeison, Maumau, and Alex who provided a comforting sense of home while I was away from my beloved Cusco, Peru. Their friendship has been a true blessing.

My sincere appreciation goes to my advisors, Anderson Rocha and Gabriel Cirac. Their exceptional guidance, profound problem-solving skills, and willingness to always listen and advise were instrumental to my growth throughout my PhD. I hold them in immense admiration. I also thank Professors Sandra Avila, Claudia Medeiros, Ricardo Torres, and Tomasz Kowaltowski for inspiring new perspectives on learning and teaching.

Finally, I am indebted to Shell Brasil for their invaluable financial support across two projects during my PhD. Special thanks are extended to Guilherme Avansi, Célio Maschio, Denis José Schiozer, and the entire staff of the Computing Institute, including Denise and Wilson, whose efforts consistently facilitated a wonderful and supportive work environment. To everyone who has positively impacted my life over the past four years, thank you for making this an incredible journey.

Resumo

Modelos proxy são representações simplificadas de modelos numéricos complexos usados para aumentar a eficiência computacional de um processo. O desenvolvimento de proxies é desafiador devido à alta dimensionalidade inerente aos espaços que eles modelam. Esta pesquisa adota uma abordagem ampla e holística para enfrentar esses desafios. À medida que o estudo dos modelos proxy evolui para uma compreensão mais profunda das características complexas dos reservatórios, nosso trabalho segue essa tendência. Começamos propondo um método supervisionado para aproximar a resposta do simulador usando uma única Rede Neural Artificial (RNA). Essa rede é projetada para prever o comportamento de múltiplas curvas de produção cumulativa simultaneamente, divergindo da abordagem convencional na literatura, que geralmente depende do treinamento de múltiplas RNAs para cada curva de produção cumulativa. O próximo passo envolve a reconstrução das incertezas geológicas originais. Em engenharia de reservatórios, retornar ao espaço original após o processo de Data Assimilation for Uncertainty Reduction (DAUR) é essencial para validar a capacidade do proxy de recuperar informações enquanto reduz eficazmente a dimensionalidade da entrada. Finalmente, abordamos o desafio da quantificação de incerteza de modelos proxy, um problema menos explorado na literatura, utilizando conjunto de Funções de Ativação Meta (FAM) para avaliar a confiança nas previsões por meio da análise de incerteza das saídas do proxy. Avaliamos os três métodos propostos usando um conjunto de dados de um campo offshore brasileiro, demonstrando a eficácia e a praticidade de nossa abordagem.

Abstract

Proxy models in reservoir engineering are simplified representations of complex numerical models used to increase the computational efficiency of a process. Developing proxies in this context is challenging due to the high dimensionality inherent in the spaces they model, and the abundance of producing wells, each associated with different objective functions, further complicates the problem. This research takes a broad and holistic approach to address these challenges. As the study of proxy models has evolved towards a deeper understanding of the complex characteristics of reservoirs, our work aligns with this trend. We begin by proposing a supervised method to integrate dimensional reduction and approximate the simulator's response using a single Artificial Neural Network (ANN). This network is designed to predict the behavior of multiple cumulative production curves simultaneously, diverging from the conventional approach in the literature, which often relies on training multiple ANNs for each cumulative production curve. The next step involves reconstructing the original geological uncertainties. In reservoir engineering, returning to the original geological space after the Data Assimilation for Uncertainty Reduction (DAUR) process is essential for validating the proxy's capability to recover information while effectively reducing the dimensionality of the input. Finally, we address the challenge of proxy model uncertainty quantification, a less-explored problem in the literature by using Meta Activation Functions (MAF) Ensemble to assess prediction confidence through uncertainty analysis of the proxy outputs. We evaluate the three proposed methods using a Brazilian offshore field dataset, demonstrating the effectiveness and practicality of our approach.

List of Figures

2.1 Architecture proposed. 26
2.2 The processing details for conducting dimensionality reduction of Porosity (POR) are illustrated in (a). it specifically involves 3D block encoders, 3D max-pooling, and a 3D residual block. This entire process is repeated for every geological uncertainty. A more in-depth explanation of the encoder block is provided in (b). Likewise, for a comprehensive explanation of the residual block, refer to (c). 27
2.3 The end-to-end solution employs the dimensionally-reduced properties to conduct regression via the utilization of Dense layers. 31
2.4 Illustrates our methodology, building on the approach described by Maschio *et al.* [1]. In a), Well P10 is presented with a set of cumulative curves spanning 2206 days. In b), we select a single curve from the set. In c), we identify five equidistant points along this curve and use them to replace the entire curve. As shown in d), this process is repeated for each cumulative curve within Well P10. We underscore that this procedure is applied not only to Well P10 but to all other wells and each respective objective function. 33
2.5 Cumulative production curves from our end-to-end 3D CNN (in green) closely match the simulator's outputs (in gray) across selected days. The analysis spans various wells and objective functions; oil for Wells p10 and p8, gas for Wells p6 and p2, pressure for Well p10, and water for Well p5, demonstrating the methodology's accuracy in replicating real-world production data. 39
2.6 Several outcomes stemming from our end-to-end 3D CNN application on the simulator are showcased for all models within the P2 producer in the test set. 40
2.7 For each image, the correlation representation for each objective function encompasses the entirety of the producers, specifically P10, P2, P3, P5, P6, P8, and P9. The illustration includes correlation metrics R with R^2 and SMAPE metric. 40
2.8 Sensitivity Analysis: The upper section delineates the first-order interactions, while the lower showcases the higher-order interactions. 42

3.1 The 3D DRVAE reconstruction module showcasing the interconnected components of the encoder, decoder, and residual connections. This framework facilitates the reconstruction of 3D geological uncertainties, validating the robustness of the dimensionality reduction technique and enabling the visualization of uncertainty reduction within the reservoir. 47

3.2 Diagram illustrating the 3D Encoder architecture. Figure (a) shows the overall structure of the encoder, including layers of encoder blocks and 3D pooling, leading to a dense layer. Figure (b) provides a detailed breakdown of the encoder block, highlighting its hierarchical processing through convolutional layers, batch normalization, ReLU activation, and dropout. These components work together to handle complex geological data, preserve high-frequency details, and efficiently reduce dimensionality. 48
3.3 The diagram depicts the 3D Decoder architecture. Figure (a) shows the details of the decoder, including dense layers, decoder blocks, 3D upsampling layers, and 3D transpose convolutions. Figure (b) provides a detailed breakdown of the decoder block, highlighting its hierarchical processing through transpose convolutional layers, batch normalization, and ReLU activation, while emphasizing skip connections from the encoder block linked to the outputs of ReLU1 and ReLU2. The decoder architecture reconstructs detailed geological features from the latent sample through a series of sequential operations. 50
3.4 Shortcut connections between the encoder and decoder blocks. Specifically, the first encoder block is connected to the last decoder block, the second encoder block is connected to the penultimate decoder block, and the third encoder block is connected to the antepenultimate decoder block. Finally, the last encoder block is connected to the fourth-to-last decoder block, ultimately reaching the latent space. These connections aim to mitigate the issue of vanishing and exploding gradients. 53
3.5 Detailed view of the residual connection. From left to right, the encoder block is shown in yellow, and the decoder block is shown in blue. The interaction between these two blocks is facilitated by two skip connections from the encoder to the decoder. First, the output of the first convolutional layer in the encoder (CONV1 3D) is connected to the output of the activation function in the decoder (RELU3). Second, the output of the second convolutional layer in the encoder (CONV2 3D) is connected to the output of the second activation function in the decoder (RELU4). 54
3.6 Visual analysis of the PCA technique proposed in Maschio et al.'s work. Utilizing 140 principal components, Figure 3.6 presents findings across four uncertainties. Each row comprises four images, with pairs corresponding to specific uncertainties. The first row depicts layer 12 (porosity) and layer 13 (net-to-gross uncertainty), while the second row represents layer 14 (type of rock uncertainty) and layer 14 (permeability uncertainty). Original uncertainties are displayed on the left, while reconstructions are on the right. 56
3.7 Ablation study on 3D geological reconstruction, assessing component impact. Panel a) shows the original 3D porosity uncertainty, b) VAE approach, c) DRVAE architecture, and d) DRVAE with grid search on L2 regularization. 58
3.8 Reconstructions of Geological Uncertainties Focusing on Layer 12. 60
3.9 Reconstructions of Geological Uncertainties Focusing on Layer 14. 61
3.10 Reconstruction Analysis of KDE Distributions for Porosity, Rock Type, Net-to-Gross Ratio, and Permeability in Layers 12 and 14. 62
3.11 Geological Property Histogram Comparison. 63

4.1 Results on the He dataset. (a) Ground-truth function (red dashed) and training data (blue). (b–c) ERN and UR-ERN underestimate uncertainty in OOD regions. (d–e) AE and RAF better capture epistemic uncertainty. 74
4.2 Results on the Forrester dataset. (a) Ground-truth function (red dashed) and training data (blue) restricted to subintervals. (b–c) ERN and UR-ERN are competitive, capturing epistemic uncertainty in $[0.85, 1]$ but underestimating it in $[0, 0.2]$. (d–e) AE and RAF underestimate uncertainty in unseen regions. 75
4.3 Overview of cumulative production profiles for seven producing wells across a 2,206-day period. (a) Cumulative oil production for well P10, with each curve representing a different simulation model, showing the variation in production over time. (b) Selection of five key time instances (t_0, t_1, t_2, t_3, t_4) for well P10, representing critical production milestones across the simulation models. (c) Production trends for each of the seven producing wells (P10, P2, P3, P5, P6, P8, P9), with well production data separated by spaces to distinguish each well's trajectory. 80
4.4 Overview of the modified end-to-end proxy architecture for geological uncertainty. a) The dimensionality reduction module processes 3D geological data using 3D block encoders, max pooling, and residual blocks. b) The encoder block, replacing ReLU with a meta-network, extracts rich feature representations. c) The residual block enhances gradient flow through shortcut connections and meta-network activations. d) The neural meta-network integrates DenseNet and Attention mechanisms, utilizing dense layers, Batch Normalization, and Dropout, with the Attention mechanism refining feature representation before the final softmax-based prediction layer. 81
4.5 Oil production uncertainty quantification for the seven producing wells (P10, P2, P3, P5, P6, P8, and P9) 83
4.6 Water production uncertainty quantification for the seven producing wells (P10, P2, P3, P5, P6, P8, and P9) 84
4.7 Gas production uncertainty quantification for the seven producing wells (P10, P2, P3, P5, P6, P8, and P9) 85
4.8 Pressure production uncertainty quantification for the seven producing wells (P10, P2, P3, P5, P6, P8, and P9) 85

A.1 3D porosity distribution in the S-field reservoir, highlighting the locations of production and injection wells, as generated by Maschio *et al.* [1]. 103
A.2 The continuous green line represents the cumulative oil production (Np), while the blue line represents the cumulative water production (Wp) over time, based on the simulation model outputs, as generated by Maschio *et al.* [1]. 104

B.1 Here, we showcase an automated uncertainty reduction pipeline proposed in the work of Maschio *et al.* [1]. The process involves the initial analysis of geological attributes (referred to as *'prior'*) and their associated multiple objective functions. Subsequently, a mathematical relationship between geological uncertainties and objective functions is established using regression techniques, with the specific implementation of a proxy. An analysis entails the evaluation of proxy estimations against a numerical simulator. If the current iteration (*it*) does not match the final iteration (it_max), DA is performed iteratively until it_max is reached. This iterative process updates each value of geological uncertainty. In parallel, a geological reconstruction is conducted concurrently to preserve geological information. Lastly, in the final iteration (it_max), a comparison is made between the uncertainty reduction achieved by the proxy and that generated by the simulator. 106
B.2 Here, we present our approach, which is a modification of the contemporary uncertainty reduction pipeline proposed by Maschio *et al.* [1]. Our innovation introduces an end-to-end deep learning framework based on proxies. 107

List of Tables

2.1 3D CNN for dimensionality reduction of a single geological uncertainty (e.g., Porosity). The network compresses a 72×48×24×1 volume into a **256D** vector while preserving key spatial patterns. 28
2.2 Comparative evaluation of our end-to-end 3D CNN with state-of-the-art methods focusing on oil and water-related objective functions. 35
2.3 Comparative evaluation of our end-to-end 3D CNN with state-of-the-art methods focusing on gas and pressure-related objective functions. 36
2.4 Comparative evaluation of the time performance of our end-to-end 3D CNN compared to state-of-the-art methods. This evaluation focuses on computing time for tasks such as uploading files, normalization, training, and predicting the replicated simulator responses across various combinations of training samples: 50, 100, and 140. 37

3.1 3D CNN encoder architecture details. 49
3.2 3D CNN decoder architecture details. 52
3.3 Performance Evaluation of the DRVAE Method for 3D Geological Uncertainties. 59

4.1 Predictive performance of different methods using NLL and 95% confidence intervals for 19 datasets. The best scores are highlighted in bold. Whenever confidence intervals overlap, multiple methods are considered statistically indistinguishable and are jointly bolded. 76
4.2 Predictive performance of different methods using SMAPE and 95% CI for 19 datasets. The best SMAPE values are highlighted in bold, allowing ties when confidence intervals overlap. 77
4.3 Comparative evaluation of our approach with state-of-the-art method focusing on oil and water-related objective functions. 82
4.4 Comparative evaluation of our approach with state-of-the-art method focusing on gas and pressure-related objective functions. 83
4.5 Predictive performance of different objective functions using NLL and 95% confidence intervals. 86

A.1 Input data: Scalar properties of the S-Field reservoir. Scalar attributes are modeled using a uniform prior distribution. 103

Contents

1 Introduction **15**
- 1.1 Motivation and Problem Overview . . . 17
- 1.2 Challenges . . . 17
- 1.3 Objective . . . 18
- 1.4 Research Questions . . . 18
- 1.5 Main Contributions . . . 19
- 1.6 Text Organization . . . 20

2 End-to-end dimensionality reduction and regression from 3D geological uncertainties to estimate oil reservoir simulations **21**
- 2.1 Introduction . . . 21
- 2.2 Related Work . . . 23
  - 2.2.1 Dimensionality Reduction . . . 23
  - 2.2.2 Replication of the Simulator's Response . . . 25
- 2.3 Proposed Method . . . 25
  - 2.3.1 Preprocessing . . . 25
  - 2.3.2 End-to-end Architecture . . . 26
- 2.4 Output Representation . . . 31
- 2.5 Results and Discussions . . . 32
  - 2.5.1 Sensitivity Analysis . . . 41
- 2.6 Final Remarks . . . 42

3 A Multimodal Approach to Reconstructing Geological Uncertainties Using a Deep Residual Variational Autoencoder **43**
- 3.1 Introduction . . . 43
- 3.2 Related Work . . . 45
- 3.3 Proposed Methodology . . . 46
  - 3.3.1 Encoder . . . 46
  - 3.3.2 Decoder . . . 49
  - 3.3.3 Residual Connection . . . 53
- 3.4 Results and Discussion . . . 54
  - 3.4.1 Loss Function and Metrics . . . 54
  - 3.4.2 Implementation and Training . . . 55
  - 3.4.3 Baseline . . . 55
  - 3.4.4 Ablation Study for 3D Geological Reconstruction . . . 57
  - 3.4.5 Comparing the Baseline with Our 3D Deep Residual Variational Autoencoder . . . 58
  - 3.4.6 Analysis and Visualization Approach . . . 59

3.4.7 KDE and Histogram Analysis . . . 60
3.5 Final Remarks . . . 63

4 Meta Activation Functions for Uncertainty Quantification in Reservoir Modeling Proxies 64
4.1 Introduction . . . 64
4.2 Related Work . . . 66
4.3 Preliminaries . . . 68
4.4 Proposed Method . . . 69
4.5 Experiments . . . 71
4.5.1 Baselines . . . 72
4.5.2 Synthetic and Real-world Datasets . . . 72
4.5.3 Evaluation Metrics . . . 73
4.6 Results and Discussions . . . 73
4.7 Case Study: Application to a Geological Reservoir . . . 78
4.7.1 End-to-End Proxy: Uncertainty Quantification . . . 82
4.8 Final Remarks . . . 86

5 Conclusions and Future Work 87
5.1 Revisiting the Research Questions . . . 88
5.2 Future Works . . . 89

Bibliography 91

Appendix A Brazilian Offshore Field Dataset 102
A.1 S-Field: Inputs and Outputs . . . 102

Appendix B Our Contribution to the Literature 105

Chapter 1

Introduction

Petroleum reservoirs are critical for supplying global energy demands. A comprehensive understanding of the intricate characteristics of these reservoirs is essential for effective extraction, reservoir management, and the advancement of sustainable energy practices [2]. By advancing our knowledge in this field, we not only ensure the efficient use of resources but also reduce environmental impacts, contribute to energy security, and foster economic growth. Moreover, enhanced reservoir management practices are crucial for transitioning to more sustainable energy systems.

A petroleum reservoir comprises hydrocarbon accumulations within porous or fractured rock formations, often located thousands of feet beneath the surface and concealed beneath multiple overlying geological layers [3]. Over the decades, significant efforts have been made to develop advanced technologies that unravel the intricacies of subterranean oil reservoirs [4, 5, 6]. However, the available sources of information remain insufficient for achieving precise measurements of all parameters for effective reservoir management. Nevertheless, there are methods to understand reservoir dynamics. Among the technologies employed in reservoir management, numerical reservoir simulation (commonly referred to as a "simulation model") plays a pivotal role. As highlighted by Thomas [7] and Ertekin *et al.* [8], simulation models are computer systems comprising data sets that aim to estimate reservoir rock and fluid properties, as well as production strategy variables, including oil production rate (Q_O) and water production rate (Q_W).

Specifically, simulation models are theoretical approximations produced by the simulator (or numerical simulator) and are inherently uncertain. Each simulation execution is referred to as a realization. These simulation models are typically high-dimensional, heterogeneous, and characterized by strong nonlinearities [9]. Despite considerable advances in numerical simulation technologies, simulation models remain a computationally intensive task, often requiring days or even weeks to generate a single reservoir simulation [1]. Some of the main reasons why numerical reservoir simulation faces computational bottlenecks and remains a challenging problem are as follows:

i. The computational time required for a single simulation model increases significantly with the complexity and heterogeneity of the reservoir model.

ii. Due to the high computational cost associated with each simulation model, only a limited number of realizations are typically available in practical reservoir studies

[1, 10, 11].

iii. Reservoir simulation models are inherently high-dimensional, involving numerous interdependent parameters and state variables. This complexity is associated with rock properties, such as porosity or permeability, which hinder the identification of relevant patterns and relationships.

Hence, expediting this process using a limited number of samples and addressing the challenge of high dimensionality are crucial steps to enhance its practicality. One potential approach to this challenge is using proxies—also known as "surrogate models," "proxy models," or "metamodels" [12]—which replace the numerical simulator in time-consuming steps.

This work investigates the application of proxy methods in various scenarios involving numerical reservoir simulation, using a dataset from a Brazilian offshore field [1] (see the details in the Appendix A). First, we propose a proxy model to reduce the dimensionality of rock properties and to approximate the simulator's response under different production strategies, such as cumulative production curves. Second, after dimensionality reduction, some original information from the geological models is inevitably lost. This loss of information complicates the validation of the proxy's ability to retain fundamental geological features. Therefore, a reconstruction process is necessary to project the data back into the original geological uncertainty space, ensuring analytical consistency and compatibility with the simulator's input requirements.

The reconstruction process is necessary for two primary reasons: (i) current data assimilation (DA) approaches [1, 2, 13] typically perform uncertainty reduction on a latent space derived from dimensionality-reduced geological models and subsequently use a proxy model to reconstruct the original geological data without uncertainty after DA; and (ii) the proxy model may exhibit generative behavior, allowing for the creation of new realizations, which could reduce dependence on the numerical simulator. The DA technique integrates diverse sources of data, including those obtained from wells or sensors, historical records, and computational models. It combines observations and simulations to iteratively refine the reservoir model and reduce uncertainty in spatial properties like permeability, porosity, and pressure. It is important to note that DA procedures are outside the scope of this thesis.

Finally, we address a frequently overlooked aspect of proxy model development: the prediction uncertainty quantification. This step is crucial for assessing the reliability and confidence of proxy model outputs. Current proxy implementations still face significant limitations, particularly their inability to guarantee the robustness of predictions. This limitation stems from a persistent gap in the literature regarding the Uncertainty Quantification (UQ) of proxies within the context of reservoir management. As emphasized by Volodina and Challenor [14], transparent reporting and communication of uncertainties are fundamental to ensuring both the credibility of proxy models and the trust placed in their results.

1.1 Motivation and Problem Overview

The characterization of numerical reservoir simulations using proxy models is required, underscored by two primary factors. Firstly, early-stage simulation models often exhibit high intrinsic uncertainty due to numerical approximations, resulting in unreliable reservoir behavior and performance forecasts [15]. DA [16], a fundamental methodology in geosciences [17, 18, 19] and particularly relevant in reservoir engineering, addresses this by integrating observational data with model simulations to estimate the reservoir's current state [16]. However, applying DA to reduce uncertainty in three-dimensional properties is computationally intensive, with single applications spanning days or weeks depending on model complexity [1]. This high computational demand emphasizes the need to optimize processes by replacing the reservoir simulator in the DA workflow. Proxy models offer a solution, providing computationally efficient approximations that mimic simulator input-output behavior, thereby substantially reducing the computational burden [20, 21, 22].

Secondly, while DA mitigates simulation model uncertainty, quantifying the inherent uncertainty in proxy models is equally vital, especially in reservoir engineering, where decisions have cascading effects. Proxy models are susceptible to various uncertainties, including observational noise, approximation errors, and limitations in their architectural generalization capacity. Consequently, assessing the reliability and robustness of their predictions is necessary before their integration into decision-making workflows. Our motivation stems from the dual need to accelerate complex simulations while rigorously quantifying the confidence in their predictive power.

1.2 Challenges

Effective reservoir management is complicated by the inherent heterogeneity and dynamic complexity of oil reservoirs [23]. This gap necessitates the construction of reliable simulation models and the accurate characterization of reservoir properties, tasks that are made challenging by limited sample availability and the need for practical efficiency.

Current methodologies for proxy model construction typically segment the problem into three independent stages: dimensionality reduction, replication of the simulator's response, and reconstruction [1, 4, 20, 24]. Each stage presents distinct challenges:

i. **Dimensionality Reduction:** Simulation models are inherently high-dimensional, involving numerous interdependent parameters that mask meaningful patterns and relationships. A key difficulty in applying dimensionality reduction to 3D geological models is the potential loss of crucial spatial context. Unlike 2D models, 3D models contain complex spatial relationships, often poorly preserved by simpler reduction techniques [10]. This fact is further complicated in our specific real reservoir analysis, which incorporates four distinct geological uncertainties (porosity, permeability, net-to-gross ratio, and rock type), requiring the capture of individual patterns and their intricate interdependencies.

ii. **Simulator Response Replication:** Following dimensionality reduction, proxy models must accurately replicate the complex, non-linear behaviors observed in

various production rates, a demanding task intensified by the multi-phase fluid interactions. While regression-based approaches are common, the large number of cumulative production curves often necessitates training multiple proxy models for different production strategies [1, 4].

iii. **Reconstruction:** The reconstruction stage is vital for verifying consistency between the original geological uncertainty and its reduced-dimensional representation. This process ensures that the fundamental geological features are adequately retained and projected back into the original uncertainty space.

A significant limitation is the independent training of these modular components. This cascaded setup allows errors from early stages to propagate and degrade the performance of subsequent steps [25]. Furthermore, the reliance on extensive feature engineering often involves laborious processes, heuristic assumptions, and brittle design choices, which undermine robustness and generalizability [26].

Finally, a frequently overlooked challenge is the robust quantification of uncertainty in proxy model predictions. Traditional ensemble methods often yield overconfident predictions, underestimating variability [27] and potentially leading to suboptimal reservoir management decisions. Developing more reliable approaches for uncertainty quantification is, therefore, critical.

To address these challenges, we organized our thesis in the format of a collection of papers. Each paper focuses on addressing the challenges previously mentioned:

i. **Dimensionality Reduction and Simulator Response Replication**: the published paper [2] in the journal *Applied Soft Computing*.

ii. **Reconstruction**: we submitted our article [28] to the journal *Computers and Geosciences*.

iii. **Uncertainty Quantification**: our article [29] was submitted to the journal *IEEE Transactions on Neural Networks and Learning Systems*.

1.3 Objective

This thesis has two main goals: (a) to develop an optimized end-to-end proxy model to replace the numerical simulator, and (b) to quantify uncertainty in an end-to-end proxy model to build trust among petroleum engineering experts.

1.4 Research Questions

To achieve the two main goals, this thesis is guided by the following research questions:

Proxy Optimization

- RQ1: Given a set of 3D geological models, how can dimensionality reduction be effectively applied and integrated into an end-to-end proxy model?
- RQ2: Given a set of dimensionally-reduced geological models and various production rates, how can the proxy replicate the simulator's response across these rates as part of an end-to-end solution?
- RQ3: Given a set of dimensionally-reduced geological models, how can the reconstruction of the original 3D reservoir model be performed in a way that preserves the maximum amount of the reservoir's spatial information, enabling its incorporation into an end-to-end proxy?

Proxy Quantification

- RQ4: Given an end-to-end proxy, how can we ensure the reliability and robustness of its predictions to build trust among petroleum engineering experts?

1.5 Main Contributions

This work introduces advancements in proxy modeling for reservoir simulation, with the following main contributions:

Development of an Optimized End-to-End Proxy Model: We propose a novel, end-to-end proxy model designed to replace the numerical simulator during computationally intensive stages of reservoir management. This integrated architecture simultaneously performs dimensionality reduction of 3D geological models, replicates the simulator's response (Chapter 3), and reconstructs the original 3D geological models (Chapter 4). By consolidating these traditionally separate tasks into a single proxy model, our approach inherently minimizes the error propagation inherent in cascaded, independently trained components.

Key features and benefits of our proposed architecture include:

i. **Integrated Heterogeneity and High-Dimensionality Management:** Our model efficiently addresses high-dimensional data and reservoir heterogeneity through parallel feature extraction and regression, effectively capturing complex spatial variations.

ii. **Multi-Task Fluid Dynamics Capture:** We employ a multi-task learning approach to efficiently model intricate fluid relationships, leveraging the inherent interconnectedness of flow dynamics within the reservoir.

iii. **Robustness with Limited Data and Computational Efficiency:** The architecture demonstrates robust performance even when trained with a limited number of simulation models, leading to substantial reductions in computational costs.

iv. **Advanced Multi-Property 3D Geological Reconstruction:** Unlike existing literature that typically handles a single reservoir rock property, our design simultaneously processes and reconstructs four distinct rock properties (e.g., porosity, permeability, net-to-gross ratio, and rock type) within the same architecture. This advancement enables a comprehensive capture and reconstruction of individual patterns, structures, and their complex interdependencies across all geological uncertainties.

Quantify Uncertainty in an End-to-End Proxy: We introduce an innovative methodology for quantifying the uncertainty in proxy models used for reservoir management (Chapter 5), addressing a critical gap in the existing literature on proxy model development.

Key features and benefits of our approach include:

i. **Improved diversity from input:** We leverage an ensemble-based approach to enhance diversity. This strategy is achieved by allowing each base learner in the ensemble to dynamically adjust its activation functions, thereby improving the flexibility and diversity of the model's response based on the input context.

ii. **Improved Scalability and Predictive Accuracy:** We demonstrate the scalability and robustness of our method in a real-world scenario. By adapting the end-to-end proxy model proposed by Farfan *et al.* [2] to incorporate a meta-network instead of traditional activation functions, we improve predictive accuracy and provide an estimate of uncertainty, marking a significant advancement in proxy models for reservoir management.

In Appendix B, we present a detailed comparison between the baseline pipeline and our approach pipeline.

1.6 Text Organization

This thesis is organized as follows. Chapters 2, 3, and 4 address the main challenges identified in this research. Chapter 2 focuses on dimensionality reduction and on reproducing the simulator's response. Chapter 3 addresses the reconstruction of the original three-dimensional geological models from the reduced representations. Chapter 4 introduces a novel methodology for quantifying uncertainty in proxy models used for reservoir management. Finally, Chapter 5 concludes the thesis by summarizing the main findings and contributions, reflecting on the challenges addressed throughout the work, and outlining directions for future research.

Chapter 2

End-to-end dimensionality reduction and regression from 3D geological uncertainties to estimate oil reservoir simulations

This work was published in the Applied Soft Computing (ASOC) Journal, with permission from "Jeanfranco Farfan, Gabriel Cirac, Guilherme Daniel Avansi, Célio Maschio, Denis José Schiozer, and Anderson Rocha. End-to-end dimensionality reduction and regression from 3D geological uncertainties to estimate oil reservoir simulations, 2024."

This chapter introduces an optimized, end-to-end deep neural network (DNN) that functions as a proxy model for reservoir simulation. The proposed model integrates dimensionality reduction and response replication into a single framework, enabling it to handle 3D geological uncertainties and replicate multiple cumulative production curves from a reservoir simulator using only one network.

2.1 Introduction

Numerous approaches have been proposed that utilize DNNs to develop proxy models [1, 3, 4, 5, 6, 20]. Contemporary methodologies for generating a proxy divide the problem into two steps [1, 4, 24, 20]: dimensionality reduction and replication of the simulator's response. Generally, a cascade approach is employed, in which the reduced output serves as input to the simulator's response replication step. Unfortunately, this can lead to significant issues. First, errors in the early stages can propagate and degrade the performance of subsequent steps [25]. Second, extensive feature engineering is often required, which involves labor-intensive processes, heuristic assumptions, and brittle design choices that undermine the robustness and generalizability of the proxy model [26]. We propose an end-to-end proxy approach designed to manage potential errors that may occur during the transformation or processing of 3D geological uncertainties. This method ensures that errors are handled by the architecture itself.

Depending on the characteristics of the reservoir model considered, running a single

reservoir application in the simulator may take days or even weeks [1]. Likewise, since acquiring each model is computationally expensive, in a real oil reservoir [1, 10, 11], there are still limited available models. Additionally, each model exhibits high dimensionality, resulting in numerous variables and parameters. This high dimensionality complicates the identification of meaningful patterns. Therefore, a practical proxy model must be efficient with a limited number of samples while effectively addressing the challenge of high dimensionality.

We introduce a proxy modeling approach that integrates dimensionality reduction and response replication within a single framework. This method employs a 3D convolutional neural network (CNN) to manage both high-dimensional geological uncertainties and complex fluid behavior in an end-to-end manner. By using a single 3D CNN for the entire process, we aim to improve computational efficiency, reduce the number of required models, and enhance prediction accuracy across multiple wells and objective functions. This new architecture has the potential to address the limitations of traditional proxy models by capturing the interconnected relationships between oil, water, gas, and pressure in a unified framework.

In current literature, the focus has shifted towards training multiple neural networks, as highlighted by Sudakov *et al.* [4] and Maschio *et al.* [1]. This approach is driven by the need to deal with diverse scenarios involving multiple producers and varying objective functions. For instance, Sudakov *et al.* [4] employed 31 ANNs to forecast oil saturation and pressure values. Similarly, Maschio *et al.* [1] trained 160 ANNs to predict oil, water, gas, and pressure. Relying on separate artificial neural networks for each well or for each objective would lead to a combinatorial explosion in the number of models to be trained and maintained. In contrast, our present study deviates from these established practices by adopting a novel approach: training a single end-to-end proxy. This comprehensive approach aims to faithfully replicate the simulator's response for each well, encompassing four distinct objective functions within the same proxy to predict oil, water, gas, and pressure.

Unlike previous studies that required numerous ANNs to predict various aspects of reservoir performance, our method leverages a single 3D CNN to predict the behavior of multiple objective functions simultaneously— oil, water, gas, and pressure—within a unified framework. By addressing four objectives in a multi-task approach, the network simultaneously captures the inherent relationships between fluids—a correlation expected in reservoir engineering due to interconnected flow dynamics.

The contributions of this chapter are summarized as follows:

1. Single 3D CNN for predicting oil, water, gas, and pressure within one framework.

2. The method captures fluid relationships through a multi-task approach, leveraging interconnected flow dynamics.

3. The approach addresses high-dimensional data and reservoir heterogeneity via parallel feature extraction and regression.

4. Integrates dimensionality reduction and regression into a single model, streamlining predictions and reducing computational costs.

5. The proxy supports robustness and accuracy even when trained with a limited number of geological simulation models.

2.2 Related Work

In this section, we review the literature on dimensionality reduction techniques and methods for replicating simulator responses.

2.2.1 Dimensionality Reduction

Researchers have developed techniques for vectorizing multidimensional data, which have been adapted by reservoir engineering experts for application in simulation models. Ma *et al.* [24] proposed a framework known as the vector-to-sequence model. Their approach to creating the vector involves utilizing principal component analysis (PCA) to reduce the dimensionality of 3D geostatistical realizations. In a comparable scenario, Da Silva *et al.* [20] investigated 18 dimensionality reduction techniques commonly utilized in the literature for 3D spatial properties, focusing on petrophysical attributes, such as porosity and permeability. Notably, PCA emerged as the optimal choice. Similarly, the methodology proposed by Maschio *et al.* [1] employs PCA to reduce the dimensionality of 3D geostatistical realizations. This reduction is facilitated by integrating an ANN into the Data Assimilation (DA) process, replacing the reservoir simulator. While these and other [1, 6] previous work have demonstrated favorable outcomes through the use of PCA, it is imperative to acknowledge that PCA operates under a linear assumption [30]. Notably, using 3D realizations entails nonlinear features, making PCA less suitable for our purposes.

Opting for a nonlinear technique is crucial when addressing nonlinear geological uncertainties, as it ensures the representation of underlying patterns and relationships. Following this concept, Zhang *et al.* [31] employed the Uniform Manifold Approximation and Projection (UMAP) [32] algorithm to reduce the dimensionality of gas data, characterized by intricate and complex nonlinear relationships. Cirac *et al.* [11, 10] integrated the UMAP to preserve the local structure of a dataset comprising 3D geological representations, which included 2,467,644 data points to represent each variable ($x = 79, y = 137, z = 228$), with a proxy-oil model based on deep hierarchical distillation. The reduced embeddings were fed into a hierarchical distillation framework built on CNNs and applied to a regression problem. The proxy utilized three distinct 1D CNN architectures, each employing slightly different regression schemes. These architectures progressively extracted features based on filter size, capturing local, intermediate, and global patterns.

Kwon *et al.* [5] harnessed the power of a CNN within an optimization framework to determine the optimal positioning of an oil production well within a petroleum reservoir. In their approach, Kwon *et al.* [5] employed a CNN to process 3D reservoir properties, such as permeability. Nonetheless, their method used a 2D CNN for feature extraction to handle the 3D uncertainty of the problem. However, a 3D approach offers an

advantage when addressing complex, multidimensional geological uncertainties, where interdependencies across three dimensions are essential for capturing intrinsic relationships, specifically depth information.

Other investigations have explored more robust methodologies to address the limitation of depth information. Notably, Ma *et al.* [24] employed a Recurrent Neural Network (RNN) [33] to construct a surrogate model for the History Matching problem. Their approach entails a multilayer RNN surrogate model built upon a vector-to-sequence framework. This multilayer RNN, equipped with a Gated Recurrent Unit (GRU) [34], is specifically designed to approximate the mapping from feature vectors of geological realizations to production data. One significant drawback is the substantial computational expense and storage demands associated with extracting spatial features through a CNN. In contrast, Wang *et al.* [35] introduced a hybrid architecture that merges CNN with Long Short-Term Memory (LSTM) networks [36]. This integrated approach serves as a solution to the challenges associated with geothermal production. It utilizes a convolutional component that gathers spatial information from reservoir properties, while the recurrent aspect manages the relationships in sequence data. This combination enables the predictive modeling of dynamic production responses, including variations in rates and temperatures over time.

The widespread use of CNNs in the oil industry can be attributed to their adoption in reservoir management studies. For instance, Cantorna *et al.* [37] investigated the detection of marine oil spills using synthetic aperture radar imaging. Their study examined segmentation algorithms, including clustering, logistic regression, and CNNs, to delineate oil spill areas in Envisat and Sentinel-1 satellite images. While logistical regression and clustering are useful, CNNs yield superior results with minimal computational requirements. Similarly, Khan [38] addressed the classification of frequency-hopping spread spectrum signals in complex electromagnetic environments by employing a modified approach using rectangular filters of different shapes and sizes, along with max pooling, to extract distinctive features through a CNN. Meanwhile, Sharifinasab *et al.* [39] aimed to develop a model for estimating the drill bit penetration rate using Deep Learning (DL) concepts. Their innovative approach combined Generative Adversarial Networks (GANs) with ANNs, specifically Multi-Layer Perceptron Neural Networks (MLP-NNs) and One-Dimensional CNNs.

Conversely, our research applies 3D CNNs to create compact and informative representations of complex three-dimensional data. In contrast to preceding methodologies that utilized techniques like UMAP or 2D CNNs [20, 21], the distinctive strength of 3D CNNs lies in their ability to effectively capture and leverage both spatial and depth relationships inherent within the data. While RNNs excel in capturing sequential dependencies, these may face challenges when confronted with the intricacies of multi-dimensional information [22]. In contrast, a 3D CNN offers a clear advantage, particularly in handling volumetric or spatial-depth data.

2.2.2 Replication of the Simulator's Response

After reducing the high dimensionality of 3D geological uncertainties, the proxy model's primary role is to replicate the complex fluid behaviors within the objective functions [3, 40]. This replication aims to replace the computationally intensive reservoir simulator by rapidly determining target responses for 3D geological realizations with accuracy comparable to that of the simulator [6]. This crucial stage is inherently a regression task.

However, current literature predominantly addresses this by training multiple Artificial Neural Networks (ANNs) to manage diverse scenarios, especially those involving numerous producers and varying objective functions. For instance, Sudakov *et al.* [4] trained 31 ANNs for forecasting oil saturation and pressure. Similarly, Ma *et al.* [41] employed 28 Hybrid Recurrent Convolutional Networks (HRCNs) to predict various reservoir properties, while Maschio *et al.* [1] trained 160 ANNs for oil, water, gas, and pressure values. Maschio *et al.* also introduced a day selection strategy, utilizing only five equidistant days from the production period, rather than all days, to reduce computational overhead while preserving essential information for each cumulative production curve.

In contrast, our study departs from these conventional methods by training a single end-to-end 3D CNN proxy. This approach aims to replicate the simulator's response for each well, incorporating four objectives within a unified framework.

2.3 Proposed Method

We introduce an end-to-end deep learning methodology as a proxy for dimensionality reduction and replicating the simulator's response in a real oil reservoir. This comprehensive approach encompasses three pivotal stages: pre-processing, handling 3D geological uncertainties, and replicating the simulator's response. The high-level architecture is shown in Figure 2.1, where porosity (Por), rock type (Rtp), net-to-gross (Ntg), permeability (Perm), and fluid properties are represented as scalar variables.

2.3.1 Preprocessing

Each entry is initially subject to normalization, making it suitable for integration into the end-to-end architecture. The use of a widely recognized technique known as quantile normalization, introduced by [42], was critical. This choice is based on our treatment of non-linear data. The application of quantile normalization allowed us to adjust the data to a more ML-friendly pattern, which streamlined the training of the architecture. We then cropped and padded the data to ensure correct dimension alignment. We transitioned from the original dimensions of $73 \times 38 \times 24$ to $72 \times 48 \times 24$.

Turning to the outputs, we have seven producer wells (P10, P2, P3, P5, P6, P8, P9) for each objective function, characterized by a historical period spanning 2,206 days. We normalized each objective function using Min-Max scaling transformation inspired by the work of [43]. This process scales the output values to a specific range, often $[0, 1]$, ensuring that all target values adhere to a consistent scale. This practice proves particularly beneficial for regression tasks as it prevents target values from disproportionately influencing

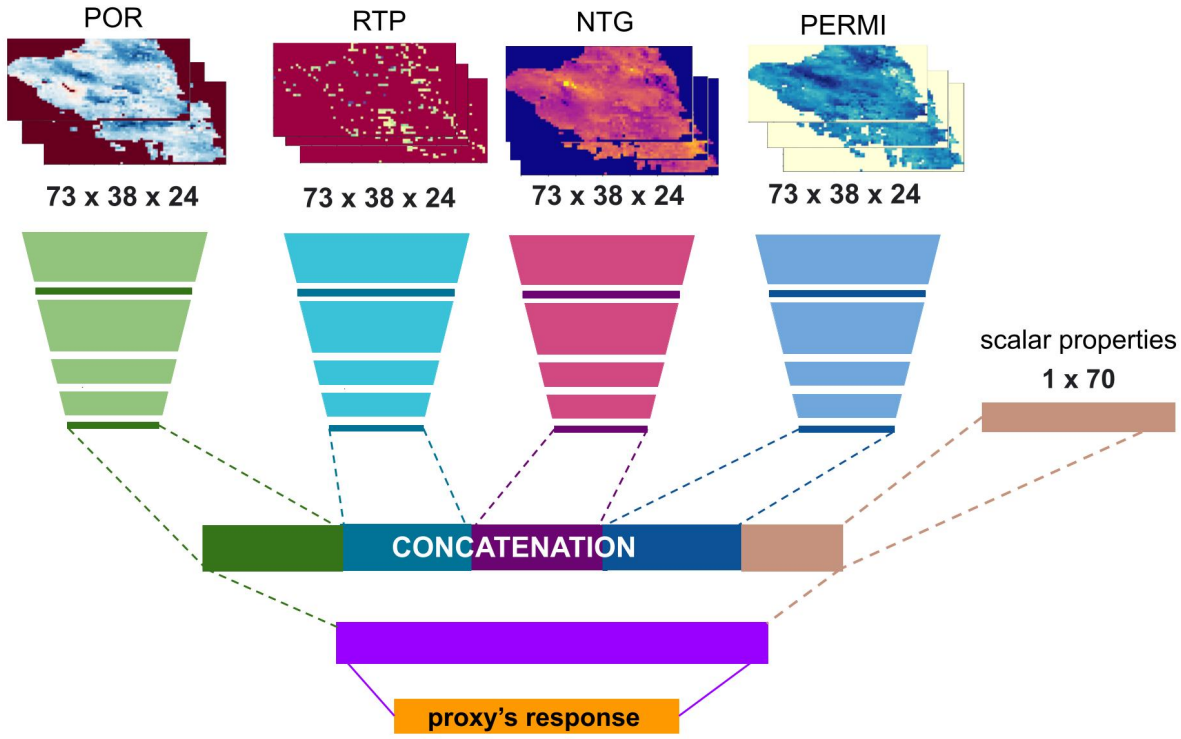


Figure 2.1: Architecture proposed.

the learning process due to their larger magnitude. This approach aligns well with our use of cumulative curves, and normalization helps ensure that observed differences are not attributed to variations in scale. In conclusion, the inputs and outputs were normalized to facilitate their integration into our proxy.

2.3.2 End-to-end Architecture

The trained end-to-end 3D CNN allows for proxy construction. The model is composed of several layers connected sequentially and in parallel. Each layer consists of several 3D convolution and pooling operations. The outcomes of these operations are consolidated through concatenation to form the feature map for subsequent layers. These operations are executed simultaneously across each 3D geological uncertainty. This attribute makes the approach much more efficient than other dimensionality reduction techniques, such as Principal Component Analysis (PCA) [44].

We have structured our solution into two complementary components. The first focuses on addressing 3D geological uncertainty, while the latter replicates the simulator's response. Together, these elements aim to enhance our approach to managing geological challenges.

3D Geological Representation

During the pre-processing stage, each 3D geological uncertainty, with dimensions of $72 \times 48 \times 24$, is fed into our specialized dimensionality reduction module. This module processes each geological realization, obtaining a compact representation vector through

a hierarchical architecture comprising 3D block encoders, 3D max-pooling, and a 3D residual block, as illustrated in Figure 2.2(a). The encoder block (Figure 2.2(b)) is a sequence of convolutional layers, batch normalization, and activation functions, designed to progressively extract and learn robust features from the input data.

The 3D residual block, depicted in Figure 2.2 (c), plays a crucial role in preserving information from earlier layers and facilitating efficient gradient flow through shortcut connections [45]. This block, inspired by Ma *et al.*'s work [41], consists of convolutional layers, batch normalization, and activation functions, followed by another convolutional layer and batch normalization for stable activations. A residual connection then adds the original input to the output of the batch normalization layer before a final activation function constrains the output. Ultimately, the feature vector for the geostatistical realizations is formed by a 256-dimensional representation obtained from the last dense layer.

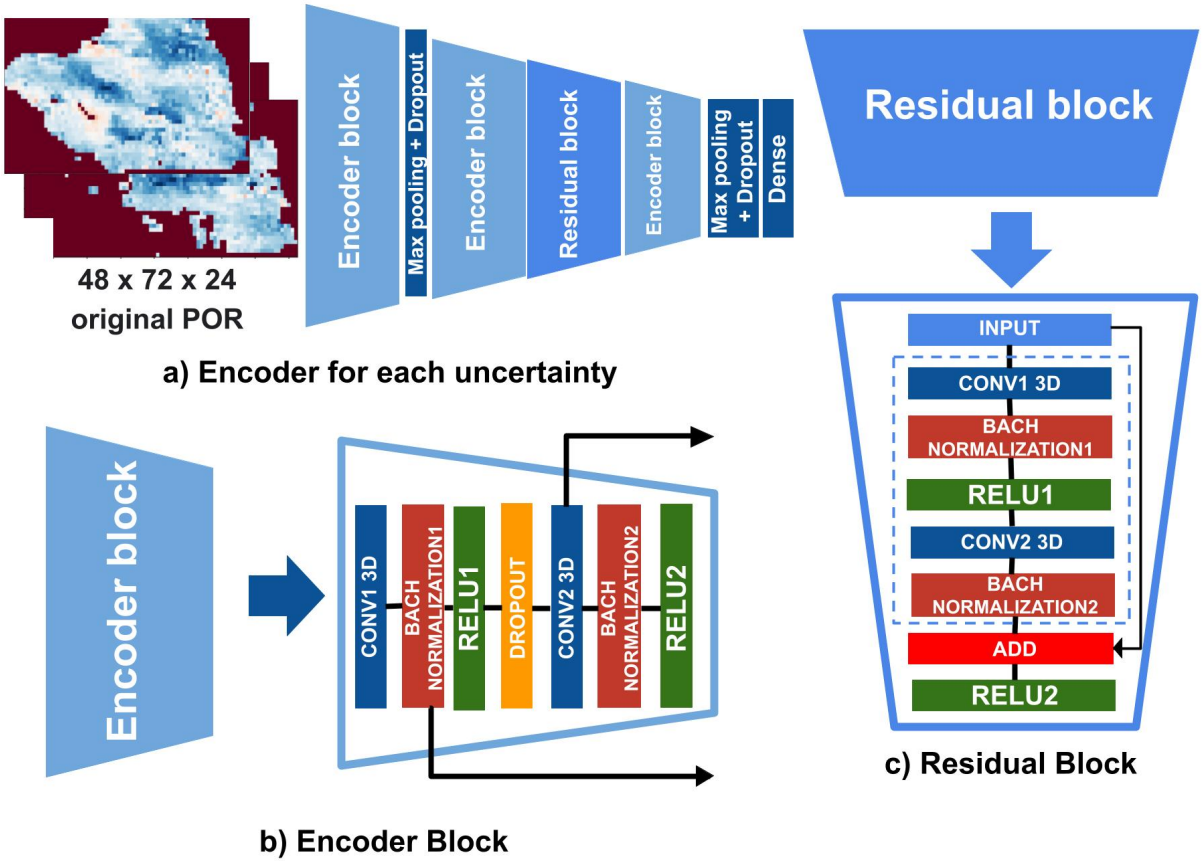


Figure 2.2: The processing details for conducting dimensionality reduction of Porosity (POR) are illustrated in (a). it specifically involves 3D block encoders, 3D max-pooling, and a 3D residual block. This entire process is repeated for every geological uncertainty. A more in-depth explanation of the encoder block is provided in (b). Likewise, for a comprehensive explanation of the residual block, refer to (c).

Table 2.1 thoroughly details the encoder phase for each 3D geological uncertainty, where *EncoderBlock*, *Pool*, *Dense*, and *ResidualBlock* denote a 3D encoder block, 3D max-pooling layer, dense layer, and 3D residual block, respectively, with each $72 \times 48 \times 24$ uncertainty expanded to a $72 \times 48 \times 24 \times 1$ tensor for 3D convolutional layer process-

ing across x, y, and z grid block counts. Specifically, each *EncoderBlock* comprises a sequence of 3D convolutional layers (with L2 normalization), batch normalization, ReLU, and dropout, followed by a 3D Max Pooling and dropout layer; concurrently, a *ResidualBlock* integrates 3D convolutional layers (with L2 normalization), batch normalization, ReLU, and a residual connection between the original input and the output of the last batch normalization, before a final ReLU, with this hierarchical sequence culminating in a 256-dimensional feature vector from a dense layer after flattening.

Table 2.1: 3D CNN for dimensionality reduction of a single geological uncertainty (e.g., Porosity). The network compresses a 72×48×24×1 volume into a **256D** vector while preserving key spatial patterns.

Stage	**Operation / Hyperparameters**	**Output**	**Purpose / Notes**
Input	**Input tensor**	**(72, 48, 24, 1)**	Single 3D cube (e.g., POR) with one channel.
Block 1	EncoderBlock1: Conv3D, 16 filters, 3×3×3, stride 2	(36, 24, 12, 16)	Initial 3D feature extraction with coarse downsampling.
	MaxPool1: 3×3×3	(12, 8, 4, 16)	Spatial summarization (×3 per axis).
	Dropout1	(12, 8, 4, 16)	Regularization.
Block 2	EncoderBlock2: Conv3D, 32 filters, 3×3×3	(12, 8, 4, 32)	Increases channel capacity at constant spatial scale.
	ResidualBlock: Conv3D, 32 filters, 3×3×3	(12, 8, 4, 32)	Residual shortcut preserves information and eases gradient flow.
Block 3	EncoderBlock3: Conv3D, 64 filters, 3×3×3, stride 2	(6, 4, 2, 64)	Second abstraction stage with further spatial reduction.
	MaxPool2: 2×2×2	(3, 2, 1, 64)	Final 3D compaction.
	Dropout2	(3, 2, 1, 64)	Additional regularization before projection.
Projection	Flatten	(384)	3×2×1×64 = 384 features.
	Dense (POR feature vector)	(256)	Final compact representation (**256D**) for the proxy/simulator.

Operation: volumetric input → hierarchical reduction (Conv+Pool) with residual refinement → compact 256D vector (compression ∼324×: 72×48×24×1=82944 → 256).

The objective of using 3D convolutional layers is to capture spatial and depth features within 3D geological uncertainty over time. Moreover, 3D pooling layers can effectively

downsample the spatial dimensions of the volumetric data, preserving crucial information. Furthermore, the encoder block in this sequence acquires and learns representations and features from the input data. The residual block facilitates direct gradient flow, improving the training efficiency. Lastly, integrating Dropout and L2 regularization techniques has been instrumental in mitigating overfitting tendencies and augmenting the network's capacity for generalization.

In this thesis, the CNN architecture adopted for each 3D geological uncertainty is outlined as follows: at the l^{th} layer of a neural network, we initially focus on transforming the output from the previous layer, E_{l-1}, into an auxiliary representation, E_{laux1}. In this context, E_l and E_{l-1} respectively represent the feature maps at layers l and $l-1$ (where E_0 represents some geological uncertainty). This transformation is crucial for capturing relevant features before applying regularization techniques. Equation (3.1) describes the process of computing E_{laux1}, incorporating L2 normalization, batch normalization, and ReLU activation.

$$E_{laux1} = \text{ReLU}(\text{BatchNorm}(\text{L2Norm}(W_l \cdot E_{l-1} + b_l))) \tag{2.1}$$

The symbols W_l and W_{laux2} represent 3D kernels with a size of $3 \times 3 \times 3$, while b_l and b_{laux2} represent the biases. The $*$ symbol signifies the convolution operation, involving the summation of element-wise products of W_l and the corresponding region of E_{l-1}, and W_{laux2} and the corresponding region of E_{laux2}. The subsequent application of *Dropout* to E_{laux1} results in E_{laux2}, as detailed by Equation 3.2, a critical step for mitigating overfitting during training.

$$E_{laux2} = \text{Dropout}(E_{laux1}) \tag{2.2}$$

Finally, E_{laux2} is transformed into the layer's output, E_l, through a second series of L2 normalization, batch normalization, and ReLU activation operations. Equation 3.3 outlines this process, emphasizing the importance of maintaining consistency and effectiveness across layers.

$$E_l = \text{ReLU}(\text{BatchNorm}(\text{L2Norm}(W_{laux2} \cdot E_{laux2} + b_{laux2}))) \tag{2.3}$$

L2 regularization is applied to mitigate overfitting, followed by a Batch Normalization layer to normalize the input of each layer l by adjusting and scaling the activations. Normalization is followed by a *ReLU* activation, which introduces non-linear features. The *Dropout* layer then helps reduce overfitting. Specifically, this describes the configuration of the encoder block.

The method incorporates 3D max pooling layers using two pooling kernels: $3 \times 3 \times 3$ and $2 \times 2 \times 2$. These layers reduce the dimensions of each feature map by a factor of 3, and in the last layer, decrease the dimensions by a factor of 2. We then introduce a Dropout layer to mitigate overfitting.

In our CNN architecture, we integrate a residual block. The residual block is structured with a shortcut connection, allowing information to propagate directly from earlier to later layers. This design preserves crucial features and streamlines gradient flow dur-

ing training. The objective is to approximate the desired underlying mapping $H(x)$ by leveraging the residual function $F(x) = H(x) - x$, where x represents the inputs. This residual function captures the disparity between the desired mapping $H(x)$ and the identity mapping x, enabling the stacked layers within the residual block to focus on learning the necessary residual features for refining the input data. Operationally, the residual block performs $F(x) + x$ through a sequence of composite functions, including convolution, batch normalization, and ReLU layers. Put together, these operations contribute to the computation of the residual features $F(x)$. The shortcut connection structure through element-wise addition seamlessly integrates the input x with the calculated residual features $F(x)$. This direct connection path facilitates efficient information flow and mitigates the risk of vanishing gradients.

As the 3D geological uncertainty input (e.g., porosity) undergoes recursive processing through the EncoderBlock-Pool-Dropout-ResidualBlock layers, the dimensions of the feature map (d_i) along the x, y, and z directions decrease, while the total number of feature maps increases. Following the final Dropout layer, we apply a flattened layer to compact the spatial dimensions of the feature map. Finally, a Dense layer generates the vector representation of the geological uncertainty in 256 features. We repeat this process for each 3D geological parameter.

Regression Block

By implementing dimensional reduction to 3D geostatistical realizations, we obtain four feature vectors with 256 dimensions each. These arrays are concatenated with the scalar properties, resulting in 1094 features. Subsequently, two fully connected layers were employed to process the compacted vector. The first layer contains 512 neurons, while the second layer consists of 140 neurons, resulting in the proxy response (Figure 2.3). The process of concatenation is demonstrated through the following equation.

$$Compacted_Vector = CONCAT(F_{POR}, F_{PERM}, F_{NTG}, F_{RockType}, F_{Scalar}) \tag{2.4}$$

Within this framework, the $Compacted_Vector$ symbolizes the concatenated array, culminating in a comprehensive assemblage of 1094 features. In precise terms, F_{POR} designates the Porosity attribute subjected to dimensional reduction, F_{PERM} corresponds to the dimensionally-condensed Permeability characteristic, F_{NTG} encapsulates the Net-to-Gross aspect following dimensional reduction, $F_{RockType}$ encapsulates the rock type trait post-dimensional condensation, and F_{Scalar} encompasses the scalar properties of the reservoir. Similarly, the subsequent equation outlines the regression task.

$$Proxy_Response = FC(FC(Compacted_Vector)) \tag{2.5}$$

In this equation, $Proxy_Response$ signifies the replication of the simulator's response applied to the $Compacted_Vector$. Specifically, this process involves utilizing two successive fully connected layers, each housing 512 and 140 neurons, respectively. The final output of this operation constitutes the simulator's response replicated for the cumulative

production across the entire spectrum of objective functions, encompassing oil, water, gas, and pressure.

The choice of loss function is crucial in determining the training process and learning dynamics of neural networks. In our proxy, we specifically aim to minimize the SMAPE loss function. The formulation of this loss function is presented below.

$$SMAPE(\Theta) = \frac{100}{n} \sum_{i=1}^{n} \frac{|F_i - A_i|}{(|F_i| + |A_i|)/2} \tag{2.6}$$

where Θ represents the trainable parameters, F_i corresponds to the simulation data, and A_i corresponds to the regression generated by the proxy for the i-th sample.

We adopt the Adam optimizer [46] with an initial learning rate of 0.0001 for the training. This learning rate was selected after testing a range of values to identify the most promising value that ensures stable convergence while avoiding issues such as overshooting or slow learning. To mitigate the risk of overfitting, we employ an early stopping technique. The proxy undergoes training over 250 iterations, leading to the derivation of weights tailored to our proposed methodology. Given the relatively modest sample size of the proxy model, a batch size of 8 is used.

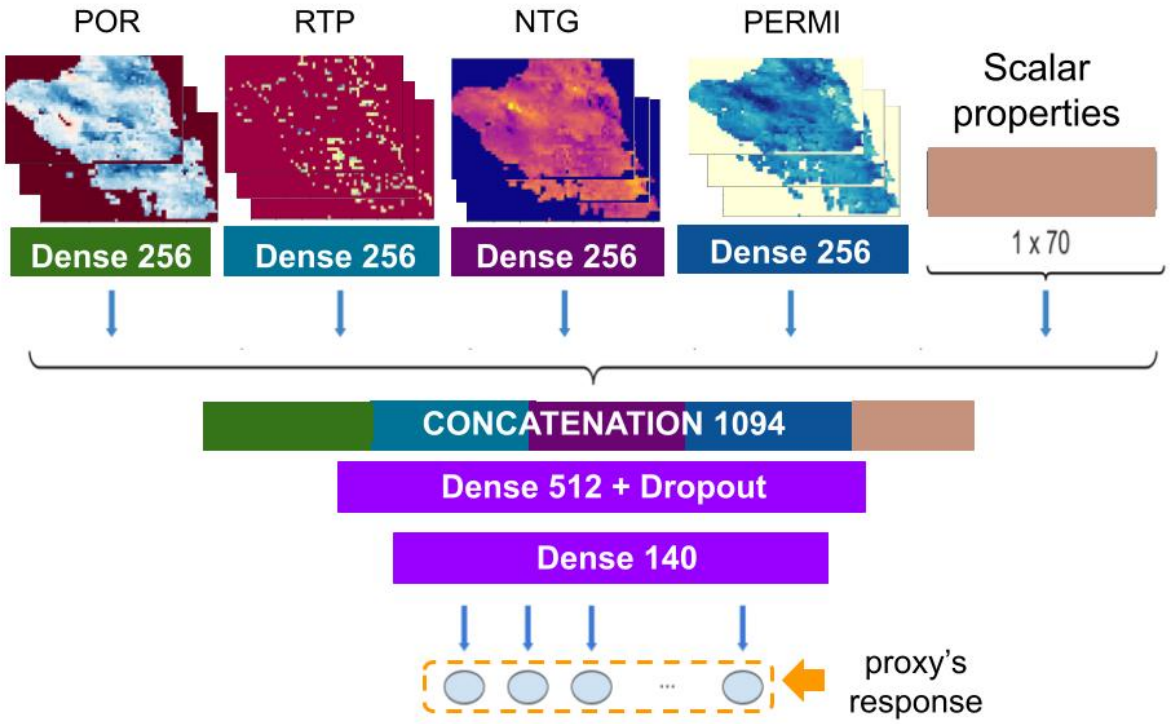


Figure 2.3: The end-to-end solution employs the dimensionally-reduced properties to conduct regression via the utilization of Dense layers.

2.4 Output Representation

We trained a single architecture with 140 output neurons. These outputs include the responses generated by our end-to-end proxy. This approach starkly contrasts conventional

methodologies, where the same architectures are trained multiple times, each corresponding to the objective function evaluated for individual wells. In other words, traditional approaches repeatedly train the same neural network. Recent research efforts by Sudakov *et al.* [4] and Maschio *et al.* [1] also involved training multiple instances of the same neural network. These networks utilize 3D geological uncertainty inputs during training to replicate the simulator's response based on the specific objective function under analysis. For example, Sudakov *et al.* [4] trained a set of 31 ANNs to predict values for oil saturation and pressure, while Maschio *et al.* [1] conducted a study with 160 ANNs. In contrast to these established practices, our present study trained a unique neural network. We employed a single end-to-end 3D CNN proxy that can replicate the simulator's response for each well and each distinct objective function.

To comprehend the significance of the 140 neurons, one must perform a detailed analysis of Figure 2.4. In part a), the oil objective function, specifically the well P10, is depicted with a series of cumulative curves spanning 2206 days. However, it is worth noting that the well was not operational throughout this period; it was active for only 323 days. Consequently, the x-axis represents these 323 days, and the y-axis represents production during this specific time frame.

For part b), a single curve is selected from this set. In part c), following a methodology inspired by the research conducted by Maschio *et al.* [1], we identified five specific time instances, denoted as t_0, t_1, t_2, t_3, and t_4, corresponding to days 506, 1,014, 1,360, 1,775, and 2,156, respectively. This selection is made possible by cumulative production curves, which demonstrate that a small set of time points can effectively represent the entire curve. As shown in (d), this process is repeated. For each cumulative curve within well P10, use these curves with just five days rather than the full 323 days. Therefore, the curves between a) and d) differ; however, d) represents a).

The previous example was employed for the well P10 within the context of the oil objective function. Given that there are seven producer wells for oil, by applying the same procedure and selecting identical days for each well, we obtained a total of 35 days. The same process was thereafter executed for each objective function, ultimately yielding a cumulative total of 140 days (35 days $\times$ 4 objective functions).

We underscore that these cumulative curves encapsulate data generated by the reservoir simulator, encompassing key metrics such as Np (cumulative oil production), Wp (cumulative water production), Gp (cumulative gas production), and BHP (pressure production) for each respective producer well.

2.5 Results and Discussions

To assess the effectiveness of the proposed methodology, we employed a Brazilian offshore field, known as the S-Field, situated within the geographical boundaries of the Campos Basin. This reservoir comprises 200 geostatistical realizations, i.e., 3D maps of porosity, permeability, net-to-gross, and rock types, each dedicated to simulating the intricate dynamics of fluid flow through porous and permeable media under the subsurface.

To verify the method's effectiveness against data limitations, we explored three combi-

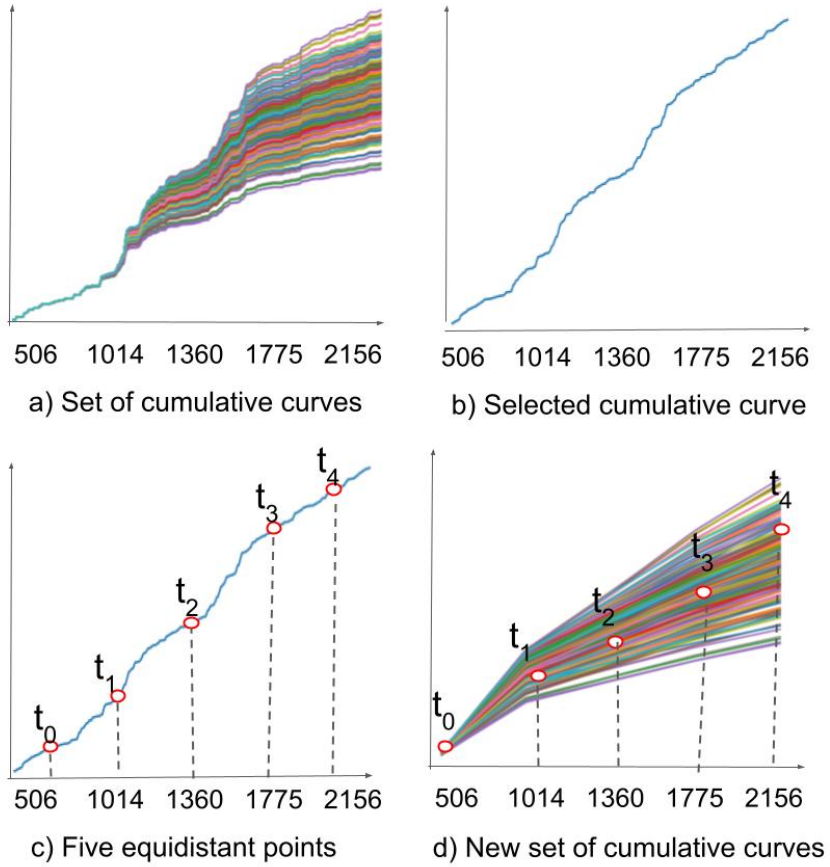


Figure 2.4: Illustrates our methodology, building on the approach described by Maschio *et al.* [1]. In a), Well P10 is presented with a set of cumulative curves spanning 2206 days. In b), we select a single curve from the set. In c), we identify five equidistant points along this curve and use them to replace the entire curve. As shown in d), this process is repeated for each cumulative curve within Well P10. We underscore that this procedure is applied not only to Well P10 but to all other wells and each respective objective function.

nations to partition the dataset. The first combination consisted of 50 samples allocated for regressor training, 20 for validation, and 130 for the subsequent testing set. In the second combination, 100 examples were used for regressor training, 40 samples for validation, and 60 samples for further testing. Lastly, the third combination involved 140 samples for training, 20 for validation, with the remaining 40 samples for further testing. Experiments were performed using two NVIDIA Quadro RTX 8000 GPUs, each featuring 48 GB of GDDR6 memory.

The study evaluated the robustness of the architecture compared to state-of-the-art methodologies. We selected recent methodologies proposed by Maschio *et al.* [1], Sudakov *et al.* [4], Ma *et al.* [41] and Kolajoobi *et al.* [6] for comparison. We replicated the methods proposed by these authors on our dataset to meticulously assess the introduced technique, adhering closely to their original methodology.

For each of these methodologies, we employed four metrics: the Correlation Coefficient (R), the Coefficient of Determination (R^2), the Symmetric Mean Absolute Percentage Error (SMAPE), and Computational Time (CTime). The R value provides insight into the degree of association between simulated and proxy values, offering information on

the strength and direction of their relationship. Meanwhile, R^2 measures the predictive capacity of the proxy. SMAPE gives a balanced assessment of forecast accuracy. Finally, CTime denotes the time required for each approach to upload files, train, and predict the proxy's response.

In Tables 4.3 and 4.4, we compare our end-to-end 3D CNN approach with state-of-the-art methodologies, including a comparison with Maschio *et al.* [1], who also analyzed the same reservoir. In contrast to the approach that utilizes Principal Component Analysis (PCA) with 150 feature vectors for dimensionality reduction, our method strategically employs 129 components. This number represents the optimal count at which the explained variance reaches an acceptable threshold. Additionally, we incorporate 70 scalar attributes, resulting in a total of 199 components for each model. We chose five points on cumulative production curves at specific times (days 506, 1,014, 1,360, 1,775, and 2,156) for each producer, resulting in 140 points across seven producers and four objective functions. Consequently, our architecture predicts 140 points; 35 related to each fluid.

To compare, Sudakov *et al.* [4] reshaped uncertainties for use in a multilayer feedforward network with two hidden layers (128 and 64 neurons), training 28 ANNs for seven producers across four objective functions. Ma *et al.* [41] introduced a Hybrid Recurrent Convolutional Network (HRCN) combining residual and encoder blocks with two bidirectional LSTM layers (64 neurons), also training 28 HRCNs. Kolajoobi *et al.* [6] applied PCA with 10 components followed by two ANNs with layers of 30 and 20 neurons. In contrast, our methodology distinctively trains a single ANN, diverging from the multiple model approach of the prior studies.

Table 4.3 compares our end-to-end 3D CNN with leading methods on oil and water prediction, presenting metrics averaged over training runs with 50, 100, and 140 simulation models. Our approach demonstrates superior performance, achieving the lowest SMAPE values for both oil (3.413) and water (19.437), which indicates high predictive accuracy. Furthermore, it yields strong correlation coefficients (R) of 0.970 for oil and 0.954 for water. However, the R^2 values suggest the model's goodness of fit could be enhanced, particularly for water predictions. In contrast, the method by Maschio *et al.* [1], while competitive, results in higher prediction errors (SMAPE of 5.370 for oil and 27.458 for water). Its slightly lower R and R^2 values also point to a less robust data fit compared to our model.

To continue the analysis of Table 4.3, Sudakov *et al.* [4] present an alternative approach that closely competes with our end-to-end 3D CNN method. Their SMAPE values are slightly higher than ours for both oil (3.615) and water (19.648), but they achieve superior R^2 values. Notably, their R^2 value of 0.963 for oil stands out as the highest among all methods, indicating a superior fit to the data. Ma's *et al.* [41] methodology yields results consistent with other approaches. While their SMAPE values are higher than ours, their R^2 (0.907, 0.857) values suggest reasonable predictive accuracy. Kolajoobi's *et al.* [6] method demonstrates moderate performance compared to other approaches. Their SMAPE values are higher for oil (4.980) and water (26.945), indicating prediction errors. Similarly, their R and R^2 values suggest a less precise fit to the data.

The results represent the average performance across three different partitioning ratios. By carefully analyzing each combination, we can recognize the significant influence that

Table 2.2: Comparative evaluation of our end-to-end 3D CNN with state-of-the-art methods focusing on oil and water-related objective functions.

	Oil			Water		
Work	**SMAPE**	**R**	R^2	**SMAPE**	**R**	R^2
end-to-end 3D CNN	**3.413**	**0.970**	0.939	**19.437**	**0.954**	**0.907**
Maschio *et al.* [1]	5.370	0.929	0.847	27.458	0.907	0.798
Sudakov *et al.* [4]	3.615	0.963	**0.963**	19.648	0.946	0.890
Ma *et al.* [41]	4.167	0.959	0.907	22.711	0.936	0.857
Kolajoobi *et al.* [6]	4.980	0.940	0.882	26.945	0.919	0.843

dataset size exerts on our findings. Employing 50 simulation models for training yielded an SMAPE of 3.528. Increasing the training dataset to 100 simulation models resulted in a slightly improved SMAPE of 3.359. Further enlarging the training dataset to 140 simulation models yielded a comparable SMAPE of 3.351 for oil prediction. For water prediction specifically, utilizing 50 simulation models for training produced an SMAPE of 19.744, while doubling the training dataset to 100 simulation models resulted in a reduced SMAPE of 19.436. Employing 140 simulation models for training led to a further reduction in SMAPE to 19.13. A notable observation is that, even with a limited dataset size of 50 samples, our approach demonstrates competitive predictive performance compared to using a larger dataset of 140 samples. This finding underscores the practical applicability of the approach in data-constrained environments.

Regarding Maschio's *et al.* [1] work, employing 50 simulation models for training yielded an SMAPE of 7.203 and 33.84 for oil and water, respectively. Utilizing 100 simulation models for training resulted in an SMAPE of 5.099 and 27.125 for oil and water, respectively. Meanwhile, employing 140 simulation models for training yielded an SMAPE of 3.809 and 21.409 for oil and water, respectively. We observe that, as the number of simulations increases, the SMAPE decreases for oil and water. Continuing with Sudakov's *et al.* [4] findings, employing 50 simulation models for training resulted in an SMAPE of 3.658 and 19.476 for oil and water, respectively. Utilizing 100 simulation models for training yielded an SMAPE of 3.576 and 19.572 for oil and water, respectively. In the meantime, employing 140 simulation models for training resulted in an SMAPE of 3.611 and 19.897 for oil and water, respectively. Even with a higher count of samples, the slight variation in SMAPE suggests the method's resilience and effectiveness, even when faced with data constraints.

The results presented in Table 4.4 provide a comparative evaluation of gas and pressure. Our end-to-end 3D CNN demonstrates promising performance across gas and pressure domains, as indicated by the lowest SMAPE values achieved: 3.461. Our approach also yields strong correlation coefficients (R) of 0.969 for gas and 0.880 for pressure, meaning significant linear relationships between predicted and actual values. However, while the R^2 values for gas and pressure suggest a reasonable fit to the data, there is still room for improvement, particularly in the pressure predictions, where the R^2 value is comparatively lower at 0.765.

In comparison, the results presented by Maschio *et al.* [1] show higher SMAPE values

for both gas (5.357) and pressure (1.562), indicating larger prediction errors compared to our approach. Additionally, their correlation coefficients and R^2 values are lower, which suggests a less robust fit to the data. Sudakov *et al.* [4] present an alternative approach with competitive performance. Their SMAPE values are slightly higher than ours for both gas (3.622) and pressure (1.136), but they achieve high correlation coefficients and R^2 values, particularly for gas predictions. Similarly, the approaches by Ma *et al.* [41] and Kolajoobi *et al.* [6] demonstrate moderate performance compared to our method. Although their SMAPE values are similar to ours, our end-to-end 3D CNN outperforms them with higher R and R^2 values in gas and pressure predictions, highlighting its superiority in accurately replicating the simulator's response.

Table 2.3: Comparative evaluation of our end-to-end 3D CNN with state-of-the-art methods focusing on gas and pressure-related objective functions.

	Gas			**Pressure**		
Work	**SMAPE**	**R**	R^2	**SMAPE**	**R**	R^2
end-to-end 3D CNN	**3.461**	**0.969**	**0.939**	**1.055**	**0.880**	**0.765**
Maschio *et al.* [1]	5.357	0.930	0.864	1.562	0.791	0.538
Sudakov *et al.* [4]	3.622	0.964	0.930	1.136	0.858	0.727
Ma *et al.* [41]	4.128	0.960	0.910	1.257	0.842	0.680
Kolajoobi *et al.* [6]	4.995	0.940	0.882	1.401	0.801	0.611

Our analysis across different training sizes (50, 100, and 140 simulation models) demonstrates a steady improvement in SMAPE values for both gas and pressure predictions, with marginal gains as the dataset size increases. Employing 50 simulation models for training yielded an SMAPE of 3.57; increasing the training dataset to 100 simulation models resulted in a slightly improved SMAPE of 3.414. Further enlarging the training dataset to 140 simulation models yielded a comparable SMAPE of 3.399 for gas prediction. Specifically for pressure prediction, employing 50 simulation models for training produced an SMAPE of 1.085, while doubling the training dataset to 100 simulation models resulted in a reduced SMAPE of 1.062. Employing 140 simulation models for training led to a further reduction in SMAPE to 1.017. Remarkably, the model shows competitive performance even with as few as 50 models, indicating its robustness in sparse data conditions. This robustness emphasizes the model's potential for practical use in situations where data is limited.

Regarding Maschio's *et al.* [1] work, employing 50 simulation models for training yielded an SMAPE of 7.297 and 2.006 for gas and pressure, respectively. Utilizing 100 simulation models for training resulted in an SMAPE of 4.958 and 1.562 for gas and pressure, respectively. Meanwhile, employing 140 simulation models for training yielded a SMAPE of 3.815 for gas and 1.119 for pressure, respectively. We observe a significant decrease in SMAPE for gas and pressure as the number of simulations increases. Continuing with Sudakov's *et al.* [4] findings, employing 50 simulation models for training resulted in an SMAPE of 3.675 and 1.123 for gas and pressure, respectively. Utilizing 100 simulation models for training yielded an SMAPE of 3.568 and 1.12 for gas and pressure, respectively. Meanwhile, employing 140 simulation models for training resulted in

an SMAPE of 3.623 and 1.165 for gas and pressure, respectively. Despite the increase in the number of samples, we observe that the change in SMAPE is almost negligible.

Our end-to-end 3D CNN approach surpassed all other methodologies in simulating responses across diverse objective functions (oil, water, gas, pressure) thanks to its ability to process volumetric data and maintain spatial patterns. Sudakov's *et al.* [4] method stands out as the second-best, effectively utilizing spatial geological data, yet simpler techniques like PCA and CNN fall short in comparison. Ma's *et al.* [41] complex HRCN architecture, despite its advanced components, struggles with stability, often complicating rather than enhancing predictive performance. Similarly, the methodology proposed by Kolajoobi *et al.* [6], which employs PCA and ANN with a shallow network, may underperform compared to 3D CNN or HRCN, primarily due to limitations in capturing complex nonlinear relationships within the data. While PCA effectively reduces input space dimensionality, it may not capture intricate data variations, particularly in high-dimensional spaces. Lastly, Maschio's *et al.* [1] methodology, relying on PCA and a simplistic neural network, was less effective in capturing data variability, resulting in modest performance against more sophisticated models.

An essential consideration in reservoir analysis pertains to the computational time (CTime) required for conducting the necessary calculations. Table 2.4 compares the computational performance of our 3D CNN approach with several state-of-the-art methods. It is worth noting that our results are based on a single neural network architecture, unlike the methodologies referenced, such as Maschio *et al.* [1], which employed 140 neural networks—equivalent to 5 days $\times$ (7 producers $\times$ 4 objective functions), while the remaining methodologies utilized 28 neural networks, which correspond to (7 producers $\times$ 4 objective functions). The assessment covered various computational tasks, including file loading, normalization, training, and replicating the simulator's response using different training samples: 50, 100, and 140.

Table 2.4: Comparative evaluation of the time performance of our end-to-end 3D CNN compared to state-of-the-art methods. This evaluation focuses on computing time for tasks such as uploading files, normalization, training, and predicting the replicated simulator responses across various combinations of training samples: 50, 100, and 140.

	CTime		
Work	**50**	**100**	**140**
end-to-end 3D CNN	**6:36**	**6:41**	**6:46**
Maschio *et al.* [1]	13:53	30:13	34:53
Sudakov *et al.* [4]	01:08:58	01:15:55	01:48:09
Ma *et al.* [41]	02:29:13	06:32:05	07:18:45
Kolajoobi *et al.* [6]	7:21	8:17	9:13

Our end-to-end 3D CNN consistently outperforms all methodologies assessed in computational time across all training sample combinations. Specifically, our approach exhibits significantly shorter computation times than Maschio *et al.* [1], who also tackled this reservoir but utilized considerably greater computational resources. Ma *et al.* [41] showcases the lengthiest computation times, especially with greater training sample sizes,

indicating potential scalability concerns with larger datasets. Similarly, Sudakov *et al.* [4] demonstrate computationally intensive results as sample sizes increase, although not as pronounced as in Ma *et al.* [41], making it impractical. Meanwhile, Kolajoobi *et al.* [6] demonstrate relatively competitive computation times, particularly with smaller sample sizes.

The time calculations were based on the fact that all methodologies execute four tasks: file loading, geological uncertainty normalization, training, and predicting the simulator's responses. Concerning the first two tasks, all methods load and normalize files in 2 minutes and 13 seconds. The discrepancy in performance becomes apparent during training and predicting the simulator's response, where the efficiency of the methodology is observed.

When using 50 simulation models for training and prediction, the total time invested was 4 minutes and 23 seconds. Including the additional 2 minutes and 13 seconds required for file loading and normalization, the total time amounted to 6 minutes and 36 seconds. Similarly, using 100 simulation models for training and prediction required 4 minutes and 28 seconds, resulting in a total of 6 minutes and 41 seconds. Meanwhile, employing 140 simulation models for training and prediction took 4 minutes and 33 seconds, resulting in a total time of 6 minutes and 46 seconds. Summarizing the findings from Table 2.4, it is clear that our end-to-end 3D CNN approach surpasses all other methodologies in terms of computational efficiency for replicating the simulator response.

A closer examination of Tables 4.3 and 4.4 reveals that the methodology closest to ours, as observed in terms of SMAPE, R, and R^2 metrics, implemented by Sudakov *et al.* [4], does not maintain consistency, as illustrated in Table 2.4. This inconsistency stems from the high computational costs incurred by training numerous networks and increasing the number of simulation models. Similarly, Table 2.4 highlights that the methodology with the closest computational efficiency to ours corresponds to Kolajoobi *et al.* [6]. However, despite this efficiency, suboptimal performance is observed in SMAPE, R, and R^2 metrics. This fact can be attributed to the limitations of the technique employed to reduce dimensional and spatial information, which may not be the most suitable approach for addressing 3D geological uncertainties.

Sudakov et al. (2019) are recognized for achieving the second-highest performance with their approach, which effectively combines spatial geological data with fluid properties. However, their methodology is associated with high computational costs, mainly due to the absence of initial dimensional reduction and the requirement to train a total of 28 networks. Kolajoobi *et al.* [6] also show commendable computational efficiency and adaptability to increased simulation model sizes despite training 28 networks. The approach under discussion relies on just 10 PCA components to represent high-dimensional data. While this method aims for simplicity, it may overlook crucial spatial details, which could lead to inaccuracies in modeling and analysis due to the possible loss of significant information.

Our method surpasses other state-of-the-art approaches in key performance metrics (SMAPE, R, and R^2), as shown in Tables 4.3 and 4.4, highlighting the aptitude of the 3D CNN for managing volumetric data to identify complex geological patterns. The effectiveness and scalability of our comprehensive approach are highlighted in Table 2.4. This table illustrates how our method can simulate responses across all wells and functions

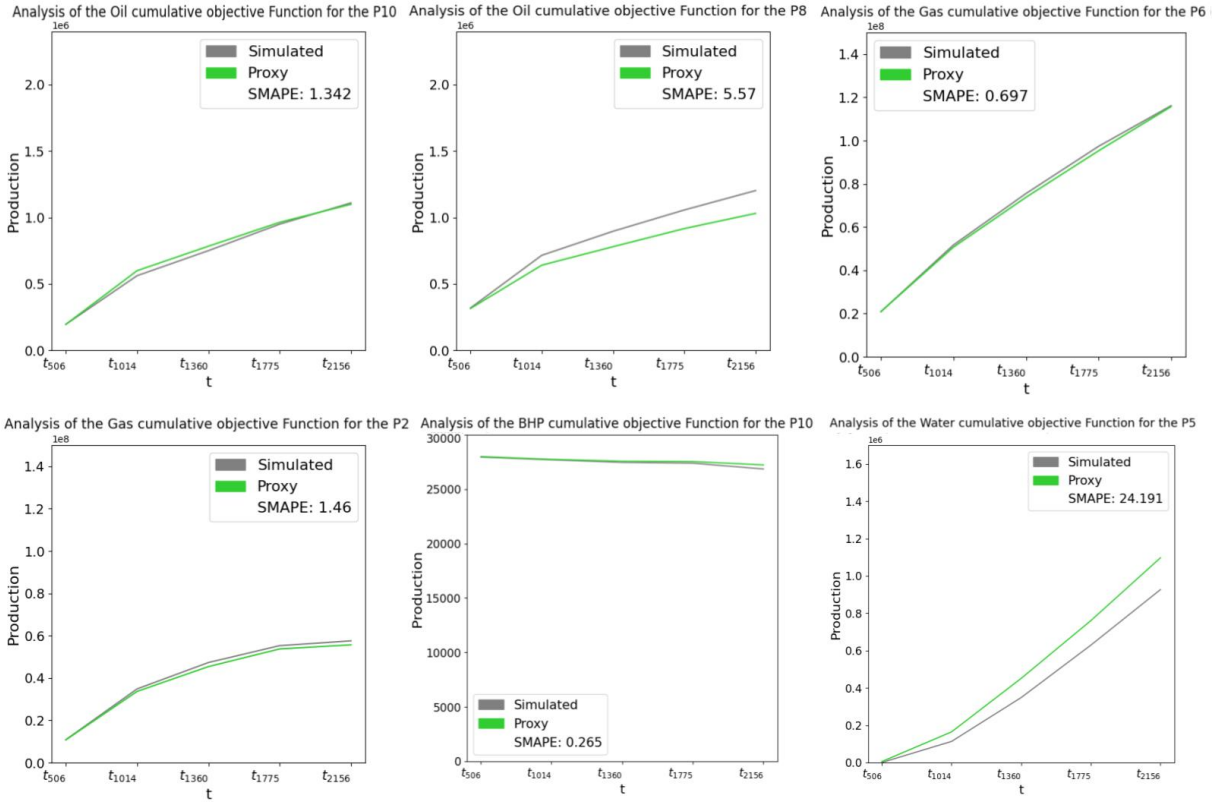


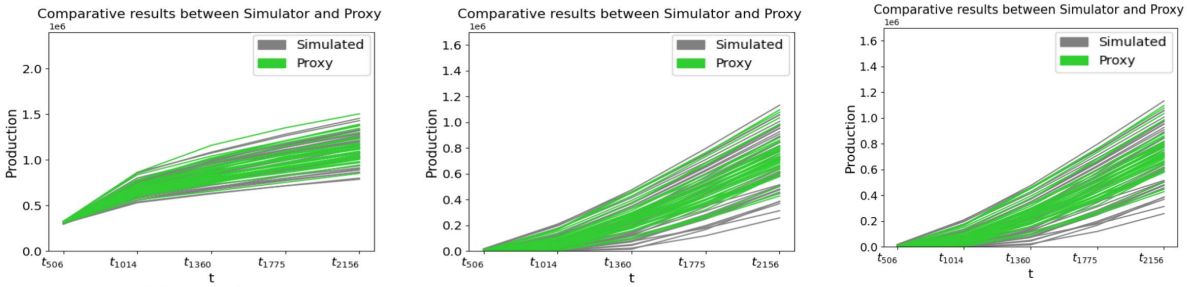
Figure 2.5: Cumulative production curves from our end-to-end 3D CNN (in green) closely match the simulator's outputs (in gray) across selected days. The analysis spans various wells and objective functions; oil for Wells p10 and p8, gas for Wells p6 and p2, pressure for Well p10, and water for Well p5, demonstrating the methodology's accuracy in replicating real-world production data.

using a single network, demonstrating its capability to handle an increasing number of simulation models efficiently. We now proceed to detail the performance metrics visually.

Figure 2.5 presents six plots of cumulative production curves from different wells and objective functions, with five points per plot, indicating the production on days 506, 1014, 1360, 1775, and 2156. The proxy results from our 3D CNN methodology (green) align closely with the simulator's outputs (gray), evidenced by the SMAPE metric, which indicates accurate proxy simulation behavior.

Figure 2.6 further corroborates these findings across all test simulation models, showing a close alignment between the proxy and simulator curves. The high values of R and R^2 indicate a strong consistency that affirms the accuracy of the proxy. This also demonstrates the effectiveness of our end-to-end 3D CNN architecture, as evidenced by the results presented in Tables 4.3 and 4.4.

To provide a comprehensive understanding of the characteristics of our proxy, Figure 2.7 elucidates the visual representation of the simulator's replicated responses in the objective functions of the test set. These figures are also accompanied by their respective metric values - namely R, R^2, and SMAPE - which indicate correlation strength, explanatory effectiveness, and precision. These results validate the accuracy of the regression model predictions and offer insights into the extent to which the end-to-end 3D CNN



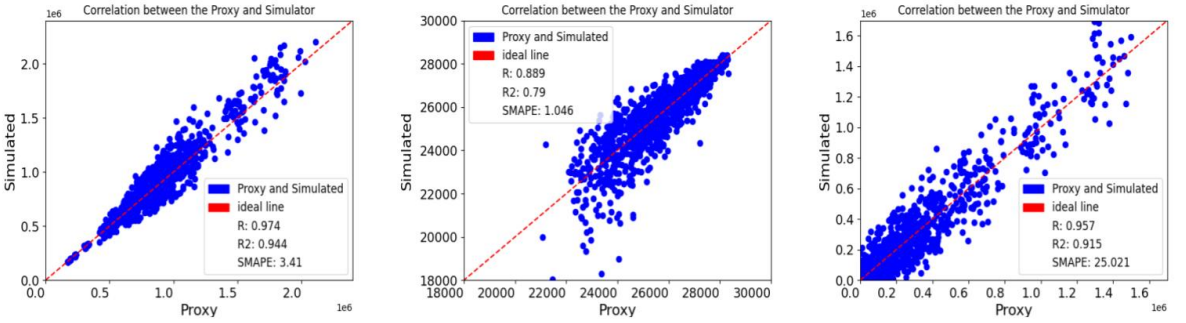
a) Analysis of the Oil Objective Function for the whole models in the P2 producer

b) Analysis of the Water Objective Function for the whole models in the P2 producer

c) Analysis of the Gas Objective Function for the whole models in the P2 producer

Figure 2.6: Several outcomes stemming from our end-to-end 3D CNN application on the simulator are showcased for all models within the P2 producer in the test set.

reflects the behavior of the simulator.



a) Correlation of the Oil Objective Function for Each Producer Well

b) Correlation of the Pressure Objective Function for Each Producer Well

c) Correlation of the Water Objective Function for Each Producer Well

Figure 2.7: For each image, the correlation representation for each objective function encompasses the entirety of the producers, specifically P10, P2, P3, P5, P6, P8, and P9. The illustration includes correlation metrics R with R^2 and SMAPE metric.

Figures 2.7 (a) and (b) present an agreement between the results, where R, R^2, and SMAPE show a directly proportional relationship. Specifically, for the oil objective function, R is 0.974, R^2 is 0.944, and SMAPE is 3.41. Meanwhile, for the pressure objective function, R is 0.889, R^2 is 0.79, and SMAPE is 1.046. However, regarding the water objective function, while the relationship between R and R^2 remains directly proportional, SMAPE exhibits an inversely proportional relationship, with our approach showing a SMAPE value of 25.021.

To understand the elevated SMAPE values for water, the study analyzed the metric's behavior. As shown in Figure 2.7 (c), many water values are close to zero. The SMAPE metric is known to be sensitive to near-zero values, as small absolute errors can yield disproportionately large percentage errors when the real and predicted values are small [47]. This artifact explains the high SMAPE score observed. To provide a more robust assessment, our analysis is augmented with scale-invariant metrics. Our approach yielded an R value of 0.957, indicating a strong linear relationship, and an R^2 value of 0.915, suggesting our model explains 91.5% of the variance in the observed data. These results offer a more comprehensive view of the model's strong performance on the water objective function.

2.5.1 Sensitivity Analysis

Explainable Artificial Intelligence (XAI) is an emerging field aimed at enhancing the transparency of ML architectures to foster trust among users. XAI frameworks generate reports elucidating the operations of ML architectures, seeking to demystify their workings. Similarly, DL networks are often characterized as "black boxes" due to the challenge of interpreting their predictions. To address this challenge, Kassem *et al.* [48] employed class activation mapping and Grad-CAM [49] to offer visual explanations of predictions made by CNNs. Grad-CAM generates class activation maps that highlight crucial areas in input images used by the CNN to predict class labels.

Understanding the mechanics behind our end-to-end 3D CNN methodology is crucial for its trustworthiness, interpretability, and generalizability [10, 48]. Explaining how our methodology works improves transparency and enables stakeholders to make informed decisions. An in-depth explanation helps identify key geological uncertainties, understand simulation models under various conditions, and pinpoint any biases or weaknesses. Given the importance of precise simulator predictions in the oil industry, elucidating our approach enhances the confidence and reliability of our results. We will further explore this through a sensitivity analysis in the subsequent subsection.

Similar to Cirac *et al.* [10], we utilize Sobol sensitivity analysis [50], a mathematical technique, to assess the influence of both geological uncertainties and fluid properties (inputs) on replicating the simulator response (output). This method decomposes the output variance, attributing portions to each input and its interactions, facilitating the identification of significant variables and their interactions. Sobol is well-suited for this task because it can capture nonlinear effects and interactions between variables, therefore providing a comprehensive understanding of the behavior of our approach.

In the context of our study, we deal with a DL model that processes high-dimensional 3D features and concatenates them with scalar properties after encoder reduction. In the analysis, we only used the encoded representations of the 3D features (each reduced to 256 dimensions). This approach allowed us to focus on the influence of the most significant features extracted by the model rather than analyzing the millions of 3D data points. However, we identified that scalar inputs, not subjected to dimension reduction, exerted a disproportionate influence on the model outputs. We believe this was due to differences in scale. We standardize all model inputs, including encoded representations of 3D and scalar features. This step was essential to ensure that Sobol's analysis was not unduly influenced by the scale of the variables, allowing for a fair comparison of the relative importance of each input.

The Figures 2.8 represent the results of Sobol's analysis. The first image shows the first-order, while the second shows the total index. Both indicate the average contribution of five groups of characteristics — Porosity (Por), Rock Type (Rtp), Net to Gross (Ntg), Permeability (Perm), and Fluid properties — to the variance in the output of our methodology. The 3D geo-realizations presented a more significant contribution to the first-order index and the total, suggesting that variations in these characteristics have a considerable direct and interactive impact on our end-to-end 3D CNN output. On the other hand, Fluid properties show a reduced influence, indicating that they have a less

direct relationship with output variability.

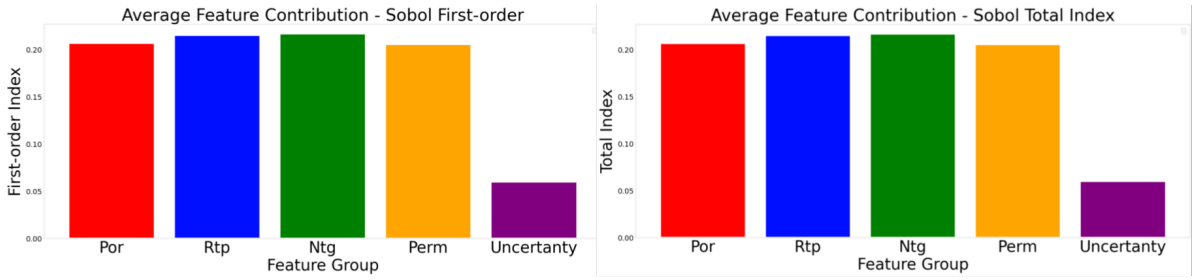


Figure 2.8: Sensitivity Analysis: The upper section delineates the first-order interactions, while the lower showcases the higher-order interactions.

2.6 Final Remarks

DA processes frequently require days or even weeks for a single reservoir application. Incorporating our end-to-end proxy model to substitute the simulator effectively tackled several challenges, including the treatment of 3D geological uncertainties and the replication of the simulator's response within oil reservoir management. Simultaneously, our proposed method diverges from prevailing practices seen in recent work [1] and [4], which involve training numerous proxies, ranging from 31 to 160 ANNs. In contrast, our study deviates from these conventional approaches by opting to train a single ANN. We present a novel end-to-end 3D CNN proxy approach that delegates any errors that might occur during various stages of 3D uncertainty processing to the architecture. By leveraging a 3D CNN for dimensionality reduction, we achieved an accurate vector representation, demonstrating a tangible advantage in handling complex, multidimensional geological uncertainties. The use of 3D geological realizations, including porosity and permeability, obtained from a live oil reservoir ($73 \times 38 \times 24$ dimensions) underscores the practical applicability of our approach in real-world scenarios.

A notable accomplishment of our study lies in the simultaneous handling of four distinct objective functions within a cutting-edge multi-objective framework. Our methodology overcomes the limitations of fragmented training and offers a comprehensive description of reservoir behavior. The successful emulation of the simulator's behavior across all fluid types underscores the versatility of our approach. Comparatively, our method demonstrates robustness when contrasted with state-of-the-art methodologies: Maschio *et al.* [1], Sudakov *et al.* [4], Ma *et al.* [41], and Kolajoobi *et al.* [6].

Chapter 3

A Multimodal Approach to Reconstructing Geological Uncertainties Using a Deep Residual Variational Autoencoder

This work was submitted to the Computers & Geosciences Journal, with permission from "Jeanfranco Farfan, Gabriel Cirac, Guilherme Daniel Avansi, Denis José Schiozer, and Anderson Rocha. Multimodal Approach to Reconstructing Geological Uncertainties Using a Deep Residual Variational Autoencoder, 2025."

In this chapter, we advance our end-to-end proxy beyond production prediction to address a second, complementary task: *reconstruction*. We introduce a unified framework that reconstructs original 3D geological uncertainties with high fidelity, integrating rock properties (e.g., porosity, permeability, rock type, net-to-gross) and reservoir-fluid descriptors. Accurate reconstruction is essential for DAUR workflows: it validates the dimensionality-reduction stage, restores simulator-compatible inputs, and preserves geologically plausible structure.

3.1 Introduction

Modern proxy models are commonly classified by Input/Output type: *image-to-image*, *image-to-sequence*, and *vector-to-sequence* [51, 52]. The Image-to-Image methods [52, 53, 54, 55, 56] employ images for both input and output. The methods use geological parameters as input (such as permeability, effective grids, cracks, and irregular boundaries) and flow field parameters as output (such as saturation or pressure maps). In contrast, Image-to-Sequence methods [41, 57] utilize geological image input data to generate output sequences often associated with the evolving dynamics of the reservoir over time. Lastly, the Vector-to-Sequence methods [24] use vectors as input data, representing reservoir properties such as relative permeability, density, and viscosity, to predict production data.

Contemporary proxy models employing a multimodal approach for predicting produc-

tion data typically tackle the problem through a two-step process [1, 4, 20, 24]: first, reducing the dimensionality of 3D geological images, and second, replicating the simulator response. Concerning the reduction of dimensionality in geological images, Ma *et al.* ([41, 57]) proposed two methods. In the initial method (Ma *et al.* [57]), they introduced a spatio-temporal convolution recurrent neural network (CRNN) surrogate model. This approach utilized a densely connected Convolutional Neural Network (CNN) to extract features from spatial parameters, followed by Spatial Pyramid Pooling (SPP) to convert spatial features into vector features. The second method (Ma *et al.* [41]) presented a Hybrid Recurrent Convolutional Network (HRCN) that employed a residual CNN to extract spatial features using Global Average Pooling (GAP) to derive vector features.

While these approaches forecast production from reduced representations [41, 57], the reduction itself discards variability that may be geologically meaningful. As a result, validating a proxy requires a *restoration* step that (i) recovers information lost during reduction, (ii) returns to simulator-compatible 3D inputs after DAUR, and (iii) checks that proxy outputs remain consistent with the original uncertainties. The restoration process also facilitates the return to the original 3D geological uncertainty after the DAUR process. This step is essential for ensuring that the proxy-generated data aligns with the original geological uncertainties, thereby maintaining the integrity of the analysis and interpretation. Additionally, reconstructing the 3D geological uncertainty is critical to ensure compatibility with the simulator's input requirements.

In the literature, Maschio *et al.* [1] employed Principal Component Analysis (PCA), incorporating the PCA inverse transform, to reconstruct geostatistical realizations after the Data Assimilation for Uncertainty Reduction (DAUR) process. Similarly, PCA was utilized by Ma *et al.* [24] to reduce the dimensionality of both 2D and 3D geological maps while generating reconstructions. However, PCA's objective is to capture maximum data variance using a reduced set of orthogonal components.

While effective for dimensionality reduction and simplifying data representation, this approach may result in the loss of crucial information, especially within the domain of 3D geological uncertainties, where detailed and nuanced information is vital for accurate reconstruction. While prior studies on geological uncertainties have demonstrated positive outcomes with PCA, such as those referenced by [1, 6], it is important to acknowledge that PCA operates under a linear assumption [30]. However, when dealing with 3D geological realizations, nonlinear characteristics become significant, thus rendering PCA less advisable for our domain. Additionally, we diverge from the approach of using independent PCA modules, as we emphasize an end-to-end framework. A single model is likely to be more robust than a multi-stage model [26], where errors from each component can compound.

Our approach differs from the literature, where the reconstruction of 3D geological uncertainty typically involves a single geological uncertainty as input [3, 4, 6, 20]. Our design handles four geological uncertainties in the same end-to-end proxy, capturing and reconstructing patterns, structures, and variations inherent to each geological uncertainty, while also accounting for the intricate relationships and interactions among them.

We leverage a 3D DRVAE to reconstruct all 3D geological uncertainties from a latent space. The reduction of 3D geological dimensionality generates the latent space, includ-

ing porosity, permeability, rock type, net-to-gross, and a vector of reservoir properties. Our end-to-end proxy solution integrates three distinct yet interconnected components: an encoder, a decoder, and residual connections. The encoder maps 3D geological uncertainties into a lower-dimensional latent space. The decoder reconstructs samples that closely resemble the original 3D uncertainties. The residual connections enable a more direct flow of gradients through the network during backpropagation. These connections preserve information from preceding layers and establish shortcuts for efficient gradient flow.

The contributions of this chapter are summarized as follows:

1. The proposed surrogate proxy reduces the computational time of the DAUR process from days to minutes.

2. Our framework integrates compression, assimilation, and reconstruction, preserving the structural integrity of 3D geological data.

3. The proposed 3D DRVAE end-to-end proxy outperforms traditional PCA, reconstructing geological features with high spatial fidelity.

4. The high-fidelity reconstructions enhance the DAUR process and help mitigate reservoir management risks.

3.2 Related Work

Researchers have applied dimensionality reduction to multimodal reservoir data to improve computational efficiency [57, 41]. However, since real-world reservoir models contain vast numbers of grid cells [11], this process risks substantial information loss. Our work is distinguished by its use of geological reconstruction to validate the fidelity of the dimensionality reduction, ensuring the reduced representation preserves the original semantics.

Previous studies have used Principal Component Analysis (PCA) for dimensionality reduction and reconstruction of geological maps [1, 24]. While PCA can be effective [6, 1], it is a linear method [30]. This linearity makes PCA less suitable for 3D geostatistical realizations, where nonlinear characteristics are significant and their preservation is vital for accurate reconstruction.

Deep learning methods offer advantages in handling such nonlinear relationships [58, 59]. Within Image-to-Image proxy approaches, convolutional encoder-decoder architectures, such as Autoencoders (AEs) [60], are commonly used to transform geological parameters into a low-dimensional representation and then generate output maps [51, 56]. For example, a Deep Convolutional Autoencoder (DCAE) was used to reduce data dimensionality for monitoring CO_2 sequestration [61], and a temporal-3D CNN with an AE design was employed for simulator surrogacy in subsurface flow problems [62]. However, AEs like the DenseNet-based encoder-decoder (DenseED) [55] do not model the latent space probabilistically, limiting the diversity and realism of reconstructed samples.

Variational Autoencoders (VAEs) [63] address this by modeling the latent space as a probabilistic distribution, enabling the generation of more diverse and realistic samples.

VAEs have been applied in reservoir characterization, often with Ensemble Smoother with Multiple Data Assimilation (ES-MDA) or other variants of this optimization strategy [64], to parameterize and reconstruct geological uncertainties [65, 66]. These studies, however, highlighted two key challenges: a noted decrease in reconstruction accuracy for 3D geological uncertainties [66], and the potential for vanishing gradients in deep VAE architectures [67].

The vanishing gradient problem can be mitigated by incorporating residual connections [45], a technique used in architectures like Unet [68] and recent geoscience applications to improve gradient flow and capture complex data patterns [69]. Inspired by these advances [70, 71], we introduce residual connections into our VAE design to create a novel network, the 3D Deep Residual Variational Autoencoder (DRVAE). In contrast to prior studies that often rely on linear PCA or handle a single geological uncertainty, our approach introduces a 3D Deep Residual Variational Autoencoder (DRVAE) capable of reconstructing multiple uncertainties simultaneously. By incorporating residual connections, the proxy overcomes vanishing gradient issues and captures complex spatial patterns more effectively. This work uniquelyreconst ensures high-fidelity reconstruction across four geological properties, setting a new standard for preserving 3D uncertainty structure in reservoir modeling workflows.

3.3 Proposed Methodology

We propose a 3D Deep Residual Variational Autoencoder (DRVAE) to reconstruct 3D geological uncertainties. Reconstruction serves two roles for DAUR: (i) it validates dimensionality reduction by comparing originals and reconstructions, and (ii) it produces simulator-ready volumes that visualize how uncertainty evolves through assimilation.

The methodology is organized in two stages. Stage 1 introduces the 3D DRVAE reconstruction module, which integrates an encoder, a decoder, and residual connections, as illustrated in Figure 3.1. Stage 2 evaluates the 3D reconstructions against the original 3D uncertainties to assess the technique's robustness.

The encoder maps the concatenated 3D uncertainties into a lower-dimensional latent space, from which the VAE generates new samples based on learned mean and variance parameters. The decoder then reconstructs samples that closely mirror the original 3D uncertainties. Residual connections, shown as dashed lines in the figure, link the encoder and decoder to fine-tune the reconstructions, ensuring alignment with the initial geological data.

The proxy utilizes five geological uncertainties: four 3D geological uncertainties (represented as grid blocks) and one vector of reservoir properties. Each 3D geological uncertainty undergoes quantile normalization [42], a critical preprocessing step for handling non-linear data and adjusting it to a format more suitable for proxy training.

3.3.1 Encoder

Following preprocessing, each 3D geological uncertainty is cropped and padded to adjust its dimensions from $73 \times 38 \times 24$ to $72 \times 48 \times 24$. This adjustment creates dimen-

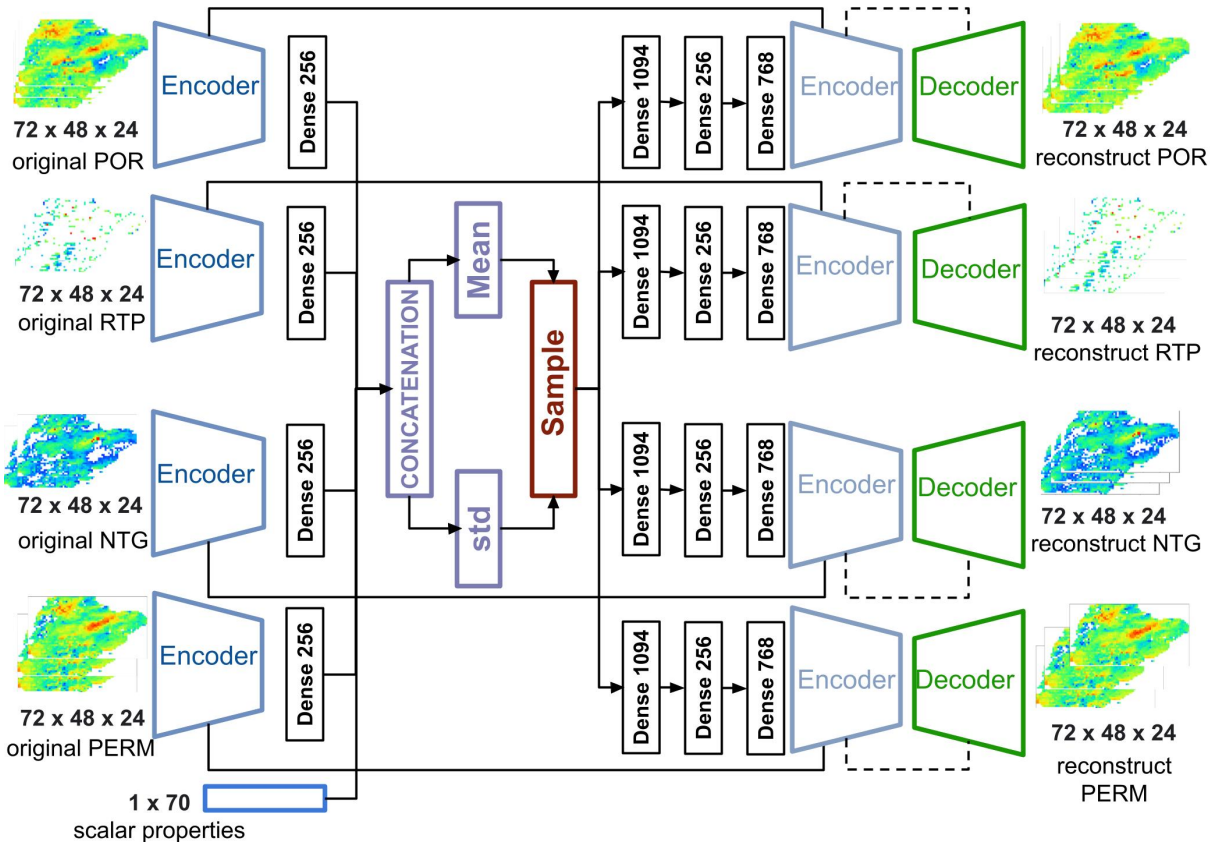


Figure 3.1: The 3D DRVAE reconstruction module showcasing the interconnected components of the encoder, decoder, and residual connections. This framework facilitates the reconstruction of 3D geological uncertainties, validating the robustness of the dimensionality reduction technique and enabling the visualization of uncertainty reduction within the reservoir.

sions that are more computationally efficient for the encoder's convolution and pooling operations, allowing for their uniform application across the data matrix while preserving spatial structure [72, 58]. The preprocessed data is then fed into the dimensionality reduction module (Figure 3.2 (a)).

As shown in Figure 3.2 (b), each 3D uncertainty is processed through hierarchical strata using 3D encoder blocks and max-pooling layers. An encoder block encapsulates convolutional layers, normalization layers, and activation functions to learn data representations. It also incorporates skip connections [45] (CONV1 3D and CONV2 3D) to capture identity information from the input data for use with the residual connections (Section 3.3.3). This reduction process yields a 256-dimensional feature vector for each of the four 3D uncertainties. These vectors are concatenated with the scalar properties to create a 1,094-dimensional feature set.

Table 3.1 summarizes the encoder for each 3D geological uncertainty. We use *TEncB*, *Pool*, and *Dense* for a 3D encoder block, a 3D max-pooling layer, and a dense layer, respectively. Each uncertainty is expanded to a 4D tensor of size $72 \times 48 \times 24 \times 1$, where 72, 48, and 24 are the grid-block counts along x, y, and z. A *TEncB* contains *Conv–Batch–ReLU–Dropout–Conv–Batch–ReLU*, with L2 regularization on the convolu-

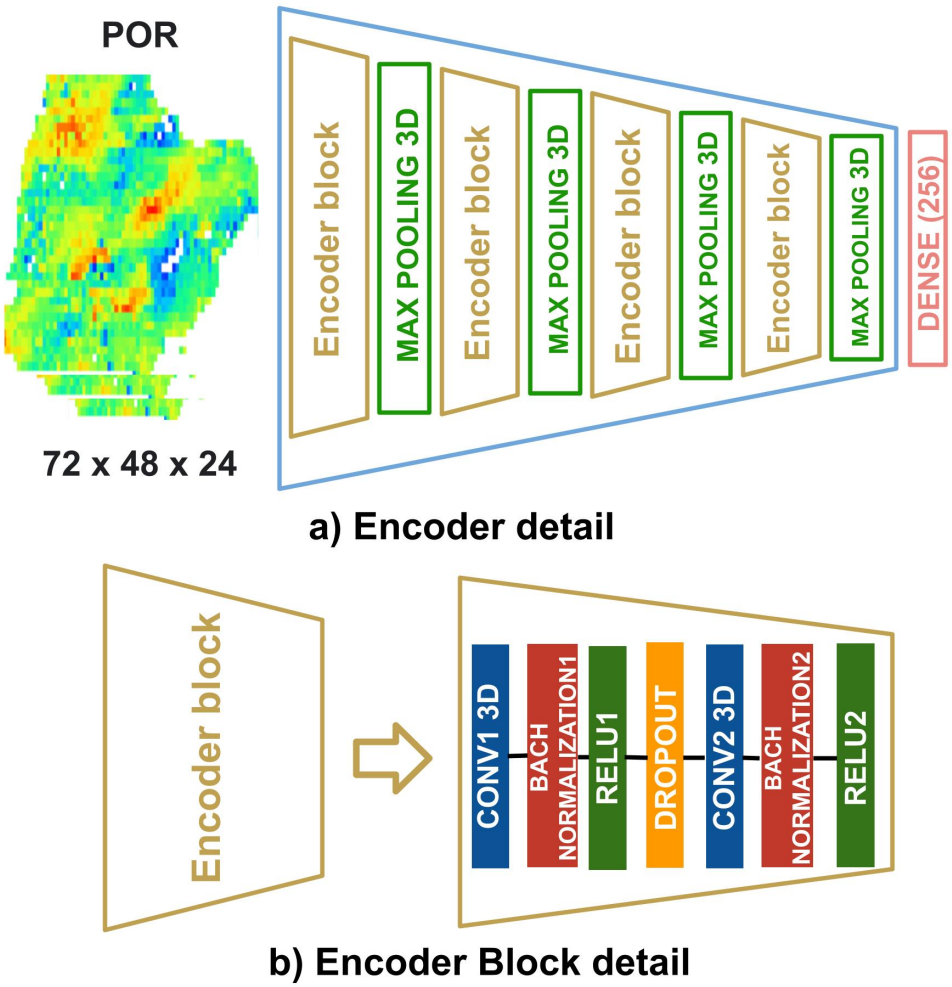


Figure 3.2: Diagram illustrating the 3D Encoder architecture. Figure (a) shows the overall structure of the encoder, including layers of encoder blocks and 3D pooling, leading to a dense layer. Figure (b) provides a detailed breakdown of the encoder block, highlighting its hierarchical processing through convolutional layers, batch normalization, ReLU activation, and dropout. These components work together to handle complex geological data, preserve high-frequency details, and efficiently reduce dimensionality.

tional weights. After each block, a 3D max-pooling layer and Dropout are applied. The hierarchy proceeds to *Flatten* and a *Dense* layer that yields a 256-D vector. The resulting encodings, E_{POR}, E_{PERM}, E_{RTP}, E_{NTG}, and E_{Scalar} (porosity, permeability, rock type, net-to-gross, and scalar properties), are concatenated to form the latent space.

In the framework's l^{th} layer, the previous output E_{l-1} (with E_0 denoting the input geological uncertainty) is mapped to an auxiliary representation E_{laux1}. Equation 3.1 summarizes this step, combining L2 regularization, batch normalization, and ReLU:

$$E_{laux1} = \text{ReLU}\big(\text{Batch}(\text{L2}(W_l * E_{l-1} + b_l))\big) \tag{3.1}$$

Here, W_l and W_{laux2} are $3 \times 3 \times 3$ kernels and b_l, b_{laux2} are biases. Dropout is then applied to E_{laux1} to obtain E_{laux2} (Equation 3.2) to mitigate overfitting:

$$E_{laux2} = \text{Dropout}(E_{laux1}) \tag{3.2}$$

Table 3.1: 3D CNN encoder architecture details.

Input	**Output Shape**
Input 1 (Porosity)	**(72, 48, 24, 1)**
TEncB1, 16 filters, $3 \times 3 \times 3$	(72, 48, 24, 16)
Pool1, $2 \times 2 \times 2$; Dropout	(36, 24, 12, 16)
TEncB2, 32 filters, $3 \times 3 \times 3$	(36, 24, 12, 32)
Pool2, $2 \times 2 \times 2$; Dropout	(18, 12, 6, 32)
TEncB3, 64 filters, $3 \times 3 \times 3$	(18, 12, 6, 64)
Pool3, $2 \times 2 \times 2$; Dropout	(9, 6, 3, 64)
TEncB4, 128 filters, $3 \times 3 \times 3$	(9, 6, 3, 128)
Pool4, $3 \times 3 \times 3$	(3, 2, 1, 128)
Flatten	(768)
E_{POR}	(256)
E_{PERM}	(256)
E_{RTP}	(256)
E_{NTG}	(256)
E_{Scalar}	(70)
$E_{POR} \oplus E_{PERM} \oplus E_{RTP} \oplus E_{NTG} \oplus E_{Scalar}$	(1,094)

Finally, E_{laux2} is transformed into the layer output E_l through a second pass of L2 regularization, batch normalization, and ReLU (Equation 3.3), ensuring consistency across layers:

$$E_l = \mathrm{ReLU}\big(\mathrm{Batch}(\mathrm{L2}(W_{laux2} * E_{laux2} + b_{laux2}))\big) \tag{3.3}$$

Within each encoder block, *L2* regularization and *Dropout* curb overfitting, *Batch* normalization stabilizes activations, and *ReLU* introduces nonlinearity.

The architecture includes 3D max-pooling layers with $2 \times 2 \times 2$ and $3 \times 3 \times 3$ kernels, each followed by Dropout, to reduce feature-map dimensions. As inputs progress through the *TEncB–Pool–Dropout* hierarchy, the spatial sizes (d_i) decrease while the number of feature maps increases. A *Flatten* layer compacts the volume, and a *Dense* layer forms a 256-dimensional feature vector for each uncertainty.

The four 256-dimensional vectors are then concatenated with the scalar properties, yielding 1,094 features:

$$E_{POR} \oplus E_{PERM} \oplus E_{RTP} \oplus E_{NTG} \oplus E_{Scalar} \tag{3.4}$$

where $\oplus$ denotes concatenation of the reduced 3D uncertainties (E_{POR}, E_{PERM}, E_{NTG}, E_{RTP}) and the reservoir scalar properties (E_{Scalar}).

3.3.2 Decoder

The decoder uses the concatenated array as the latent space within the VAE framework. This distribution is updated using the mean and variance (Figure 3.1), introducing stochasticity that enables diverse output generation during training. The sampled data

is then processed hierarchically through 3D transpose CNNs, 3D upscaling, and decoder blocks to reconstruct the 3D geological uncertainty, as shown in Figure 3.3 (a).

The decoder block (Figure 3.3 (b)) employs transpose convolutional layers, batch normalization, and activation functions to reconstruct the 3D uncertainty data. It integrates two residual connections, $ADD1$ and $ADD2$, from the encoder block to mitigate the vanishing gradient issue (detailed in subsection 3.3.3).

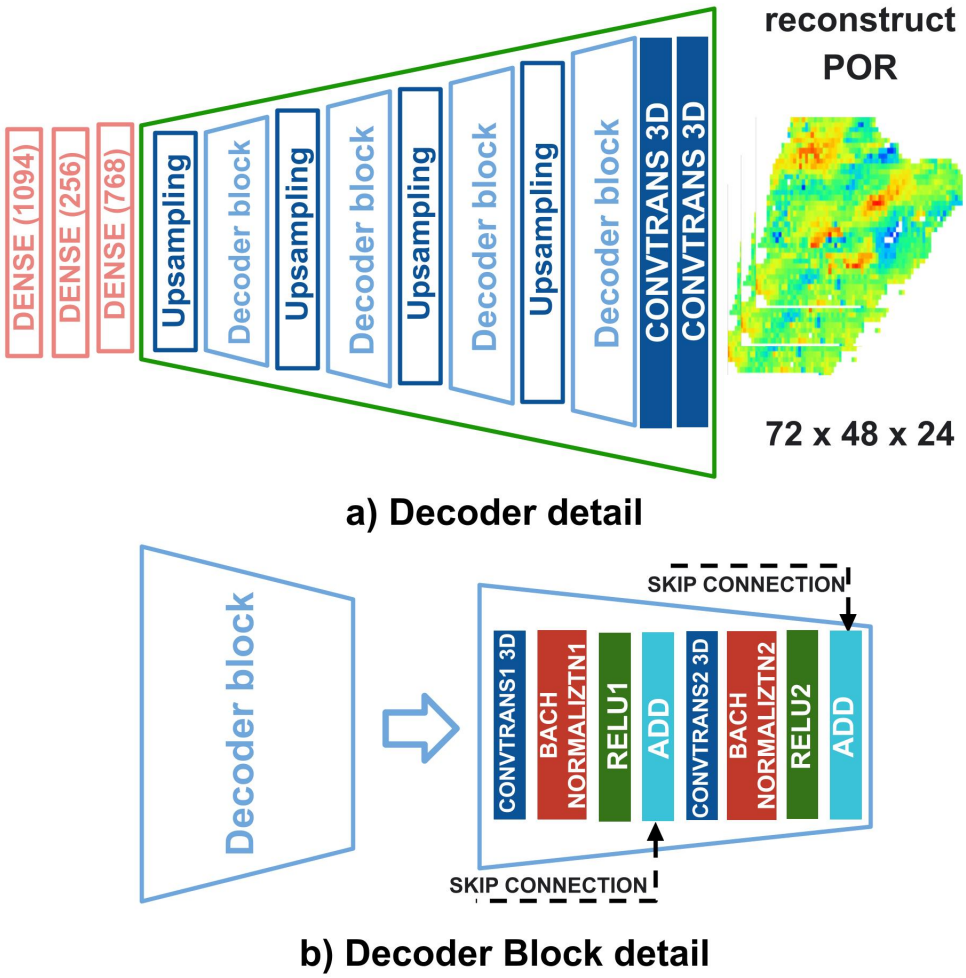


Figure 3.3: The diagram depicts the 3D Decoder architecture. Figure (a) shows the details of the decoder, including dense layers, decoder blocks, 3D upsampling layers, and 3D transpose convolutions. Figure (b) provides a detailed breakdown of the decoder block, highlighting its hierarchical processing through transpose convolutional layers, batch normalization, and ReLU activation, while emphasizing skip connections from the encoder block linked to the outputs of ReLU1 and ReLU2. The decoder architecture reconstructs detailed geological features from the latent sample through a series of sequential operations.

In the encoder, each 3-D uncertainty enters as a $72 \times 48 \times 24$ voxel block and passes through four strata where a binary reduction of spatial resolution ($72 \rightarrow 36 \rightarrow 18 \rightarrow 9 \rightarrow 3$) is matched by a doubling of channel depth ($1 \rightarrow 16 \rightarrow 32 \rightarrow 64 \rightarrow 128$). The decoder mirrors this pathway in reverse: a variational re-sampling of the latent vector is successively expanded, transposing convolutions that double spatial extent at each step

$(3 \rightarrow 6 \rightarrow 12 \rightarrow 24 \rightarrow 48 \rightarrow 72)$ while reducing channel depth, culminating in four fully restored geological volumes. Residual links from the encoder are injected at homologous depths, so the network learns corrections rather than entire reconstructions. This strategy stabilises gradients and preserves fine-scale geological structure. This compact yet expressive hierarchy, therefore, captures, compresses, and faithfully regenerates complex 3-D uncertainty fields within a single, coherent framework.

Table 3.2 presents the decoder architecture, where *TDecB*, *UP*, *TConv*, and *Dense* denote a 3D decoder block, a 3D upscaling layer, a 3D transpose convolutional layer, and a dense layer, respectively. The variational layer (VAR) processes the concatenated array ($\oplus$) by applying two dense layers to compute the mean and log-variance of the latent space. Reparameterization then generates a sampled representation by adding random noise to these parameters. The sequence consists of a Dropout layer, two Dense layers, and a reshaping step into a 4D tensor with dimensions $3 \times 2 \times 1 \times 128$, where 3, 2, and 1 represent the grid block counts along the x, y, and z axes, and 128 denotes the number of feature maps. Each 3D decoder block is composed of a transpose convolution (`TConv`), *Batch*, a *ReLU*, and an addition (*Add*) operation. The sequence includes two 3D transpose convolutional layers with intermediate L2 regularization, batch normalization, ReLU activation, and element-wise additions with the outputs from the encoder's *CONV2* and *CONV1*.

After completing the 3D decoder block, the architecture applies a 3D upscaling layer and a Dropout layer, progressively enlarging the dimensions to 72, 48, and 24 along the x, y, and z directions. This sequence is followed by two additional 3D transpose convolutional layers, with the final layer's filters capturing the key features required for 3D geological reconstruction. The outputs correspond to the reconstructions of the four 3D geological properties.

Equation 3.5 represents the fundamental components of a VAE. The VAE framework encodes the input 3D geological uncertainty (x) into a latent space (z), represented by the concatenated vector ($\oplus$). From this latent space, the mean (μ) and log variance ($\log(\sigma^2)$) are computed. A new sample ($\hat{z}$) is then generated using reparameterization, which integrates random noise (ϵ) from a normal distribution, $\mathcal{N}(0, 1)$.

$$\begin{aligned} z &= \text{Encoder}(x) \\ \mu &= \text{Mean}(z) \\ \sigma &= \log(\sigma^2(z)) \\ \epsilon &\sim \mathcal{N}(0, 1) \\ \hat{z} &= \mu + \sigma \cdot \epsilon \end{aligned} \tag{3.5}$$

The architecture of the CNN decoder designed for 3D geological reconstruction is predicated on sequential operations. The decoder architecture reconstructs detailed geological features from this latent sample through a series of sequential operations. The process begins with feature map enhancement, as described in Equation 3.6. This step combines learned features with those from the encoding phase.

$$D_{laux} = \text{ADD}(\text{ReLU}[\text{Batch}\left(\text{L2}(W_l^T * D_{l-1} + b_l)\right)], E_l) \tag{3.6}$$

Table 3.2: 3D CNN decoder architecture details.

input	**output size**
$E_{POR} \oplus E_{PERM} \oplus E_{RTP} \oplus E_{NTG} \oplus E_{Scalar}$	**(1,094)**
VAE, Dropout	(1,094)
Dense1, L2, ReLU	(1,094)
Dense2, L2, ReLU, Dropout	(256)
Dense3, L2, ReLU	(768)
Reshape, Dropout	(3, 2, 1, 128)
UP1, upsampling of size $2 \times 2 \times 2$	(6, 4, 2, 128)
TDecB1, 128 filters of size $3 \times 3 \times 3$	(6, 4, 2, 128)
UP2, upsampling of size $2 \times 2 \times 2$	(12, 8, 4, 128)
TDecB2, 64 filters of size $3 \times 3 \times 3$	(12, 8, 4, 64)
UP3, upsampling of size $3 \times 3 \times 3$	(36, 24, 12, 64)
TDecB3, 32 filters of size $3 \times 3 \times 3$	(36, 24, 12, 32)
UP4, upsampling of size $2 \times 2 \times 2$	(72, 48, 24, 32)
TDecB4, 16 filters of size $3 \times 3 \times 3$	(72, 48, 24, 16)
TConv1	(72, 48, 24, 8)
TConv2	(72, 48, 24, 1)
POR_R	(72, 48, 24, 1)
$PERM_R$	(72, 48, 24, 1)
NTG_R	(72, 48, 24, 1)
RTP_R	(72, 48, 24, 1)

In this operation, D_{l-1} is the previous layer's feature map, W_l^T is a 3D transpose convolution kernel, and E_l is the feature map from the corresponding encoder layer. The process involves 3D transpose convolution, L2 regularization, batch normalization, and ReLU activation. A subsequent refinement step, shown in Equation 3.7, further improves the output by integrating earlier encoder features (E_{laux1}).

$$D_l = \text{ADD}(\text{ReLU}[\text{Batch}(\text{L2}(W_{laux2}^T * D_{laux} + b_{laux2}))], E_{laux1}) \tag{3.7}$$

Here, Equation 3.7 extends the transformation process, employing another set of transpose convolution operations (W_{laux2}^T, b_{laux2}) along with the reintroduction of earlier encoder features (E_{laux1}) to further refine the output. This step underscores the synthesis of detailed textural and structural attributes that are essential for accurate 3D reconstruction.

The decoder's architecture, through Equations 3.6 and 3.7, leverages convolutional operations for feature extraction and spatial translation, enhanced by regularization and normalization techniques to counteract overfitting and improve training stability. Moreover, the encoder-derived features via residual connections exemplify a strategy to preserve and augment critical geological information throughout the decoding process, ensuring that the reconstructed outputs faithfully represent the complexity and nuance of 3D geological formations.

In this study, we utilize 3D upsampling layers with $3 \times 3 \times 3$ and $2 \times 2 \times 2$ pooling

kernels, implementing a “same" padding strategy. Following the dense layers with 1,094, 256, and 768 neurons, we apply a reshape layer to handle multidimensional data. The data then undergoes recursive processing through the UP-DROPOUT-TDecB layers until it reaches three transpose convolutional layers, with the final layer utilizing an activation function to output the 3D reconstructed geological uncertainty. This process happens four times until it reaches the shape of the 3D geological uncertainties. Subsequently, two 3D transpose convolutional layers are employed. The final filter in the second 3D transposition convolutional layer captures the essential features for 3D geological reconstruction.

3.3.3 Residual Connection

Residual connections, incorporated into our DRVAE network depicted in Figure 3.4, aim to mitigate the issue of vanishing/exploding gradients [45]. We employ shortcut connections that involve identity mapping, which links the encoding layer to its corresponding mirrored decoding layer, enabling the implementation of residual networks in a nested manner, extending from the outermost to the innermost layers, as depicted in Figure 3.4. We propose this methodology inspired by the work of Unet [68], Segnet [73], and Wang *et al.* [71].

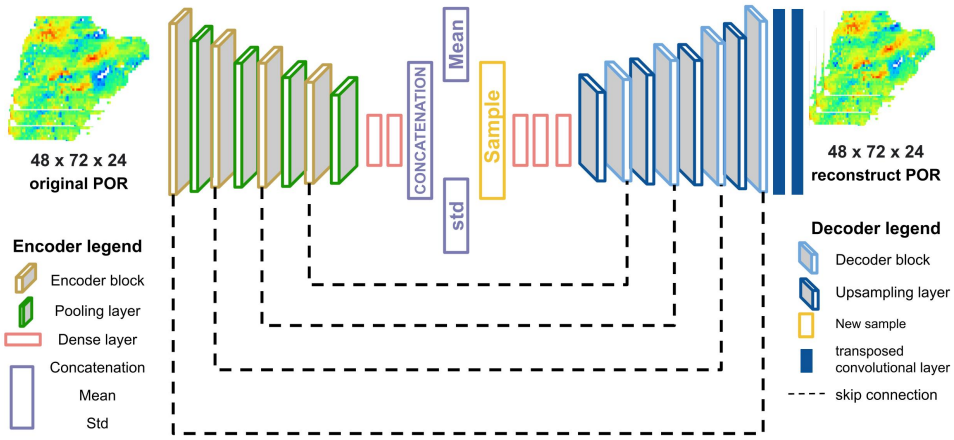


Figure 3.4: Shortcut connections between the encoder and decoder blocks. Specifically, the first encoder block is connected to the last decoder block, the second encoder block is connected to the penultimate decoder block, and the third encoder block is connected to the antepenultimate decoder block. Finally, the last encoder block is connected to the fourth-to-last decoder block, ultimately reaching the latent space. These connections aim to mitigate the issue of vanishing and exploding gradients.

In Figure 3.5, we illustrate the encoder block in yellow-brown and the decoder block in blue, providing detailed views in the right part. These two blocks collaborate to establish the residual connection. For the encoder, we selected two 3D convolutional layers, namely CONV1 3D and CONV2 3D, which are interconnected through the residual connection, forming a skip connection. This connection involves element-wise addition with RELU3 and RELU4 in the decoder block.

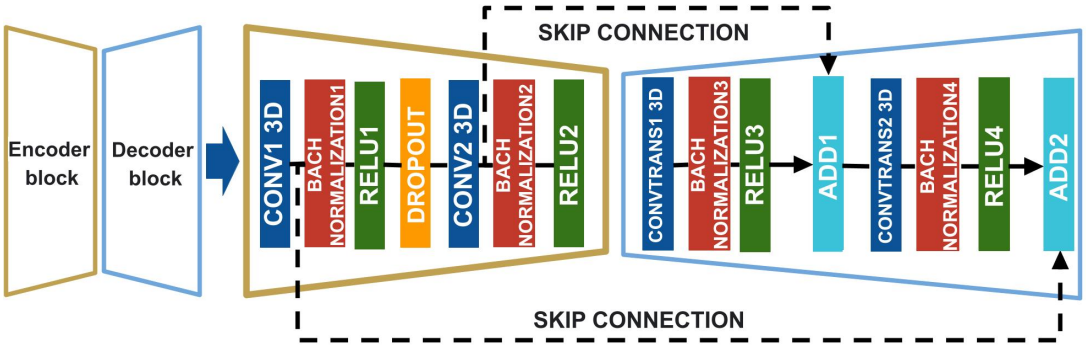


Figure 3.5: Detailed view of the residual connection. From left to right, the encoder block is shown in yellow, and the decoder block is shown in blue. The interaction between these two blocks is facilitated by two skip connections from the encoder to the decoder. First, the output of the first convolutional layer in the encoder (CONV1 3D) is connected to the output of the activation function in the decoder (RELU3). Second, the output of the second convolutional layer in the encoder (CONV2 3D) is connected to the output of the second activation function in the decoder (RELU4).

3.4 Results and Discussion

We illustrate and examine the outcomes from the comparative analysis between the baseline approach, an ablation study, and the optimal configuration unearthed through the ablation study. For both the baseline and Ablation Study, our analysis focuses on particular layers along the Z-axis, specifically visualizing selected layers and comparing their original and reconstructed uncertainties. After identifying the best performance within an approach, we further investigate it by analyzing various metrics. This allows to assess different elements related to reconstruction accuracy and fidelity in greater detail.

3.4.1 Loss Function and Metrics

In Equation 3.8, the DRVAE loss ($\mathcal{L}_{\text{DRVAE}}$) is calculated as the mean loss over a batch of data samples, with N representing the batch size. Each MSE term quantifies the match between original and reconstructed data for a specific property, for instance, $PERM_i$ and $PERMr_i$. This loss formulation aims to guide the training toward accurate reconstruction, which is crucial for effectively handling 3D geological uncertainty.

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} \text{MSE}(POR_i, \text{PORr}_i)+ \\ \text{MSE}(RTP_i, \text{RTPr}_i)+ \\ \text{MSE}(NTG_i, \text{NTGr}_i)+ \\ \text{MSE}(PERM_i, \text{PERMr}_i)+ \\ \text{KL}(\mathcal{N}(\mu_i, \sigma_i^2)||\mathcal{N}(0,1)) \end{pmatrix} \tag{3.8}$$

In this context, y denotes the original 3D geological uncertainty, while $\hat{y}$ signifies the reconstructed 3D geological uncertainty. The variable n corresponds to the total number of simulation models. Similarly, the Kullback–Leibler (KL) divergence can be illustrated

using Eq. 3.9.

$$\mathrm{KL}\big(\mathcal{N}(\mu, \sigma^2) \,\|\, \mathcal{N}(0,1)\big) = -\frac{1}{2}\sum_{i=1}^{n}\big(1 + \log(\sigma_i^2) - \mu_i^2 - \sigma_i^2\big) \tag{3.9}$$

The expression $\mathrm{KL}(\mathcal{N}(\mu, \sigma^2)||\mathcal{N}(0,1))$ denotes the KL divergence, which measures the disparity between a normal distribution characterized by mean μ and variance σ^2, and the standard normal distribution. It quantifies the dissimilarity between these two distributions. In the formula, $\mathcal{N}(\mu, \sigma^2)$ represents a normal distribution with mean μ and variance σ^2, while $\mathcal{N}(0,1)$ signifies a standard normal distribution with a mean of 0 and variance of 1. The variable n corresponds to the number of dimensions in the distribution, while μ represents the mean of the distribution in dimension i, and σ^2 represents the variance in dimension i.

We assess reconstruction quality using eight metrics: the Correlation Coefficient (R) for similarity, Mean Squared Error (MSE) for discrepancy, Histograms for distributional similarity, the Kernel Density Estimation (KDE) provides a smooth, continuous estimate of the probability density function of the data, the Structural Similarity Index (SSIM) for spatial feature preservation, Fréchet Inception Distance (FID), a metric used to evaluate the quality of 3D geological uncertainty generated by generative proxy, and Peak Signal-to-Noise Ratio (PSNR) for evaluating the clarity and sharpness of the reconstructed image, and Visual Comparison (VC) for overall visual agreement with the original uncertainty. Each metric evaluates a distinct reconstruction accuracy and fidelity aspect, which offers a comprehensive view of the proxy's performance.

3.4.2 Implementation and Training

The DRVAE network was implemented using the TensorFlow and Keras libraries [74]. The 200 simulation models from the dataset were randomly partitioned into a training set (70%), a validation set (10%), and a testing set (20%).

The end-to-end proxy was trained by minimizing the combined MSE and KL-divergence loss function (Equation 3.8) using the Adam optimizer with an initial learning rate of 0.0001. Due to the limited sample size, a batch size of 8 was used. The training was run for 250 iterations, with an early stopping mechanism to prevent overfitting.

3.4.3 Baseline

A baseline for comparison was established by evaluating the Principal Component Analysis (PCA) reconstruction technique from Maschio *et al.* [1], using 140 principal components on the test dataset. A visual analysis of specific layers for four uncertainties (porosity, net-to-gross, rock type, and permeability), as shown in Figure 3.6, reveals significant discrepancies between the original and reconstructed data.

Figure 3.6 illustrates that the reconstructed uncertainties only loosely resemble the original ones regarding porosity, net-to-gross ratio, rock type, and permeability, revealing several differences. This discrepancy suggests that the reconstruction process inadequately captures the spatial patterns and distributions of geological uncertainties, indicating low

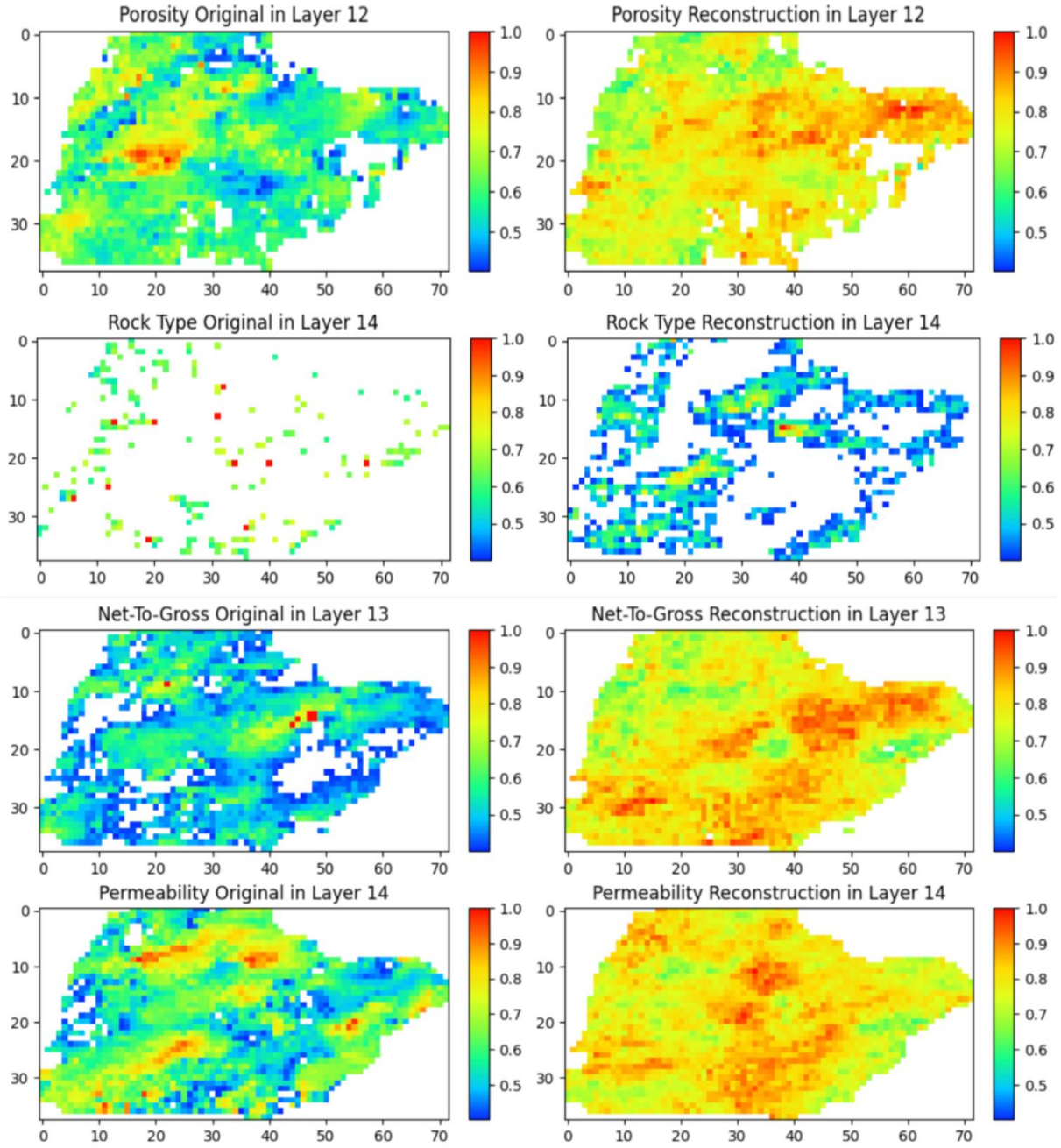


Figure 3.6: Visual analysis of the PCA technique proposed in Maschio et al.'s work. Utilizing 140 principal components, Figure 3.6 presents findings across four uncertainties. Each row comprises four images, with pairs corresponding to specific uncertainties. The first row depicts layer 12 (porosity) and layer 13 (net-to-gross uncertainty), while the second row represents layer 14 (type of rock uncertainty) and layer 14 (permeability uncertainty). Original uncertainties are displayed on the left, while reconstructions are on the right.

fidelity and precision in the reconstruction effort. Specifically, the outputs for porosity, net-to-gross ratio, and permeability exhibit similar patterns, with high heterogeneity (depicted in red) observed in several parts of the figure. In contrast, low heterogeneity (depicted in blue) is absent compared to the original image. Concerning rock type, the

reconstruction introduces several components that do not align with the original uncertainty. The issue arises because PCA aims to maximize data variance by using a set of orthogonal components. However, in this process, it can overlook critical information related to spatial distributions and subtle relationships within geological uncertainties. Consequently, reconstructions based on PCA often fail to maintain the fine-grained details necessary for accurate 3D geological uncertainties, resulting in the loss of essential features during the reconstruction process.

3.4.4 Ablation Study for 3D Geological Reconstruction

Based on the baseline results, this method clearly needs to be more accurate when applied with a limited number of simulation models for the reconstruction task. Specifically, a lower visual similarity between the original and reconstructed uncertainties implies that the reconstructed data only loosely resembles the original, with several discrepancies. This disparity implies that the reconstruction process using PCA proposed by Maschio *et al.* [1] fails to accurately capture the spatial patterns and distributions of geological uncertainties, thus demonstrating low fidelity and precision in the reconstruction effort. Therefore, adopting a deep learning approach, such as our DRVAE, becomes essential to construct a more effective system for reconstructing 3D geological uncertainties.

An ablation study involves evaluating the performance of a system by systematically removing one or more of its components to assess the relative contribution of the omitted elements [75]. In our investigation of the contribution of each component to the 3D geological reconstruction, we conducted an ablation study, with the results presented in Figure 3.7. For consistency, we selected 2D slices extracted from the same geological uncertainty and at the same depth. Specifically, our analysis focuses on the twelfth layer of porosity. To facilitate interpretation, we performed a side-by-side comparison between the reconstructed outputs and the original uncertainty (Figure 3.7a illustrates the original 3D porosity uncertainty), enabling specialists to assess the preservation or distortion of spatial patterns. To minimize bias, we applied a uniform color scale across both the reconstructed and original uncertainty.

The study demonstrated that a standard VAE architecture (Figure 3.7b) was insufficient, showing a limited ability to represent the reservoir's shape and heterogeneity. The introduction of residual skip connections in the DRVAE (Figure 3.7c) significantly improved performance, enabling the proxy to discern general features such as heterogeneity and geobody direction. However, this configuration struggled with an opaque visual quality and had difficulty rendering low-heterogeneity regions. To resolve this, a grid search was performed to optimize the L2 regularization parameter. An optimal value of 0.0001 was found to effectively enhance the representation of these low-heterogeneity regions, leading to the final, superior end-to-end proxy configuration (Figure 3.7).

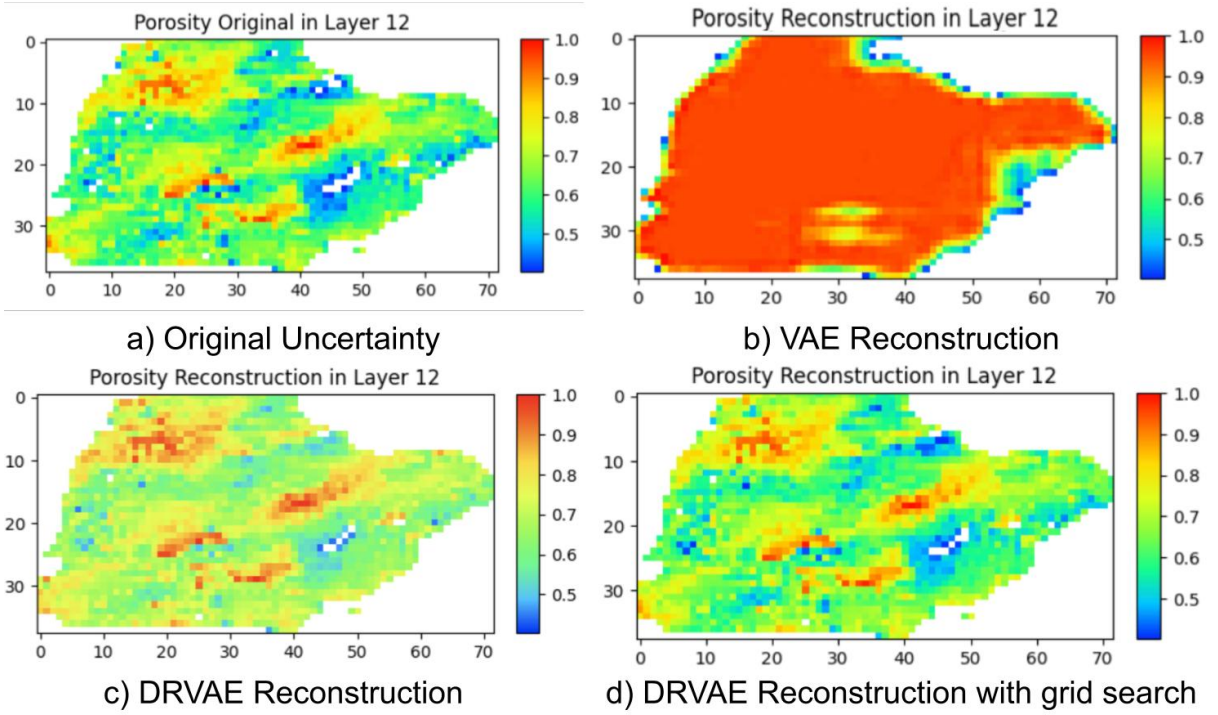


Figure 3.7: Ablation study on 3D geological reconstruction, assessing component impact. Panel a) shows the original 3D porosity uncertainty, b) VAE approach, c) DRVAE architecture, and d) DRVAE with grid search on L2 regularization.

3.4.5 Comparing the Baseline with Our 3D Deep Residual Variational Autoencoder

To effectively demonstrate the superiority of our proposed DRVAE method, we rigorously compare it against a conventional baseline approach. Unlike the DRVAE, the baseline lacks a residual variational encoding strategy. The results, presented in Table 3.3, reveal that the baseline consistently falls short, exhibiting lower correlation coefficients (R), higher reconstruction errors (MSE), and diminished structural similarity (SSIM) for all geological features. This is particularly pronounced in complex categorical attributes such as rock type. Moreover, the baseline's performance in generative fidelity (FID) and visual clarity (PSNR) is markedly inferior, highlighting its inability to capture high-level geological patterns and maintain spatial coherence. These comparisons not only underscore the limitations of the conventional approach but also illustrate the advantages of the DRVAE framework in the geological reconstruction task.

In Table 3.3, the results of our approach are presented for both the test and validation sets. Specifically, the study evaluates the performance of the DRVAE method across the reconstruction task of four 3D geological uncertainties, including porosity, net-to-gross ratio, permeability, and rock type. To assess the effectiveness of our approach, we utilized eight metrics: R, MSE, SSIM, FID, PSNR, Histograms, KDE, and a VC. The first three metrics are displayed in Table 3.3. These metrics offer insights into correlation, accuracy, fidelity, and structural preservation between the reconstructed and original uncertainties.

Our evaluation relies on five complementary criteria: R gauges linear correlation, MSE measures absolute error, SSIM quantifies structural fidelity, FID assesses distributional

Table 3.3: Performance Evaluation of the DRVAE Method for 3D Geological Uncertainties.

Approach/Set	**Feature**	**R**	**MSE**	**SSIM**	**FID**	**PSNR**
Our Approach/Test	Porosity	**0.995**	**0.081**	**0.895**	**2.444**	**10.916**
	Rock type	**0.984**	**0.076**	**0.908**	**2.561**	**11.178**
	Net-to-gross	**0.999**	**0.020**	**0.970**	**0.587**	**16.980**
	Permeability	**0.998**	**0.024**	**0.961**	**0.754**	**16.123**
Baseline/Test	Porosity	0.972	0.399	0.695	9.624	3.985
	Rock type	0.139	3.036	0.434	60.301	-4.823
	Net-to-gross	0.968	0.450	0.685	11.088	3.470
	Permeability	0.971	0.406	0.696	9.736	3.917
Our Approach/Validation	Porosity	**0.994**	**0.080**	**0.897**	**2.314**	**10.974**
	Rock type	**0.984**	**0.077**	**0.909**	**2.552**	**11.112**
	Net-to-gross	**0.999**	**0.019**	**0.969**	**0.595**	**17.109**
	Permeability	**0.998**	**0.025**	**0.959**	**0.723**	**16.075**
Baseline/Validation	Porosity	0.970	0.416	0.695	9.382	3.805
	Rock type	0.141	3.047	0.435	62.281	-4.839
	Net-to-gross	0.970	0.412	0.687	10.387	3.848
	Permeability	0.970	0.412	0.696	8.953	3.851

similarity, and PSNR captures visual clarity.

Table 3.3 demonstrates consistently strong performance across all metrics, particularly for Net-to-gross, which achieves nearly perfect linear correlation ($R = 0.999$), minimal reconstruction error (MSE ≈ 0.02), and outstanding structural fidelity (SSIM ≈ 0.97) in both test and validation sets. Distributional similarity is also highest for Net-to-gross (FID ≈ 0.59), matched by superior visual clarity (PSNR ≈ 17.0). Permeability closely follows, with high R, low MSE, strong SSIM (≈ 0.96), low FID (≈ 0.75), and high PSNR (≈ 16.1), indicating robust reconstruction.

In contrast, porosity and rock type exhibit slightly lower R (≈ 0.995 for porosity), higher reconstruction error (MSE ≈ 0.08), and more modest structural preservation (SSIM ≈ 0.90). Their higher distributional discrepancies (FID ≈ 2.4–2.6) and reduced visual clarity (PSNR ≈ 11) further reflect moderate but consistent deviations from the original uncertainties.

Taken together, these results demonstrate that Net-to-gross and permeability consistently yield the most accurate, structurally faithful, and visually coherent reconstructions, while porosity and rock type display slightly increased error and reduced similarity across all evaluation metrics.

3.4.6 Analysis and Visualization Approach

To assess reconstruction accuracy, we analyze specific layers along the Z-axis. First, selected layers are visualized to directly compare original and reconstructed uncertainties. Next, kernel density estimation (KDE) is used to compare their distributions within the same layers. Finally, histograms quantify global discrepancies across all Z-axis layers, providing a comprehensive evaluation of reconstruction fidelity.

For detailed visual assessment, we focus on layers 12 and 14 (see Figures 3.8 and 3.9),

presenting the original and reconstructed uncertainties side-by-side for all four principal properties.

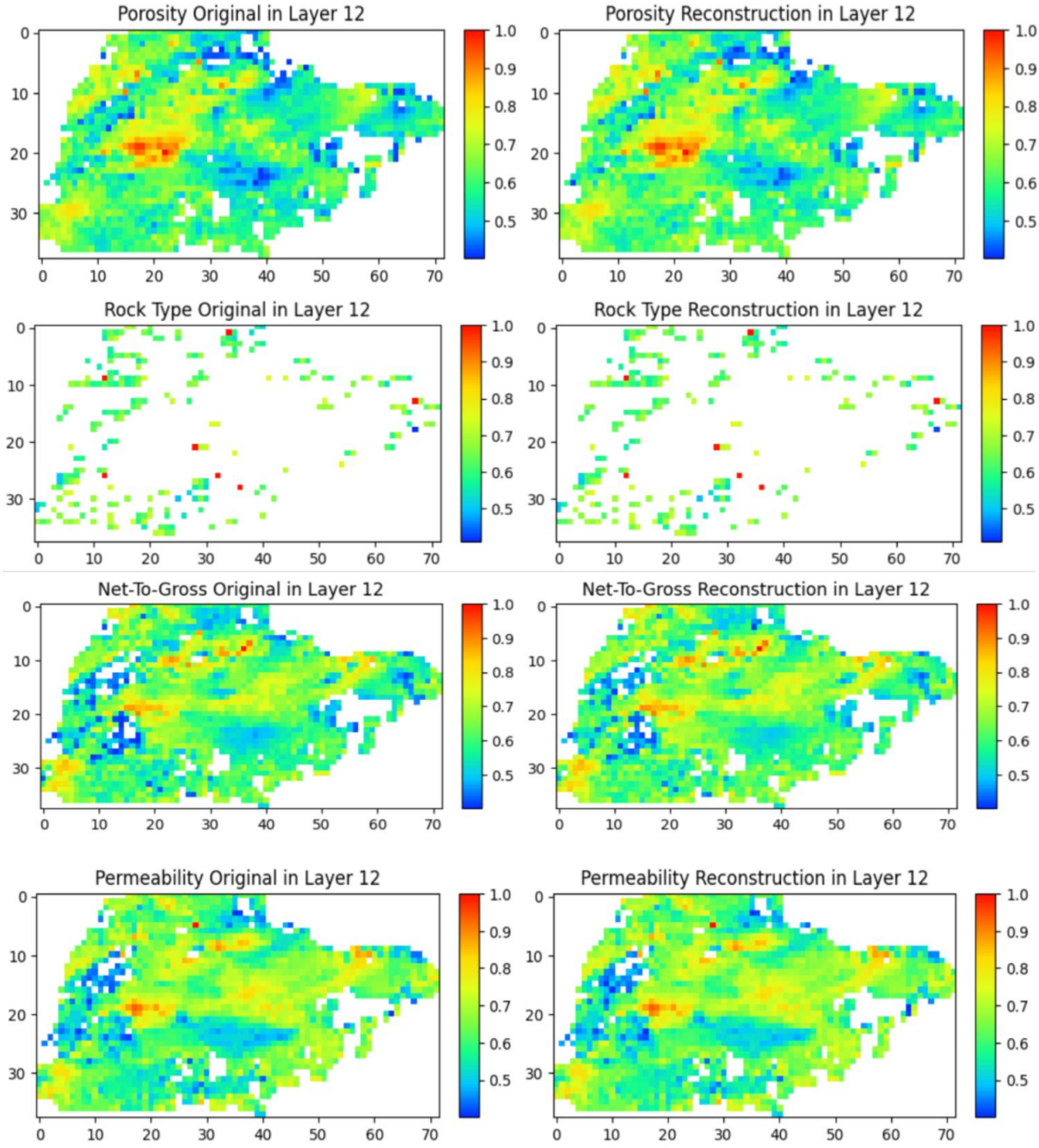


Figure 3.8: Reconstructions of Geological Uncertainties Focusing on Layer 12.

3.4.7 KDE and Histogram Analysis

Figures 3.8 and 3.9 show that the reconstructed uncertainties closely match the original ones regarding porosity, net-to-gross ratio, rock type, and permeability, with minimal difference. This resemblance suggests that the reconstruction process accurately captures the spatial patterns and distributions of geological uncertainties, demonstrating high fidelity and precision in the reconstruction effort.

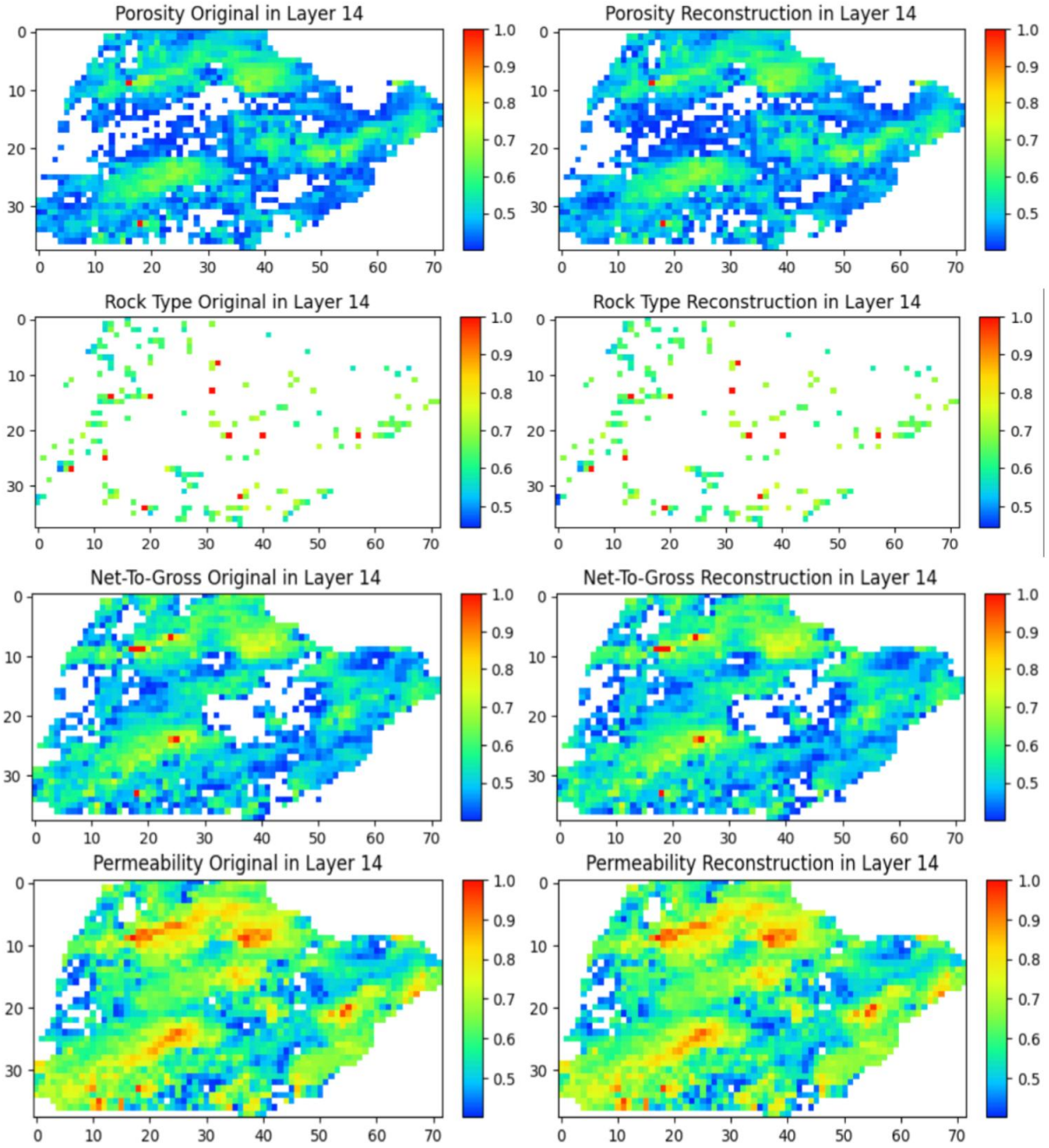


Figure 3.9: Reconstructions of Geological Uncertainties Focusing on Layer 14.

Figure 3.10 illustrates our research findings. Each plot, arranged from left to right and top to bottom, presents a KDE plot corresponding to a specific geological property and layer: porosity, rock type, net-to-gross, and permeability. The plots first display data for Layer 12, followed by Layer 14 for each scenario. The original data distribution is represented in red in each KDE plot, while its corresponding reconstructed KDE distribution is shown in blue. Across all scenarios, the reconstructed KDE distributions overlap with the original distributions, indicating the effectiveness of the reconstruction method in preserving the overall pattern or shape of the original data distribution.

Figure 3.11 shows histogram plots for porosity, rock type, net-to-gross ratio, and permeability comparing original and reconstructed distributions. Each plot pairs an original

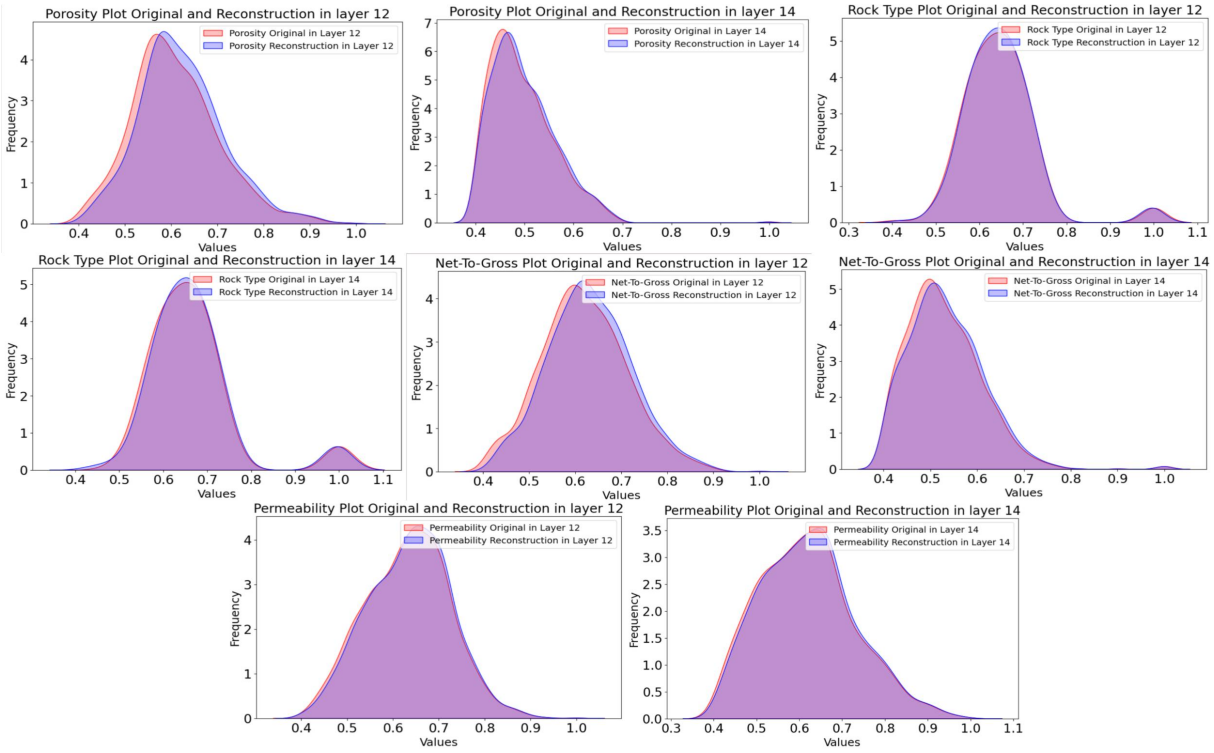


Figure 3.10: Reconstruction Analysis of KDE Distributions for Porosity, Rock Type, Net-to-Gross Ratio, and Permeability in Layers 12 and 14.

geological property with its reconstruction, arranged from left to right and top to bottom for porosity, rock type, net-to-gross ratio, and permeability. The x-axis represents property values, and the y-axis shows reservoir levels. By overlaying the reconstruction on the original distributions, we find a close match in the histograms, which means that the reconstruction process successfully captures and replicates the original data's variability and structure.

The findings demonstrate a robust correlation among geological properties, with high R values indicating effective alignment between the reconstructed and original uncertainties. The mean squared error results suggest precise reconstruction accuracy, although variations were observed among different properties. Structural similarity was notably good, though porosity and rock type exhibited slightly lower values due to their inherent complexity. This still reflected significant resemblance overall.

Moreover, the FID metric reinforced the reconstruction's effectiveness by evaluating the distributional similarity between the original and reconstructed data. The PSNR contributed insights into the sharpness and clarity of the reconstructed images. Visual analyses, including Kernel Density Estimation (KDE) and histogram plots, further confirmed the close match between the reconstructed and original uncertainties. Collectively, these results underscore the accuracy, fidelity, and reliability of the reconstructed uncertainties, providing a solid foundation for geological analysis in intricate subsurface environments.

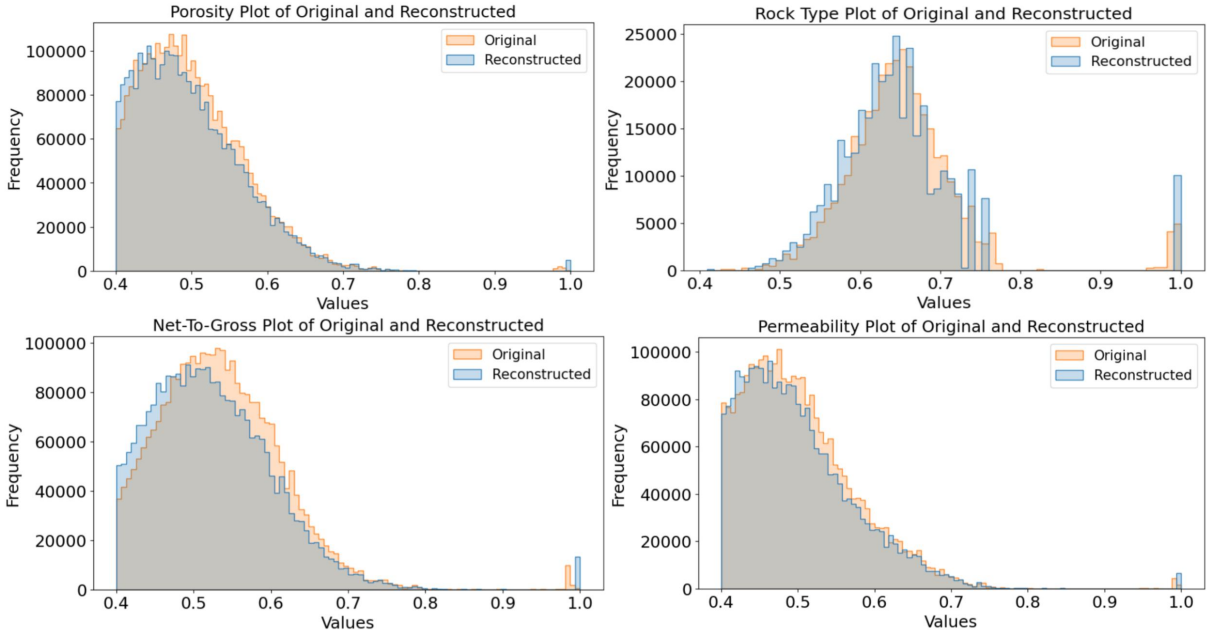


Figure 3.11: Geological Property Histogram Comparison.

3.5 Final Remarks

Our methodology demonstrated its potential by effectively reproducing the original geological uncertainties. The ablation study, which relied on the visual interpretation of our experts through side-by-side comparison of the original and reconstructed uncertainties using the same color scale, focused on the twelfth porosity layer. This approach enabled the identification of preserved and distorted spatial patterns. The results indicate that integrating residual connections and applying normalization with an L2 regularization value of 0.0001 are essential for achieving accurate reconstructions. Remarkably, high correlation values indicate a significant relationship between the reconstructed and original data, underscoring the accuracy of our approach. Despite minor discrepancies in porosity and rock-type reconstructions because of their complex nature, our method maintained a high level of structural fidelity. Visual analyses further validated the resemblance of the reconstructed uncertainties to the originals, highlighting our model's capability to capture essential spatial patterns and distributions.

Chapter 4

Meta Activation Functions for Uncertainty Quantification in Reservoir Modeling Proxies

This work was submitted to the IEEE Transactions on Neural Networks and Learning Systems Journal, with permission from "Jeanfranco Farfan, Gabriel Cirac, Alexandre Ferreira, and Anderson Rocha. Meta Activation Functions for Uncertainty Quantification in Reservoir Modeling Proxies, 2025."

In this chapter, we address a gap in the existing literature on proxy model development. Specifically, to enhance the reliability and credibility of our end-to-end proxy predictions, which generate approximate responses to the reservoir simulator, we have developed a novel approach to estimate uncertainty in the proxies. As highlighted by Volodina and Challenor [14], transparent reporting and communication of uncertainties are crucial for ensuring the credibility of proxy models and maintaining trust in their predictions.

4.1 Introduction

Proxy-based neural networks provide an effective solution by offering efficient approximations that replicate the simulator's input–output behavior while substantially reducing computational requirements [20, 21, 22]. In recent years, DNN-based approaches have been widely explored for this purpose. Maschio *et al.* [1] proposed a deep learning method as a proxy for the reservoir simulator, employing Principal Component Analysis (PCA) for dimensionality reduction.

Farfan *et al.* [2] proposed an end-to-end DNN proxy capable of handling 3D geological uncertainties and replicating the simulator's response across several cumulative production curves. This approach leverages 3D Convolutional Neural Networks (CNNs) to process spatial and depth dimensions within heterogeneous reservoirs. Similarly, Yan *et al.* [76] trained and tested a CNN-based proxy model to accurately predict oil and water production. Their model utilized six channels to encode fracture information, including fracture shape, time, matrix porosity, matrix permeability, fracture porosity, and fracture permeability. Bayazitova *et al.* [77] proposed a proxy for real-time anomaly detection

in offshore natural flow wells. Their methodology integrates Long Short-Term Memory (LSTM) networks and Gated Recurrent Unit (GRU) backbones with genetic algorithms to improve the monitoring of oil and gas flow.

However, current implementations of neural network–based proxies face important limitations, most notably their inability to guarantee the reliability of predictions. This shortcoming stems from a gap in the literature regarding the Uncertainty Quantification (UQ) of proxies in reservoir management. As emphasized by Volodina1 and Challenor [14], transparent reporting and communication of uncertainties are essential to ensure the credibility of proxy models and the trust placed in them. This form of uncertainty is called epistemic uncertainty, which arises from limited knowledge about the parameters of neural network-based proxies.

Several approaches to UQ in neural networks (NNs) have been proposed, including ensemble-based methods [78, 79, 80], Bayesian neural networks (BNNs) [81, 82], hybrids that combine both (Bayesian ensemble) [27, 83], and non-Bayesian approaches [84, 85]. Incorporating these techniques into NN-based proxies enhances the credibility of surrogate predictions and strengthens confidence when approximating high-fidelity reservoir simulators. In this work, we focus on Bayesian ensembles, which extend classical ensembles by introducing priors over the parameters of the proxy models. Whereas classical ensembles capture uncertainty through diversity induced by random initialization, Bayesian ensembles draw proxy model weights from posterior distributions shaped by prior assumptions. The choice of prior is therefore critical, as it encodes beliefs about parameter distributions and directly influences the quality of the posterior predictive distribution [27]. To this end, Pearce *et al.* [83] proposed the Anchored Ensemble (AE) method, which replaces conventional weight decay (regularizing parameters toward zero) with a scheme that regularizes weights toward their initialization values (anchors). Because these initialization values are sampled from a prior distribution, the resulting ensemble admits a bayesian interpretation.

In this work, we propose a novel method, called Meta Activation Functions Ensemble (MAFs Ensemble), designed to improve the robustness of uncertainty estimation in DNNs. MAFs Ensemble builds upon two existing techniques: Anchored Ensembles (AE) and Random Activation Functions Ensembles (RAFs Ensemble), originally introduced by Pearce *et al.* [83] and Stoyanova *et al.* [27], respectively. Following the approach of Pearce *et al.* [83], we adopt the AE framework to handle uncertainty via Bayesian anchoring ensemble. However, we depart from the RAFs Ensemble method of Stoyanova *et al.* [27], in which ensemble members are initialized with randomly chosen activation functions that remain fixed during training. Instead, in MAFs Ensemble, we propose that activation functions should be input-dependent rather than static or aleatoric.

To this end, we introduce an adaptive computation mechanism inspired by principles of meta-learning [86, 87, 88]. The key idea is to increase ensemble diversity by allowing each neural network to dynamically modulate its activation functions via a convex combination of multiple base activations, with mixture weights determined by the input. This mechanism enables the activation behavior to adapt continuously based on the input context, thereby improving both flexibility and representational capacity. Ensemble diversity is widely recognized as a critical factor in ensemble performance. As highlighted

by Zhang and Ma [89], lack of diversity often results in degraded performance. MAFs Ensemble addresses this by introducing activation-level diversity as a new axis of variation within ensemble members.

We evaluate MAFs Ensemble against four state-of-the-art methods: AE [83], RAF Ensemble [27], Evidential Regression Network (ERN) [84], and Uncertainty-Regularized ERN (UR-ERN) [85]. Experiments are conducted on both synthetic and real-world regression datasets, with performance measured via negative log-likelihood (NLL) and symmetric mean absolute percentage error (SMAPE). We evaluate the proposed method and the baseline on a comprehensive collection of synthetic and real-world regression datasets. Finally, we tested the robustness and scalability of our approach in the context of our Brazilian offshore field, specifically by modifying the end-to-end proxy model proposed by Farfan *et al.* [2] to incorporate the meta-network in place of the ReLU activation function. The results show that our approach not only maintains high predictive accuracy but also introduces a reliable estimation of uncertainty, marking a significant advancement in the proxy models used for reservoir management.

The contributions of this chapter are summarized as follows:

1. The proposed method improves uncertainty quantification in DNN-based proxy models for reservoir simulation by dynamically adjusting activation functions based on input data, enhancing ensemble diversity and robustness.

2. The Meta-Learning mechanism embedded in MAFs Ensemble allows each ensemble member to adapt its activation function contextually, improving flexibility and representational capacity, and strengthening uncertainty estimation.

3. The MAFs Ensemble method outperforms or matches state-of-the-art uncertainty quantification methods in both synthetic and real-world regression tasks, demonstrating superior predictive accuracy and uncertainty estimation capabilities.

4. The proposed methodology was successfully applied to our Brazilian offshore field, quantifying uncertainty in proxy-based models that replicate simulator behavior, thereby improving decision-making in reservoir management.

4.2 Related Work

Over the past decade, the need to quantify the uncertainty of predictions generated by DNNs has gained increasing attention. Several approaches to UQ in NNs have been proposed, including Bayesian neural networks (BNNs), ensemble-based methods, hybrids that combine both (Bayesian ensembles), and non-Bayesian approaches. Among these, BNNs have been the most extensively studied, as they are robust to overfitting and can be trained on both small and large datasets [90]. Puzyrev and Duuring [91] applied UQ to geochemical data using the widely adopted Monte Carlo Dropout method. Their work, validated on the WACHEM and WAMEX geochemical databases, demonstrated a favorable balance between predictive accuracy and uncertainty estimation. Similarly, Borquez *et al.* [92] employed Monte Carlo Dropout for classifying HER2-overexpressing

breast cancer tissues, successfully quantifying predictive uncertainty. In another application, Wang *et al.* [93] combined Markov chain Monte Carlo (PMCMC) with factorial polynomial chaos expansion (FPCE) to improve UQ in hydrological predictions, where a Gaussian anamorphosis transform was used to link parameter inference with uncertainty propagation. Beyond these, Bayesian methods have also been integrated with deep active learning (AL) to identify informative samples. For instance, Chen *et al.* [94] proposed a probabilistic Bayesian active learning framework for machine fault diagnosis under scarce labeled data, leveraging BNNs to guide sample selection. While Bayesian approaches provide a rigorous theoretical foundation, they typically require substantial modifications to training procedures and are computationally demanding.

Ensemble-based methods represent another prominent direction, valued for their scalability and ease of implementation in real-world applications. Lakshminarayanan *et al.* [78] introduced deep ensembles, a simple yet effective technique that combines probabilistic ensemble training with adversarial training. Their approach delivered competitive or superior performance compared to Bayesian methods such as MC Dropout, and proved efficient even for large-scale tasks like ImageNet [95]. Osband *et al.* [96] proposed a reinforcement learning strategy inspired by statistical bootstrapping, where an ensemble of models trained on perturbed datasets captures predictive uncertainty. Wenzel *et al.* [97] extended this concept by incorporating variability not only in network weights but also in hyperparameters. Their Hyperdeep ensembles, constructed via random hyperparameter search combined with multiple initializations, highlighted the benefits of diversity in both parameters and configurations. Nonetheless, ensemble methods also face limitations. Stoyanova *et al.* [27] noted that ensembles often produce overly confident predictions, while Pearce *et al.* [83] emphasized that capturing the full posterior distribution would require a prohibitively large number of ensemble members.

Non-Bayesian approaches to UQ have recently gained attention, particularly the evidential regression network (ERN), which frames learning as an evidence acquisition process and integrates deep learning with Dempster–Shafer theory to jointly predict outcomes and quantify uncertainty [84, 85]. Extensions of ERN include Wong *et al.* [98] modeled data distributions using a scale mixture distribution (SMD), offering a simpler alternative to the normal-inverse-gamma (NIG) prior by avoiding explicit priors on latent mean and variance parameters. Ye *et al.* [85] further addressed the zero-gradient issue in evidential regression networks by introducing an additional regularization term beyond the total evidence penalty, allowing the model to capture and generalize patterns across the entire training dataset. Similarly, Deng *et al.* [99] incorporated the Fisher Information Matrix (FIM) to mitigate the excessive penalization of mislabeled classes by weighting class-specific contributions for each sample in classification tasks. Despite these innovations, several studies have shown that epistemic uncertainty is not reliably captured by evidential methods, as their theoretical foundations remain incomplete and suffer from issues of identifiability, interpretability, and lack statistical meaning. Conversely, these techniques should be recognized as energy-based algorithms for detecting out-of-distribution instances[100, 101, 102].

Although the Bayesian framework offers a principled and theoretically grounded approach to uncertainty quantification, its application to NNs is challenging due to the high

computational cost and parameter space. NN ensembles offer a scalable alternative, but are often criticized for overconfidence in predictions. To address this, hybrid approaches have led to the development of Bayesian ensembles. A notable example is the anchored ensemble introduced by Pearce *et al.* [83], which extends standard ensembles by regularizing each network's parameters around values sampled from an anchoring distribution, typically chosen to match the Bayesian prior. By repeating this randomized procedure across networks, the ensemble approximates sampling from the full Bayesian posterior. Efforts to further enhance ensemble diversity have been explored by Stoyanova *et al.* [27], who proposed diversifying ensemble members through heterogeneous activation functions. Rather than initializing each NN with the same activation function, their method assigns a different (random) activation function to each prior, thereby introducing distinct nonlinear properties and improving ensemble diversity. Extending this line of work, we introduce diversity at training time by employing a convex combination of multiple activation functions, where the contribution of each function is adaptively weighted according to its relevance for the input.

4.3 Preliminaries

Following the notation of [27, 83, 103], aleatoric and epistemic uncertainties are denoted by σ_a^2 and σ_e^2, respectively. Aleatoric uncertainty captures the inherent variability in the data, while epistemic uncertainty reflects the lack of knowledge about model parameters [85]. The predictive (total) uncertainty is defined as the sum of these two components:

$$\sigma_p^2 = \sigma_a^2 + \sigma_e^2. \tag{4.1}$$

Let S_{train} denote a training dataset consisting of N independently and identically distributed (i.i.d.) samples:

$$S_{\text{train}} = (x_i, y_i); i = 1, ..., N, \tag{4.2}$$

where $x_i \in \mathbb{R}^d$ represents a d-dimensional feature vector and $y_i \in \mathbb{R}$ is the corresponding scalar output. The test dataset is denoted by S_{test}. The design matrix is represented by X and the output vector by y, such that $(S_{\text{train}}.X, S_{\text{train}}.y)$ and $(S_{\text{test}}.X, S_{\text{test}}.y)$ correspond to training and test data, respectively. A general regression task is formulated as

$$y_i = f(x_i) + \epsilon_i, \tag{4.3}$$

where ϵ_i is a normally distributed noise term,

$$\epsilon_i \sim \mathcal{N}(0, \sigma_a^2), \tag{4.4}$$

representing the aleatoric uncertainty inherent in the observations.

Consider a neural network with parameters θ, H hidden units, and N training samples.

Assuming a Gaussian prior over the parameters,

$$P(\theta) = \mathcal{N}(\theta_{\text{prior}}, \Sigma_{\text{prior}}), \tag{4.5}$$

the Maximum a Posteriori (MAP) estimate of θ is obtained as

$$\theta_{\text{MAP}} = \text{argmax}_\theta \log(P(S_{\text{train}} \mid \theta)) - \frac{1}{2}\left|\Sigma_{\text{prior}}^{-1/2}(\theta - \theta_{\text{prior}})\right|_2^2, \tag{4.6}$$

In the ensemble setting with M neural networks, stochastic anchoring is introduced by replacing θ_{prior} with a random variable $\theta_{\text{anc}} \sim \mathcal{N}(\theta_{\text{prior}}, \Sigma_{\text{prior}})$ for each ensemble member. This formulation ensures that each network is anchored to a distinct prior draw, encouraging diversity and improving estimates of epistemic uncertainty.

Assuming homoskedastic Gaussian noise, the practical training objective for the j-th ensemble member is expressed as

$$\mathcal{L}_j(\theta_j) = \frac{1}{N}\left|y - \hat{y}_j\right|_2^2 + \frac{1}{N}\left|\Gamma^{1/2}(\theta_j - \theta_{\text{anc},j})\right|_2^2, \tag{4.7}$$

where Γ is a diagonal regularization matrix, whose j-th diagonal element is the ratio of data noise of the target variable to prior variance for parameter θ_j, $\text{diag}(\Gamma)_j = \sigma_a^2/\sigma_{\text{prior},j}^2$, and $\theta_{\text{anc},j}$ is the stochastic anchor for the j-th network. This formulation explicitly links the MAP estimation with the anchored ensemble framework, while remaining consistent with the uncertainty decomposition in Equation (4.1).

However, similar to Stoyanova *et al.* [27], we emphasize the limitations of the approach proposed by Pearce *et al.* [83]. Specifically, it only models epistemic uncertainty, while treating aleatoric (random) uncertainty as a constant. As highlighted by Stoyanova [27], this is a limitation because, in practice, it is not always possible to clearly distinguish between different sources or types of uncertainty.

Finally, the ensemble mean prediction over M networks is computed as

$$\bar{y} = \frac{1}{M}\sum_{j=1}^{M} \hat{y}_j. \tag{4.8}$$

4.4 Proposed Method

This work introduces Meta-Activation Function (MAFs) Ensemble, a framework for estimating predictive uncertainty by extending ensemble learning with dynamic, data-dependent activation functions. Unlike conventional ensemble approaches, which assign a fixed activation function to each model, as in AE [83], or randomly allocate them, as in RAFs [27], the proposed method leverages a meta-learning mechanism to adaptively select and weight activation functions for each input. This adaptive selection enhances the expressive capacity of the ensemble, increases diversity among members, and improves uncertainty quantification.

At the core of MAFs Ensemble lies the adaptive activation mechanism. For the j-th base model NN_j, the activation function $\phi_j(x)$ is expressed as a convex combination of k

candidate activation functions $\{a_1, a_2, \ldots, a_k\}$:

$$\phi_j(x) = \sum_{i=1}^{k} \alpha_i^{(j)}(x)\, a_i(x), \qquad \sum_{i=1}^{k} \alpha_i^{(j)}(x) = 1. \tag{4.9}$$

The combination coefficients $\alpha_i^{(j)}(x)$ are dynamically generated by a meta-network conditioned on the input. This design enables each ensemble member to adapt its activation behavior to the local structure of the data, thereby capturing complex nonlinearities and promoting diversity in each base learner.

The meta-network architecture (see Figure 4.4 to get more details) consists of a DenseNet [104] block, which facilitates feature reuse and stable gradient propagation, followed by an attention [105] module that assigns higher weights to the most relevant features for each prediction. This design ensures that the meta-network not only selects activation functions but also learns when and how much each function should contribute. Importantly, the activation adaptation is learned jointly with the base model parameters via backpropagation under the anchor loss (Eq. 4.7).

After training, uncertainty is estimated from the variance of predictions across the ensemble members. The key innovation of MAFs Ensemble is the meta-learning strategy for adaptive activation functions, which enhances both ensemble expressiveness and diversity, two crucial factors for reliable uncertainty estimation. By enabling each base learner to condition its activation dynamics on the input, the proposed method demonstrates superior ability to capture complex data patterns compared to ensembles with static activations. This property directly addresses the observation by Stoyanova *et al.* [27] that ensemble diversity is central to robust predictive performance.

Following the notation of Stoyanova *et al.* [27], let μ_0 and Σ_0 denote the prior mean and covariance, $\hat{\sigma}_a^2$ the aleatoric variance (uncertainty), m denote the number of base-learners, and NN_j the j-th ensemble. The training and prediction steps of MAFs Ensemble are summarized in Algorithm 1. The procedure begins by constructing a regularization matrix and defining the set of candidate activation functions (lines 1–2). Each ensemble member is then initialized from the prior distribution, while a meta-learning block generates adaptive activation weights, and both are trained jointly using stochastic gradient descent (lines 3–7). During prediction (lines 8–13), outputs from all ensemble members are aggregated to compute the predictive mean and epistemic variance, from which the total predictive uncertainty is obtained.

Algorithm 1 MAFs Ensemble

Require: $S_{\text{train}}, S_{\text{test}}$, priors μ_0, Σ_0, ensemble size m, aleatoric variance $\hat{\sigma}_a^2$
Ensure: Predictive mean $\hat{y}$ and variance $\hat{\sigma}_p^2$

$\Gamma \leftarrow \hat{\sigma}_a^2 \Sigma_0^{-1}$ ▷ Regularization matrix
$\mathcal{A} \leftarrow \{a_1, \ldots, a_k\}$ ▷ Set of k activation functions
for $j \leftarrow 1$ to m **do** ▷ Train the ensemble
    Initialize $\theta_{j,0} \sim \mathcal{N}(\mu_0, \Sigma_0)$
    Create NN_j with adaptive activation $\phi_j(x)$ ▷ Defined in Eq. (4.9)
    $NN_j.\textsc{Train}(S_{\text{train}}, \Gamma, \theta_{j,0}, \mathcal{A})$ ▷ Use loss in Eq. (4.7)
end for
for $j \leftarrow 1$ to m **do** ▷ Predict with the ensemble
    $\hat{y}_j \leftarrow NN_j.\textsc{Predict}(S_{\text{test}}.X)$
end for
$\hat{y} \leftarrow \frac{1}{m} \sum_{j=1}^{m} \hat{y}_j$ ▷ Mean predictions
$\hat{\sigma}_e^2 \leftarrow \frac{1}{m-1} \sum_{j=1}^{m} (\hat{y}_j - \hat{y})^2$ ▷ Epistemic variance
$\hat{\sigma}_p^2 \leftarrow \hat{\sigma}_e^2 + \hat{\sigma}_a^2$ ▷ Total variance (Eq. 4.1)
return $\hat{y}, \hat{\sigma}_p^2$

Distinct from AE and RAFs, the proposed MAFs Ensemble meta-learns input-conditioned mixtures of activation functions within anchored networks, boosting ensemble diversity and calibration to deliver more reliable predictive uncertainty.

4.5 Experiments

In our experiments, the base learners of the MAF ensemble are multilayer perceptrons (MLPs) with a single hidden layer comprising 100 neurons. The ensemble size, m, is set to five, a consistent choice across all methods evaluated in this study. We employ an ensemble of six non-saturating activation functions due to their faster convergence during training with gradient descent [106]. These functions include: (i) SELU (Scaled Exponential Linear Unit) [107], (ii) Leaky ReLU (Leaky Rectified Linear Unit) [108], (iii) ReLU (Rectifier Activation Function) [109], (iv) ERF (Element-wise Gauss Error Function) [110], (v) SiLU (Sigmoid Linear Unit) [111], and (vi) GELU (Gaussian Error Linear Unit) [112].

Following the methodology of Stoyanova *et al.* [27], we conduct experiments on both synthetic and real-world datasets. For each dataset, a fixed random seed is used to ensure reproducibility. We deliberately select a larger test set than the training set to densely probe the predictive behavior (e.g., He/Forrester datasets: 20 training points versus 50 test points). To capture epistemic uncertainty, the training support is strictly narrower than the test support. Specifically, for one-dimensional synthetic datasets (He and Forrester), training inputs are sampled uniformly at random from two disjoint intervals, leaving a central gap. In contrast, test inputs span a substantially larger domain. For higher-dimensional datasets, training inputs are drawn from a bounded hyperrectangle, which is then expanded for the test set.

4.5.1 Baselines

We include four state-of-the-art methods as baselines for empirical comparison with the MAFs Ensemble: (i) Anchored Ensemble (AE) [83], (ii) Random Activation Functions (RAF) Ensemble [27], (iii) Evidential Regression Network (ERN) [84], and (iv) Uncertainty-Regularized ERN (UR-ERN) [85]. All methods are evaluated on both synthetic and real-world datasets spanning different input dimensionalities.

4.5.2 Synthetic and Real-world Datasets

To ensure a comprehensive evaluation, we evaluate the proposed method and the baseline on a comprehensive collection of synthetic and real-world regression datasets. The synthetic suite is designed to probe complementary aspects of predictive uncertainty under controlled conditions, while the real datasets provide empirical validation in practical, noisy environments.

Synthetic Data. We generate multiple synthetic datasets that span several categories: (i) benchmark optimization functions, (ii) physical and engineering models, and (iii) high-dimensional nonlinear functions. Benchmark functions include Schaffer, Griewank, Rastrigin, Styblinski–Tang, Roos–and Arnold, which are known for their high nonlinearity, multimodality, and the presence of many local minima. These datasets are characterized by a predominance of epistemic uncertainty, since small perturbations in the input can lead to highly variable outputs.

The second category is formed by datasets generated from physics-inspired models such as the double pendulum, planar arm torque, piston simulation, robot arm dynamics, wing weight estimation, and environmental models. These tasks resemble real-world engineering systems, where the model dynamics are complex and uncertainty is largely epistemic due to the wide generalization regions.

Finally, nonlinear benchmark functions such as Ishigami, Forrester, He's function, Friedman, and Adapted Welch are used to evaluate robustness under strongly non-additive effects and interaction terms. Across functions, we vary dimensionality from 1 to 10 inputs and construct training sets from restricted subdomains while sampling test sets over broader ranges. This split intentionally creates out-of-distribution (OOD) evaluation to stress epistemic calibration. Additive Gaussian noise is injected at generation time to control the aleatoric component.

Real-world Data. In addition to synthetic datasets, we employ several real-world regression tasks widely used in uncertainty quantification benchmarks. These include: (i) *Abalone* [113] (age prediction of abalone shells), (ii) *Naval propulsion plant* [114] (predicting turbine degradation indicators), (iii) *Parkinson's disease* dataset [115] (predicting UPDRS scores from biomedical features).

Together, the synthetic and real-world datasets provide complementary testbeds: synthetic datasets allow controlled exploration of uncertainty properties under known dynamics, while real-world data demonstrate the applicability and reliability of the proposed method in practical predictive tasks.

4.5.3 Evaluation Metrics

Our evaluation prioritizes the quality of uncertainty estimates, with predictive accuracy serving as a complementary measure. Following established practice [116, 27, 83, 85], we adopt the **Negative Log-Likelihood (NLL)** as our principal metric. NLL assesses how well the entire predicted distribution aligns with the observed data, directly rewarding models that produce accurate and well-calibrated uncertainty; lower values are better [117]. To verify point-prediction performance, we also report the **Symmetric Mean Absolute Percentage Error (SMAPE)**. Together, these metrics provide a robust framework that emphasizes the fidelity of uncertainty quantification without neglecting predictive accuracy.

4.6 Results and Discussions

The performance of different baselines compared to our proposed method is illustrated in Figures 4.1 and 4.2 using two one-dimensional benchmark datasets: the He function $y = x \cdot \sin(x) + \epsilon$ [118] and the Forrester function $y = (6x - 2)^2 \cdot \sin(12x - 4) + \epsilon$ [119].

For the He dataset (Figure 4.1a), the red dashed line denotes the ground-truth function over the interval $x \in [-6, 6]$, while the blue points indicate the sampled training data. The function exhibits nonlinear oscillatory behavior, with training data covering only a limited portion of the input domain. Figures 4.1b and 4.1c compare our method (MAF) with ERN and UR-ERN, respectively. Both ERN and UR-ERN display overly narrow predictive uncertainty bands in regions without training coverage, suggesting overconfident predictions in out-of-distribution (OOD) areas. This behavior is undesirable, especially in safety-critical applications, where reliable uncertainty estimation requires high epistemic uncertainty when extrapolating beyond observed data. By contrast, Figures 4.1d and 4.1e compare our method with AE and RAF, both of which produce wider predictive intervals in unobserved regions. These intervals more accurately capture epistemic uncertainty, thereby improving robustness against OOD inputs. Notably, ensemble MAFs provide predictive intervals that are on par with RAFs while being wider than AE, confirming that MAF remains competitive in this setting.

For the Forrester dataset (Figure 4.2a), the red dashed curve represents the true function over the interval $x \in [0, 1]$, while the blue points correspond to sparsely sampled training observations. The training data are restricted to two subintervals, $[0.2, 0.4]$ and $[0.65, 0.85]$, leaving significant regions of the input space unobserved. Figures 4.2b and 4.2c compare MAF with ERN and UR-ERN, respectively. In the region $[0, 0.2]$, both ERN and UR-ERN again produce narrow uncertainty bands, reflecting an overestimation of confidence in OOD areas. Interestingly, in the region $[0.85, 1]$, these methods exhibit uncertainty levels comparable to MAF, correctly indicating epistemic uncertainty due to lack of knowledge. Conversely, Figures 4.2d and 4.2e show that AE and RAF yield narrower uncertainty intervals relative to MAF, underestimating epistemic uncertainty in unseen regions. Importantly, while RAF remains competitive on the He dataset, it underperforms here, whereas MAF maintains robust and well-calibrated uncertainty estimates across both datasets.

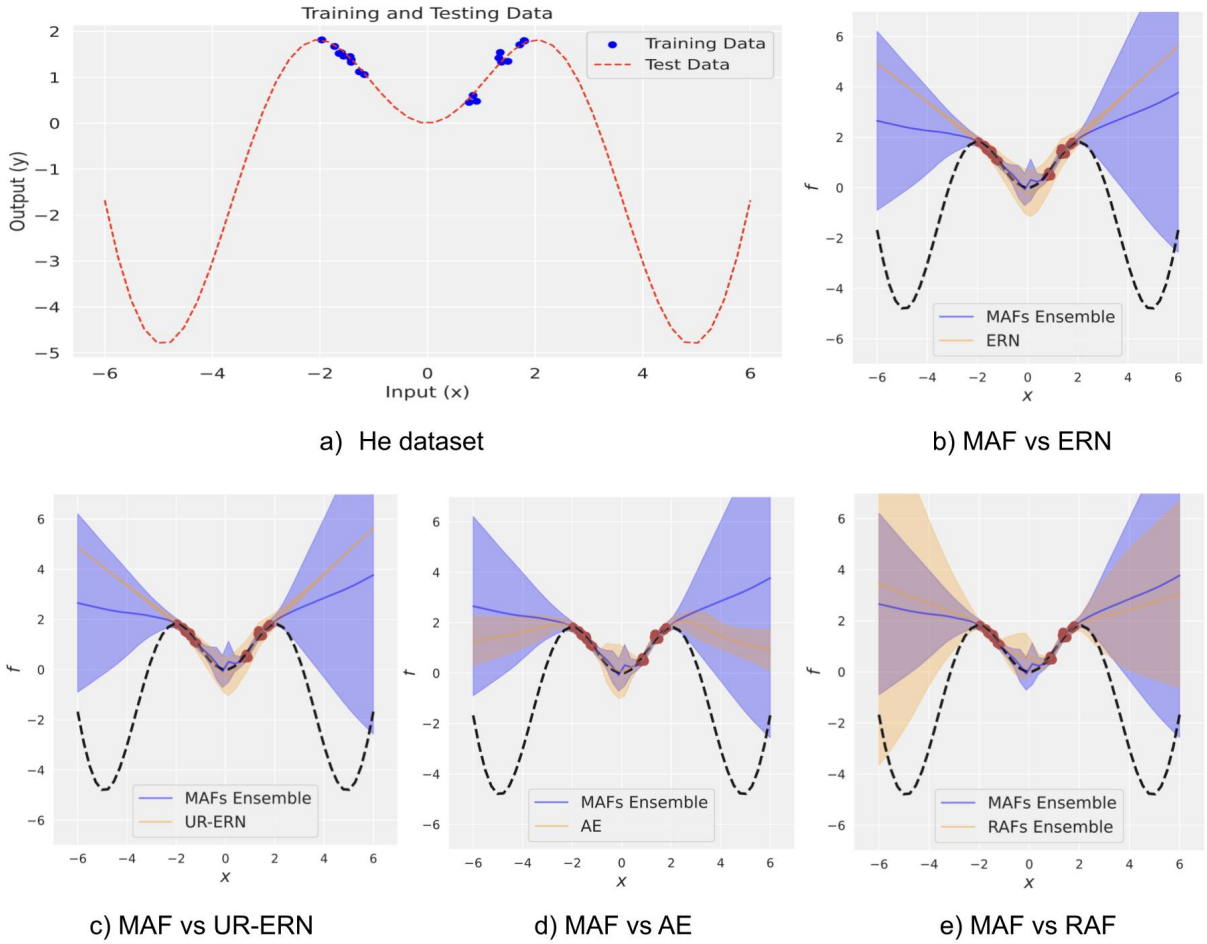


Figure 4.1: Results on the He dataset. (a) Ground-truth function (red dashed) and training data (blue). (b–c) ERN and UR-ERN underestimate uncertainty in OOD regions. (d–e) AE and RAF better capture epistemic uncertainty.

Quantitative results in Table 4.1 and Table 4.2 confirm these findings: ensemble MAFs achieve the lowest NLL, demonstrating not only better-calibrated predictive uncertainty and more accurate probabilistic predictions overall, but also stronger consistency across different datasets compared to ERN, UR-ERN, AE, and RAF. This consistency across heterogeneous datasets underscores the robustness of MAF as a general-purpose uncertainty quantification method.

We evaluated the overall performance of all methods using NLL and SMAPE. Tables 4.1 and, Table 4.2 present the comparative results of the Ensemble MAFs against four benchmark methods across 16 synthetic and 3 real-world datasets. The proposed MAF method consistently demonstrates competitive performance and, in many cases, outperforms the baselines, providing clear advantages in predictive accuracy and uncertainty estimation.

The results in Table 4.1 present NLL values with 95% confidence intervals. To ensure fair comparison, the best outcomes are highlighted in bold whenever confidence intervals overlap, indicating statistical equivalence. Accordingly, some datasets feature more than one leading method.

On low-dimensional synthetic datasets, MAF clearly outperforms the baselines. For

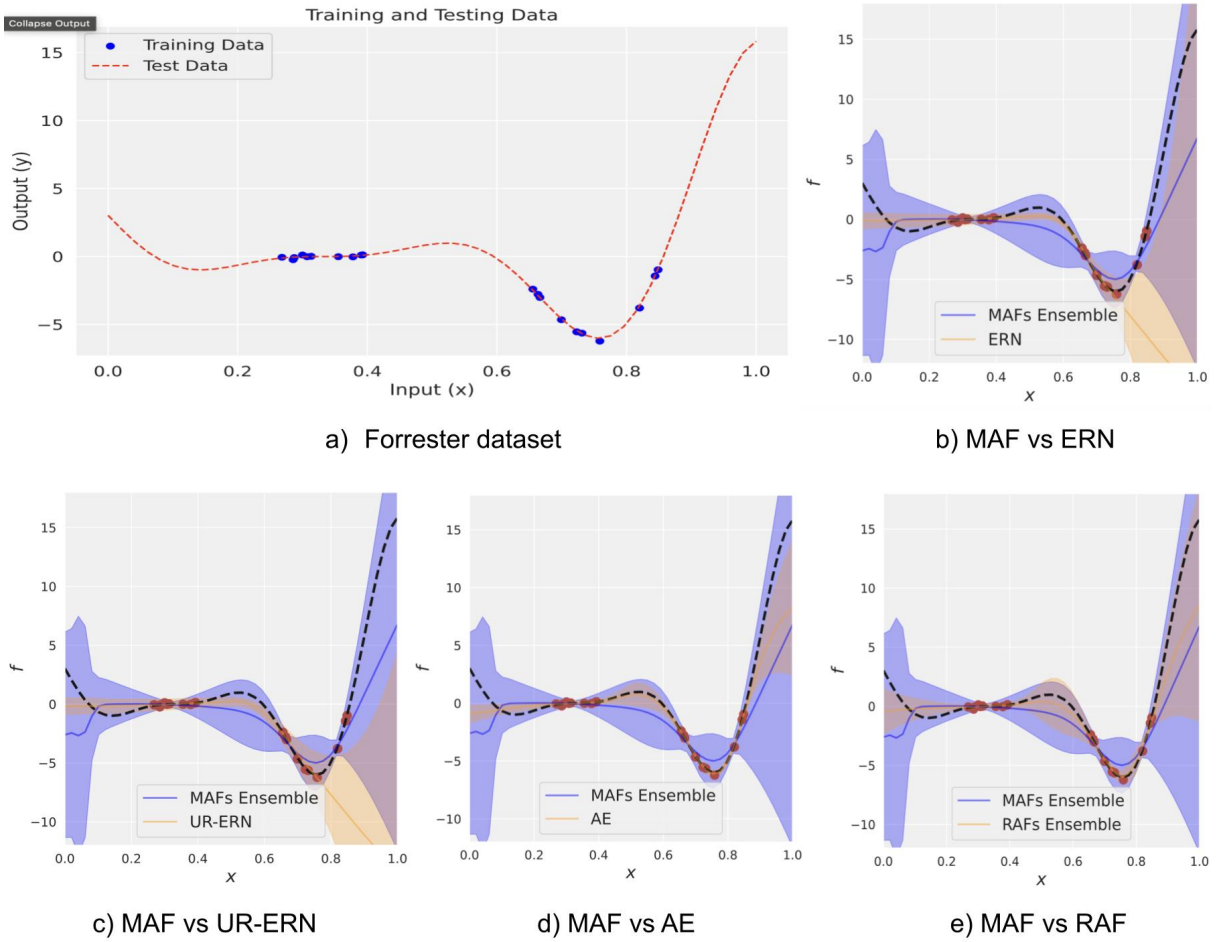


Figure 4.2: Results on the Forrester dataset. (a) Ground-truth function (red dashed) and training data (blue) restricted to subintervals. (b–c) ERN and UR-ERN are competitive, capturing epistemic uncertainty in $[0.85, 1]$ but underestimating it in $[0, 0.2]$. (d–e) AE and RAF underestimate uncertainty in unseen regions.

example, on He 1D, MAF (1.51 ± 0.90) surpasses RAF (4.05 ± 0.47) and the evidential models (both above 90). Similarly, on Double Pendulum 2D, MAF (11.09 ± 0.05) provides substantial improvements over AE (50.71 ± 0.04) and UR-ERN (25.84 ± 0.06).

For datasets with smoother predictive structures, MAF often ties with RAF. In Forrester 1D, both RAF (1.01 ± 0.73) and MAF (2.12 ± 1.90) are statistically indistinguishable, because their 95% confidence intervals overlap $[0.28, 1.74] \cap [0.22, 4.02] \neq \emptyset$. As a result, the difference between their means cannot be declared significant at the 95% confidence level, while the remaining methods yield much higher NLL values. On Schaffer 2D, MAF achieves the lowest NLL (-0.72 ± 0.01), with RAF (-0.54 ± 0.01) and AE (-0.53 ± 0.01) remaining competitive. These results confirm that MAF is robust even in scenarios where other methods also perform well.

In medium and high-dimensional synthetic tasks, MAF again outperforms. On Friedman 5D, both MAF (1.63 ± 0.40) and RAF (2.18 ± 0.44) outperform ERN (9.46 ± 0.47) and AE (22.60 ± 0.30). On Planar Arm Torque 6D, MAF (1.19 ± 0.15) and RAF (1.34 ± 0.10) deliver the best results. Conversely, datasets such as Rastrigin 3D and Styblinski–Tang 9D highlight scenarios where ERN and UR-ERN remain competitive.

Table 4.1: Predictive performance of different methods using NLL and 95% confidence intervals for 19 datasets. The best scores are highlighted in bold. Whenever confidence intervals overlap, multiple methods are considered statistically indistinguishable and are jointly bolded.

Dataset	ERN	UR-ERN	AE	RAF	MAF
He 1D	96.71 ± 0.54	94.51 ± 0.53	23.51 ± 0.14	4.05 ± 0.47	$\mathbf{1.51 \pm 0.90}$
Forrester 1D	4.11 ± 0.58	4.50 ± 0.62	5.85 ± 0.58	$\mathbf{1.01 \pm 0.73}$	$\mathbf{2.12 \pm 1.90}$
Schaffer 2D	-0.14 ± 0.01	-0.10 ± 0.01	-0.53 ± 0.01	-0.54 ± 0.01	$\mathbf{-0.72 \pm 0.01}$
Double pendulum 2D	20.94 ± 0.06	25.84 ± 0.06	50.71 ± 0.04	42.64 ± 0.05	$\mathbf{11.09 \pm 0.05}$
Rastrigin 3D	$\mathbf{3.87 \pm 1.28}$	$\mathbf{3.84 \pm 1.24}$	11.19 ± 0.78	12.45 ± 1.07	59.68 ± 1.32
Ishigami 3D	3.64 ± 0.08	4.55 ± 0.08	0.81 ± 0.08	$\mathbf{-0.20 \pm 0.09}$	0.38 ± 0.0
Environmental 4D	4.21 ± 0.41	4.75 ± 0.41	4.23 ± 0.37	4.87 ± 0.42	$\mathbf{2.85 \pm 0.40}$
Griewank 4D	$\mathbf{6.13 \pm 0.84}$	$\mathbf{6.33 \pm 0.69}$	97.67 ± 0.64	$\mathbf{3.30 \pm 3.04}$	$\mathbf{6.28 \pm 3.58}$
Roos & Arnold 5D	-1.02 ± 0.01	-1.04 ± 0.01	$\mathbf{-1.85 \pm 0.01}$	-1.78 ± 0.01	$\mathbf{-1.85 \pm 0.01}$
Friedman 5D	9.46 ± 0.47	11.10 ± 0.47	22.60 ± 0.30	$\mathbf{2.18 \pm 0.44}$	$\mathbf{1.63 \pm 0.40}$
Planar arm torque 6D	40.39 ± 0.10	18.83 ± 0.10	5.60 ± 0.05	$\mathbf{1.34 \pm 0.10}$	$\mathbf{1.19 \pm 0.15}$
Piston simulation 7D	9.69 ± 0.04	23.77 ± 0.06	$\mathbf{-0.69 \pm 0.02}$	-0.04 ± 0.04	$\mathbf{-0.67 \pm 0.03}$
Robot arm 8D	8.12 ± 0.04	12.89 ± 0.04	1.14 ± 0.02	$\mathbf{0.73 \pm 0.05}$	2.74 ± 0.01
Styblinski-Tang 9D	$\mathbf{8.66 \pm 0.43}$	12.21 ± 0.30	$> 100 \pm 2.41$	42.21 ± 4.93	77.70 ± 5.29
Adapted Welch 10D	$\mathbf{14.52 \pm 0.49}$	18.96 ± 0.48	$> 100 \pm 0.36$	16.29 ± 0.49	62.30 ± 0.59
Wing weight 10D	$\mathbf{7.35 \pm 0.91}$	8.53 ± 0.86	$> 100 \pm 2.13$	$\mathbf{4.45 \pm 2.87}$	10.82 ± 2.99
Abalone	3.15 ± 0.09	3.18 ± 0.10	$> 100 \pm 0.10$	69.50 ± 0.10	$\mathbf{1.72 \pm 0.12}$
Naval propulsion	-0.92 ± 0.00	-1.23 ± 0.00	$\mathbf{-2.29 \pm 0.00}$	-2.26 ± 0.00	-0.29 ± 0.00
Parkinson	$\mathbf{4.35 \pm 0.07}$	$\mathbf{4.35 \pm 0.07}$	$> 100 \pm 0.11$	$> 100 \pm 0.12$	45.13 ± 0.07

On real-world datasets, MAF shows consistent reliability. It achieves the lowest NLL on Abalone (1.72 ± 0.12), while on Parkinson, ERN and UR-ERN perform best, but MAF still improves over AE and RAF. In contrast, on Naval Propulsion, AE (-2.29 ± 0.00) and RAF (-2.26 ± 0.00) outperform MAF (-0.29 ± 0.00), suggesting that simpler models suffice in low-uncertainty regimes.

Overall, MAF achieves the best performance on 11 of 19 datasets and frequently ties with RAF or AE. These results establish MAF as a versatile and reliable method for uncertainty estimation, while underscoring the complementary roles of ensemble and evidential approaches across diverse predictive settings.

Following the same rationale used in the NLL evaluation, we report SMAPE values with their corresponding 95% confidence intervals, as adopted in previous studies [27, 116, 83]. The results in Table 4.2 provide a more balanced view compared to the NLL analysis. Across the 19 datasets, MAF achieves the best or joint-best performance in 9 cases, with AE attaining a similar number of wins. Notably, ties occur frequently due to overlapping confidence intervals, underscoring that performance differences among methods are often subtle rather than absolute.

On low-dimensional synthetic tasks, AE demonstrates strong performance. For instance, in Forrester and Schaffer, AE achieves the lowest SMAPE (0.86 ± 0.58 and 0.26 ± 0.01, respectively), outperforming all other methods with non-overlapping intervals. Similarly, AE shares the best performance in the Double Pendulum dataset together with RAF and MAF, reflecting its competitiveness in capturing predictive accuracy in non-linear dynamical systems. By contrast, MAF exhibits robust performance on Griewank

Table 4.2: Predictive performance of different methods using SMAPE and 95% CI for 19 datasets. The best SMAPE values are highlighted in bold, allowing ties when confidence intervals overlap.

Dataset	ERN	UR-ERN	AE	RAF	MAF
He	**1.18 ± 0.54**	1.23 ± 0.53	1.23 ± 0.14	1.25 ± 0.47	1.38 ± 0.90
Forrester	1.39 ± 0.58	1.42 ± 0.62	**0.86 ± 0.58**	1.13 ± 0.73	1.33 ± 1.90
Schaffer	0.32 ± 0.01	0.32 ± 0.01	**0.26 ± 0.01**	0.31 ± 0.01	0.39 ± 0.01
Double pendulum	1.38 ± 0.06	1.33 ± 0.06	**1.29 ± 0.04**	**1.29 ± 0.05**	**1.29 ± 0.05**
Rastrigin	**0.14 ± 1.28**	**0.14 ± 1.24**	0.28 ± 0.78	0.16 ± 1.07	**0.14 ± 1.32**
Ishigami	0.12 ± 0.08	0.12 ± 0.08	**0.11 ± 0.08**	0.15 ± 0.09	0.22 ± 0.09
Environmental	0.25 ± 0.41	0.25 ± 0.41	**0.19 ± 0.37**	0.20 ± 0.42	0.30 ± 0.40
Griewank	0.75 ± 0.84	0.79 ± 0.69	0.45 ± 0.64	0.16 ± 3.04	**0.12 ± 3.58**
Roos & Arnold	**2.00 ± 0.01**	**2.00 ± 0.01**	**2.00 ± 0.01**	**2.00 ± 0.01**	**2.00 ± 0.01**
Friedman	0.22 ± 0.47	0.25 ± 0.47	0.28 ± 0.30	0.16 ± 0.44	**0.15 ± 0.40**
Planar arm torque	**1.03 ± 0.10**	1.07 ± 0.10	1.09 ± 0.05	1.06 ± 0.10	1.06 ± 0.15
Piston simulation	0.71 ± 0.04	0.94 ± 0.06	0.79 ± 0.02	0.74 ± 0.04	**0.67 ± 0.03**
Robot arm	0.77 ± 0.04	0.79 ± 0.04	**0.72 ± 0.02**	0.81 ± 0.05	0.87 ± 0.02
Styblinski-Tang	1.85 ± 0.43	1.86 ± 0.30	1.80 ± 2.41	1.44 ± 4.93	**1.39 ± 5.29**
Adapted Welch	1.04 ± 0.49	1.04 ± 0.48	1.17 ± 0.36	1.06 ± 0.49	**0.98 ± 0.59**
Wing weight	0.75 ± 0.91	0.79 ± 0.86	0.28 ± 2.13	**0.22 ± 2.87**	0.27 ± 2.99
Abalone	**0.14 ± 0.09**	**0.14 ± 0.10**	0.15 ± 0.10	0.15 ± 0.10	0.18 ± 0.12
Naval propulsion	**0.01 ± 0.00**	**0.01 ± 0.00**	**0.01 ± 0.00**	**0.01 ± 0.00**	0.12 ± 0.00
Parkinson	0.37 ± 0.07	0.35 ± 0.07	**0.33 ± 0.11**	**0.33 ± 0.12**	**0.33 ± 0.07**

(0.12 ± 3.58), Friedman (0.15 ± 0.40), and Piston simulation (0.67 ± 0.03), where it either attains the best results outright or remains statistically indistinguishable from RAF.

At higher dimensions, the relative advantages shift among methods. MAF achieves the lowest SMAPE on Styblinski–Tang and Adapted Welch, while RAF attains its strongest results on Wing Weight (0.22 ± 2.87). Interestingly, in Planar Arm Torque, ERN (1.03 ± 0.10) delivers the best outcome, suggesting that simpler evidential methods may retain competitive accuracy in moderately complex scenarios. Likewise, both ERN and UR-ERN dominate the He dataset and tie on Abalone, highlighting that their regularization strategy remains effective under certain conditions.

The real-world datasets further illustrate the diversity of strengths. On Abalone, both ERN variants are statistically tied for the best performance, while in Naval Propulsion all methods except MAF converge to nearly identical, minimal SMAPE values (0.01 ± 0.00). For Parkinson dataset, AE, RAF, and MAF yield indistinguishable results, each achieving 0.33 with overlapping confidence intervals. These findings suggest that, in practical domains, performance differences among methods diminish, and several approaches can be considered equally reliable within statistical margins of error.

Overall, the analysis shows that no single method dominates across all datasets. MAF and AE stand out as the most competitive, each achieving the best performance in nearly half of the datasets. RAF and ERN also demonstrate notable strengths in specific contexts, while UR-ERN, although effective, less frequently achieves the top rank. MAF's primary strength is its consistent performance among the top methods across diverse scenarios, underscoring its robustness in predictive accuracy.

Taken together, the experimental evidence demonstrates that MAF consistently de-

livers strong performance across both synthetic and real-world settings. By adaptively tuning activation functions through a meta-learning mechanism, MAF ensembles achieve an effective balance between predictive accuracy and reliable uncertainty estimation. In contrast to ERN and UR-ERN, which tend to underestimate epistemic uncertainty in out-of-distribution regions, and AE or RAF, which show dataset-dependent variability, MAF remains robust across diverse tasks and dimensionalities. Its repeated ability to match or surpass competing methods in both NLL and SMAPE evaluations highlights its versatility.

Our approach, MAF introduces activation-level diversity as a new axis of variation within ensemble members. The key idea is to enhance ensemble diversity by allowing each neural network to dynamically modulate its activation functions through a convex combination of multiple base activations, with mixture weights determined by the input. This mechanism enables activation behavior to adapt continuously to the input context, thereby improving both flexibility and representational capacity.

Overall, these findings establish MAF as a competitive, general-purpose framework that advances ensemble methods by mitigating their typical limitations, particularly the challenge of insufficient diversity. By embedding activation-level diversity at the core of the ensemble design, MAF not only strengthens uncertainty estimation but also emerges as the most reliable technique among the compared approaches.

4.7 Case Study: Application to a Geological Reservoir

Building on insights from experiments with both synthetic and real-world datasets, we next evaluate the proposed methods in a realistic, high-dimensional application. Comparative analyses across these testbeds indicate that the MAF ensemble provides the most reliable balance between predictive accuracy and well-calibrated uncertainty estimates. Motivated by this evidence, we adopt the MAF ensemble as the primary approach for application to a real 3D geological reservoir. This case study not only validates the method under practical subsurface conditions but also demonstrates its potential to support decision-making in reservoir engineering, where predictive precision and robust uncertainty quantification are both critical.

Regression techniques are crucial for capturing the relationship between the simulator's inputs and outputs. The proxy model is designed to replicate the simulator's response for each producing well across all objective functions. Building on the approach proposed by Farfan *et al.* [2] and Maschio *et al.* [1], five representative moments are identified over a 2206-day production period: t_0, t_1, t_2, t_3, and t_4, corresponding to days 506, 1014, 1360, 1775, and 2156, respectively (Figure 4.3). These moments are strategically selected to capture key stages in the production process, representing both the early and later phases of the well's production.

This procedure is applied systematically to the 200 simulation models, ensuring that the entire production period is considered for each well. For each well, the analysis is conducted using four cumulative production objective functions: oil, water, gas, and pressure, providing a comprehensive assessment of well performance.

Figure 4.3-a) illustrates the cumulative oil production for well P10, shown across the entire 2206-day production period. Each curve represents the total cumulative production over time. Figure 4.3-b) shows the selection of five points (t_0, t_1, t_2, t_3, and t_4) for each of the 200 simulation models for well P10. These points, marked with red circles, are chosen to highlight significant milestones in the production process, reflecting the well's performance at these specific times.

Figure 4.3-c) displays the production profiles of seven producing wells (P10, P2, P3, P5, P6, P8, and P9). The x-axis represents the production times at t_{506} and t_{2156}, while the y-axis shows the cumulative production at these times. The production curves for each well are separated by spaces to emphasize their individual trajectories, clearly distinguishing the performance of each well.

In this work, we modify the end-to-end proxy originally proposed by Farfan *et al.* [2]. Specifically, the activation function is altered to incorporate a meta-network architecture, as described in Section 4.4. As shown in Figure 4.4, the module dedicated to dimensionality reduction takes as input a 3D geological uncertainty with dimensions of 72 $\times$ 48 $\times$ 24. This dimensionality reduction module is applied in parallel to the four geological uncertainties. In Figure 4.4-a), an overview of the different hierarchical levels is presented, which specifically includes 3D block encoders, 3D max pooling, and a 3D residual block.

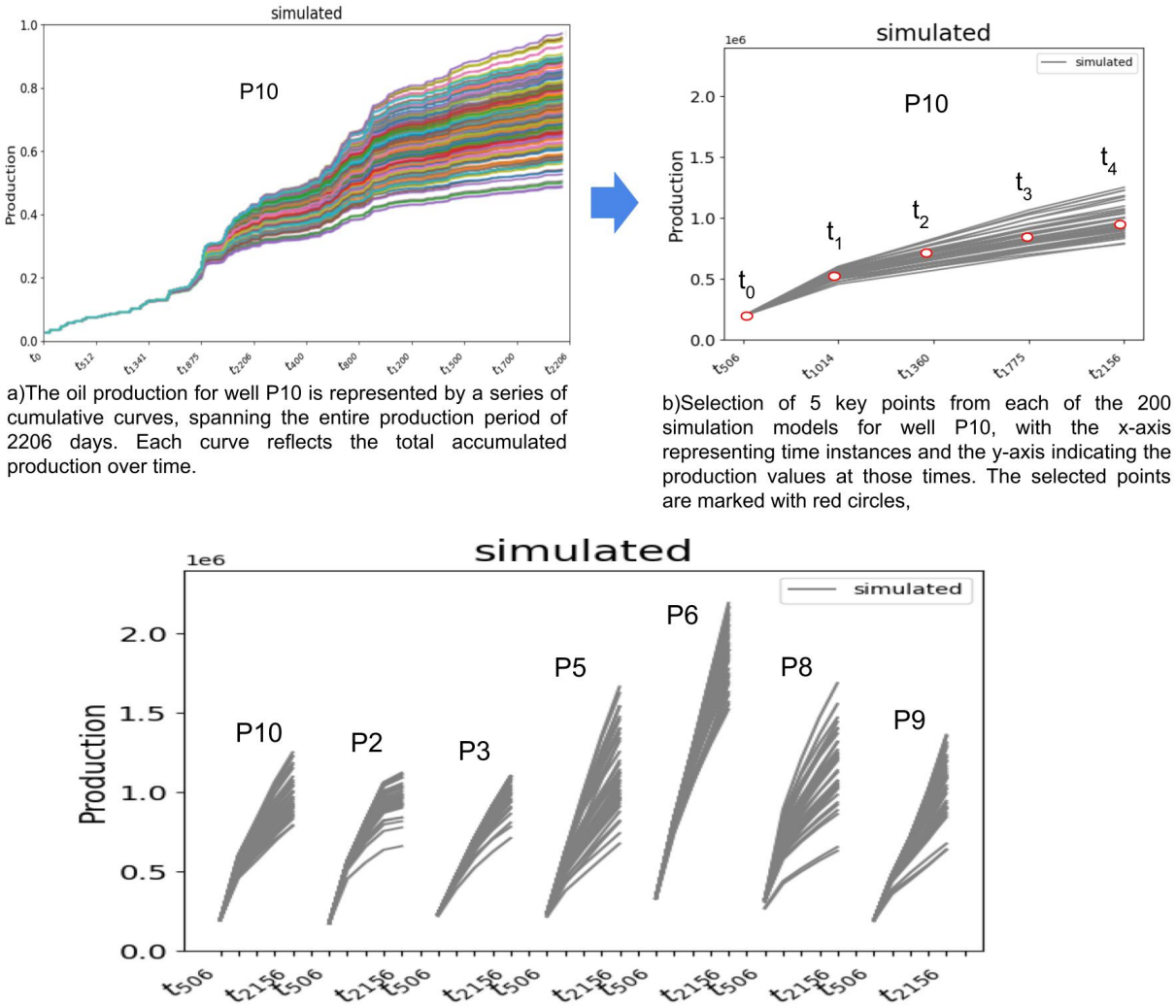
Regarding the encoder block, Figure 4.4-b) shows that this module consists of a series of convolutional layers, batch normalization, and additional operations. Notably, the ReLU activation function originally proposed by Farfan *et al.* [2] is replaced with a meta-network. The primary purpose of this sequence is to extract and learn rich representations and features from the input data.

In Figure 4.4-c), the residual block preserves information from previous layers and introduces shortcut connections to facilitate efficient gradient flow. This block includes convolutional layers and batch normalization, and the activation functions are similarly replaced with a meta-network to enhance adaptability.

Figure 4.4-d) also provides a detailed breakdown of the neural meta-network that integrates DenseNet and Attention mechanisms. The architecture begins with a dense layer of 512 neurons applied to the input. Each block starts with Batch Normalization to standardize the inputs and stabilize learning, followed by Dropout to regularize the network and prevent overfitting through random deactivation of neurons. The outputs of these operations are concatenated with the original 512-neuron input layer and passed through a dense layer of 256 neurons, ensuring the propagation of feature-rich representations throughout the network.

Subsequently, Batch Normalization and Dropout are applied again. The resulting output is concatenated with the original input and all previous outputs. A dense layer of 128 neurons is then applied, followed by another Batch Normalization and Dropout. Finally, the outputs of these operations are concatenated with the original 512-neuron input, the 512-neuron output, and the 256-neuron output, maintaining a rich flow of information across the network.

Following the DenseNet block, an Attention mechanism is applied to further enhance feature representation. The Attention Layer implements scaled dot-product attention, where the input is projected into query (Q), key (K), and value (V) vectors through



a)The oil production for well P10 is represented by a series of cumulative curves, spanning the entire production period of 2206 days. Each curve reflects the total accumulated production over time.

b)Selection of 5 key points from each of the 200 simulation models for well P10, with the x-axis representing time instances and the y-axis indicating the production values at those times. The selected points are marked with red circles,

c)Production trends for each of the producing wells (P10, P2, P3, P5, P6, P8, P9) are depicted from left to right, with spaces between each well to distinguish their individual production curves. The x-axis represents the time frame of the production period, marked by time steps (t506 and t2156), while the y-axis illustrates the production levels at each corresponding time step

Figure 4.3: Overview of cumulative production profiles for seven producing wells across a 2,206-day period. (a) Cumulative oil production for well P10, with each curve representing a different simulation model, showing the variation in production over time. (b) Selection of five key time instances (t_0, t_1, t_2, t_3, t_4) for well P10, representing critical production milestones across the simulation models. (c) Production trends for each of the seven producing wells (P10, P2, P3, P5, P6, P8, P9), with well production data separated by spaces to distinguish each well's trajectory.

learned dense layers. Attention weights are computed as the softmax of the scaled dot product of Q and K, capturing dependencies across all input features. These weights are then used to combine the value vectors, producing a context-aware representation that emphasizes the most informative features and relationships. In this model, the attention mechanism receives the output of the DenseNet block as query, key, and value, enabling the network to learn which features are most relevant for generating the final predictions.

Finally, the output of the attention mechanism is passed through a dense layer with six neurons and a softmax activation to produce the final predictions. These predictions

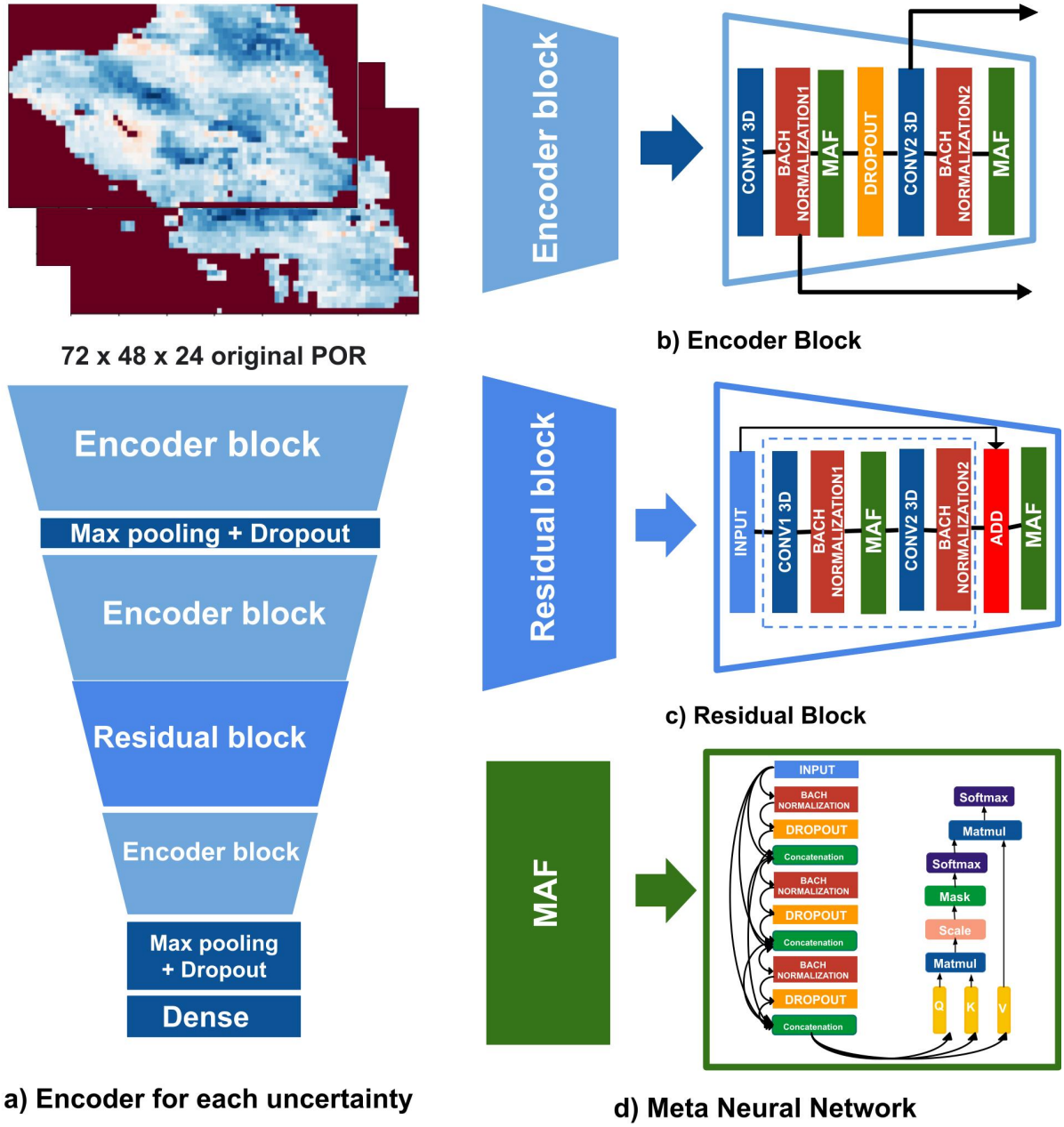


Figure 4.4: Overview of the modified end-to-end proxy architecture for geological uncertainty. a) The dimensionality reduction module processes 3D geological data using 3D block encoders, max pooling, and residual blocks. b) The encoder block, replacing ReLU with a meta-network, extracts rich feature representations. c) The residual block enhances gradient flow through shortcut connections and meta-network activations. d) The neural meta-network integrates DenseNet and Attention mechanisms, utilizing dense layers, Batch Normalization, and Dropout, with the Attention mechanism refining feature representation before the final softmax-based prediction layer.

correspond to relevance scores or probabilities, depending on the application, and can guide downstream tasks such as adaptive activation function selection or other meta-network objectives.

4.7.1 End-to-End Proxy: Uncertainty Quantification

In this part, we evaluate the performance of our proposed model against the state-of-the-art approach proposed by Farfan *et al.* [2]. Specifically, we focus on the prediction of oil, water, gas, and bottom-hole pressure (BHP) within the context of reservoir simulation. The model's predictive capability is assessed in terms of its ability to reproduce the simulator responses. To ensure consistency, we employ the same evaluation metrics used in [2], namely the Correlation Coefficient (R), the Coefficient of Determination (R^2), and the Symmetric Mean Absolute Percentage Error (SMAPE). Finally, after establishing the robustness of our method, we evaluate its uncertainty reliability using the Negative Log-Likelihood (NLL) and through the visualization of its uncertainty bands.

According to the Table 4.3, our method demonstrates competitive performance in predicting oil production, achieving an SMAPE of 3.119, a R of 0.975, and an R^2 value of 0.951. These results indicate a strong linear relationship between the predicted and actual values, with the model able to capture the characteristic dynamics of oil production. In comparison, the state-of-the-art 3D CNN model achieved an SMAPE of 3.413, an R of 0.970, and an R^2 of 0.939. Our method slightly outperforms this CNN model in terms of accuracy, suggesting that our model provides more precise oil production predictions with lower error.

Table 4.3: Comparative evaluation of our approach with state-of-the-art method focusing on oil and water-related objective functions.

	Oil			Water		
Work	SMAPE	R	R^2	SMAPE	R	R^2
Our approach	**3.119**	**0.975**	**0.951**	**18.326**	**0.965**	**0.930**
Farfan *et al.* [2]	3.413	0.970	0.939	19.437	0.954	0.907

Similarly, in the Table 4.3, for water production, our model yields an SMAPE of 18.326, an R of 0.965, and an R^2 of 0.930, which indicates a higher error than the oil predictions, yet still offers a strong correlation. In contrast, the state-of-the-art method reports an SMAPE of 19.437, an R of 0.954, and an R^2 of 0.907. Although our model shows a lower SMAPE, it performs better in capturing water production trends compared to the leading approach.

For gas production in Table 4.4, our model achieves an SMAPE of 3.163, an R of 0.976 and an R^2 of 0.951, which is comparable to the state-of-the-art model that achieved an SMAPE of 3.461, an R of 0.969, and an R^2 of 0.939. Both models demonstrate high correlation values and R^2, reflecting robust predictive performance. The slight differences in SMAPE and R^2 suggest that both approaches perform similarly, though our model offers marginally better predictions.

In terms of BHP predictions in the Table 4.4, our model reports an SMAPE of 0.99, an R of 0.891, and an R^2 of 0.792. This indicates a strong correlation but less precise fitting compared to the other objective functions. While predicting BHP remains a challenge for both models, our approach still performs favorably compared to the state-of-the-art

Table 4.4: Comparative evaluation of our approach with state-of-the-art method focusing on gas and pressure-related objective functions.

	Gas			Pressure		
Work	SMAPE	R	R^2	SMAPE	R	R^2
Our approach	**3.163**	**0.976**	**0.951**	**0.99**	**0.891**	**0.792**
Farfan *et al.* [2]	3.461	0.969	0.939	1.055	0.880	0.765

method. This suggests that while the model effectively captures the overall trends in BHP production, further refinement is necessary to improve prediction accuracy.

The UQ of the proxy model predictions are illustrated in Figure 4.5, Figure 4.6, Figure 4.7 and Figure 4.8, where the mean of the proxy prediction ensembles (green) is compared with the simulator outputs (gray) for the seven producing wells: P10, P2, P3, P5, P6, P8, and P9. The blue regions represent the uncertainty bands generated by the ensemble using the MAF method, which capture the variability of the model predictions over time.

In Figure 4.5, the proxy model accurately reproduces cumulative oil production across all wells, demonstrating strong agreement with the simulator. The uncertainty bands are generally narrow and stable, indicating high confidence in the proxy's predictions. However, wells P10, P3, P5, and P8 exhibit slightly wider bands during specific production phases, suggesting increased uncertainty under dynamic flow conditions. In contrast, wells P2, P6, and P9 display more compact uncertainty intervals that remain well-aligned with the simulated curves, reflecting more predictable production behaviors.

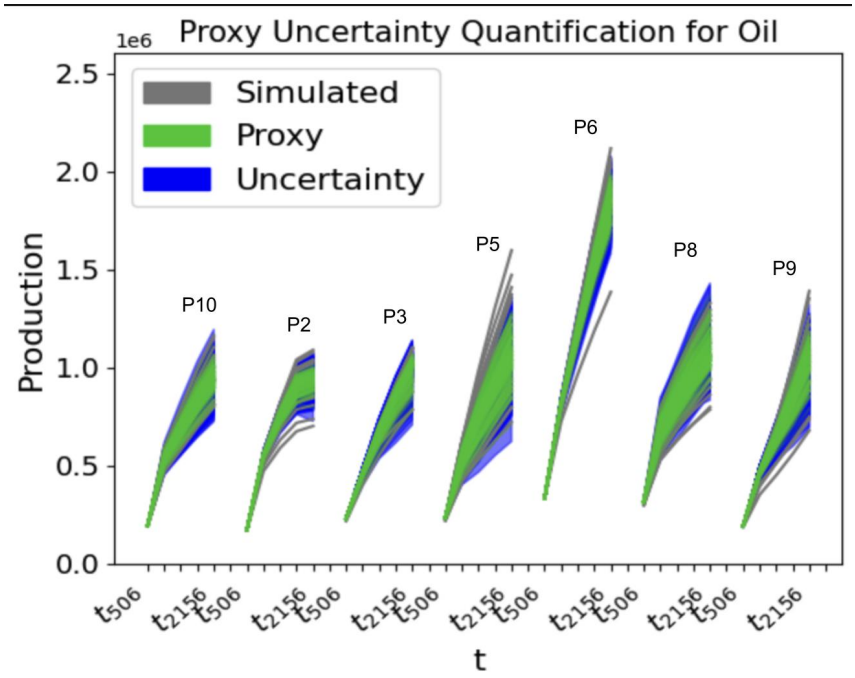


Figure 4.5: Oil production uncertainty quantification for the seven producing wells (P10, P2, P3, P5, P6, P8, and P9)

In Figure 4.6, the water production profiles show greater variability, particularly for wells P10, P3, P5, and P8, where the uncertainty bands are noticeably wider. This

pattern is consistent with reservoir heterogeneity, which introduces stronger nonlinearities. Nevertheless, the proxy effectively captures the overall production trends, maintaining reliable estimates despite the larger uncertainty ranges.

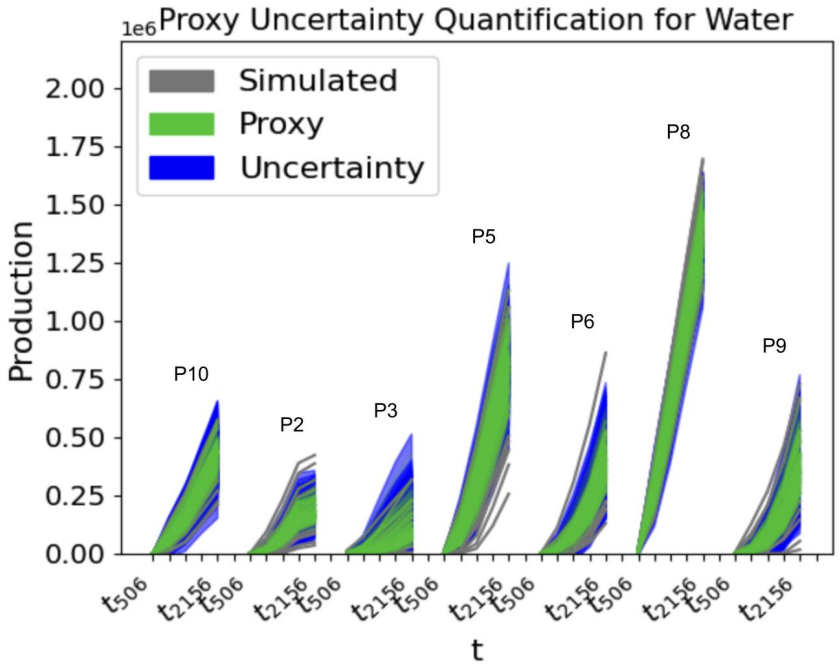


Figure 4.6: Water production uncertainty quantification for the seven producing wells (P10, P2, P3, P5, P6, P8, and P9)

The gas production results (Figure 4.7) also show a close match between the proxy and simulator across all wells. Slightly wider uncertainty bands are observed in wells P10, P3, P5, and P8, indicating sensitivity to gas expansion and compressibility effects. Even so, the proxy maintains strong predictive capability and captures global gas production trends with robustness.

In Figure 4.8, the BHP curves exhibit more pronounced variability and local fluctuations compared to the production variables. Wells P5, P6, and P9 present wider uncertainty bands, reflecting the proxy's higher uncertainty in modeling transient pressure responses under complex reservoir conditions. Despite this, the proxy ensemble successfully reproduces the overall BHP dynamics, validating its capacity to quantify pressure-related uncertainty.

Across the oil, water, and gas production objective functions, wells P10, P3, P5, and P8 consistently show higher uncertainty, which can be attributed to their higher production rates and more dynamic reservoir interactions. Conversely, wells P2, P6, and P9 display narrower uncertainty intervals, indicating more stable and predictable behaviors. For BHP, the highest uncertainty is observed in wells P5, P6, and P9 due to transient effects, while wells P10, P3, P5, and P8 exhibit smoother uncertainty bands, suggesting steadier pressure responses.

The predictive performance of the UQ, assessed using the NLL and its 95% confidence intervals, is summarized in Table 4.5. The absolute magnitude of the NLL values (ranging from 38.03 to 119.42) is significantly higher than the low single-digit scores typically

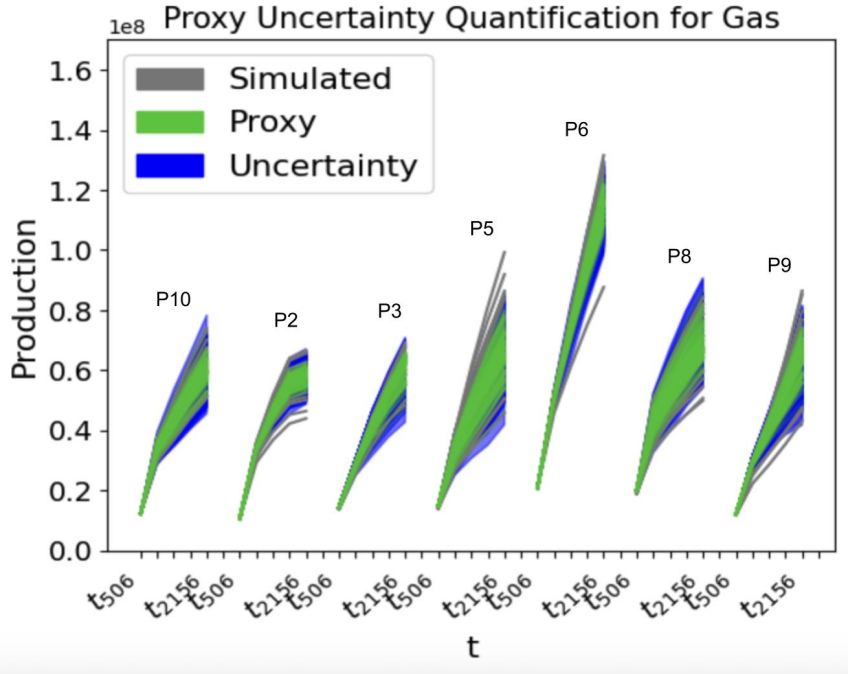


Figure 4.7: Gas production uncertainty quantification for the seven producing wells (P10, P2, P3, P5, P6, P8, and P9)

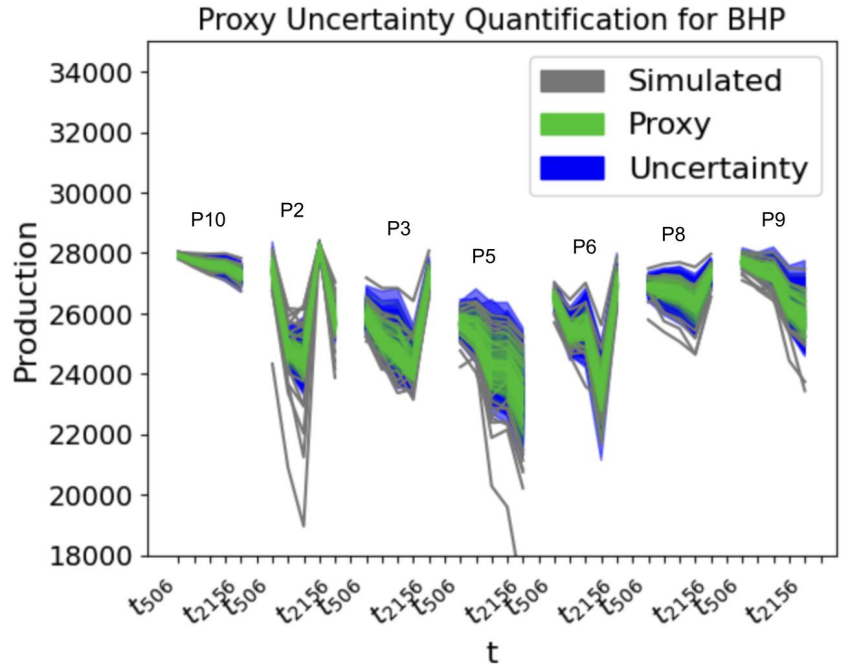


Figure 4.8: Pressure production uncertainty quantification for the seven producing wells (P10, P2, P3, P5, P6, P8, and P9)

reported for the synthetic datasets explored, as shown in Table 4.1. This increase is directly attributable to the calculations performed on the raw, unscaled production period. As evidenced by the production scale on the y-axis of Figure 4.5 (reaching values up to 1.5×10^6), the absolute variance (σ^2) of the production period is substantial. A high value of σ^2 results in a large positive term for $\log(\sigma^2)$ in the NLL formulation, which

mathematically raises the overall metric well above the normalized reference value seen in the synthetic datasets.

Table 4.5: Predictive performance of different objective functions using NLL and 95% confidence intervals.

Oil	**Water**	**Gas**	**BHP**
119.42 ± 0.10	99.19 ± 0.03	53.93 ± 1.02	38.03 ± 0.08

In summary, the baseline end-to-end proxy model proposed by Farfan *et al.* [2] was enhanced by introducing an ensemble-based MAF approach. This improvement allowed for more accurate replication of the simulator response, as evidenced by superior performance in terms of SMAPE, R, and R^2. Furthermore, the reliability of the results was assessed by estimating the uncertainty of the MAF-based end-to-end proxy using the NLL and visualizing the uncertainty bands for the oil, water, gas, and BHP objective functions across the wells P10, P2, P3, P5, P6, P8, and P9. These results demonstrate the model's capacity to handle the complexity of the reservoir system while effectively capturing both the mean responses and the inherent variability of the data.

4.8 Final Remarks

We introduced MAF Ensemble, a novel approach to model uncertainty estimation for neural networks. By enhancing the diversity within the ensemble, each neural network is allowed to adapt through a convex combination of multiple activation functions dynamically, weighted according to their relevance to the input. This adaptation mechanism is firmly grounded in meta-learning principles and is successfully implemented through the integration of a meticulously designed meta-network.

Our extensive empirical study demonstrates that the proposed method outperforms four state-of-the-art uncertainty quantification methods on a range of regression tasks, utilizing 19 datasets (16 synthetic and 3 real-world). Furthermore, we tested the robustness of our approach in the context of a Brazilian offshore field, specifically by modifying the end-to-end proxy model proposed by Farfan *et al.* [2] to incorporate the meta-network in place of the ReLU activation function. The results show that our approach not only maintains high predictive accuracy but also introduces a reliable estimation of uncertainty, marking a significant advancement in the proxy models used for reservoir management.

In comparison to the original Farfan *et al.* [2] method, our approach achieves improved performance in terms of SMAPE, R, and R^2 metrics. More importantly, it provides a detailed uncertainty quantification, which is crucial for the practical application of proxies in reservoir management, thus contributing a novel layer of reliability to the existing literature on proxy-based models.

Chapter 5

Conclusions and Future Work

Simulation models play a crucial role in understanding and managing oilfields. Their importance lies in their ability to simulate reservoir behavior over time. These models are typically high-dimensional, heterogeneous, and characterized by strong nonlinearities. Despite significant advances in numerical simulation technologies, simulation models remain computationally intensive, often requiring days or even weeks to generate a single reservoir simulation. Consequently, only a limited number of simulation models are available in practical reservoir studies. One approach to address this challenge is the use of proxy models that replace the numerical simulator in time-consuming steps.

In this thesis, we presented our proxy-based research, focusing on various phases within the reservoir management process that can be optimized, including replicating the simulator's response. We began by reducing the dimensionality of 3D geological uncertainties and replicating the complex, nonlinear behavior observed in various cumulative production curves across different producing wells. We also explored methods for reconstructing the data back to the original 3D geological uncertainty space, ensuring analytical consistency and compatibility with simulator input requirements. Finally, we addressed an often overlooked aspect of proxy model development: the quantification of prediction uncertainty. By dividing our work into these three studies, we gained a comprehensive understanding of proxy models and their application in reservoir management. Moreover, our results advance the state of the art on several fronts.

In our first approach, we introduced an end-to-end deep neural network that can handle 3D geological uncertainties and replicate the simulator response for various cumulative production curves. This approach leveraged 3D convolutions to process spatial and depth dimensions in heterogeneous reservoirs, incorporating various geostatistical realizations. The comprehensive solution was successfully implemented, accurately replicating the simulator's behavior and outperforming state-of-the-art methods. Our results showed correlations greater than 0.95 between the proxy response and the numerical simulator. Furthermore, our approach demonstrated robustness even when trained with a limited number of simulation models, significantly reducing computational costs. These findings highlighted our approach's ability to integrate dimensionality reduction and regression analysis into a unified framework, effectively predicting different fluid behaviors in the reservoir while remaining robust to high-dimensional and sparse data.

In our second approach, we presented a proxy reconstruction method for complex

geological uncertainties used a three-dimensional Deep Residual Variational Autoencoder (DRVAE). This technique uses multimodal properties and a multitasking approach to integrate 3D rock properties (such as porosity and permeability) and fluid properties into a single architecture. Experimental results demonstrated the method's high accuracy in reconstructing geological uncertainties with robust geological consistency. Evaluation metrics, including SSIM greater than 0.95, FID of 0.587, and PSNR of 16.980, confirmed the high quality of the reconstruction.

In our third approach, we introduced a method for uncertainty quantification (UQ) based on ensembles and meta-learning, called Meta Activation Functions (MAF) Ensemble. The key idea was to enhance ensemble diversity by enabling each neural network to adapt dynamically through a convex combination of multiple activation functions, weighted by their relevance to the input. This design allowed every neuron in the ensemble to adjust its nonlinearity contextually, improving robustness in predictive uncertainty. In an extensive empirical study, MAF achieved lower negative log-likelihood (NLL) than state-of-the-art UQ methods on both synthetic and real-world regression tasks.

Through the experiments presented in this thesis, we can answer the research questions introduced in Chapter 1:

5.1 Revisiting the Research Questions

This thesis focused on two primary objectives, as outlined in the research questions presented in Chapter 1, Section 1.4. In light of the challenges, proposed solutions, results, and findings, we now revisit each question to illustrate how they have been addressed.

RQ1: Given a set of 3D geological models, how can dimensionality reduction be effectively applied and integrated into an end-to-end proxy model?

To address this research question, we introduced an integrated end-to-end proxy model in Chapter 2 that leverages a 3D convolutional neural network (CNN) to handle the dimensionality reduction of complex geological data. Our approach not only efficiently reduced dimensionality but also preserved critical spatial information, ensuring that the model could directly process and output the reduced data in a seamless end-to-end workflow. This integration helps overcome the high-dimensional challenges inherent in geological models while maintaining the integrity of the data.

RQ2: Given a set of dimensionally-reduced geological models and various production rates, how can the proxy replicate the simulator's response across these rates as part of an end-to-end solution?

In Chapter 2, we demonstrated how our end-to-end proxy effectively replicated the responses of traditional simulators across multiple production rates (oil, water, gas, and pressure) using a single Neural Network (NN). By adopting an integrated framework, this end-to-end model captured the complex interdependencies between these fluids, ensuring that the predictions were consistent with the actual reservoir behavior. This unified approach not only streamlined the modeling process but also ensured that the proxy could

seamlessly integrate dimensionality reduction and simulation response replication into a cohesive and efficient system.

RQ3: Given a set of dimensionally-reduced geological models, how can the reconstruction of the original 3D reservoir model be performed in a way that preserves the maximum amount of the reservoir's spatial information, enabling its incorporation into an end-to-end proxy?

To address this research question, we developed a reconstruction module in Chapter 3 that projects the reduced-dimensional data back into the original 3D geological space. This process preserved the spatial relationships between critical geological properties, ensuring that the reconstructed model remained as accurate as possible for real-world applications. By integrating this reconstruction step into our end-to-end proxy, we ensured that the entire workflow—from dimensionality reduction to model reconstruction—was seamlessly connected, enabling efficient processing and maintaining the integrity of the geological data throughout the pipeline.

RQ4: Given an end-to-end proxy, how can we ensure the reliability and robustness of their predictions to build trust among petroleum engineering experts?

In Chapter 4, we presented an innovative uncertainty quantification methodology using Meta Activation Functions (MAF) Ensemble to assess the confidence in our end-to-end proxy model predictions. By quantifying uncertainty, we ensured that the proxy model's outputs could be trusted and effectively used for decision-making in reservoir management.

5.2 Future Works

Several directions can be pursued to extend and strengthen the contributions of this thesis. First, future studies may explore alternative evaluation metrics to address known limitations of SMAPE, particularly its sensitivity to low-magnitude values and asymmetric error penalization. Incorporating complementary metrics could provide a more robust and nuanced assessment of reconstruction performance.

Second, the reconstruction task can be further investigated by evaluating other deep neural network architectures beyond those considered in this work. In particular, architectures with different inductive biases—such as attention-based models, hybrid convolutional–recurrent networks, or diffusion-inspired approaches—may offer improvements in accuracy, robustness, or uncertainty representation.

Another important direction is the validation of the proposed methodology within a data assimilation framework. Integrating the reconstructed data into sequential or variational data assimilation schemes would allow a more comprehensive assessment of its impact on state estimation, uncertainty reduction, and decision-making in reservoir management.

Finally, the MAF ensemble approach can be extended beyond reconstruction to address classification and forecasting problems in oil reservoirs. Potential applications in-

clude regime classification, anomaly detection, and production forecasting under uncertainty, enabling a broader evaluation of the methodology in operationally relevant scenarios.

Bibliography

[1] Célio Maschio, Guilherme Daniel Avansi, and Denis José Schiozer. Data assimilation using principal component analysis and artificial neural network. *SPE Reservoir Evaluation & Engineering*, pages 1–18, 2023.

[2] Jeanfranco Farfan, Gabriel Cirac, Guilherme Daniel Avansi, Célio Maschio, Denis José Schiozer, and Anderson Rocha. End-to-end dimensionality reduction and regression from 3d geological uncertainties to estimate oil reservoir simulations. *Applied Soft Computing*, 162:111799, 2024.

[3] Hosein Zanbouri and Karim Salahshoor. Development of robust surrogate model for economic performance prediction of oil reservoir production under waterflooding process. *Journal of Petroleum Science and Engineering*, 165:496–504, 2018.

[4] Oleg Sudakov, Dmitri Koroteev, Boris Belozerov, and Evgeny Burnaev. Artificial neural network surrogate modeling of oil reservoir: a case study. In *Advances in Neural Networks–ISNN 2019: 16th International Symposium on Neural Networks, ISNN 2019, Moscow, Russia, July 10–12, 2019, Proceedings, Part II 16*, pages 232–241. Springer, 2019.

[5] Seoyoon Kwon, Gayoung Park, Youngho Jang, Jinhyung Cho, Min-gon Chu, and Baehyun Min. Determination of oil well placement using convolutional neural network coupled with robust optimization under geological uncertainty. *Journal of Petroleum Science and Engineering*, 201:108118, 2021.

[6] Rasool Amiri Kolajoobi, Hosein Haddadpour, and Mohammad Emami Niri. Investigating the capability of data-driven proxy models as solution for reservoir geological uncertainty quantification. *Journal of Petroleum Science and Engineering*, 205: 108860, 2021.

[7] GW Thomas. The role of reservoir simulation in optimal reservoir management. In *SPE International Oil and Gas Conference and Exhibition in China*, pages SPE–14129. SPE, 1986.

[8] Turgay Ertekin, Jamal Hussein Abou-Kassem, and Gregory R King. Basic applied reservoir simulation. *(No Title)*, 2001.

[9] Masoud Maleki, Shahram Danaei, Felipe Bruno Mesquita da Silva, Alessandra Davolio, and Denis José Schiozer. Stepwise uncertainty reduction in time-lapse

seismic interpretation using multi-attribute analysis. *Petroleum Geoscience*, 27(3): petgeo2020–087, 2021.

[10] Gabriel Cirac, Jeanfranco Farfan, Guilherme Daniel Avansi, Denis José Schiozer, and Anderson Rocha. Cross-domain feature learning and data augmentation for few-shot proxy development in oil industry. *Applied Soft Computing*, page 110972, 2023.

[11] Gabriel Cirac, Jeanfranco Farfan, Guilherme Daniel Avansi, Denis José Schiozer, and Anderson Rocha. Deep hierarchical distillation proxy-oil modeling for heterogeneous carbonate reservoirs. *Engineering Applications of Artificial Intelligence*, 126:107076, 2023.

[12] Sameer Salasakar, Sabyasachi Prakash, and Ganesh Thakur. Recent trends in proxy model development for well placement optimization employing machine learning techniques. *Modelling*, 5(4):1808–1823, 2024.

[13] Célio Maschio, Guilherme Daniel Avansi, Felipe Bruno Mesquita da Silva, and Denis José Schiozer. Data assimilation for uncertainty reduction using different fidelity numerical models. *Journal of Petroleum Science and Engineering*, 209:109851, 2022.

[14] Victoria Volodina and Peter Challenor. The importance of uncertainty quantification in model reproducibility. *Philosophical Transactions of the Royal Society A*, 379(2197):20200071, 2021.

[15] Cristina Celia Barros Cavalcante. *Learning-from-data approaches to the history-matching problem= Abordagens de aprendizado a partir de dados para o problema de ajuste de histórico*. PhD thesis, [sn], 2020.

[16] Alexandre A Emerick and Albert C Reynolds. History-matching production and seismic data in a real field case using the ensemble smoother with multiple data assimilation. In *SPE Reservoir Simulation Conference?*, pages SPE–163675. SPE, 2013.

[17] Daiane Rossi Rosa, Denis Jose Schiozer, and Alessandra Davolio. Impact of model and data resolutions in 4d seismic data assimilation applied to an offshore reservoir in brazil. *Journal of Petroleum Science and Engineering*, 216:110830, 2022.

[18] Daiane Rossi Rosa, Denis Jose Schiozer, and Alessandra Davolio. Evaluating the impact of 4d seismic data artefacts in data assimilation. *Petroleum Geoscience*, 29 (3):petgeo2022–069, 2023.

[19] Daiane Rossi Rosa, Denis José Schiozer, and Alessandra Davolio. Data assimilation of production and multiple 4d seismic acquisitions in a deepwater field using ensemble smoother with multiple data assimilation. *SPE Reservoir Evaluation & Engineering*, pages 1–13, 2023.

[20] Luciana Maria Da Silva, Leandro Mendes Ferreira, Guilherme Daniel Avansi, Denis José Schiozer, and Solange Nice Alves-Souza. Selection of a dimensionality reduction method: An application to deal with high-dimensional geostatistical realizations in oil reservoirs. *SPE Reservoir Evaluation & Engineering*, 26(02):392–410, 2023.

[21] Joonyi Kim, Hyungjun Yang, and Jonggeun Choe. Robust optimization of the locations and types of multiple wells using cnn based proxy models. *Journal of Petroleum Science and Engineering*, 193:107424, 2020.

[22] Md Asif Bin Syed and Imtiaz Ahmed. A cnn-lstm architecture for marine vessel track association using automatic identification system (ais) data. *arXiv preprint arXiv:2303.14068*, 2023.

[23] Liping Wang. Study on reservoir heterogeneity in block s. In *IOP Conference Series: Earth and Environmental Science*, volume 770, page 012020. IOP Publishing, 2021.

[24] Xiaopeng Ma, Kai Zhang, Hanjun Zhao, Liming Zhang, Jian Wang, Huaqing Zhang, Piyang Liu, Xia Yan, and Yongfei Yang. A vector-to-sequence based multilayer recurrent network surrogate model for history matching of large-scale reservoir. *Journal of Petroleum Science and Engineering*, 214:110548, 2022.

[25] Wenjie Luo, Bin Yang, and Raquel Urtasun. Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 3569–3577, 2018.

[26] Yuxuan Wang, RJ Skerry-Ryan, Daisy Stanton, Yonghui Wu, Ron J Weiss, Navdeep Jaitly, Zongheng Yang, Ying Xiao, Zhifeng Chen, Samy Bengio, et al. Tacotron: Towards end-to-end speech synthesis. *arXiv preprint arXiv:1703.10135*, 2017.

[27] Yana Stoyanova, Soroush Ghandi, and Maryam Tavakol. Toward robust uncertainty estimation with random activation functions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 15152–15160, 2023.

[28] Guilherme Daniel Avansi Denis José Schiozer Jeanfranco Farfan, Gabriel Cirac and Anderson Rocha. Multimodal approach to reconstructing geological uncertainties using a deep residual variational autoencoder. *Computers Geosciences Journal*, 2025.

[29] Alexandre Ferreira Jeanfranco Farfan, Gabriel Cirac and Anderson Rocha. Meta-learned activation functions for uncertainty quantification in reservoir modeling proxies. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

[30] Dong Dong and Thomas J McAvoy. Nonlinear principal component analysis—based on principal curves and neural networks. *Computers & Chemical Engineering*, 20 (1):65–78, 1996.

[31] Lunning ZHANG, RONG Ming, HE Wei, ZHU Keying, and ZHAO Qing. Study on gas concentration prediction of tunnel engineering based on umap-bilstm. 2022.

[32] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[33] Jeffrey L Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.

[34] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[35] Nanzhe Wang, Haibin Chang, Xiang-Zhao Kong, and Dongxiao Zhang. Deep learning based closed-loop well control optimization of geothermal reservoir with uncertain permeability. *Renewable Energy*, 211:379–394, 2023.

[36] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[37] Diego Cantorna, Carlos Dafonte, Alfonso Iglesias, and Bernardino Arcay. Oil spill segmentation in sar images using convolutional neural networks. a comparative analysis with clustering and logistic regression algorithms. *Applied Soft Computing*, 84: 105716, 2019.

[38] Muhammad Turyalai Khan. A modified convolutional neural network with rectangular filters for frequency-hopping spread spectrum signals. *Applied Soft Computing*, 150:111036, 2024.

[39] Mohammad Hassan Sharifinasab, Mohammad Emami Niri, and Milad Masroor. Developing gan-boosted artificial neural networks to model the rate of drilling bit penetration. *Applied Soft Computing*, 136:110067, 2023.

[40] Lian Wang, Yuedong Yao, Xiaodong Luo, Caspar Daniel Adenutsi, Guoxiang Zhao, and Fengpeng Lai. A critical review on intelligent optimization algorithms and surrogate models for conventional and unconventional reservoir production optimization. *Fuel*, 350:128826, 2023.

[41] Xiaopeng Ma, Kai Zhang, Jinding Zhang, Yanzhong Wang, Liming Zhang, Piyang Liu, Yongfei Yang, and Jian Wang. A novel hybrid recurrent convolutional network for surrogate modeling of history matching and uncertainty quantification. *Journal of Petroleum Science and Engineering*, 210:110109, 2022.

[42] Stephanie C Hicks, Kwame Okrah, Joseph N Paulson, John Quackenbush, Rafael A Irizarry, and Héctor Corrada Bravo. Smooth quantile normalization. *Biostatistics*, 19(2):185–198, 2018.

[43] SGOPAL Patro and Kishore Kumar Sahu. Normalization: A preprocessing stage. *arXiv preprint arXiv:1503.06462*, 2015.

[44] Nikita Singhal, Vaishali Ganganwar, Menka Yadav, Asha Chauhan, Mahender Jakhar, and Kareena Sharma. Comparative study of machine learning and deep learning algorithm for face recognition. *Jordanian Journal of Computers and Information Technology*, 7(3), 2021.

[45] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[46] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[47] Hansika Hewamalage, Klaus Ackermann, and Christoph Bergmeir. Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*, 37(2):788–832, 2023.

[48] Mohamed A Kassem, Soaad M Naguib, Hanaa M Hamza, Mostafa M Fouda, Mohamed K Saleh, Khalid M Hosny, et al. Explainable transfer learning-based deep learning model for pelvis fracture detection. *International Journal of Intelligent Systems*, 2023, 2023.

[49] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.

[50] Jiri Nossent, Pieter Elsen, and Willy Bauwens. Sobol'sensitivity analysis of a complex environmental model. *Environmental Modelling & Software*, 26(12):1515–1525, 2011.

[51] Yulong Zhao, Ruike Luo, Longxin Li, Ruihan Zhang, Deliang Zhang, Tao Zhang, Zehao Xie, Shangui Luo, and Liehui Zhang. A review on optimization algorithms and surrogate models for reservoir automatic history matching. *Geoenergy Science and Engineering*, page 212554, 2023.

[52] Kai Zhang, Wenhao Fu, Jinding Zhang, Wensheng Zhou, Chen Liu, Piyang Liu, Liming Zhang, Xia Yan, Yongfei Yang, Hai Sun, et al. Inversion framework of reservoir parameters based on deep autoregressive surrogate and continual learning strategy. *SPE Journal*, pages 1–17, 2023.

[53] Zhi Zhong, Alexander Y Sun, Yanyong Wang, and Bo Ren. Predicting field production rates for waterflooding using a machine learning-based proxy model. *Journal of Petroleum Science and Engineering*, 194:107574, 2020.

[54] Meng Tang, Yimin Liu, and Louis J Durlofsky. A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems. *Journal of Computational Physics*, 413:109456, 2020.

[55] Yinhao Zhu and Nicholas Zabaras. Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. *Journal of Computational Physics*, 366:415–447, 2018.

[56] Kai Zhang, Xiaoya Wang, Xiaopeng Ma, Jian Wang, Yongfei Yang, Liming Zhang, and Jun Yao. The prediction of reservoir production based proxy model considering spatial data and vector data. *Journal of Petroleum Science and Engineering*, 208: 109694, 2022.

[57] Xiaopeng Ma, Kai Zhang, Jian Wang, Chuanjin Yao, Yongfei Yang, Hai Sun, and Jun Yao. An efficient spatial-temporal convolution recurrent neural network surrogate model for history matching. *SPE Journal*, 27(02):1160–1175, 2022.

[58] Jeanfranco D Farfan-Escobedo, Lauro Enciso-Rodas, and John E Vargas-Muñoz. Towards accurate building recognition using convolutional neural networks. In *2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, pages 1–4. IEEE, 2017.

[59] Jeanfranco D Farfan-Escobedo, Kelly Lopes, and Julio C Dos Reis. Active learning approach for intent classification in portuguese language conversations. In *2021 IEEE 15th International Conference on Semantic Computing (ICSC)*, pages 227–232. IEEE, 2021.

[60] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. Learning internal representations by error propagation, 1985.

[61] Mingliang Liu and Dario Grana. Petrophysical characterization of deep saline aquifers for co2 storage using ensemble smoother and deep convolutional autoencoder. *Advances in Water Resources*, 142:103634, 2020.

[62] Gege Wen, Catherine Hay, and Sally M Benson. Ccsnet: a deep learning modeling suite for co2 storage. *Advances in Water Resources*, 155:104009, 2021.

[63] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[64] Adailton José do Nascimento Sousa, Malu Grave, Renan Vieira Bela, Thiago MD Silva, Sinesio Pesco, and Abelardo Borges Barreto Junior. Efficient reservoir characterization using dimensionless ensemble smoother and multiple data assimilation in damaged multilayer systems. *Computers & Geosciences*, 195:105777, 2025.

[65] Youngbin Ahn and Jonggeun Choe. Reliable channel reservoir characterization and uncertainty quantification using variational autoencoder and ensemble smoother with multiple data assimilation. *Journal of Petroleum Science and Engineering*, 209:109816, 2022.

[66] Smith WA Canchumuni, Alexandre A Emerick, and Marco Aurélio C Pacheco. Towards a robust parameterization for conditioning facies models using deep variational autoencoders and ensemble smoother. *Computers & Geosciences*, 128:87–102, 2019.

[67] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[68] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[69] Xin-Yu Zhuang, Wen-Dong Wang, Yu-Liang Su, Zhen-Xue Dai, and Bi-Cheng Yan. Deep learning-assisted optimization for enhanced oil recovery and co2 sequestration considering gas channeling constraints. *Petroleum Science*, 2025.

[70] Lianfa Li, Ying Fang, Jun Wu, and Jinfeng Wang. Autoencoder based residual deep networks for robust regression prediction and spatiotemporal estimation. *arXiv preprint arXiv:1812.11262*, 2018.

[71] Han Wang, Ruiliu Fu, Xuejun Zhang, and Jun Zhou. Rvae-lamol: Residual variational autoencoder to enhance lifelong language learning. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. IEEE, 2022.

[72] Jared Leon-Malpartida, Jeanfranco D Farfan-Escobedo, and Gladys E Cutipa-Arapa. A new method of classification with rejection applied to building images recognition based on transfer learning. In *2018 IEEE XXV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, pages 1–4. IEEE, 2018.

[73] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[74] Francois Chollet. *Deep learning with Python*. Simon and Schuster, 2021.

[75] Sokratis J Anagnostopoulos, Juan Diego Toscano, Nikolaos Stergiopulos, and George Em Karniadakis. Residual-based attention in physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 421:116805, 2024.

[76] Bingyang Yan, Zhi Zhong, and Bin Bai. A convolutional neural network-based proxy model for field production prediction and history matching. *Gas Science and Engineering*, 122:205219, 2024.

[77] Guzel Bayazitova, Maria Anastasiadou, and Vitor Duarte dos Santos. Oil and gas flow anomaly detection on offshore naturally flowing wells using deep neural networks. *Geoenergy Science and Engineering*, 242:213240, 2024.

[78] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.

[79] William H Beluch, Tim Genewein, Andreas Nürnberger, and Jan M Köhler. The power of ensembles for active learning in image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9368–9377, 2018.

[80] Remus Pop and Patric Fulop. Deep ensemble bayesian active learning: Addressing the mode collapse issue in monte carlo dropout via ensembles. *arXiv preprint arXiv:1811.03897*, 2018.

[81] Vikram Mullachery, Aniruddh Khera, and Amir Husain. Bayesian neural networks. *arXiv preprint arXiv:1801.07710*, 2018.

[82] Meire Fortunato, Charles Blundell, and Oriol Vinyals. Bayesian recurrent neural networks. *arXiv preprint arXiv:1704.02798*, 2017.

[83] Tim Pearce, Felix Leibfried, and Alexandra Brintrup. Uncertainty in neural networks: Approximately bayesian ensembling. In *International conference on artificial intelligence and statistics*, pages 234–244. PMLR, 2020.

[84] Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. Deep evidential regression. *Advances in neural information processing systems*, 33:14927–14937, 2020.

[85] Kai Ye, Tiejin Chen, Hua Wei, and Liang Zhan. Uncertainty regularized evidential regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16460–16468, 2024.

[86] Sonia Corraya, Shamim Al Mamun, and M Shamim Kaiser. Learning meta-learning (lml) dataset: Survey data of meta-learning parameters. *Data in Brief*, 51:109777, 2023.

[87] Yao Ma, Shilin Zhao, Weixiao Wang, Yaoman Li, and Irwin King. Multimodality in meta-learning: A comprehensive survey. *Knowledge-Based Systems*, 250:108976, 2022.

[88] Yingjie Tian, Xiaoxi Zhao, and Wei Huang. Meta-learning approaches for learning-to-learn in deep learning: A survey. *Neurocomputing*, 494:203–223, 2022.

[89] Cha Zhang and Yunqian Ma. *Ensemble machine learning*, volume 144. Springer, 2012.

[90] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U Rajendra Acharya, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information fusion*, 76:243–297, 2021.

[91] Vladimir Puzyrev and Paul Duuring. Uncertainty quantification of geochemical data imputation using monte carlo dropout. *Journal of Geochemical Exploration*, 272:107695, 2025.

[92] Sebastian Borquez, Raquel Pezoa, Luis Salinas, and Claudio E Torres. Uncertainty estimation in the classification of histopathological images with her2 overexpression using monte carlo dropout. *Biomedical Signal Processing and Control*, 85:104864, 2023.

[93] Shuo Wang, GH Huang, BW Baetz, and BC Ancell. Towards robust quantification and reduction of uncertainty in hydrologic predictions: Integration of particle markov chain monte carlo and factorial polynomial chaos expansion. *Journal of Hydrology*, 548:484–497, 2017.

[94] Yuejian Chen, Liang Zhao, and Te Han. Uncertainty-guided bayesian active learning for cost-effective fault diagnosis with minimal labeled data. *Engineering Applications of Artificial Intelligence*, 160:111945, 2025.

[95] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[96] Ian Osband, John Aslanides, and Albin Cassirer. Randomized prior functions for deep reinforcement learning. *Advances in neural information processing systems*, 31, 2018.

[97] Florian Wenzel, Jasper Snoek, Dustin Tran, and Rodolphe Jenatton. Hyperparameter ensembles for robustness and uncertainty quantification. *Advances in Neural Information Processing Systems*, 33:6514–6527, 2020.

[98] Steven YK Wong, Jennifer SK Chan, and Lamiae Azizi. Quantifying neural network uncertainty under volatility clustering. *Neurocomputing*, 614:128816, 2025.

[99] Danruo Deng, Guangyong Chen, Yang Yu, Furui Liu, and Pheng-Ann Heng. Uncertainty estimation by fisher information-based evidential deep learning. In *International conference on machine learning*, pages 7596–7616. PMLR, 2023.

[100] Nis Meinert, Jakob Gawlikowski, and Alexander Lavin. The unreasonable effectiveness of deep evidential regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 9134–9142, 2023.

[101] Mira Juergens, Nis Meinert, Viktor Bengs, Eyke Hüllermeier, and Willem Waegeman. Is epistemic uncertainty faithfully represented by evidential deep learning methods? *arXiv preprint arXiv:2402.09056*, 2024.

[102] Maohao Shen, Jongha Jon Ryu, Soumya Ghosh, Yuheng Bu, Prasanna Sattigeri, Subhro Das, and Gregory Wornell. Are uncertainty quantification capabilities of evidential deep learning a mirage? *Advances in Neural Information Processing Systems*, 37:107830–107864, 2024.

[103] Tim Pearce. *Uncertainty in neural networks; bayesian ensembles, priors & prediction intervals*. PhD thesis, 2020.

[104] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[105] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[106] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[107] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. *Advances in neural information processing systems*, 30, 2017.

[108] Andrew L Maas, Awni Y Hannun, Andrew Y Ng, et al. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3. Atlanta, GA, 2013.

[109] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

[110] Larry C Andrews. *Special functions of mathematics for engineers*, volume 49. Spie Press, 1998.

[111] Stefan Elfwing, Eiji Uchibe, and Kenji Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural networks*, 107:3–11, 2018.

[112] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

[113] Warwick J Nash, Tracy L Sellers, Simon R Talbot, Andrew J Cawthorn, and Wes B Ford. The population biology of abalone (haliotis species) in tasmania. i. blacklip abalone (h. rubra) from the north coast and islands of bass strait. *Sea Fisheries Division, Technical Report*, 48:p411, 1994.

[114] Andrea Coraddu, Luca Oneto, Aessandro Ghio, Stefano Savio, Davide Anguita, and Massimo Figari. Machine learning approaches for improving condition-based maintenance of naval propulsion plants. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, 230(1): 136–153, 2016.

[115] Max Little, Patrick Mcsharry, Stephen Roberts, Declan Costello, and Irene Moroz. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Nature Precedings*, pages 1–1, 2007.

[116] Jakob Heiss, Jakob Weissteiner, Hanna Wutte, Sven Seuken, and Josef Teichmann. Nomu: Neural optimization-based model uncertainty. *arXiv preprint arXiv:2102.13640*, 2021.

[117] Simon Kristoffersson Lind, Ziliang Xiong, Per-Erik Forssen, and Volker Krüger. Uncertainty quantification metrics for deep regression. *Pattern Recognition Letters*, 186:91–97, 2024.

[118] Bobby He, Balaji Lakshminarayanan, and Yee Whye Teh. Bayesian deep ensembles via the neural tangent kernel. *Advances in neural information processing systems*, 33:1010–1022, 2020.

[119] Alexander Forrester, Andras Sobester, and Andy Keane. *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, 2008.

[120] M Correia, J Hohendorff, AT Gaspar, and D Schiozer. Unisim-ii-d: Benchmark case proposal based on a carbonate reservoir. In *SPE Latin America and Caribbean Petroleum Engineering Conference*, page D031S020R004. SPE, 2015.

[121] CE Romero and JN Carter. Using genetic algorithms for reservoir characterisation. *Journal of Petroleum Science and engineering*, 31(2-4):113–123, 2001.

[122] David K Ponting. Corner point geometry in reservoir simulation. In *ECMOR I-1st European conference on the mathematics of oil recovery*, pages cp–234. European Association of Geoscientists & Engineers, 1989.

Appendix A Brazilian Offshore Field Dataset

The main objective of this appendix is to introduce the fundamental elements used in the reservoir simulation process for real fields, with a focus on the S-Field reservoir [13].

A.1 S-Field: Inputs and Outputs

Although synthetic reservoirs have been explored in previous studies [120, 121], the complexity and dynamic nature of real oil reservoirs present greater challenges. Oil reservoirs are inherently heterogeneous [23], exhibiting varying spatial and depth properties. For this study, we utilize a real oil reservoir, the S-Field [13], an offshore turbidite reservoir located in the Campos Basin, Brazil, with a production period spanning 2,206 days. The field is produced through 7 production wells and 4 water injection wells, which predominantly follow long horizontal trajectories.

The S-Field reservoir consists of 200 simulation models, each incorporating two types of input data: 3D and scalar properties. The 3D properties are focused on four reservoir attributes: porosity, permeability, net-to-gross ratio, and rock type. These attributes each exhibit inherent uncertainties, collectively referred to as 3D geological uncertainties. The values for these 3D properties are derived from nodal analysis results provided by the operating company, with a $\pm 30\%$ uncertainty range applied to the base values. This uncertainty range reflects the natural variability in the reservoir, which is a crucial aspect of understanding its heterogeneity. Alongside the 3D properties, there are 70 scalar properties, some of which are detailed in Table A.1. The table emphasizes the variability of these scalar properties, further highlighting the intrinsic heterogeneity of the reservoir.

The 3D properties, referred to as 3D geological uncertainties or 3D georealizations, are divided into several interconnected gridblocks (referred to as grid cells or nodes) with geometric dimensions x, y and z in a classic Cartesian coordinate system. As indicated by Ponting [122], the most common coordinate system for more complex structural reservoir simulation models is the corner-point geometry. Specifically, the S-Field reservoir has dimensions of $73 \times 38 \times 24$. Figure A.1 shows the 3D porosity distribution, along with the locations of all wells over the historical period.

Once the inputs for the numerical simulator are understood,each simulation model has multiple outputs, which form a complex multiphase flow system. Specifically, each simulation model is designed to estimate several objective functions that represent the dynamics of fluid production within the reservoir. These objective functions include the

Scalar Properties		
Uncertain Attribute	Minimum	Maximum
Pore Compressibility	3×10^{-7} psi^{-1}	3×10^{-5} psi^{-1}
Depth of Water-Oil Contact	9,609.25 ft	9,642.06 ft
Aquifer Productivity Index	4,000 ft^3/d/psi	60,000 ft^3/d/psi
Initial Aquifer Volume of Water	1×10^9 ft^3	1×10^{13} ft^3
Well Productivity Index (Injectivity)	-30%	+30%

Table A.1: Input data: Scalar properties of the S-Field reservoir. Scalar attributes are modeled using a uniform prior distribution.

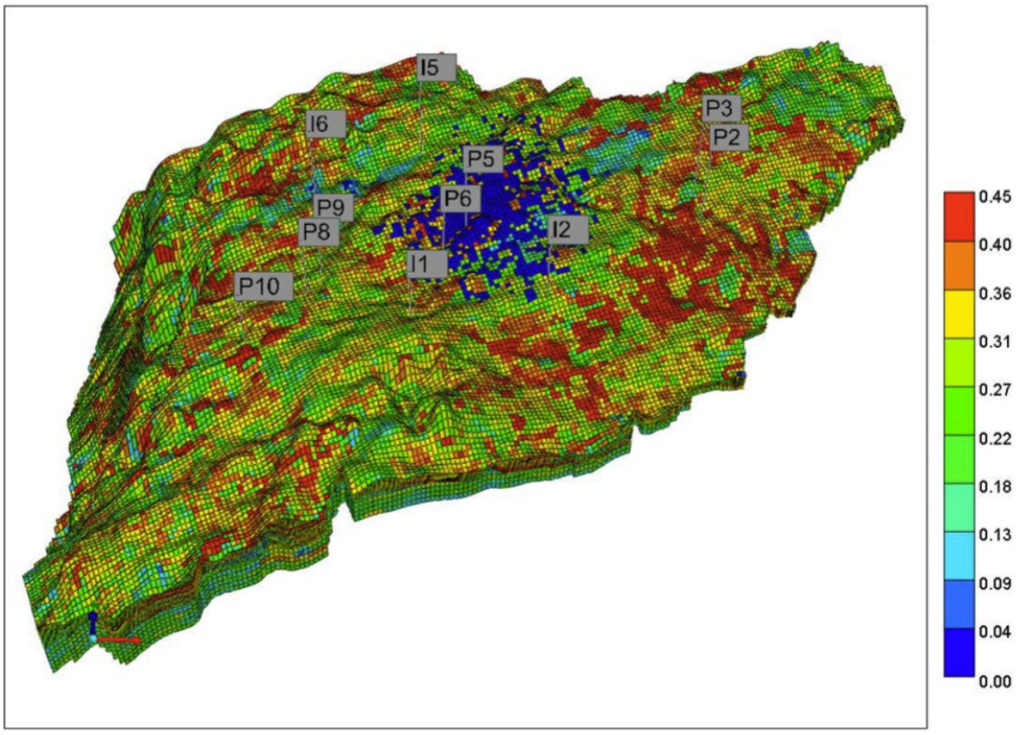


Figure A.1: 3D porosity distribution in the S-field reservoir, highlighting the locations of production and injection wells, as generated by Maschio *et al.* [1].

cumulative production of oil (Np), water (Wp), and gas (Gp), as well as the evolution of the reservoir pressure (P) over time. The cumulative production curves are computed at each time step of the simulation, which represents the duration of each step. These cumulative production curves, or objective functions, are derived from the production well outputs (P10, P2, P3, P5, P6, P8, P9). Each well has been operating for a historical period of 2,206 days, and their production behavior is tracked through these objective functions. Figure A.2 provides a visual representation of some of these objective functions, illustrating the temporal evolution of fluid production across the wells and their associated pressures.

In this work, the S-Field reservoir simulation incorporates multiple inputs and outputs. The inputs consist of 3D properties, including porosity, permeability, net-to-gross ratio, and rock type, all of which come with inherent uncertainties. Additionally, there are 70 scalar properties, such as pore compressibility and aquifer productivity, which are modeled

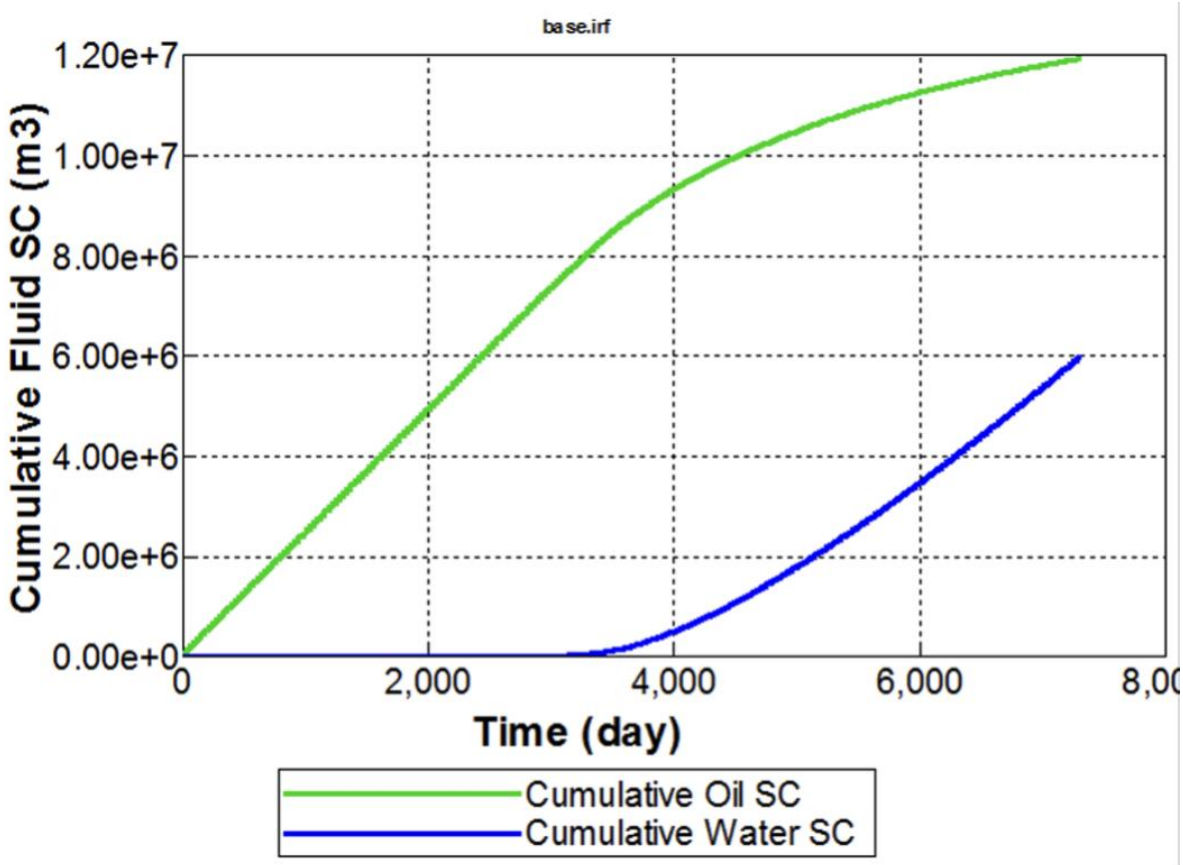


Figure A.2: The continuous green line represents the cumulative oil production (Np), while the blue line represents the cumulative water production (Wp) over time, based on the simulation model outputs, as generated by Maschio *et al.* [1].

using uniform prior distributions. On the output side, the simulation provides cumulative production data for oil, water, and gas, along with reservoir pressure, all calculated over time from seven production wells.

Appendix B Our Contribution to the Literature

For an overview of the work, we present two figures; Figure B.1 and Figure B.2. Figure B.1 involves a modern uncertainty reduction pipeline proposed by Maschio *et al.* [1], which illustrates a modern proxy designed to reduce uncertainties across four distinct steps. In Step 1, geological attributes characterize each reservoir model, denoted as X, which exhibit uncertainties (e.g., porosity or permeability) and serve multiple objective functions (e.g., oil and gas production), referred to as Y. Complex fluid behavior affects each of these objective functions. Data dimensionality reduction is performed for each uncertainty to obtain a vector, denoted as $X_0^{'}$. In Step 2, the objective of the proxy approach is to establish a mathematical relationship between the reduced geological uncertainty ($X_0^{'}$) and the objective functions (Y). Regression techniques typically represent this relationship in an attempt to mimic the simulator's response. Subsequently, the proxy estimates against the numerical simulator are compared. Then, if the current iteration (it) does not equal the last iteration (it_max), a DA is performed. This process continues until a predefined number of iterations, denoted as it_max, is reached, where the values of geological uncertainty ($X_0^{'}$) are updated ($X_i^{'}$), where $i = 1, ..., it_max$. Concurrently, in Step 3, a geological reconstruction is conducted in parallel to ensure the preservation of geological information. Finally, in Step 4, when the current iteration reaches the last iteration, a comparison is made between the reduced uncertainty generated by the proxy and the simulator concerning the observed data. Steps 3 and 4 fall outside the scope of this study and we will prioritize them in future research.

Figure B.2 illustrates our methodology, which incorporates specific modifications to a modern uncertainty reduction pipeline proposed by Maschio *et al.* [1]. In particular, we introduced an end-to-end deep learning framework designed to function as a proxy. This framework effectively addresses the challenge of high dimensionality, replicates the complex responses of the simulator in the realistic context of an oil reservoir, and streamlines the process by combining vector generation, proxy steps, 3D reconstruction, and a new approach to uncertainty quantification, as shown in Figure B.1.

Rather than using separate steps, we input the original geological uncertainty X directly into our end-to-end proxy. The proxy seamlessly integrates with the uncertainty reduction workflow, enabling cohesive collaboration across the entire pipeline. This integration simplifies the process, ensuring that all components within the end-to-end framework interact directly and efficiently with the system.

Including uncertainty quantification in this framework offers several key advantages.

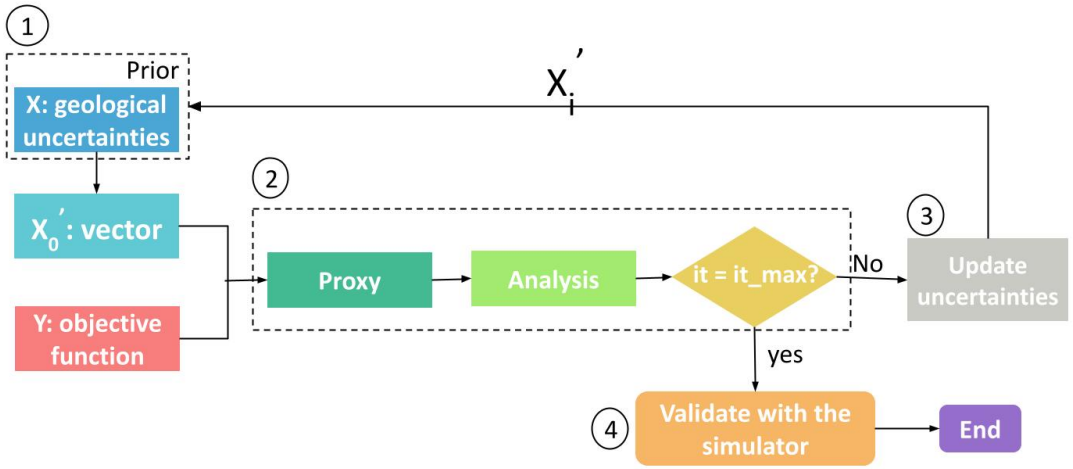


Figure B.1: Here, we showcase an automated uncertainty reduction pipeline proposed in the work of Maschio *et al.* [1]. The process involves the initial analysis of geological attributes (referred to as *'prior'*) and their associated multiple objective functions. Subsequently, a mathematical relationship between geological uncertainties and objective functions is established using regression techniques, with the specific implementation of a proxy. An analysis entails the evaluation of proxy estimations against a numerical simulator. If the current iteration (*it*) does not match the final iteration (it_max), DA is performed iteratively until it_max is reached. This iterative process updates each value of geological uncertainty. In parallel, a geological reconstruction is conducted concurrently to preserve geological information. Lastly, in the final iteration (it_max), a comparison is made between the uncertainty reduction achieved by the proxy and that generated by the simulator.

It allows for a more reliable and transparent estimation of the confidence in the proxy's predictions, which is crucial for decision-making in reservoir management. Additionally, it enables better risk management by providing insights into the variability of predictions, helping experts make more informed choices. Uncertainty quantification also facilitates model validation by comparing predicted uncertainty with actual outcomes, ensuring the robustness and reliability of the proxy model.

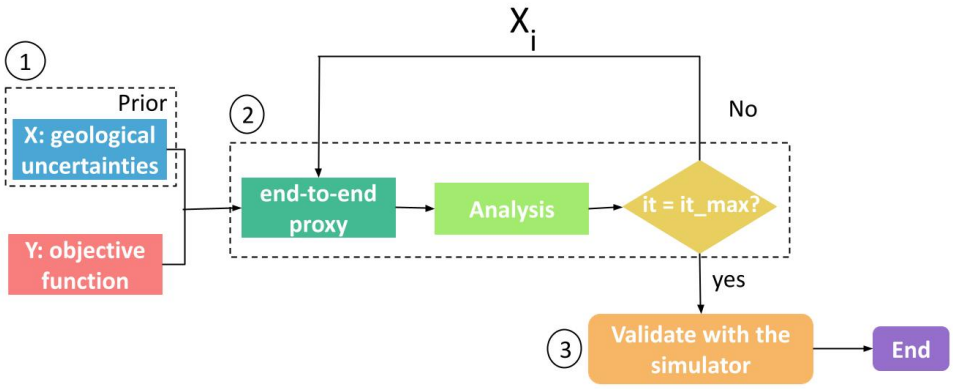


Figure B.2: Here, we present our approach, which is a modification of the contemporary uncertainty reduction pipeline proposed by Maschio *et al.* [1]. Our innovation introduces an end-to-end deep learning framework based on proxies.