

History Matching and Uncertainty Quantification of Reservoir Performance with Generative Deep Learning and Graph Convolutions

Gleb Shishaev

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Heriot-Watt University
Institute of GeoEnergy Engineering
School of Energy, Geoscience, Infrastructure and Society
July 2024

The copyright in this thesis is owned by the author. Any quotation from the thesis or use of any of the information contained in it must acknowledge this thesis as the source of the quotation or information.

Abstract

This research tackles the challenges faced in geological modelling under uncertainty to flow profiling and history matching. Geological uncertainty encompasses various interpretations that may be consistent with the available data. When it comes to modelling, these interpretations necessitate different modelling approaches and configurations. As a result, it becomes challenging to effectively define the space of models and their parameters, as the problem's dimensionality constantly changes. To overcome this challenge, the study proposes an innovative approach that involves parameterisation through implicit low-dimensional hidden spaces.

The study recognises the need to consider uncertainty in geological scenarios, structural uncertainty, and petrophysical dependencies. These factors play a crucial role in accurately representing and predicting the behaviour of geological objects. The chosen methodology is the Graph Variational Autoencoder approach, which allows for the parameterisation of the prior set of geological representations while considering various uncertainty.

The main idea behind this approach is to utilise an Encoder to map the original prior set into a latent space that implicitly describes the prior. The latent space and the Decoder act as a generator that can search for realisations that meet specific requirements. This methodology enables the estimation of uncertainty in dynamic response and history matching, enhancing the overall understanding of geological systems.

The study justifies and presents a transition to graph-based generative modelling. I will show that geometric deep learning, in particular graph convolutions, is the most convenient method to account for geological representations with generative models. This transition is motivated by the need to handle non-Euclidean data types, specifically those lacking a strict structure. This enables the consideration of the structural and spatial features of the reservoir by moving away from the classical lattice representation. The transition expands the applicability of generative models to a broader range of geological objects and enhances the realism of the generated representations, as conventional approaches have limitations and cannot describe the complex structural features of reservoirs or irregularities in flow behaviour.

The study employs advanced analytics tools to gain a deeper understanding of the hidden spaces within the generative models. These tools provide valuable insights into the internal structure of hidden spaces, allowing for a more informed analysis of the generative models' capabilities and limitations.

Moreover, the study introduces a geodesic metric for efficient navigation in high-dimensional hidden spaces. This metric enables more effective exploration and interpolation within the hidden space, resulting in a more predictable behaviour than the standard Euclidean metric. The geodesic metric also serves as the foundation for controlling the geological realism in the latent space, ensuring that the generated realisations maintain their geological coherence.

To test the generative capabilities of a graph-based generative model, the study develops three prior datasets of 3D geological objects, focusing on the uncertainty of geological scenarios, structural uncertainty, and the semi-synthetic Brugge field dataset, which represents four different stratigraphic zones. These datasets serve as test cases to evaluate the effectiveness and limitations of the proposed generative models.

Acknowledgements

So much has happened since I started this project... First of all, I want to thank the scientific community, which has dramatically influenced me through articles, books, conferences, and seminars... everything I have done during this project is just a small step forward (I hope), built upon the enormous scientific foundation that was laid before me.

I want to thank Vasily Demyanov for his immense contribution and help. I'm sure I was one of the "toughest" students, but your work style, patience, and endless support helped me complete the project and grow as a person. Thanks to Dan Arnold and Andy Gardiner for their valuable advice, which enriched my work. Thank you for the wonderful time I spent in Edinburgh. It will remain one of the best memories in my mind.

I thank Valery Rukavishnikov for making my education possible. I still remember our first conversation when I asked how feasible my PhD project was, and it seemed impossibly difficult at the time. I'm glad it all worked out in the end. Also, a big thank you to the entire Tomsk team. Each of you personally influenced this project through consultations, advice, help with document processing, and IT support. This is very important to me, and I greatly appreciate it.

Thank you to Roman, who contributed significantly by constantly helping me with machine learning questions, coding, and various engineering tricks.

Thank you to the entire GeoDataScience Group, Bastian, Quentin, Chao, and Atos; even though I didn't interact much with you, your successes were an example and motivation for me to continue. Special thanks to Rock Flow Dynamics for providing the simulator used in this work.

This project will be remembered as a constant roller coaster, where everything works out or nothing works and everything falls apart. Lonely bike rides through the beautiful landscapes of Scotland. A huge amount of coffee was consumed (thanks, Oleg), classical music, and thoughtful walks.

Thank you to everyone who reads this message for your interest in my work. I hope you find pleasure in it and learn something new for yourself.

The following message in Russian to my family.

Мама и Папа, спасибо, что всегда меня поддерживают. Спасибо моей жене за бесконечный позитив и поддержку. Ты прошла со мной весь этот путь и сделала его легким и веселым ❤️

Saint Petersburg, November 2023

Gleb

Inclusion of Published Works Form

Declaration

This thesis contains one or more multi-author published works. I hereby declare that the contributions of each author to these publications is as follows:

Citation details	Shishaev G., Demyanov V., Arnold D., Vygon R. History Matching and Uncertainty Quantification of Reservoir Performance With Generative Deep Learning and Graph Convolutions. ECMOR 2022, Sep 2022, Volume 2022, p.1 - 9
Author 1	Contribution to method development in discussions, code implementation, wrote manuscript
Author 2	Supervision, contribution to method development in discussion
Author 3	Supervision, contribution to method development in discussion
Author 4	Supervision, contribution to method development in discussion

Citation details	Shishaev G., Demyanov V., Arnold D., Vygon R. Application of Graph Variational Autoencoders for History Matching Problem of Brugge Field. Fifth EAGE Conference on Petroleum Geostatistics, Nov 2023, Volume 2023, p.1 - 5
Author 1	Contribution to method development in discussions, code implementation, wrote manuscript
Author 2	Supervision, contribution to method development in discussion
Author 3	Supervision, contribution to method development in discussion
Author 4	Supervision, contribution to method development in discussion

Contents

Abstract	i
Acknowledgements	iii
Contents	viii
1 Introduction	1
1.1 Uncertainty	5
1.2 How to model uncertainty	6
1.3 What Machine Learning is offering?	8
1.4 Challenges of geological modelling and thesis outline	10
2 Review of recent advances in generative deep learning and problem statement	13
2.1 Geostatistical approaches in reservoir modelling	13
2.1.1 Variogram-based algorithms	14
2.1.2 Object-based algorithms	17
2.1.3 Multi-point statistics	18
2.1.4 Limitations of geostatistics	19
2.2 Machine Learning approaches in reservoir modelling	20
2.2.1 Generative Adversarial Network	22
2.2.2 Variational Autoencoder	24
2.2.3 Flow-Based generative model	26
2.2.4 Diffusion Models	28
2.3 Recent advances of generative modelling in geoscience	30
2.3.1 Structural Uncertainty in generative modelling	36
2.4 General workflow of AHM and UQ	39
2.5 Research objectives	42
3 Methodology	46
3.1 VAE description	46
3.2 From KL to Wasserstein Loss	50

3.2.1	Optimal Transport basics	51
3.2.2	From VAE to WAE	52
3.2.3	MMD regulariser	53
3.3	Importance of Geometric Deep Learning	54
3.3.1	CNN for Euclidean data types	56
3.3.2	Convolutions on non-Euclidean data	59
3.3.3	Graph Wasserstein Variational Autoencoder (GWAE)	62
3.4	Optimisation theory	64
3.4.1	Covariance-Matrix Adaptation Evolution Strategy (CMA-ES)	66
3.5	Why it is important to keep LS small enough	68
3.6	Inner Geometry of a Latent Space	70
3.7	Objective function in HM process	73
3.7.1	Flow loss function	75
3.7.2	Static loss function	76
3.7.3	Realism loss function	77
3.7.4	Optimisation process	78
3.8	Tools for analysing multidimensional spaces	79
3.8.1	Principal Components Analysis	80
3.8.2	t-Distributed Stochastic Neighbour Embedding	83
3.8.3	Topological Data Analysis	86
3.9	Conclusion	88
4	Modelling uncertainty and datasets construction process	89
4.1	Set up for various geological scenarios case study (<i>FLUVIAL</i>)	91
4.2	Set up for structural uncertainty case study (<i>STRUCTURAL</i>)	96
4.3	Set up for Brugge field case study	101
4.3.1	Brugge benchmark	101
4.3.2	<i>BRUGGE</i> dataset construction for GWAE	103
4.3.3	The process of converting corner point grid into a graph representation	112
4.4	Conclusion	114
5	Graph Wasserstein Autoencoders and channelized synthetic case	116
5.1	Generative quality of GWAE with <i>FLUVIAL</i> dataset	119
5.1.1	Latent Space inner structure	121
5.1.2	Variation of models along the shortest path trajectories in the LS	124
5.2	Uncertainty of dynamic response and history matching	126
5.2.1	Uncertainty of dynamic response	126
5.2.2	History matching through the latent space	128
5.2.3	Objective Function efficiency study	129

5.3	Variational Autoencoders and 3D reservoir representations	132
5.4	Architecture of Generative Models	133
5.5	Results of training Generative Models	134
5.6	Conclusion	139
6	Graph Wasserstein Autoencoders and fault synthetic case	140
6.1	Structural Uncertainty	140
6.2	Generative quality of GWAE with <i>STRUCTURAL</i> dataset	141
6.2.1	Latent Space inner structure with structural uncertainty	144
6.2.2	Interpolation in the LS	147
6.3	Uncertainty of dynamic response and history matching under structural uncertainty	149
6.3.1	Uncertainty of dynamic response	150
6.3.2	History matching through the hybrid latent space	150
6.4	Conclusion	156
7	Graph Wasserstein Autoencoders for AHM Brugge case	157
7.1	GWAE under uncertainty of modelling approach	157
7.1.1	Combined LS analysis	158
7.1.2	Generative quality of GWAE with combined dataset	159
7.2	Generative quality of GWAE with Brugge dataset	160
7.2.1	LS analysis	161
7.3	HM of Brugge field with GWAE	163
7.4	Uncertainty of Prediction	171
7.5	Conclusion	173
8	Conclusion and future work	174
8.1	Conclusion	174
8.2	Recommendations	178
	List of Acronyms	179
	References	182

Chapter 1

Introduction

If I have seen further it is by standing on the
shoulders of Giants

Isaac Newton letter to Robert Hooke, 1675

In geoscience, scientific knowledge is driven by a fundamental objective – to uncover a concise set of equations and parameters that can comprehensively describe the intricate workings of nature and convert it into a theory. This knowledge holds a multitude of complex interrelationships between various scientific domains and their interactions with the physical world. However, the quest for simplicity in representation often conflates the theory with the reality it represents, blurring the distinction between the two.

In our eagerness to understand the subject, we frequently try to relate the theory with what is represented. This intellectual misstep hinders our comprehension and prevents us from truly understanding the subject. In other words, if all we have is a theory, everything becomes an object of study. This paradox can lead us to the point where the theory we are using can distort our view of reality. To navigate this conceptual challenge, researchers in geoscience must differentiate between theories, tools, abstractions, and concepts and the phenomena they describe.

The same object of study can be examined through the prism of different theories, providing its own representation. In this regard, the researcher needs to integrate representations from different theories into a unified framework to understand the real object of study better. However, building such “bridges” between theories can lead to contradictions, while excluding certain concepts can result in an excessive simplification of the inherent complexity of Earth’s systems.

Consider, for instance, the concept of a geological object. It can be examined through various theories: sedimentology, lithology, petrophysics, geophysics, flow modelling, well-test, data science, economics and so on. Each perspective offers a different representation of the object, but which truly represents nature?

These epistemological questions can have a profound impact on the field of geoscience. It is crucial to recognise the limitations of our scientific methodologies and resist the temptation to entirely replace an understanding of the full complexity of the object with purely theoretical constructs. Instead, we must see the gaps as opportunities for further exploration and allow our theories to remain representations, distinct from the complex realities they seek to capture. Science can be viewed as a supplement to describe an object, as it provides a specialised language and toolkit.

With its specialised language and toolkit, science can complement the project design process by providing the necessary tools and knowledge to describe the object of study. Making the right project design is a challenging high-impact problem and requires revisiting throughout the reservoir's lifetime. Multiple system elements, uncertainty interdependencies, dynamic data, and multiple pursuing goals are updated along the reservoir life cycle, subject to the decision's consequences. Examples of such decisions include drilling a new well, implementing hydraulic fracturing or shutting an existing one. Generally, every project design has a sequence of interventions, which is assumed to maximise the expected outcome and get new information.

Reservoir intervention decisions require understanding how a particular well would behave in the reservoir under uncertainty. Often, a lack or incomplete understanding of the interaction between the well and the field or incomplete knowledge of the field leads to unsatisfactory or sad results. At every stage of a field's life (exploration, appraisal, early production, late production, abandonment), it is necessary to consider uncertainty, the complexity of a sequence of decisions to be made, multiple objectives and time components. We can draw the general project design scheme in a Reinforcement Learning style, where we try to maximise Rewards in time, taking some Actions under Objects in the Environment and getting new information as a response, figure 1.1.

Uncertainty usually has a paramount impact on development decisions. For example, ignoring such features as muddy deposits around channel bodies, which dramatically affect the vertical flow, leads to an overestimation of the recovery, as more oil would remain trapped, and further interventions and drilling decision updates may be needed. Therefore, an expert requires a reliable (with consideration to his aims) model of a reservoir and its interaction with a well regarding uncertainty — the range of factors not known explicitly to expect, figure 1.2.

In this case, modelling plays a crucial part in the project design. Geologists and engineers develop a conceptual model of a field with their limited knowledge. Then, they try to translate this mental concept into a digital model. The decisions and subsequent actions are made based on conclusions drawn from such a model. Jay Wright Forrester, the father of system dynamics, described a mental model as follows:

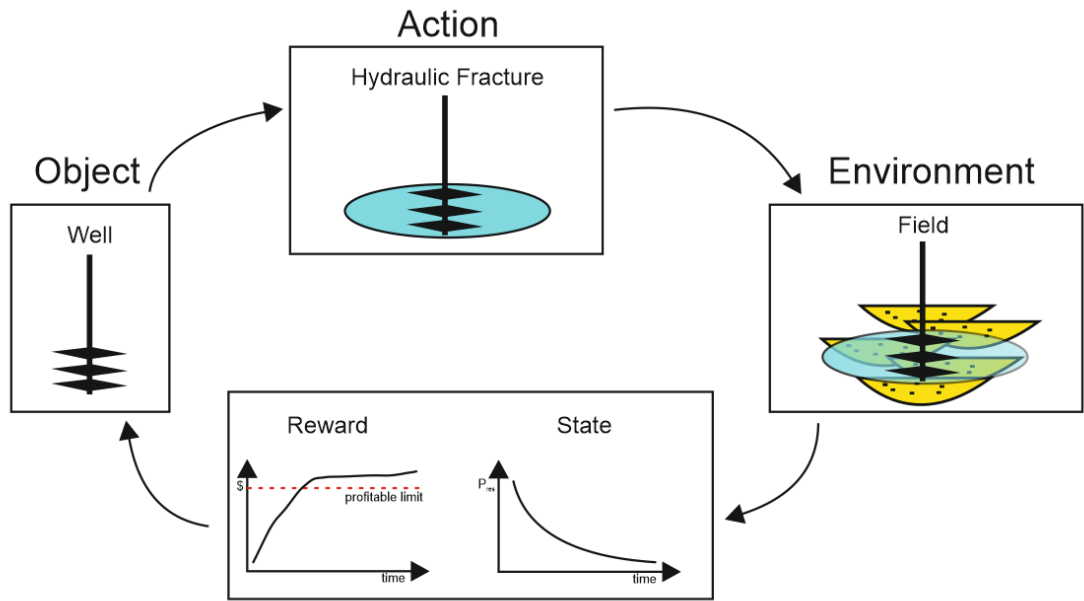


Figure 1.1: Every time step, we should make an Action in respect of an Object to maximise the Reward from Environment response and get a new State.

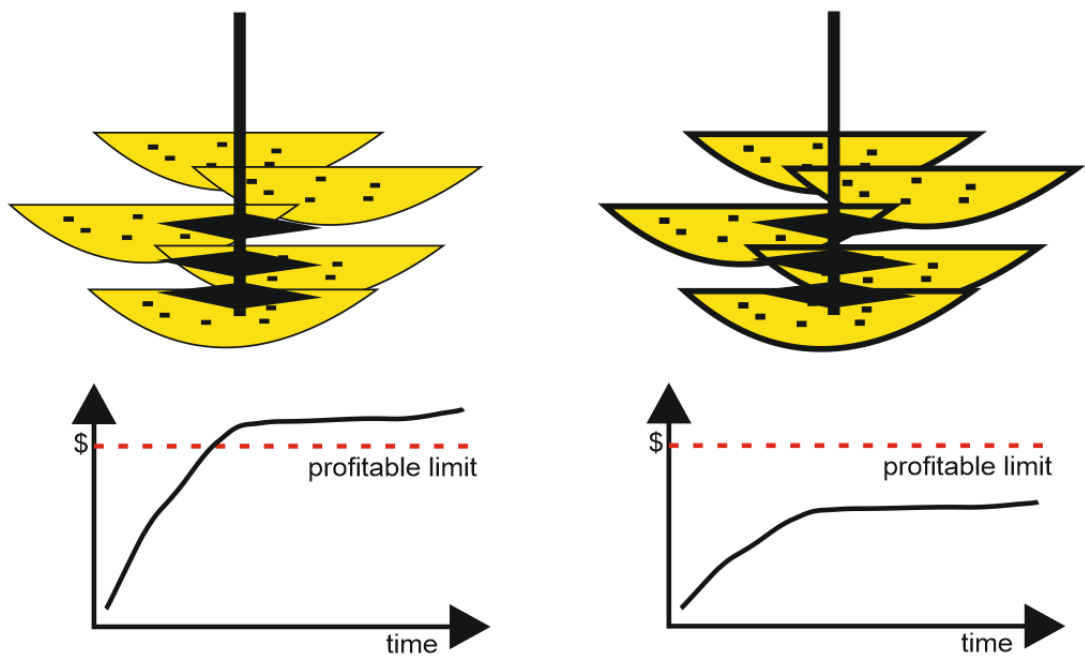


Figure 1.2: Schematic representation of well and channel-fill sand bodies interaction. On the left picture, muddy deposits are omitted, which leads to flow overestimation and false decisions. On the right, the reservoir model is presented in more detail, which led to a more pessimistic cash flow.

“The image of the world around us, which we carry in our head, is just a model. Nobody in his head imagines all the world, government, or country. He has only selected concepts, and relationships between them, and uses those to represent the real system.” (Forrester, 1971)

Modelling of natural systems is one of the major challenges in geology. Experts can mostly describe particular tech object characteristics, such as well design, grade of steel, quality of cement, x-mas tree and so on. So, there is not much uncertainty about it, and it can be modelled quite determined in most cases, although many options are possible here, and often those options would exclude each other. The steel grade of a well varies and depends on the manufacturing process and bespoke design. On the other hand, specialists hold only limited knowledge about an object, such as a field. That is why he is forced to model its initial and current states due to the interaction with an object under uncertainty, with some assumptions and simplifications. The main problem is the lack of data describing the particular geological object. The principal understanding of the depositional processes can help reduce the variety of cases. Still, the modelling is usually ill-posed – many solutions to the problem can be found.

To better understand and characterise a reservoir, geoscientists create 3D geological models that incorporate various data and properties. However, these models can only capture a limited level of complexity in nature. Based on these models, reservoir engineers use numerical methods to simulate fluid flow through the reservoir. However, relying on a single model to predict reservoir performance does not adequately account for the uncertainty in the subsurface.

Furthermore, scientists frequently use simplified proxy geological modelling, which entails the construction of a geological representation by focusing solely on the most crucial reservoir characteristics. This approach diminishes the demand for computational resources, which is particularly advantageous in the context of optimisation processes. There are two main approaches Physics-Based (May et al., 2021; Ren et al., 2019) and Data-Driven (Shuku et al., 2023; L. Zhang et al., 2023).

Ideally, a large ensemble of reservoir models would be used to analyse the impact of uncertainty on fluid flow behaviour. However, due to time constraints and practical difficulties, a representative subset of models is chosen for flow simulation. This often results in collapsing the uncertainty into a single base case, which is not ideal as the predictive capabilities of a single model are questionable (Ringrose et al., 2015).

This ensemble helps to understand possible well/reservoir interactions so optimal scenarios and decisions can be made. That is why it is essential to understand sources of uncertainty and ways of their quantification.

1.1 Uncertainty

There is no unique definition of uncertainty. It often depends on the application discipline and modelling approach. Uncertainty refers to the lack of complete knowledge or predictability about a particular object, event, or outcome.

There are three general sources of uncertainty (I. Goodfellow et al., 2016):

1. Inherent stochasticity of the system being modelled. The system is inherently variable and unpredictable. In practical terms, no matter how much we study and understand a natural system, we can not model with certainty or predict its future response.
2. Incomplete observability. Incomplete observability uncertainty refers to the uncertainty that arises due to the inability to fully observe or measure all the variables or states of a system. It occurs when certain aspects or components of the system cannot be directly observed or measured, leading to an incomplete understanding of the system's behaviour. Incomplete observability uncertainty can arise from limitations in measurement techniques, incomplete data collection, or other factors that prevent a complete observation of the system.
3. Incomplete modelling. Incomplete modelling uncertainty refers to the uncertainty that arises due to the lack of complete knowledge or understanding of all the variables and factors that influence a system's behaviour. It occurs when the model used to describe or predict the system cannot capture or account for all the relevant variables, leading to an incomplete system representation. As a result, the system's behaviour may appear stochastic or unpredictable, even though it is fundamentally deterministic. Incomplete modelling uncertainty can arise from limitations in data availability, measurement errors, simplifications, assumptions made in the model, or other sources of incomplete information about the system.

Caers (2011) in his book “Modelling Uncertainty in the Earth Sciences” pointed out five sources of uncertainty which can be applicable to geoscience problems. Let's try to connect them with previous, more general classes:

1. Uncertainty of determining the type of geological object, which arises from the interpretation of data, or either based on physical models, which, in turn, also have uncertainty – class 1;
2. Uncertainty associated with the error or inaccuracy of measurements, or the processing of measurements – class 2;
3. Uncertainty of interpreting raw data. Uncertainty regarding the fact that the process of interpretation also requires the construction of some model — class 2;

4. Spatial uncertainty. Typically, the data either does not cover the entire modelled volume, or the data resolution does not match the model resolution – class 3;
5. Uncertainty of the response, since geological uncertainty transitions into the uncertainty of dynamic response – class 3.

A common approach in modelling under uncertainty considers explicit relative dependencies of parameters. This approach initially has some drawbacks, such as ignorance of implicit dependencies and simplification. Moreover, it is usually impossible to identify such implicit dependencies in case of multi-parametric problems. So, the additional source of uncertainty should be noted:

6. Relative dependency: includes uncertainty in the dependency between reservoir characteristics (geological parameters such as porosity/permeability), commonly represented explicitly in models – class 3.

The two types of uncertainty are complementary — uncertainty in modelling concepts / parameters and intrinsic uncertainty due to natural stochasticity / randomness.

The problem is that this may lead to over-determined / over-constrained models that ignore some uncertainty. These dependencies may be far more complex and cannot be modelled explicitly. For example, there is a common approach to model poro/perm dependency through the logarithmic relationship, but this is far from reality and just a simplification. First, this relationship is not as deterministic as it is usually modelled and can depend on the connectivity of pores, tortuosity, lamination, etc. In other words, it is a function of the shape of the capillary pore spaces and the degree to which pores are connected, but it is too complex to model in field-scale models.

1.2 How to model uncertainty

It is often more feasible to employ a straightforward yet uncertain rule instead of a complex yet certain one, even when the actual rule is deterministic and our modelling system can handle a complex rule (I. Goodfellow et al., 2016).

Philip Ringrose and Mark Bentley, in the book “Reservoir Model Design” (Ringrose et al., 2015) provide three approaches to modelling:

1. Best Guess – in which a single model is chosen as a base case – which actually does not account for any uncertainty
2. Multiple Deterministic – which avoids making a single best guess or choosing a preferred base-case model. This approach involves constructing a small number of deterministic models, ideally reflecting the space of uncertainty

3. Multiple Stochastic (geostatistics realisations) – a large number of models are stochastically built

These three cases do not cover uncertainty quantification (UQ) when models (deterministic or stochastic) are inferred from data through inverse modelling, i.e. inferring parameter values (for deterministic case) or the entire state variables fields (ensemble approach). This usually involves Bayesian inference. However, this is often constrained to explicitly determined parametric model definition, which is uncertainty in its own — class 1.

Each case has its advantages and disadvantages. While using multiple stochastic models may seem like the best choice when dealing with uncertainty, these models are often based on available data. They may make similar assumptions as a simple forecast but with added ranges around the most likely prediction. However, the accuracy of the outcome may be compromised by giving more weight to specific input data and relying on best guesses for each variable. This can result in the most probable probabilistic outcome being close to the initial best guess.

Another problem of modelling under uncertainty is finding a link between static and dynamic data in geology. The relationship between these two data types may not be easily captured or represented empirically or analytically. Static data, such as seismic surveys, well logs, and core interpretations, typically represent the spatial distribution of geological features at a given point in space. On the other hand, dynamic data, such as production profiles and well test analysis, capture temporal responses under well/reservoir interaction and refer to some bulk integral volumes. Usually, a history matching (HM) process is needed to define the link between those data types. It is a waterfall-like process where the reservoir model is built with the static data, and after that, it is tuned to represent a dynamic response.

These two data types have inherently different characteristics and may not have a direct relationship. Dimension reduction techniques aim to reduce the complexity of high-dimensional data by identifying the most important features or variables that explain most data variability. However, in the case of static and dynamic geological data, it is challenging to determine which features or variables should be considered as crucial for capturing the relationship between them.

Additionally, dimension reduction techniques may not adequately capture the temporal aspect of dynamic data. While static data can be easily represented in a lower-dimensional space, dynamic data often require additional dimensions to represent temporal variations accurately. Ignoring the temporal aspect of dynamic data during dimension reduction can lead to the loss of crucial information and potentially misleading results.

1.3 What Machine Learning is offering?

Recently, machine learning (ML) related approaches attract scientists' interest in modelling and reconstruction problems. Developed data science approaches mostly focused on fundamental issues such as data heterogeneity, complex dependencies of physical parameters, high dimensionality, non-Gaussianity, and non-linearity. The principal philosophy in resolving such problems can be in implementing machine learning algorithms, which can consider all available data, identify and use implicit dependencies of parameters omitted by humans and combine them into a model. Neural networks perform a non-linear mapping between input and output. Non-linearity is crucial as it allows capturing complex dependencies between parameters. That is why they should be used instead of simpler (linear) models.

For the problem of modelling under uncertainty, ML algorithms can be divided into two groups: parametric and non-parametric. Parametric models learn a function described by a parameter vector whose size is finite and fixed before any data is observed. An algorithm implementation provides a model of some object, event, or process in quite a determined way by tuning weights, which stays constant after the learning process. Non-parametric models have no such limitation; all available data is mapped into compressed latent space (LS), which describes all variability and keeps implicit. This latent space can be considered a 'cloud' of a probabilistic representation of an uncertain object, environment, or instance and can be used as a source for sampling under some conditions. Such class of ML algorithms, usually named "Generative", aim to create new instances which meet pre-assigned requirements. In addition to this, there is a wide range of uncertain parameters, which can be decreased by utilising dimensionality reduction techniques.

The ideal model would be an oracle with knowledge of the probability distribution generating the data we observe. However, even such a model would still encounter some errors in many problems due to potential noise in the distribution (I. Goodfellow et al., 2016). Even if nature is deterministic, the probability distribution is our convenient imaginary statistical concept of how the truth may look with certain probability. Within such a probabilistic concept, we can accommodate the observations and realisations of the random process and evaluate each possible concept's probability.

There are benefits of ML could be emphasised:

1. No over-constraining with assumptions – Assumptions are inevitable in every model containing uncertainty. Generative class of ML algorithms may free an expert from making solid assumptions about concepts and available data, providing additional flexibility to the modelling process.
2. Additional realism – As mentioned earlier, the 'reasonableness' of the automated

combination of variables is hard to verify. Therefore, ML may add some realism to variable combinations, sequences, proportions and so on because it only “sees” realistic representations during the learning process.

3. Avoid bias towards data and prior assumptions, which may not be correct – Anchoring is one of the drawbacks of geostatistics modelling, which the utilisation of non-parametric approaches in ML can overcome.
4. ML approaches can find new data interpretations that are still consistent with the domain.

Limitations of ML:

1. Data dependency – ML algorithms are data-dependent and very sensitive to features. Therefore, the “garbage in – garbage out” principle is always valid. But it is not just about having the data but learning the right things from it. So, if we deal with data which is a product of some interpretations, it could be “bias in – bias out” (Michael Pyrcz)
2. There is still unclear how to evaluate uncertainty through latent space mapping.
3. The “no free lunch theorem” there is no universal algorithm that can perform than any other for any problem.
4. Interpretability – “black box” problem. Mostly based on correlation elicited from that, which may not represent the causality.

In the context of geoscience, ML can help address the outlined issues by utilising machine learning algorithms to analyse and model complex dependencies, data heterogeneity, and high dimensionality. It can capture non-linear relationships between parameters, making predictions and interpretations more accurate. ML also offers the advantage of flexibility, as it does not heavily rely on assumptions and can find new interpretations in the data.

In summary, while making decisions without having all the necessary information is possible, it is crucial to acknowledge and account for the uncertainty involved. In subsurface situations, where direct observations may be limited, decision-makers rely on predicted quantities based on measured data and utilise techniques such as sensitivity analysis and risk analysis to address uncertainty. Expert judgment and experience also play a significant role in making decisions in such scenarios. It is worth noting that the dimensionality of the parameter space describing the model is usually much larger than the dimensionality of the parameters that need to be predicted. Thus, generative neural networks are a good candidate as a tool to reduce the dimensionality of the original space without losing internal dependencies, which leads to more efficient solutions for the inverse prediction

problem in conditions of non-uniqueness of solutions.

1.4 Challenges of geological modelling and thesis outline

Céline Scheidt (2018) outlines several areas where research is needed to develop optimal solutions and best practices: fundamental sciences, development of predictive models, data science, economic models and models of human behaviour. This thesis mainly prioritises the second and third points.

Training generative models is a modern approach to building models that represent the object of interest using the training data provided. However, as a relatively new approach to geological modelling, it brings not only new perspectives, but also a number of challenges that need to be addressed by the scientific community:

- How to leverage generative modelling in geology? Data science and generative machine learning, among others, are successfully implemented in computer vision, medicine, etc. How can we use this experience in geological modelling and history-matching processes?
- How to account for the inner physics of a reservoir concerning uncertainty of a different nature?
- How to define implicit dependencies between geological parameters or geological parameters and dynamic response? Keeping every parameter in its realistic range is not the same as keeping geological realism. In reality, most dependencies are not linear; sometimes, we can't find a reliable description. Ideally, a geologically reliable model should hold dependencies between parameters and dynamic response.
- How to change/update a model to fit the data without losing geological reliability? It is a crucial problem for every model, which should be history-matched. If a model history-matched but is not geological reliable, or geological reliable but does not perform a historical dynamic response, we can't rely on prediction. So, the tuning process should cover both issues.

The thesis will walk through the modern advances of generative modelling in geosciences to the specific method of realistic geological model update through the latent space of Variational Autoencoder based on Graph Convolutions, which demonstrates the ability to implicitly parameterise unstructured data into a latent space of reduced dimensionality and provides ways to balance reservoir model conditioning to both static well and dynamic production data.

The thesis is divided into eight chapters and structured as follows:

Chapter 2: Review of recent advances in generative deep learning and problem statement

Most decision-making approaches try to create an accurate model representation; here, we explore techniques of replacing actual models with generated ones as it brings the most uncertainty. I focus on the application related explicitly to reservoir geological modelling with generative algorithms, such as Generative Adversarial Networks and Variational Autoencoders. I will close the chapter with the problem statement of how to utilise generative machine learning through the concept of latent space that provides encoding of initial geological data into manifolds (latent spaces) of decreased dimensionality. This gives a way to implicitly link the static and dynamic model response without parametrisation constraints into a manifold to optimal HM and uncertainty quantification.

Chapter 3: Methodology starts with the concept that I laid down as the foundation for reservoir modelling using generative methods of machine learning - specifically, graph variational autoencoders, which will be used throughout the thesis. The second part of the chapter is dedicated to methods for analysing latent spaces, which are necessary for more efficient work with generators. Latent space is an effective representation of uncertainty — learned from data. After reading this chapter, the reader should understand the mathematical framework used in the work.

Chapter 4: Modelling uncertainty and datasets construction process. The process of constructing training datasets will be described, including the inherent uncertainty and the hypotheses tested through experiments on these datasets. Three large datasets of varying complexity and geological settings were prepared during the work. In addition to describing standard approaches to dataset construction through workflows, the transformation of lattice models into a graph representation, an adequate way of unstructured descriptions, will be described for further machine learning model training.

Chapter 5: Graph Wasserstein Autoencoders and channelized synthetic case describes a basic training pipeline of a generative network and an optimisation process using latent space and decoder for model history matching. Using a dataset that includes spatial uncertainty, the generative capabilities of the trained network and the hidden space will be tested and analysed. Then, the results of model performance will be presented. The chapter also compares the generative capabilities of the newly introduced geomodelling graph variational autoencoder and “classic” generators (namely Convolutional Neural Networks-based variational autoencoder and Wasserstein autoencoder) used in earlier applications. All three compared generators will be trained on the same synthetic dataset and with similar hyper-parameters, experimentally demonstrating the advantage of the graph-based approach.

Chapter 6: Graph Wasserstein Autoencoders and fault synthetic case extends the application of the newly designed generative network with graph convolution to consider structural geological uncertainty. During the analysis of the hidden space, the first limiting factor of the decoder's generative capabilities, which directly affects the model HM process, will be identified. Based on the results of this chapter, the first recommendations for constructing a prior are provided.

Chapter 7: Graph Wasserstein Autoencoders for AHM Brugge case consolidates the findings from the comparative studies in the earlier chapters and applies them to a full field benchmark history-matching application — the Brugge case study. The chapter will begin by justifying the effectiveness of using a family of generative models compared to a single universal model. The results of a model history matching using a hybrid latent space and a set of decoders will be shown. Additionally, the chapter will describe recommendations for constructing a universal generator and factors that limit the construction of such a generator.

Chapter 8: Conclusion and future work The last chapter of this thesis focuses on the main findings and results developed, relating them to the overall aim of the thesis. It also provides an outlook on future investigations and suggests potential improvements.

Chapter 2

Review of recent advances in generative deep learning and problem statement

The purpose of this work is to consider generative machine learning methods for solving problems in the field of geoscience like UQ and HM. Therefore, the first part of this chapter will present a literature review to provide a general understanding of the geostatistical methods used in practice for reservoir modelling and identify their weaknesses and limitations. The geomodelling UQ methods overview will describe the trends and highlight the recent advances in machine learning.

Conclusions from the review will outline the main challenges to be tackled in the thesis. I will describe how this thesis's findings and methodology can help address the task of HM and UQ.

2.1 Geostatistical approaches in reservoir modelling

Reservoir modelling aims to accurately represent the spatial variation of important geological properties, which can be influenced by various underlying phenomena such as sedimentological, diagenetic, or tectonic processes. Geostatistical methods simulate stochastic realisations of reservoir properties in areas away from well locations, assuming certain spatial correlation dependencies. However, this becomes more challenging when dealing with high heterogeneity as the influence of non-linear, non-stationary property distributions increases. The heterogeneity and uncertainty of facies distributions with contrast permeabilities significantly impact flow, as it is difficult to describe how geological properties change between observation locations.

The level of variability in natural spatial phenomena suggests that changes in properties occur in an ordered or continuous manner, reflecting the continuous nature of the under-

lying geological processes. According to Demyanov et al. (2018), there are three main questions for geostatistics which should be answered:

- How to describe spatial continuity of porous media?
- What is the best function which can describe relationships of properties?
- How to account for the uncertainty in the spatial distribution of properties?

A mathematical description of porous media is necessary to represent geological properties and facilitate further flow simulation digitally. Several approaches are available to model the spatial distribution of porous properties in a reservoir, and selecting a particular approach determines how data and knowledge are integrated into the model, which affects the flow. These approaches can be classified into three types:

- Variogram-based algorithms.
- Multi-point statistics (MPS), which is a training-image-based method. The model is built using a representative image of the geological structures.
- Object-based approaches. The objects are defined by geometrical parameters (such as width, height, and sinuosity) and object proportions that are used to create the required geological scenario according to some placement rule.

2.1.1 Variogram-based algorithms

The family of variogram-based algorithms is presented by Direct Sequential Simulation (DSS), Sequential Gaussian Simulation (SGS), Truncated Gaussian Simulation (TGS), Sequential Indicator Simulation (SIS). Algorithms are based on the (co)kriging approach, which was first developed by Danie Krige, who applied mathematical statistics to the spatial evaluation of orebodies (Krige, 1951). The processes he developed were named after him, becoming known in the industry as ‘kriging’. This is also a family of spatial statistical models of propagating mean and variance away from known data points, i.e. wells. Variogram-based models have a general assumption about a spatial process under consideration, such as second-order stationarity and linearity.

NOTE: Second-order stationarity implies stationarity of the mean and covariance, which means the similarity in spatial dependence with the distance across the whole reservoir. In the case of stationarity, the distribution of properties in space is invariant to space. This is an idealistic assumption rarely met in practice.

Kriging

Kriging is a spatial statistical technique used to estimate the mean and variance of a spatial distribution. In the context of reservoirs, it allows for the interpolation of property values between wells by considering the spatial continuity, or variogram (Demyanov et al., 2018; Pyrcz, 2014).

Different flavours of kriging can be used like Simple Kriging (W. Lee et al., 2011; Madani et al., 2022; Matheron, 1963), Ordinary Kriging (Kumar et al., 2023; Matheron, 1963), Universal Kriging (Cheng, 2013), Indicator Kriging (Demyanov et al., 2018; Journel, 1983; F. Liu et al., 2012), depending on the complexity of the problem and the relevant assumptions. However, it is important to note that kriging can only estimate the mean and variance to a certain extent and cannot fully capture the uncertainty in the reservoir, i.e. it gives a single result, and all properties are always mean.

To address the uncertainty in reservoir properties, it is necessary to generate multiple realisations of reservoir models by sampling from the estimated statistical distributions. This allows for considering various possible scenarios while maintaining the spatial correlation in the reservoir properties.

The following subsections will briefly describe methods that utilise different spatial statistics descriptions. Stochastic simulation and techniques like kriging help to capture and quantify the uncertainty in reservoir properties, providing a more comprehensive understanding of the reservoir and aiding in decision-making processes.

Sequential Gaussian Simulation

Sequential Gaussian Simulation is a specific stochastic simulation method used in geostatistics for modelling spatially correlated variables (Demyanov et al., 2018; Pyrcz, 2014). It is a popular technique for generating multiple realisations of reservoir models by sampling from estimated statistical distributions.

SGS assumes that the simulated variable follows a multi-Gaussian distribution in every location and is independent. It involves simulating values at unsampled locations sequentially, one at a time, based on the values already simulated at neighbouring locations. The spatial correlation structure derived from the available data guides the simulation process.

By simulating values sequentially and conditioning each simulation on previously simulated values, SGS preserves the spatial correlation structure of the data. It allows for generating multiple realisations that honour the observed data while also capturing uncertainty and providing insights into different possible scenarios.

SGS and its modifications are widely used in reservoir characterisation and modelling to

address uncertainty in reservoir properties, such as porosity, permeability, and saturation concerning normal-score transformation (Bai et al., 2022).

Truncated Gaussian Simulation

Truncated Gaussian Simulation is a variation of SGS that incorporates truncation limits into the simulation process (Mariethoz et al., 2009). Truncation limits are thresholds that restrict the range of values that can be simulated for a categorical variable, for example, facies.

TGS involves modifying the simulation process to ensure the simulated values fall within specified truncation limits. Hence, simulation results depend on truncation limits (Fachri et al., 2013; Kyriakidis et al., 1999).

Pluri-Gaussian Simulation

The limitation of SGS and TGS is overcome by Pluri-Gaussian Simulation (Yarus et al., 2012). Pluri-Gaussian simulation is a geostatistical method that allows for modelling multiple Gaussian distributions with different parameters. Unlike SGS, which assumes a single Gaussian distribution throughout the simulation process, pluri-Gaussian simulation recognises that different regions within a spatial domain may exhibit different statistical characteristics.

In pluri-Gaussian simulation, the spatial domain is divided into subregions or zones, each representing a distinct geological or geophysical unit with its own set of statistical parameters. These parameters typically include the Gaussian distribution's mean, variance, and spatial correlation structure within each zone.

Pluri-Gaussian simulation allows for representing complex geological or geophysical scenarios where different regions have distinct statistical characteristics. It is particularly useful in situations with abrupt changes in spatial variability or where there is a need to capture multiple geological units with different properties (Sebacher et al., 2017; Silva et al., 2017).

Sequential Indicator Simulation

Sequential Indicator Simulation is a geostatistical simulation method commonly used to model categorical or discrete binary variables (Demianov et al., 2018; Pyrcz, 2014). Unlike Gaussian-based simulation methods, SIS focuses on capturing the spatial patterns and transitions between different categories or classes within a spatial domain.

In SIS, the spatial domain is divided into subregions or zones, similar to pluri-Gaussian simulation, but in a non-parametric way. However, instead of modelling continuous vari-

ables with Gaussian distributions, SIS models discrete variables with categorical indicators. These indicators represent the presence or absence of a specific category or class within each zone.

The simulation starts with an initial random assignment of indicator values within each zone. Then, the indicator values are updated sequentially based on the conditioning data and transition probabilities for each zone. This sequential updating process ensures that the simulated indicators honour the observed data and capture the spatial patterns and transitions between different categories.

Once all the indicator values have been simulated for each zone, they can be combined to create a final simulated field representing the categorical variable of interest. This combination can be done using various approaches, such as majority voting or weighted summation based on the proportion of each category in the overall domain.

SIS is particularly useful when the variable of interest exhibits distinct categories or classes with sharp boundaries and spatial dependencies. It allows for creating simulated fields that capture the complex spatial patterns and transitions between different categories, providing a more realistic representation of the categorical variable (C.V. Deutsch, 2006). TGS can also do that, but it assumes the fixed sequence of facies.

2.1.2 Object-based algorithms

Variogram-based techniques can replicate spatial continuity and the proportion of facies, making them appropriate for conditioning various data types as they generate realisations on a pixel-by-pixel basis. Nevertheless, the variogram only captures basic spatial correlations, making it challenging for two-point statistics to reproduce complex connectivity patterns like meandering channels. Consequently, simulated realisations may not accurately capture the true connectivity, which controls the flow, in such cases.

In contrast, object-based algorithms distribute categorical variable values based on pre-defined geometric shapes (C. Deutsch et al., 2002; Haque et al., 2018). The choice of geometric shape is tailored to the geology being modelled and can capture complex and realistic geological features. Spatial correlation is determined by the arrangement of these shapes rather than the variogram. These shapes are placed stochastically within the modelling region using a stochastic optimisation technique conditioned to the data like wells or seismic.

Optimisation algorithms maximise or minimise the value of an Objective Function (OF) based on data. In object modelling, the OF typically aims to minimise the error in model properties, such as the proportions of different facies. Objects are added or removed iteratively over multiple steps until all conditioning criteria are met within an acceptable error.

Object models can be conditional or unconditional to well data.

Uncertainty in the dimensions and proportions of facies bodies can be represented in object modelling by sampling from distributions for object parameters. The geometric properties of objects can vary within certain intervals, such as dimensions, proportions, orientations, sinuosity, and amplitudes, to control their behaviour.

One of the main advantages of object-based algorithms is their interpretability, as the choice of realistic object shapes reflects the nature of the modelled phenomenon. As a result, object-based models find extensive application in domains where pre-existing knowledge of patterns is accessible.

However, there are some weaknesses to consider with object-based approaches (Demianov et al., 2018). The choice of shapes assumes knowledge of the pattern structure and is subject to uncertainty. The computational costs may increase when optimisation techniques necessitate iterative adjustments of spatial patterns to fit the data. Data conditioning can be challenging and numerically complex to achieve through iterative optimisation. Additionally, conditioning object-based realisations to data from different scales require a complex OF and increases computational requirements.

2.1.3 Multi-point statistics

Multi-point statistics (MPS) in geoscience refers to a set of techniques to model and simulate spatial patterns and structures based on the statistical relationships between multiple data points. MPS goes beyond traditional two-point statistics, such as variograms, by considering the relationships among multiple data points within a given neighbourhood (Guardiano et al., 1993; Strebelle, 2002).

MPS algorithms use training images or patterns to capture geological features' spatial correlation and connectivity. These training images contain examples of the desired spatial patterns and are used to guide the generation of simulated realisations. By analysing the patterns and relationships in the training images, MPS algorithms can reproduce complex geological structures and capture the heterogeneity and connectivity of different facies or geological units.

The key idea behind MPS is that the probability distribution of a specific pattern occurring at a given location is conditioned on the patterns observed in neighbouring locations. This allows for generating realistic and geologically plausible models that honour local and global spatial constraints. It is a powerful tool for capturing the complex spatial variability and connectivity of geological features, which is crucial for accurate and realistic simulations in geoscience (Caers et al., 2023).

Like in the aforementioned geostatistical algorithms, the basic MPS algorithm performs poorly with non-stationary data. Therefore, several modifications have been proposed that either add a trend model (Straubhaar et al., 2011) or divide the modelled area to achieve local stationarity (Strebelle, 2002). However, a drawback of MPS is that it does not offer a convenient parameterisation, which can be problematic for certain geostatistical approaches that rely on parameter optimisation, such as the history matching. Furthermore, when considering a prior model, it is important to note that MPS heavily relies on the selection of a training image, which significantly impacts the outcome (Céline Scheidt, 2018).

2.1.4 Limitations of geostatistics

To conclude this section on standard geostatistics, let's highlight the general limitations of the presented algorithms:

- **Stationarity assumption:** Many geostatistical methods assume stationarity, which means that the statistical properties of the data do not change over space. However, many natural phenomena exhibit non-stationarity, where the statistical properties vary across different regions or scales. This can lead to inaccuracies in the modelling and prediction of spatial data.
- **Geostatistics mainly operates within a predominantly linear framework.** While there are exceptions, most algorithms employed in geomodelling exhibit linear characteristics. This holds true even in the context of multivariate conditioning within methodologies such as Multiple-Point Statistics and Sequential Indicator Simulation. The linear nature of these approaches limits geostatistical analyses and modelling, bounding a foundation for understanding and predicting spatial relationships within subsurface reservoirs.
- **Lack of flexibility (linearity):** Traditional geostatistical methods often rely on simple parametric models, such as variograms or kriging, which may not adequately capture the geological complexity of spatial patterns. This lack of flexibility can result in poor representation of spatial variability and limited ability to handle complex data structures.
- **Current geostatistical approaches cannot perform under uncertainty in various geological scenarios.**
- **Data conditioning:** Hard data conditioning allows no space for interpretational uncertainty, e.g. petrophysical interpretation. Soft data conditioning is limited to linear dependency on the aggregated probability attribute. In reality, many secondary soft

information have different non-linear relations with the target property (e.g. seismic).

- **Data sparsity:** Geostatistical methods require a sufficient number of data points to estimate the model parameters accurately. However, in many cases, spatial data are sparse, especially in remote or inaccessible areas. This can lead to unreliable predictions and uncertainty estimates.
- **Computational challenges:** Geostatistical methods can be computationally intensive, particularly when dealing with large datasets or complex models. The computational burden can limit the application of these methods in practice or require simplifications that may compromise the accuracy of the results.
- **Assumption of Gaussianity:** Many geostatistical methods assume that the underlying data follows a Gaussian distribution. However, this assumption may not hold for many real-world datasets, which often exhibit non-Gaussian or heavy-tailed distributions. Failing to account for non-Gaussianity can lead to biased estimates and unreliable predictions.
- **Lack of integration with other data sources:** Geostatistical methods often focus solely on the spatial component of the data and may not effectively incorporate additional sources of information, such as remote sensing data or auxiliary variables. Integrating multiple data sources can enhance prediction accuracy and provide more robust uncertainty estimates.
- **In the context of MPS,** the lack of a convenient parameterisation means that it may be challenging to directly incorporate MPS-generated patterns into history matching workflows. The complex nature of MPS models and the absence of easily adjustable parameters make it difficult to optimise the model's performance against observed data.

All this leads to the development of more flexible approaches which can deal with the limitations above, such as machine learning techniques. They offer a powerful set of tools for geostatistics, enabling improved data analysis, prediction, optimisation, and decision-making in various geospatial applications.

2.2 Machine Learning approaches in reservoir modelling

This section introduces the generative machine learning (generative models) and their widespread applications in geosciences and beyond. Thus, the first subsection introduces general components of deep generative models.

In just a few years, generative models have become a highly effective method for unsupervised learning, enabling to learn various data distributions with great success. The primary objective of generative models is to understand the actual data distribution from the training set, allowing the generation of new data points with some degree of variation. However, it is not always feasible to precisely learn data's explicit or implicit distribution. Therefore, we strive to create a model distribution that closely resembles the actual data distribution by utilising the capabilities of neural networks. Generative models have shown promising results in various fields, including:

- **Computer Vision:** Variational Autoencoder (VAE) and Generative Adversarial Network (GAN) have been used for tasks like image synthesis, image inpainting, and image super-resolution.
- **Natural Language Processing:** Recurrent Neural Network (RNN) and Transformers have been used for tasks like language translation, text generation, and sentiment analysis.
- **Recommender Systems:** Generative models have been used for collaborative filtering and personalised recommendation systems to reduce the dimensionality of user-item interactions.
- **Bioinformatics:** To analyse genetic data, protein structure prediction, and drug discovery.
- **Finance:** Generative models have been used in financial time series analysis, portfolio optimisation, and risk management.
- **Anomaly Detection:** To detect anomalous patterns in various domains such as cybersecurity, fraud detection, and network intrusion detection.
- **Robotics:** To robot motion planning, object recognition, and scene understanding.

These are just a few examples, and generative models have also found applications in many other fields. Generative machine learning in geosciences refers to using machine learning models and algorithms to generate new data that follows the statistical patterns and characteristics of existing geoscientific data. It aims to learn the underlying distribution of the data and generate new samples similar to the observed data. In the context of geosciences, generative machine learning can be used for the following tasks:

- **Data Augmentation.** It can be trained on existing data to generate additional synthetic samples that can be used to augment the training dataset. This helps improve the performance and generalisation of other machine learning models.
- **Simulation and modelling:** To simulate realistic geoscientific data, such as geolog-

ical formations, weather patterns, or ocean currents. These simulated datasets can be valuable for testing hypotheses, conducting sensitivity analyses, or generating scenarios for decision-making.

- **Uncertainty quantification:** To estimate the uncertainty associated with geoscientific predictions. By generating multiple plausible samples from the learned distribution, it is possible to quantify the range of possible outcomes and assess the confidence in the predictions.
- **Data synthesis and completion:** To fill in missing or incomplete data in geoscientific datasets. By learning the statistical patterns of the available data, the model can generate plausible values for the missing observations.

Generative methods prove advantageous over geostatistical approaches in addressing geological uncertainty within the history matching frameworks due to their ability to capture complex, non-linear relationships between input parameters and observed production data. Unlike geostatistical methods that often assume deterministic relationships, generative methods operate within a probabilistic framework, providing multiple plausible scenarios. This is vital in history matching, where the challenge lies in finding a model that fits observed data and accounts for the inherent uncertainty in geological parameters.

The pivotal strength of generative models lies in their ability to produce diverse and realistic scenarios, encapsulating the spectrum of potential outcomes. This contrasts conventional methods like Convolutional Neural Networks and Support Vector Machines, as generative models inherently operate within a probabilistic framework. This probabilistic nature facilitates the generation of varied scenarios and the quantification of associated uncertainty for each scenario. Furthermore, the probabilistic framework of generative methods enhances decision-making processes by providing insights into potential outcomes and their respective uncertainty.

The family of generative machine learning techniques includes Generative Adversarial Networks, Variational Autoencoders, flow-based deep generative models and diffusion models. Due to its recent appearance, the last two are not represented in geosciences yet. These models learn the underlying distribution of the data and generate new samples by sampling from this learned distribution. Specific examples of generative learning methods in geostatistics and geological modelling will be discussed in **Section 2.3: Recent advances of generative modelling in geoscience**.

2.2.1 Generative Adversarial Network

Generative Adversarial Network (I.J. Goodfellow et al., 2014) has demonstrated impressive results in generating realistic content such as images, language, and music (Engel et

al., 2019; Karras et al., 2021; Zhu et al., 2019). It is inspired by game theory, where two models, a generator and a discriminator, compete and improve each other simultaneously.

A GAN consists of two models: a generator G and a discriminator D . The discriminator estimates the probability of a given sample $x \in X$ being real or fake, while the generator produces synthetic samples x based on noise input $z \in Z$. The goal of the generator is to create samples that are as realistic as possible, fooling the discriminator.

During training, the generator and the discriminator are engaged in a competitive game. The generator aims to deceive the discriminator, while the discriminator strives to distinguish real from fake samples accurately. This dynamic motivates both models to improve their performance, figure 2.1.

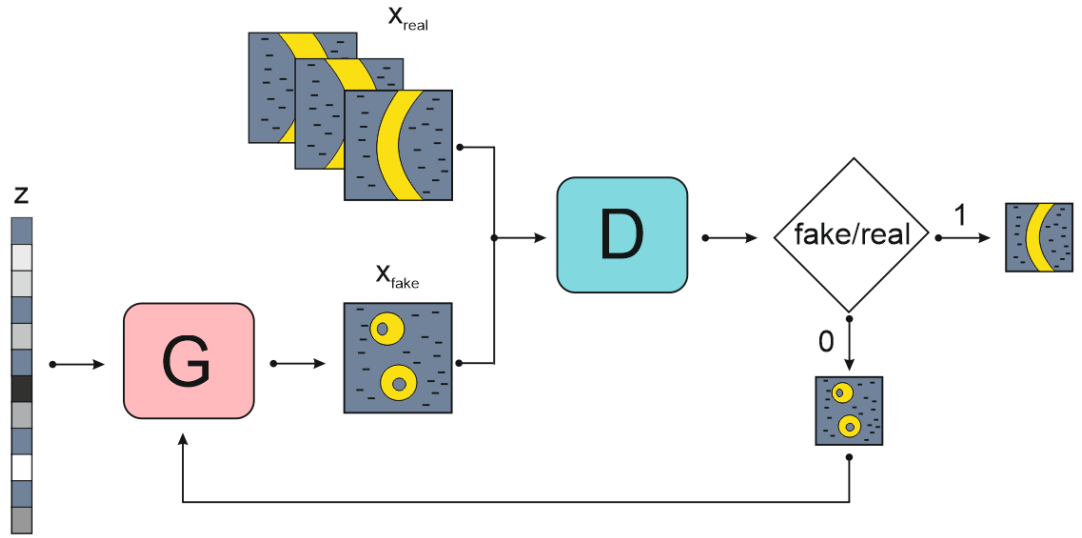


Figure 2.1: Architecture of a GAN. A generator G outputs synthetic samples given a noise variable z input. A discriminator D estimates the probability of a given sample coming from the real dataset

The training process involves optimising a loss function that consists of two components. The first component maximises the ability of the discriminator to sort fake objects from real by maximising $\mathbb{E}_{X \sim P_{real}(X)} [\log D(X)]$. At the same time, the discriminator is expected to provide a probability $D(G(Z))$ as close as possible to zero, hence maximising $\mathbb{E}_{X \sim P_{fake}(X)} [\log(1 - D(G(Z)))]$. The generator is trained to minimise the same component $\mathbb{E}_{X \sim P_{fake}(X)} [\log(1 - D(G(Z)))]$ to produce high probability fake objects which are resembling real objects. Combining two components, the loss function of GAN is as follows:

$$\min_G \max_D L(D, G) = \mathbb{E}_{X \sim P_{real}(X)} [\log D(X)] + \mathbb{E}_{X \sim P_{fake}(X)} [\log(1 - D(G(Z)))] \quad (2.1)$$

Despite its success in generating realistic objects, the training process can be slow and un-

stable. Salimans et al. (2016) discussed issues with GAN's gradient descent-based training procedure. The independent updates of the generator and discriminator can hinder convergence.

When the discriminator is perfect, the loss function becomes zero, resulting in no gradient for further learning. This creates a dilemma in GAN training: if the discriminator performs poorly, the generator lacks accurate feedback, and if the discriminator performs well, learning becomes slow or stagnant (Arjovsky et al., 2017).

Another common problem in GAN training is mode collapse, where the generator produces limited and repetitive outputs. This hinders the generator from capturing the complexity of real-world data distribution (Arjovsky et al., 2017). It should be mentioned that even in trained GAN (Mosser et al., 2017b) indicates less variation in the generated samples than in the training samples, which decreases further uncertainty quantification.

Several modifications to the vanilla GAN have been proposed to address these challenges to enhance its performance (Arjovsky et al., 2017; Karras et al., 2018; Salimans et al., 2016).

2.2.2 Variational Autoencoder

Autoencoder

I will start with a general concept of Autoencoder (AE), developed to reconstruct high-dimensional data using a neural network model with a narrow bottleneck layer in the middle. The autoencoder was initially introduced by Kramer (1991) as a Nonlinear Principal Components Analysis (NLPCA), an extension of Principal Components Analysis (PCA). The NLPCA method is used to identify and eliminate non-linear correlations in data, as opposed to PCA. Similar to principal component analysis, it can be used to reduce the dimensionality of the data by removing redundant information. It was stated that the approach can effectively eliminate non-linear correlations in the data by utilising basis functions that accurately model any bounded continuous function. The NLPCA methodology involves training a neural network with a bottleneck layer to perform the identity mapping. The network incorporates three hidden layers into the architecture to achieve non-linear fitting. So we can see an early idea of a modern VAEs.

AE has also been referred to as the auto-associator (DeMers et al., 1992; Japkowicz et al., 2000; Oja, 1991), and its earliest applications can be traced back to the 1980s (Rumelhart et al., 1986). While its traditional use was for dimensionality reduction or feature learning, it has become widely employed for learning generative data models through VAE (Cetin et al., 2023; Pham et al., 2022).

One beneficial outcome of AE is dimension reduction, where the bottleneck layer captures a latent encoding of decreased dimensionality. The latent encoding can serve multiple purposes, including facilitating search operations, assisting in compressing data, and unveiling the latent factors responsible for generating the data.

AE consists of two networks: an Encoder Q parameterised by ϕ and a Decoder P parameterised by θ . The low-dimensional vector learned for input x in the bottleneck layer is denoted as z , and the reconstructed input is denoted as \hat{x} .

The parameters ϕ and θ are optimised to produce a reconstructed data sample \hat{x} that matches the original input x . Different measures can be used to assess the difference between two vectors, including cross-entropy or MSE.

The encoder network effectively achieves dimensionality reduction, similar to PCA. Additionally, the autoencoder is explicitly optimised for reconstructing the data from the vector z .

Variational Autoencoder

The concept of AE was further developed by Kingma et al. (2022). The structure of the VAE resembles an AE, where instead of mapping the input x to a fixed vector z , it is mapped to a distribution labelled as $P(Z)$, parameterised by θ . The input data x and the latent vector z can be fully defined by the prior $P(Z)$, the likelihood $P(X|Z)$, and the posterior $Q(Z|X)$ parameterised by ϕ .

The conditional probability $P(X|Z)$ defines a generative model, similar to the decoder mentioned in the AE section. The approximation function $Q(Z|X)$ parameterised by ϕ is the probabilistic Encoder, figure 2.2.

The loss function of the VAE consists of two terms: the negative log-likelihood with a regulariser. Since we don't know the global representations shared by all data points, the loss function can be broken down into terms that only rely on individual data points. The total loss is the sum of individual losses (equation 2.2) for each data point.

$$Loss_i(\theta, \phi) = -\mathbb{E}_{Q_{\theta(z|x)}} [\log P_{\phi}(x|z)] + KL [Q_{\theta}(z|x) || P(z)] \quad (2.2)$$

The first component (*reconstruction loss*) evaluates how well the decoder reconstructs the data for each data point. If the reconstruction is poor, it indicates that the decoder's likelihood distribution does not assign much probability to the true data, resulting in a high cost in the loss function.

The regulariser term, on the other hand, measures the Kullback-Leibler (KL) divergence

between the encoder’s distribution and $P(z)$, which is usually a standard Normal distribution with mean zero and variance one. This divergence quantifies the amount of information lost when using the encoder’s distribution to represent the data. It serves as a measure of how close the encoder’s distribution is to the standard Normal distribution.

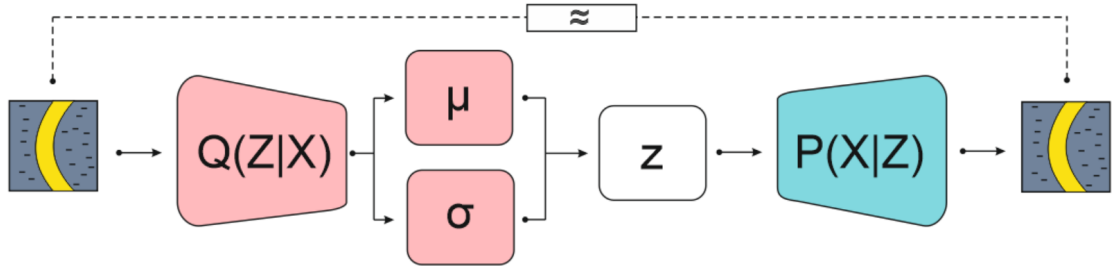


Figure 2.2: Architecture of VAE model with the multivariate Gaussian assumption. An Encoder $Q(Z|X)$ converts the initial object into a latent vector representation z . Decoder $P(X|Z)$ transforms the latent vector back into the initial object. If the Decoder and Encoder are properly trained, the restored object should be similar to the initial

To train the VAE, gradient descent is used to optimise the loss by adjusting the parameters of both the Encoder and Decoder (θ and ϕ).

Some limitations of VAEs include difficulty capturing complex data distributions. VAEs assume that the data distribution can be approximated by a simple Gaussian distribution. This assumption may not hold for complex and multimodal data distributions, leading to sub-optimal reconstructions. VAEs tend to produce blurry reconstructions, especially when dealing with high-resolution images. This blurriness is due to the smoothness assumption made by the VAE’s decoder, which can lead to loss of fine details. VAEs are sensitive to the choice of hyperparameters, such as the dimensionality of the latent space. While VAEs can learn a low-dimensional representation of the data, it may not always disentangle the underlying factors of variation. This means that different dimensions in the latent space may not correspond to separate and interpretable features of the data.

There is a family of VAE modifications, presented by Gregor et al. (2019), Higgins et al. (2017), Oord et al. (2018), and Tolstikhin et al. (2017) to overcome mentioned VAE limitations.

2.2.3 Flow-Based generative model

A flow-based generative model is a type of model that is constructed using a sequence of invertible transformations (Rezende et al., 2016). Unlike other models like GAN and VAE, this model explicitly learns the data distribution and uses the negative log-likelihood as the loss function.

In deep learning models, the embedded probability distribution (posterior $P(Z|X)$) must

be simple enough to calculate derivatives efficiently for backpropagation. This is why Gaussian distributions are often used in latent variable generative models, even though real-world distributions are more complex, figure 2.3.

A Normalizing Flow (NF) model is introduced to address this limitation. Based on the change of variables theorem, NF transforms a simple distribution into a complex one by applying a sequence of invertible transformations (Kaplan, 2002). The model can obtain a probability distribution of the final target variable by flowing through these transformations.

NOTE: A change of variables is a mathematical technique used to simplify problems by replacing the original variables with functions of other variables. The purpose is to make the problem easier to solve or to transform it into a problem that is better understood.

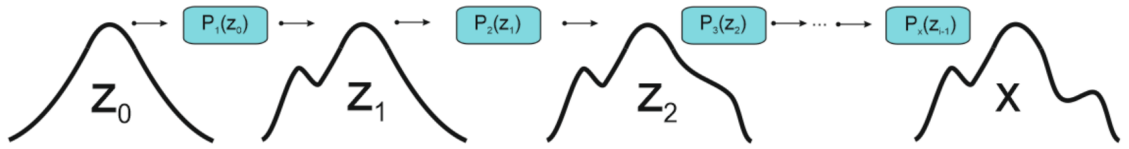


Figure 2.3: Illustration of a Normalizing Flow model, transforming a simple distribution z_0 to a complex one x step by step by $P_i(z_{i-1})$

The connection between every successive pair of variables in the probability density function sequence is determined. The equation of the output x can be expanded step by step until tracing back to the initial distribution z_0 . This chain of successive distributions is called a NF.

For the computation of the equation, the transformation function must possess two essential properties: it should be invertible, and the calculation of its Jacobian determinant should be computable in a manageable amount of time.

With NFs, the exact log-likelihood of input data becomes tractable, making the training criterion of flow-based generative models simply the negative log-likelihood over the training dataset (equation 2.3).

$$Loss = -\frac{1}{|X|} \sum_{x \in X} \log(P(x)) \tag{2.3}$$

In Variational Autoencoders, if we want to model the posterior $Q(Z|X)$ as a more complex distribution than a simple Gaussian, we can use NFs to transform the Gaussian enhancing the accuracy of density approximation (Su et al., 2018). The Encoder then predicts scale

and shift terms that are functions of the input.

Like any other generative model, the flow-based model has its limitations. Training flow-based models can be computationally expensive, especially for high-dimensional data. This is because the model needs to invert the transformation from the data space to the latent space during training, which requires calculating the determinant of the Jacobian matrix. As the dimensionality of the data increases, this calculation becomes more time-consuming. Flow-based models typically assume a simple latent space structure, such as a Gaussian distribution. This assumption can limit their ability to capture complex and multimodal data distributions. In cases where the true data distribution deviates significantly from the assumed latent distribution, flow-based models may struggle to generate realistic samples. Flow-based models require storing the entire dataset during training to compute the Jacobian determinant. This can be memory-intensive, especially for large datasets.

Additionally, as the model complexity increases, the memory requirements also increase, making it challenging to scale flow-based models to handle big data problems. Like other generative models, it can suffer from mode collapse, failing to capture all modes in the data distribution. This means that the generated samples may lack diversity and exhibit repetitive patterns. Flow-based models are primarily designed for continuous data and may struggle to efficiently handle discrete or categorical data. Although some extensions have been proposed to address discrete variables, they often involve additional complexity and can still face challenges in capturing the full complexity of discrete data distributions.

There are many examples of flow-based models utilisation in data science (Dinh et al., 2015; 2017; Kingma et al., 2018). However, it is worth noting that the technology is not widely used to solving geostatistics problems. Guan et al. (2021) presented flow-based model for reconstructing porous media by training on 2D greyscale images to reconstruct 3D representations. Therefore, I will not consider flow-based generative models further, as there are not enough examples of successful use.

2.2.4 Diffusion Models

Diffusion models are influenced by non-equilibrium thermodynamics and are based on a Markov chain of diffusion steps. These steps gradually add random noise to the data and then learn to reverse the diffusion process, ultimately generating desired data samples from the noise (Ho et al., 2020; Sohl-Dickstein et al., 2015; Song et al., 2020). Diffusion models follow a fixed procedure and have the same latent variable dimensionality as the original data, as opposed to VAE or flow models.

To explain the forward diffusion process, let's consider a data point $x_0 \in X$, where X is some real data distribution. We add a small amount of Gaussian noise in S steps, resulting

in a sequence of noisy samples x_1, \dots, x_S . The variance schedule $\{\beta_s \in (0, 1)\}_{s=1}^S$ controls the step sizes during this process.

$$Q(x_s|x_{s-1}) = \mathcal{N}(x_s; \sqrt{1 - \beta_s}x_{s-1}, \beta_s \mathbf{I}) \quad (2.4)$$

$$Q(x_{1:S}|x_0) = \prod_{s=1}^S Q(x_s|x_{s-1}) \quad (2.5)$$

NOTE: If we add up two Gaussian distributions $\mathcal{N}(0; \sigma_1^2 \mathbf{I})$ and $\mathcal{N}(0; \sigma_2^2 \mathbf{I})$ we will get $\mathcal{N}(0; \sigma_1^2 + \sigma_2^2 \mathbf{I})$

As the step size S increases, the data sample x gradually loses its distinguishable features. Eventually, when S is big, it is equivalent to an isotropic Gaussian distribution

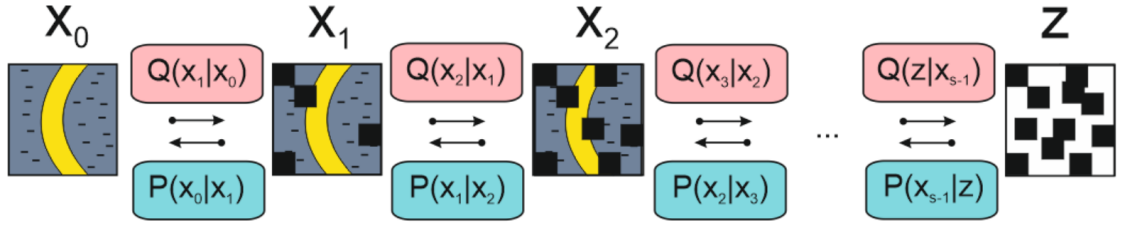


Figure 2.4: Architecture of Diffusion model. A series of the Markov chain of forward diffusion processes $Q(x_s|x_{s-1})$ add a small amount of Gaussian noise for a previous object x . A reverse series $P(x_{s-1}|x_s)$ transforms latent vector z back into initial object x .

The main objective of training a diffusion model is to learn the reverse process. By traversing backwards along this chain, we can generate new data, starting from a pure Gaussian noise $P(x_S) = \mathcal{N}(x_S; 0, \mathbf{I})$, figure 2.4

$$P(x_{0:S}) = P(x_S) \prod_{s=1}^S P(x_s|x_{s-1}) \quad (2.6)$$

The training of a Diffusion Model involves finding the reverse Markov transitions that maximise the likelihood of the training data. This is achieved by minimising the variational upper bound on the negative log-likelihood.

$$L_{total} = \mathbb{E}_Q \left[-\log \frac{P(x_{0:S})}{Q(x_{1:S}|x_0)} \right] \geq \mathbb{E} \left[-\log(P(x_0)) \right] \quad (2.7)$$

We are particularly interested in formulating the L_{total} in terms of KL divergences because the transition distributions in our Markov chain are Gaussians. It has been shown by Sohl-Dickstein et al. (2015) that L_{total} can be rewritten almost entirely regarding KL

divergences.

$$L_{total} = L_0 + L_1 + \dots + L_S \quad (2.8)$$

$$L_0 = -\log P(x_0|x_1) \quad (2.9)$$

$$L_{s-1} = KL(Q(x_{s-1}|x_s, x_0)||P(x_{s-1}|x_s)) \quad (2.10)$$

$$L_S = KL(Q(x_S|x_0)||P(x_S)) \quad (2.11)$$

Diffusion generative models have limitations similar to flow-based, such as computational complexity and memory requirements. It does not explicitly model latent variables. While this can be advantageous in some scenarios, it can also limit the model's ability to capture complex dependencies and generate diverse samples. Diffusion models tend to struggle to capture long-term dependencies in the data.

Various improvements were presented to overcome these limitations, such as Ho et al. (2020), Nichol et al. (2021), Rombach et al. (2022), and Song et al. (2020).

As for geoscience, the first mentions of diffusion model utilisations were proposed by Durall et al. (2023). Their study utilised diffusion models for three seismic applications: demultiple, denoising, and interpolation. They conducted experiments on synthetic and real data and compared the performance of diffusion models with established algorithms. The problem was defined as an image-to-image transformation, where certain modifications were provided under an input image so that the output result belongs to the target domain.

The second publication is also dated 2023 by Mosser (2023), where he investigated the use of deep learning techniques for discrete indicator simulations, specifically focusing on the newly developed order-agnostic diffusion models (Hoogeboom et al., 2022) for facies modelling. However, one limitation of this approach is that it relies on an autoregressive method, which involves a single forward pass by the U-Net and sampling from the generated probability distribution for each grid block in the model.

2.3 Recent advances of generative modelling in geoscience

After a general description of the classical geostatistical algorithms and their drawbacks, as well as various generative models, let's examine examples of using generative models to solve specific tasks in geological modelling.

As noted earlier, generative models can potentially improve the quality of reproducing geological objects, considering their uncertainty. However, each type has its features, which can be both advantageous and disadvantageous. Therefore, in the conclusion of

this section, I will focus on one type of generative neural network, which will serve as the basis for further practical research.

The Heriot-Watt University team has published several papers (Chan et al., 2018; 2019a,b), that utilise Generative Adversarial Networks to generate geological representations. These papers successfully capture initial conditioning under uncertainty. The GAN technique trains a neural network to sample from an unknown and intractable distribution, using only samples from that distribution. The result is a generator network that can generate geological realisations with a number of latent parameters = 30. This approach preserves the training data’s visual realism and flow statistics, accurately reproducing both geological structures and flow properties. The team conducted numerical experiments using conceptual images of channels of sand with constant permeability values embedded in a clay background. They tested different conditioning scenarios, such as conditioning on the presence of high permeability channels or low permeability background material in specific locations within the domain.

Dupont et al. (2018) conducted a study on utilising GANs in geomodelling, similar to the research done by the Heriot-Watt University team. They proposed an approach using deep generative models to generate realistic geological models replicating geological patterns and accommodating physical measurements. The experiments involved a training set of 5000 images of fluvial patterns, and the generated samples met the constraints while being diverse and of high quality. The samples exhibited a range of channel widths and curvatures, surpassing the current state of the art, figure 2.5.

The team from Imperial College London conducted a study on utilising GANs to reconstruct micro-structures of porous media (Mosser et al., 2017a,b). They created a

GAN model based on an oolitic Ketton limestone and showed that GANs can represent the statistical and effective properties of segmented representations of the pore space. The produced images successfully replicated the textural characteristics present in the original training image. However, the generated samples had a reduced diversity compared to the

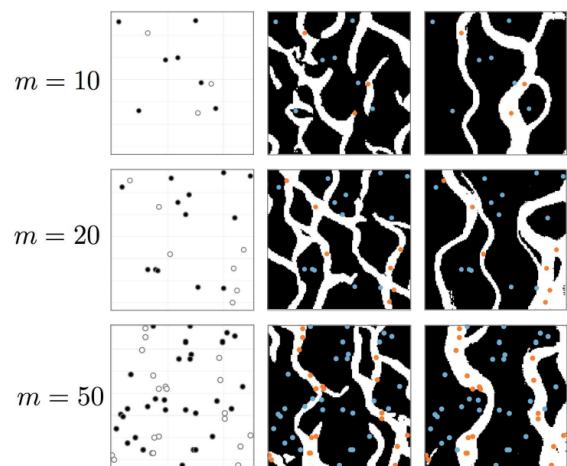


Figure 2.5: Comparison between snesim (Y. Liu, 2006) and GAN-based approach. The left column shows the sparse measurements of different facies. Left: m - number of measurements. Middle: samples generated by snesim. Right: samples generated by the GAN-based. As opposed to the GAN-based approach, snesim produces disconnected and unrealistic channels. From Dupont et al. (2018).

training samples.

The team also solved the inverse problem of reservoir history matching using a deep generative model, specifically for a synthetic 2D channelised reservoir system (Mosser et al., 2019). They provide an optimisation process in the latent space. The loss function was considered: $L(z) = L_{flow}(z) + L_{well}(z) + L_{prior}(z)$. Where $L_{well}(z)$ corresponds to the binary cross-entropy between the observed facies indicator, $L_{flow}(z)$ is a difference between the observed pressures and rates at the well locations averaged over the total duration T of the observed data and $L_{prior}(z)$ is a prior loss.

Studies by Laloy et al. (2018; 2017) attempted to reduce the complexity of an inverse problem with many unknowns using a Spatial Generative Adversarial Network (SGAN). They trained the SGAN using binary channelised training images and were able to generate 2D and 3D realisations quickly. The SGAN had a low-dimensional parameterisation, which allowed for efficient probabilistic inversion using advanced Monte Carlo methods. The inversion results for the 3D case closely matched the target data and visually resembled the true model. The SGAN captured the training image patterns, although there were some slight differences in broken channels and isolated patches. The statistics of the training image were also well-matched by the SGAN.

Compared to their previous work with VAE (Laloy et al., 2017), the SGAN had several advantages. It only required one training image, it was fully 3D, it could handle both binary and multi-categorical data, each dimension of the latent space had a specific influence on the high-dimensional domain, and it resulted in a more compact representation of the domain with a significant dimensionality reduction.

Compared to existing geostatistical methods, the SGAN was advantageous because it generated model realisations that captured the statistics of the training image using a low-dimensional representation of the original model domain.

The study by Sun et al. (2023) demonstrates that Generative Adversarial Networks have the ability to learn and replicate complex meandering fluvial patterns generated by a process-based model (Cojan et al., 2004) called FLUMY. FLUMY simulates various geological processes to create realistic facies models with diverse and intricate geometries, such as channel migration, avulsion, sediment deposition, and levee breach. This research explores whether deep generative models like GANs can accurately reproduce geological patterns at the level of complexity achieved by process-based models. Incorporating geological knowledge, concepts, and interpretations into geo-modelling is a key objective of extending GANs in facies modelling. A new variant of GAN, Fluvial GAN, has been developed, which can effectively replicate the 2D geometrical features of meandering fluvial facies and capture facies transitions along the channel.

A recent study by Fossum et al. (2024) has introduced an approach leveraging the conditional Generative Adversarial Networks with Spatially Adaptive Denormalization (SPADE) to establish an ensemble-based workflow that effectively captures complex geological patterns. The Ensemble Randomized Maximum Likelihood method assimilates data into an ensemble of coarse-scale maps, interpreted as the channel proportions, that serve as SPADE-GAN input. This data assimilation conditions the ensemble of GAN-geomodels to a combination of well data and flow data, thus improving pretrained SPADE-GANs. The method's capacity to replicate previously unseen geological configurations beyond GAN's training data has been convincingly demonstrated in numerical experiments.

Arauco et al. (2017) conducted a study on using Variational Autoencoders for parametrisation in geology. Using ensemble methods, they developed a robust parametrisation method based on deep learning for history matching of facies models. The process involved training a deep learning network on prior facies realisations to identify the main features of the images and construct a reduced parametrisation of the models. This parametrisation was then transformed to follow a Gaussian distribution and updated using Ensemble Smoother with Multiple Data Assimilation (ES-MDA) (Emerick, 2016) to account for observed data. After each data assimilation, the deep learning network was used to reconstruct the facies models based on the initial learning. The process was tested on a synthetic history matching problem.

The results showed that the implementation of the VAE had great potential in the parametrisation problem, with well-defined channels observed in the posterior realisations. However, some "broken" channels were also observed. Overall, the final realisations preserved the main geological characteristics of the prior models and had a good match with the observed production data.

In further investigations by Arauco et al. (2021) and Smith et al. (2019), the authors used deep neural networks to construct a continuous parameterisation of facies for data assimilation with ensemble methods. They conditioned the VAE on facies data and production data in 2D channelised cases. The posterior realisations showed reasonably well-preserved geological characteristics of the prior models and a good match with the observed production data.

However, in a 3D model with fluvial channels generated using object-based simulation, the posterior model was distinguishable from the prior and reference models, with some discontinuous channels and oddly shaped features. The authors attributed this lower performance in the 3D case to the difficulty of training 3D convolutions and suggested that larger training sets may be needed for 3D cases.

In another paper by Ahn et al. (2018), an alternative architecture of inverse modelling

called Artificial Neuron Network — Stacked AutoEncoder (ANN-SAE) is presented. The ANN-SAE workflow for data-driven inverse modelling is shown in figure 2.6. The performance of ANN-SAE were tested for inverse problems under uncertainty of geological scenarios. One such task was waterflooding, using a 2D geological model of a channel reservoir as an example. Dynamic production data were generated using a black-oil simulator. Static data, in the form of grids of petrophysical properties, were obtained from generated reservoir models and compressed using Stacked AutoEncoder (SAE). The study analyses five cases with different numbers of hidden neurons (10, 30, 50, 100, and 300) in Artificial Neuron Network (ANN) and compares them with ANN without SAE.

Finally, the study concludes that ANN-SAE with unsupervised pre-training improves static and dynamic data integration. It shows better performance in terms of computation time and predictability compared to the optimised ANN system. The proposed workflow based on ANN-SAE can investigate non-linear relationships among dynamic data and static reservoir properties, providing geological realism of the reservoir under investigation.

In a different study by K. Lee et al. (2018), Variational Autoencoders were used in a Distance-Based Clustering (DBC) approach to assess uncertainty in reservoir models efficiently. The DBC method grouped similar reservoir models based on their performance. An uncertainty range estimation could be obtained without simulating all models by simulating a representative model from each group. The key factor for the success of DBC was the definition of distance, which measured the dissimilarity between models. This research used a Stacked AutoEncoder to extract the main information from a reservoir model. Instead of using the original vector (PIXEL), the feature vector of VAE was used to compare the initial models and form clusters with similar models. It is important to note that this method was tested on 2D synthetic channelised reservoirs with high geological uncertainty. Two cases were presented: without observed dynamic data and with observed dynamic data.

Jiang et al. (2021) propose the use of a Variational Autoencoder, for efficient parameter-

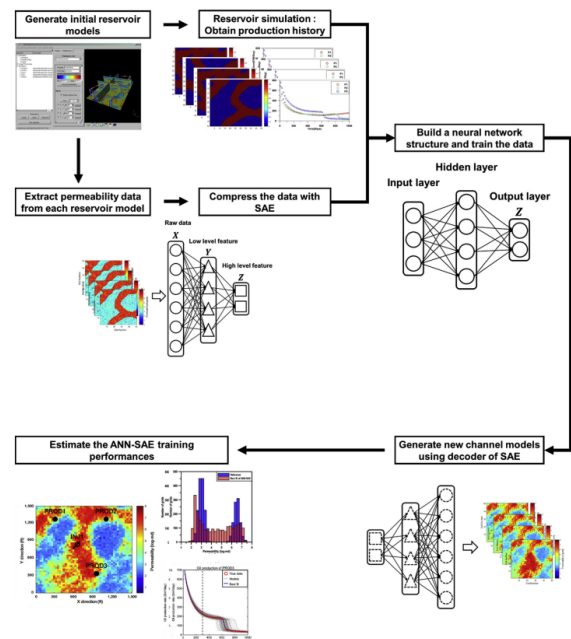


Figure 2.6: ANN-SAE based inverse modelling approach for characterising a heterogeneous channel reservoir. From Ahn et al. (2018)

ising aquifer properties in model calibration problems. These aquifer properties, such as hydraulic conductivity, are spatially distributed and subject to uncertainty in geostatistical models.

Convolutional auto-encoders demonstrate their effectiveness in tasks of non-linear parameterisation, such as the propagation of reservoir properties under uncertainty of geological scenarios, which affect fluid flow. The obtained low-dimensional parameters represent key patterns of property propagation essential for geological modelling. Numerical experiments were conducted to showcase the robustness of convolutional neural networks in tasks of low-dimensional parameterisation for subsequent model history matching. These experiments consider various plausible geologic continuity models, demonstrating the versatility and effectiveness of the proposed approach.

The following paper does not focus on alternative approaches to geological modelling. However, it provides insights into how specific realisations can be identified for future production planning and uncertainty quantification. Carneiro et al. (2018) introduce a geostatistical-based multi-objective history matching methodology based on the Pareto front.

The methodology begins by generating an initial ensemble of subsurface petrophysical properties through stochastic sequential simulation. After fluid flow simulation, each model is ranked based on its match with observed historical production data.

This method enables the generation of regionalised Pareto fronts and provides an ensemble of optimal subsurface models. These models successfully reproduce all observed production data while preserving the uncertainty space, figure 2.7.

Unlike traditional approaches, this approach ranks all models in terms of Pareto dominance for multiple objectives per influence area. It can also consider global objectives, such as total liquid, water and oil production. As observed in testing, one limitation of this methodology is the simplistic use of geostatistical modelling, which relies on two-point spatial statistics.

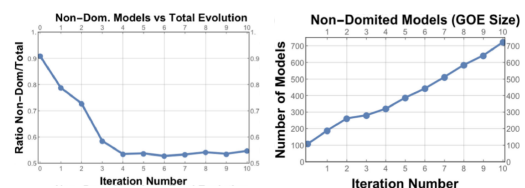


Figure 2.7: Left: the ratio of non-dominated solutions to total solutions generated up to a given iteration. Right: the evolution of the number of non-dominated models per iteration. From Carneiro et al. (2018)

Similar approaches to the history matching problem have been presented by Maschio et al. (2008), who integrate geostatistical modelling and history matching using genetic algorithms and direct search optimisation methods, and Chang et al. (2016), who propose an integrated workflow for updating facies fields while preserving realistic geology and matching production data using the adaptive

Gaussian mixture filter in data assimilation.

According to the research findings, some authors summarise the prospects for applying generative networks to solve geostatistics problems. For example, Mosser et al. (2019) concluded that GANs can solve this type of problem. Still, it is important to consider the evaluation of the generated models in terms of quality, mode-collapse, and spatial statistics representation. However, due to their challenging training and complex relationship between latent and parameter space, GANs may not be the optimal choice for reparametrisation in ill-posed inverse problems.

Bao et al. (2022) provided an extensive comparison of GAN and VAE for identifying geological structures using flow and transport data assimilation. The research findings indicate that GAN has better-generating quality, reproducing channel structures with fewer measured data. However, mode collapse may lead to underestimation of uncertainty. On the other hand, VAE performs better in data assimilation, especially with more measured data, allowing for better localisation of actual channels in the inverse problem. VAE also accurately predicts flow and contaminant transport because its reconstructed structures are close to the reference field. Both VAE and GAN face challenges with continuous variables, but GAN's training process is particularly unstable. In summary, GAN is superior for data generation and produces higher-quality, more realistic images. However, when integrating dynamic data into groundwater models, VAE outperforms GAN in terms of conductivity generated by deep learning methods.

Thus, it can be concluded that despite the high generative capabilities of GANs, VAE technology is more prioritised for tasks of model parameterisation under uncertainty and optimisation through latent spaces, as it allows for direct work in the latent space, stable to train and manage to cover all modes of the data distribution.

2.3.1 Structural Uncertainty in generative modelling

In many cases, grid-based models are constructed after generating structural models. However, it is important to perform joint inversion on both structural and property model uncertainty. Ignoring their interaction can overlook their impact on physical / chemical responses studied, such as fluid flow models or geophysical models (Céline Scheidt, 2018). Although geoscientists recognise the significance of structural uncertainty in reservoir modelling, deterministic structural frameworks are typically used in simulation and history-matching workflows. Due to the time-consuming nature of modifying the structural model and corresponding reservoir grid, it is often assumed that structural uncertainty can be disregarded or only the worst, median, and best-case scenarios are considered for simplicity. Most of the presented works are focused on solving the problem of modelling reservoir property distribution. However, this is only one type of parameter that affects

the dynamic response. The second family of parameters is structural uncertainty.

Faults play a crucial role in engineering operations. However, the location of faults is uncertain due to various factors such as inaccurate geophysical imaging, imprecise representation of vertical surfaces, limited well data, and subjective interpretation.

Computer modelling has advanced to address this uncertainty and allow for the digital representation of complex fault surfaces. This evolution has enabled the incorporation of geometric and topological perturbations in fault modelling. Efforts should be made to develop efficient techniques for generating multiple fault models or perturbing existing models, allowing for a more comprehensive assessment of uncertainty.

By addressing these issues, we can improve our understanding of fault networks and enhance the accuracy and reliability of engineering operations that rely on this knowledge.

The paper by Suzuki et al. (2008a) presents a workflow for history-matching reservoir structures using production data while considering geological and geophysical constraints. The process involves two main steps: first, modelling preliminary structural uncertainty using geological and geophysical data, and second, calibration of reservoir geometry using previously established structural models. The initial structural uncertainty is created by introducing stochastic variations of horizons and faults and considering multiple structural interpretations. The Assisted history matching (AHM) is then put into action. This approach minimises structural uncertainty from insufficient seismic quality by incorporating well production data.

Numerous structural models are presented to capture prior structural uncertainty. Each model gives rise to a population of structural variations, accounting for uncertainty in horizon correlation. The uncertainty in the top horizon position is represented by adjusting the top horizon of each stratigraphic grid. Additionally, the gross thickness of each stratigraphic grid is perturbed through a continuous perturbation field, and perturbations are introduced to the fault surfaces as well.

The prior set of structural models created represents a parameter space for HM. By incorporating this methodology, the authors aim to reduce uncertainty in fault networks and improve the accuracy and reliability of engineering operations.

The paper by Seiler et al. (2010b) presents a reservoir characterisation workflow that continuously updates the structural reservoir model. To address the challenges associated with structural-surface updating, the EnKF is utilised. The method involves creating an ensemble of reservoir models to express the uncertainty of seismic interpretation and time-to-depth conversion explicitly. The top and bottom reservoir-horizon uncertainty are considered as parameters for AHM, which is updated through sequential assimilation of pro-

duction data. An elastic-grid approach is proposed to avoid grid modification challenges. The method's effectiveness is demonstrated through a synthetic example, which shows promising outcomes. During the AHM process, a corner-point grid deformation is provided to reflect alternative structural realisations in the top and bottom reservoir horizons. The adjustment of grid nodes is performed stepwise by updating corner points to match the top and bottom horizons while keeping each other fixed. The algorithm can maintain the original depth of perforated cells by adjusting the perturbation parameter to zero at the corresponding grid nodes. The grid's topology remains constant so that geological properties can be populated without changing property vector dimensions. The authors emphasise the importance of a prior ensemble that incorporates uncertainty and note that no methodology currently exists that handles uncertainty related to faults. They also point out difficulties caused by the framework of structure updating, like making sure the updated surfaces follow horizontal well zones and match observed equilibrium regions.

The uncertainty associated with faults can significantly impact connected volumes, production performances, and final recovery (Lescoffit et al., 2005). This uncertainty is mainly due to poor seismic resolution and a noisy seismic signal around the faults. In HM, the fault network is commonly fixed to a single interpretation, and the focus is on estimating flow properties using fault transmissibility multipliers. To address this issue, the authors propose an extension of the elastic grid algorithm to handle fault geometric uncertainty in the reservoir model and HM workflow (Seiler et al., 2010a). The algorithm updates the existing base-case grid's pillar lines to reflect alternative fault positions and assimilates production data using the EnKF to update fault position and throw. While fault position may not be significant when the fault is not recognised as a flow barrier, fault throw can impact fluid flow and overall field communication. Therefore, fault uncertainty should be considered in addition to the horizon for a complete structural uncertainty assessment of any reservoir.

The study by Hillier et al. (2021b) introduces an approach using graph neural networks for generating three-dimensional structural geomodels. These models are important for various scientific and societal purposes, but current implicit approaches have limitations in handling complex structural settings. The proposed approach utilises unstructured meshes as graphs to perform coupled implicit and discrete geological unit modelling. This allows for incorporating interpolation constraints and simultaneous representation of continuous and discrete properties. The effectiveness of the approach is demonstrated through two case studies. However, a potential limitation is the reproducibility challenge due to the random initialisation of the network's parameters. In some cases, the resulting models may differ significantly if the modelling capacity is insufficient or if there are multiple plausible solutions given the constraints. Future research should focus on integrating additional geological constraints into the modelling process.

2.4 General workflow of AHM and UQ

In geological modelling, specialists encounter the challenge of uncertainty in geological concepts or scenarios. Geological concept is a composition of attributes and dependencies between them concerning spatial distribution. Geological concepts can be modelled by various statistical algorithms that describe the spatial distribution of reservoir attributes or characteristics. These statistical algorithms vary by the choice of governing equations and their parameters. They also bear certain assumptions like linearity, Gaussianity, stationarity, etc. It is necessary to identify all plausible configurations of the system's architectures to quantify the prior uncertainty in the model description of a conceptual geological scenario and its variations. Subsequently, these configurations, deemed geologically feasible, should be specified along with their frequencies in terms of probability density to capture the uncertainty accurately. According to Céline Scheidt (2018), this density estimation is usually made on data available and some prior expert knowledge. Figure 2.8 provides an overview of the different combinations of assumptions in modelling. It visualises the relationship between the data variables and the model.

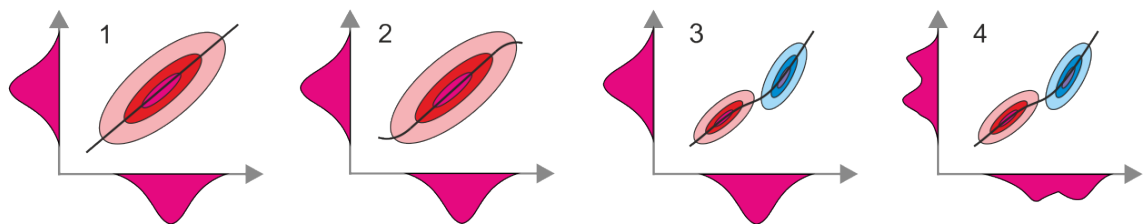


Figure 2.8: Different scenarios of modelling concerning the general assumptions, which are made to data (●, ●) and model (∩): 1. Gaussian—Linear; 2. Gaussian—Linear in the vicinity of the solution region; 3. Gaussian—non-Linear, 4. non-Gaussian—non-Linear

There are four distinct classes of combinations which can be identified:

1. The relationships between variables are linear and follow a Gaussian distribution, e.g. kriging.
2. A model which can be approximated as linear (under certain conditions) in the vicinity of a solution point, kriging would also meet this one.
3. A non-linear model, but Gaussian assumptions are made about the variables involved, e.g. kriging with external drift or multiple facies cases.
4. A non-linear model where the modelling assumptions deviate from Gaussian characteristics. The variables are not assumed to follow a Gaussian distribution. The relationship between porosity and permeability can be largely non-linear, especially in a multi-facie case.

To quantify prior uncertainty effectively, it's essential to describe a prior — all possible system configurations and the probability density that characterises their likelihood of occurrence.

Philip Ringrose and Mark Bentley describe in their book (Ringrose et al., 2015) three main ways of accounting for uncertainty in the process of geological modelling and present them as polar approaches, figure 2.9:

- best guess - this approach involves building a single base model, whose parameters are varied within predetermined boundaries, to model uncertainty.
- multiple stochastic - involves building a set of models probabilistically using geo-statistical algorithms.
- multiple deterministic - involves building several alternative models, usually fewer than in multiple stochastic.

Therefore, to encompass the entire space of uncertainty, it is essential to construct multiple geological models — an ensemble. This raises the question: How many models are needed to describe the space of uncertainty effectively? By effectiveness of description, I mean the minimum number of models required to capture uncertainty. Is it the number of models, or does their nature reflect diversity? For instance, 100 SGS realisations will not cover the uncertainty in total but only its fraction, which is stochastic uncertainty in property distribution. Are there tools that allow for parameterising geological models of different concepts? In other words, is there a universal set of parameters that can describe various geological concepts and be used as a universal set, i.e., a single space for finding solutions that meet specific tasks? It's worth noting once again that in addition to the uncertainty in geological concepts, there is also uncertainty related to the petrophysical and structural parameters of the model, thus significantly expanding the uncertainty space.

However, even if a geologist manages to construct an ensemble of geological models that most comprehensively describe the geological representation, taking uncertainty into account, this ensemble needs to be calibrated to the production history (assuming such history is already available). The model calibration process is an inverse problem with multiple solutions. To tackle inverse problems, it is necessary to assimilate data measurements (log data, production data) with relevant prior states of information related to physical objects. Typically, it is a time-consuming process that requires adjusting initial property

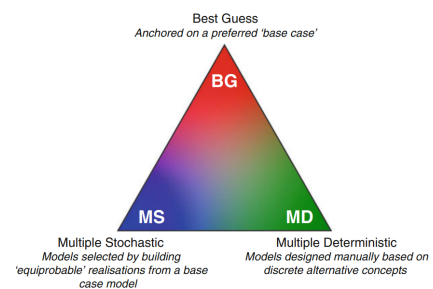


Figure 2.9: Approaches for accounting for uncertainty in geological modelling. From Ringrose et al. (2015)

fields, such as porosity or permeability, so that production parameters, as a result of simulations, closely match historical (real) measurements. At the same time, the geological realism of the finally matched model should not be compromised and must be somehow controlled.

Suppose we project the entire process from constructing a prior ensemble of geological models to the final calibration of such a family to production history. In that case, we have a highly labour-intensive process that demands significant computational and time resources and is generally challenging to implement in real projects. The idea is that current approaches to geological and simulation modelling are primarily based on sequential steps, which is a limitation of the whole approach.

The issues above lead us to the concept of a hidden/latent space, which can be used as a tool to describe various geological scenarios and other sources of uncertainty in a single “encoded” parameter space. Furthermore, the dimensionality of this space can be much smaller than the original. This generative models could be referred to 3^{rd} or even 4^{th} class from figure 2.8, mainly because neural networks serve as non linear models with the inner regularisation of Gaussianity of model’s parameters.

NOTE: An interesting and counterintuitive fact is that in high dimensions, Gaussian distributions are practically indistinguishable from uniform distributions on the unit sphere. Therefore, the process of sampling, distance measuring and so on in such spaces becomes non-trivial. (Aggarwal et al., 2002). The only question that remains is, what are high dimensions?

The parameterisation of a model through some reduced-dimensional spaces is not new. Caers et al. (2008) presented an approach to quantify reservoir uncertainty by measuring the distances between different reservoir model realisations rather than treating each realisation as a point in a complex high-dimensional space. This is achieved by constructing a distance matrix that can be customised for specific problems, such as production profiling or remaining reserves estimation, while keeping the realisations constant. The proposed technique, metric EnKF, is also discussed as an application for updating multiple non-Gaussian realisations using production data. The later research of Suzuki et al. (2008a) introduced the distance metric of similarity of a model geometry. Jung et al. (2013) provided an UQ of complex DFN models with representative scenarios employing distance-based clustering. To provide control of geological realism Kuznetsova (2016) explored three different classifiers, such as k-means, Support Vector Machines and Random Forest, while searching in a metric space for the HM. Demyanov et al. (2018) presented a workflow that addresses the uncertainty of a conceptual geological model and measures its impact on reservoir production profiling. The process employs adaptive stochastic sampling and

Bayesian inference to estimate uncertainty from several HM models, each serving as a solution to an inverse problem. This problem is addressed by sampling a unified space of geological model parameters and reservoir model descriptions. This approach captures uncertainty across diverse modelling concepts derived from multiple geological interpretations.

For us, the latent space of a generative model can serve as a kind of source for geological realisations, with the solution to the inverse problem being reduced to searching for regions of the LS that meet the specified conditions. Essentially, the calibration process is reduced to an optimisation problem in the parameter space of the generative model, which is the latent space.

2.5 Research objectives

The literature review section showed that classical geostatistical modelling tools have significant limitations that restrict their ability to model geological objects, considering modern needs, such as uncertainty quantification and the ability to effectively provide history matching (solving the inverse problem).

Modern generative machine learning models are considered by the scientific community as one of the promising directions for development, as indicated by numerous publications. It is worth noting that most publications focus on the use of generative models to solve the following types of problems: parameterisation of geological representations using latent spaces of decreased dimensionality; finding implicit (non-linear) relationships between parameters that determine the model; solving the inverse problem via latent space to search for optima, considering various types of uncertainty more efficiently. Thus, I have outlined the range of issues that researchers are currently addressing.

The choice of generative methods is appropriate to tackle the problem of inferring deterministic scenario uncertainty in the history matching problem because it can capture complex relationships between input parameters and output responses. In the context of history matching, this means that generative models can effectively link non-linear and non-stationary relationships between uncertain input parameters and observed production data. These methods can generate realistic and diverse scenarios that capture the range of possible outcomes. Using generative models, researchers can explore different scenarios and quantify the uncertainty associated with each scenario. This allows for a more comprehensive understanding of the uncertainty in the HM problem and facilitates better decision-making.

It should be noted that practically all the reviewed publications only consider a single type of uncertainty: the uncertainty of the spatial or structural features. Oliver et al. (2011)

emphasised that there is no universal approach to taking into account uncertainty; rather, it depends on the object of consideration. It is evident that certain significant sources of uncertainty still cannot be effectively managed. The most crucial issues are reservoirs with complex geology. Current methods either oversimplify the spatial relationships of facies or cannot consistently update properties. While initial progress has been made in integrating uncertainty into the geometry of reservoir surfaces, complex faulting patterns are currently beyond our capabilities. It is important to understand that the spatial distribution of properties is a key parameter controlling flow, but not the only one, e.g. structural, fluids, etc.

Céline Scheidt (2018) emphasises that a reservoir model consists of both gridded (spatial) and non-gridded components (fluid properties) parameters. The gridded component includes the structural and stratigraphic model, which describes the location and geometry of bounding layers, horizons, and faults in the reservoir. Therefore, it is important to consider various types of uncertainty in reservoir modelling, including reservoir charge, reservoir structure, petrophysical properties, and reservoir fluid properties. In this regard, generative machine learning methods are of particular interest, as the technology leverages the notion of latent space, which can implicitly combine the listed parameters, taking uncertainty into account. Such latent space is a convenient tool for UQ or Assisted history matching process.

It is important to understand why the scientific community mainly focuses on approaches considering uncertainty in property distribution (petrophysical properties). Still, structural uncertainty (reservoir structure) is represented to a lesser extent. What technological limitations currently exist. In addition to the stated research goals, the development of a tool capable of combining geological and structural uncertainty to solve the inverse problem is one of the distinguishing features of the thesis.

New technologies related to generative models also require new tools that can improve calculation efficiency and convenience of analysis for the modeller. The construction of geological models of a new type is based on parameterisation and tuning through hidden representations, which are almost always not explicitly interpretable. Therefore, the development of new analytical approaches is required. In my work, I describe and test a series of ideas that can serve as tools for auxiliary analytics.

To be more specific, this work will focus on developing a generative machine learning model that technically allows for the combination of various types of uncertainty, such as uncertainty of geological concept and structural uncertainty, and generates 3D models. The model should also allow for the calibration on production history without losing geological realism. Advanced analytics tools will also be presented, allowing for control and analysis of the generative model's performance.

In summary, the objectives of the thesis may be stated as follows:

- Utilisation of generative machine learning methods to describe geological features of different natures in a single parametric space and identification dependencies implicitly
- Consideration the uncertainty of different geological scenarios in the process of history matching and uncertainty quantification
- Combination of depositional and structural uncertainty as single process of generating geological representations
- Development of tools for implicit control of geological realism through the latent variables of a generative model
- Development of analytical tools for the efficient utilisation of generative models

The brief outline of the work follows this sequence:

- Methodological development and justification of VAE approach to parameterising uncertainty of various geological natures (depositional, structural) through the concept of a lower-dimensional latent space
- Introduction and justification of the Graph-Based approach as a more suitable method for modelling volumetric geological objects, allowing for better capture of dynamic characteristics of the geological model without compromising geological realism
- Definition of Three-Component Objective Function for AHM, aiming to achieve a static well data match, dynamic data match, and geological realism constraint in the latent space
- Development of tools to increase the efficiency of the generative model's performance and analysis
- Demonstration of advantages of the Graph-Based approach compared to conventional VAE using a simple synthetic example
- Generalisation of methodologies and approaches for analysing the performance of the generative network on synthetic examples and a full-field benchmark AHM study.

In my work, I do not focus on computational power and calculation speed up because the power changes so quickly that what seemed impossible to calculate five years ago is now a common task, especially with the advent of quantum computing (Dukalski et al., 2023; Mato et al., 2021; Sleeman et al., 2020). The primary focus of my work is the quality of

reproducing objects from a geological and dynamic response perspective, as well as the additional possibilities that arise when using the technologies described in the following chapter.

Chapter 3

Methodology

3.1 VAE description

This chapter describes the fundamental technologies that will be utilised in the subsequent chapters. I will begin with the mathematical formulation of the basic concept of a Variational Autoencoder, highlighting its primary technological limitation and then transitioning to an enhanced version, the Wasserstein Autoencoder. Next, the graph version of Wasserstein Autoencoder will be introduced, allowing for a more natural description of geological objects and their inherent uncertainty, such as depositional and structural uncertainty. To address the inverse problem, I will introduce the concept of an objective function and an optimiser. The objective function in this work consists of three components: the reproduction error of well-log data, production data, and geological realism. The first two components are mathematically straightforward, defined by mean squared error. In contrast, the geological realism error is represented using a latent space density metric, which, in turn, requires justification for the nonlinear structure of the latent space. At the conclusion of the chapter, I will present several methods for analysing the structure of multidimensional spaces, aiding in the evaluation of the generative model's performance. Ultimately, the reader should gain an understanding of the tools employed in this work and the reasons behind their utilisation.

In geoscience, a real geological object can be considered to have infinite dimensionality because it can be examined at various scales, ranging from macro (seismic) to micro (core) scales, changes over time, and has uncertainty. However, its representation needs to be discretised during its modelling regarding volume and time. But even in this case, the dimensionality of the model is usually very high (Céline Scheidt, 2018).

$$\dim(\text{geomodel}) = N_{\text{timesteps}} \cdot (N_{\text{gridcells}} \cdot N_{\text{properties}} + \dim(\text{prop})) = n \quad (3.1)$$

Where $\dim(prop)$ — dimensionality of non-gridded model variables (e.g. PVT, relative permeability).

NOTE: Here I denote some basic notation: calligraphic letters are used for sets - \mathcal{A} ; capital letters for random variables - A ; lower case letters for values - a

Each geological model is described using thousands or even millions of dimensions as a gridded property distribution. In this regard, models are highly redundant due to the spatial continuous nature of geology. This opens the room in modern geological modelling to reduce the dimensionality of the space that describes the geological media. One potential solution to the problem of describing high-dimensional spaces is generative ML. The Variational Autoencoder approach is considered, where the model can be represented as a distribution $P(X)$, defined over data X in a space \mathcal{X} of dimensionality n .

Suppose we have examples from the unknown distribution P_{real} , and our goal is to find a probability distribution function (PDF) $P(X)$, which is similar to P_{real} , $P(X) \sim P_{real}$. In other words, we want to find some distribution that will be as close as possible to nature. We assume that our initial X , which we observe during exploration and subsequent interpretation, “lives” in some *manifold* \mathcal{Z} (I will provide a comprehensive definition of this term later). Dimensionality \mathcal{Z} much less than \mathcal{X} ($m \ll n$). Formally, we assume there is a vector of latent variables Z , which we can sample according to some $P(Z)$ to construct $P(X) \sim P_{real}$.

We have a family of deterministic functions $f(Z; \theta)$, parameterised by θ in some space Θ , where $f : \mathcal{Z} * \Theta \rightarrow \mathcal{X}$. We want to find such f by optimisation over θ . In other words, we want to maximise the probability of each X in the training set (Doersch, 2021). Let’s redefine $f(Z; \theta)$ by $P(X|Z; \theta)$.

$$P(X) = \int P(X|Z; \theta)P(Z)dZ \quad (3.2)$$

$P(X)$ — probability distribution of model parameters X , usually unknown

$P(Z)$ — probability distribution of some latent variables Z , embedded into X

If we know $P(X)$, we assume that we know the “real” process about an object, for example, spatial arrangement of geological properties and structure. Hence, we should maximise

$P(X)$ by optimisation over θ . In VAE formulation, $P(X|Z; \theta)$ is usually Gaussian.

$$\begin{aligned}
 P(X; \theta) &= \int P(X|Z; \theta)P(Z)dZ \\
 P(X|Z; \theta) &= \mathcal{N}(X|f(Z; \theta), \sigma^2 I) \\
 P(X; \theta) &= \int \mathcal{N}(X|f(Z; \theta), \sigma^2 I)P(Z)dZ \xrightarrow{\theta} \max
 \end{aligned} \tag{3.3}$$

Let's define $Q(Z|X)$ - *Encoder*, which maps \mathcal{X} to \mathcal{Z} . So $Q(Z|X)$ is some artificial distribution; $P(Z|X)$ is a real and unknown distribution.

We aim to minimise the difference between $P(Z|X)$ and $Q(Z|X)$. In KL formulation (Kullback et al., 1951), the difference between two distributions could be defined in the following way:

$$KL[Q(Z|X)||P(Z|X)] = \mathbb{E}[\log Q(Z|X) - \log P(Z|X)] \tag{3.4}$$

The equation 3.4 indicates similarity of $Q(Z|X)$ and $P(Z|X)$. If the metric's value is small, the distributions are similar.

However, as previously noted, the distribution $P(Z|X)$ is unknown to us; therefore, we must define it differently. For this, we will need the Bayesian formulation:

$$P(Z|X) = \frac{P(X|Z) \cdot P(Z)}{P(X)} \tag{3.5}$$

$$\log P(Z|X) = \log P(X|Z) + \log P(Z) - \log P(X) \tag{3.6}$$

Let's define the equation 3.4 as follows:

$$KL[Q(Z|X)||P(Z|X)] = \mathbb{E}[\log Q(Z|X) - \log P(X|Z) - \log P(Z)] + \log P(X) \tag{3.7}$$

$$KL[Q(Z|X)||P(Z|X)] = KL[Q(Z|X)||P(Z)] - \mathbb{E}[\log P(X|Z)] + \log P(X) \tag{3.8}$$

$$\log P(X) - KL[Q(Z|X)||P(Z|X)] = \mathbb{E}[\log P(X|Z)] - KL[Q(Z|X)||P(Z)] \tag{3.9}$$

The equation 3.9 is a general formulation of VAE. Let's have a deeper look at it. We want to maximise the left-hand side because it reflects our beliefs about $P(X)$ reconstruction, which tells us about the nature of the object. Plus an error term, $KL[Q(Z|X)||P(Z|X)]$ which is the difference between *Encoder* $Q(Z|X)$ and $P(Z|X)$.

The right-hand side of equation 3.9 is called the (variational) lower bound. This is something we want to optimise via stochastic gradient descent:

$\mathbb{E}[\log P(X|Z)]$ — reconstruction cost, i.e., quality of X reconstruction from Z , which is our *Decoder* P . So we pick z from \mathcal{Z} , and *Decoder* should reconstruct it to x with a minimum information loss.

$KL[Q(Z|X)||P(Z)]$ — regulariser, penalised discrepancy between $P(Z)$ and distribution induced by the *Encoder* Q .

The representation of the input data, X , is captured by the latent distribution. From this distribution, the *Decoder* creates an output by selecting points z through sampling. However, directly sampling these points introduces randomness, which is incompatible with backpropagation, which depends on deterministic operations. Backpropagation requires computing gradients of non-random operations. Hence, a method is needed to introduce the required randomness for sampling while maintaining differentiability. This is where the reparameterisation trick is employed. It functions by isolating the deterministic and stochastic elements of the sampling process. This is achieved by defining $z = \mu + \sigma * \epsilon$, where ϵ provides the required randomness. Consequently, the z is differentiable for backpropagation.

The standard choice for $Q(Z|X) = \mathcal{N}(Z|\mu(X;\theta), \sigma(X;\theta))$, where μ and σ are deterministic functions with parameter θ , that can be learned from data. $P(Z)$ is usually specified as a standard Normal distribution $P(Z) = \mathcal{N}(0, I)$. If the *Encoder* outputs representations Z different from those from a standard normal distribution, it will receive a penalty in the loss. This regulariser keeps the LS \mathcal{Z} sufficiently normally distributed, providing no 'holes' in the LS \mathcal{Z} , so we can sample every z , which were not seen before. Both Q and P are parameterised by deep neural networks.

Variational Autoencoders can simultaneously encode and reconstruct / generate various objects, are stable in training, and are generally not restricted to normal distribution. At the same time, they can generate objects that differ significantly from the training dataset if we sample regions of low probability in the latent space. This aspect will be further examined in later chapters dedicated to studying the LS.

As a result, the *Encoder* and *Decoder* of the VAE can be used for geological modelling in the following way. In the first stage, we train the *Encoder* to encode training samples X into a representation Z in the latent space, which typically has a much lower dimensionality than the original. Assuming that our training dataset has high variability and covers all the uncertainty space, we can assume that our representation of the possible geological conditions is complete, allowing us to reconstruct the geological representation of the desired object considering the available information about it. In the second step, we need to train the *Decoder* to reproduce the reverse process of reconstruction / generation according to samples from the manifold \mathcal{Z} . In reality, both the training processes of the *Encoder* and *Decoder* occur simultaneously, figure 3.1.

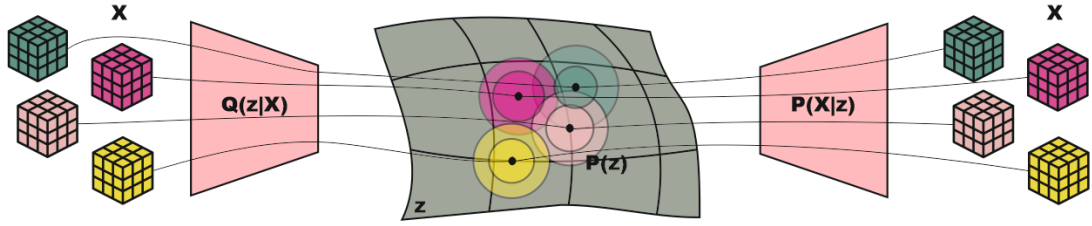


Figure 3.1: The training process of the Variational Autoencoder consists of two main components: Training the *Encoder* $Q(Z|X)$, which converts the original training images X into a representation of the LS Z . And training the *Decoder* $P(X|Z)$, which performs the reverse action of reconstruction from Z to the original representation X .

Putting equation 3.9 into geoscience formulation, we want to maximise our belief $P(X)$, which represents the "true" geological description of an object, by minimising the difference between the *Encoder* $Q(Z|X)$, which can be interpreted as a model of unobservable physical, chemical, biological, etc. processes Z , and the "real" processes in the form of $P(Z)$. In other words, $P(Z)$ is a combination of unobservable processes that resulted in the formation of the geological object $P(X)$, figure 3.2.

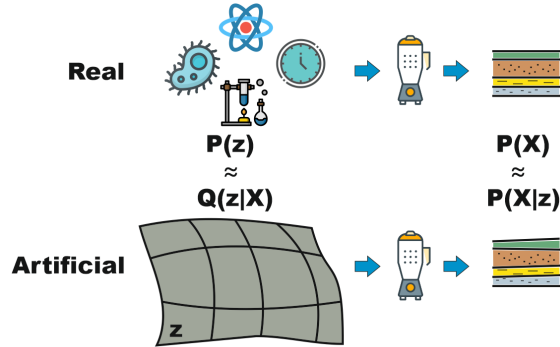


Figure 3.2: In reality, many processes govern the formation of a geological object, such as physical, chemical, biological, etc., which occur over time. The VAE replicates these processes through hidden variables Z to reconstruct the geological object as accurately as possible based on the available data.

To provide the object reconstruction process, we need to determine a *Decoder* for which Z is informative enough to reconstruct X accurately.

3.2 From KL to Wasserstein Loss

During the optimisation process, VAE maximise a variational bound, which is composed of the reconstruction cost $\mathbb{E}[\log P(X|Z)]$ and regulariser $KL[Q(Z|X)||P(Z)]$. VAE forces the $Q(Z|X)$ to match $P(Z)$ for every sample x from \mathcal{X} , which is not guaranteeing that the overall encoded distribution $Q(Z|X)$ matches $P(Z)$. This problem leads us to

the main drawback of VAE, such as the blurriness of the generated objects. If we sample z from the intersection area of different encoded x 's in \mathcal{Z} , then the *Decoder* reconstructs blurry objects. This process is depicted in figure 3.3.

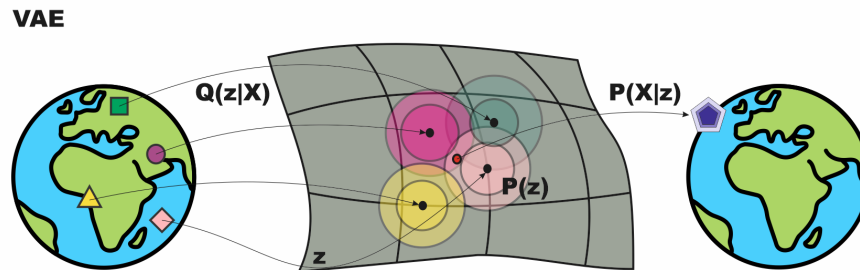


Figure 3.3: VAE forces $Q(Z|X)$ to match $P(Z)$ (coloured regions in \mathcal{Z}) for every single input (●, ▲, ■, ◆), drawn from $P(X)$. Coloured regions in \mathcal{Z} starts to intersect, which leads to unreliable reconstructions (◆) due to blurriness problem if the sample (●) was picked from the region of intersections.

In data analysis, probability measures are commonly compared using divergences such as Euclidean, total variation, Hellinger, and Kullback-Leibler. These measures are used to quantify fitting errors or losses in parameter estimation problems. Unlike other metrics, Optimal Transport is based on a physical approach of mass displacement and geometry to compare measures. Thus, using the Wasserstein metric, which is based on Optimal Transport theory, as a loss function becomes very appealing (Peyré et al., 2018).

In this section, we will transform the regulariser to match the overall training distribution and the prior distribution $P(Z)$. To achieve this, we need to incorporate Optimal Transport (OT) theory into the generative approach. Optimal Transport provides a cost metric that measures the similarity between probability distributions (Villani et al., 2003).

3.2.1 Optimal Transport basics

Let's first put some intuition into the Optimal Transport metric. Assume we have a distribution of geological properties, denoted as $P(X \in \mathcal{X})$. We want to move the mass of this distribution to another "location" in the same space \mathcal{X} and obtain $P(Y \in \mathcal{X})$, i.e., generate a geological object with similar properties.

OT is a geometry for measuring probability, initially developed by French mathematician Monge in 1781. OT can be conceptually formulated as follows: what is the minimum cost of moving a pile of sand to a different space with a predetermined shape? Later, Kantorovich and Koopmans received the Nobel Prize for their work on the OT problem, highlighting the significance of the theory for logistics and economics. Currently, OT has evolved into a tool in mathematical analysis and machine learning as a metric for comparing different distributions.

To evaluate the cost of this transportation, we need to estimate $c(X, Y) \rightarrow [0, \infty)$. c gives us an estimate of the cost of transporting $P(X)$ to $P(Y)$. The transportation plan can be described as $\gamma(X, Y)$, and the cost c of this plan is minimised.

To provide that transportation, we need the following properties:

$$\int \gamma(X, Y) dY = P(X) \quad (3.10)$$

$$\int \gamma(X, Y) dX = P(Y) \quad (3.11)$$

These equations tell us that the mass being transported from X should be equal to the total mass at X , and the mass at Y should be transported to its final location. Therefore, the total mass from X being transported to Y is equal to $\gamma(X, Y) dX dY$, and the cost of such transportation is $c(X, Y) \gamma(X, Y) dX dY$. The overall cost of transportation is determined as follows:

$$\iint c(X, Y) \gamma(X, Y) dX dY = \int c(X, Y) d\gamma(X, Y) \quad (3.12)$$

The transportation plan $\gamma \in \Gamma$ is not unique, so we need one, across all in Γ , with the minimum cost

$$C = \inf_{\gamma \in \Gamma} \int c(X, Y) d\gamma(X, Y) \quad (3.13)$$

Now, we can return to the OT and Wasserstein distance when we understand the basics of VAE formulation.

3.2.2 From VAE to WAE

The family of regularised Auto-Encoders based on OT is called Wasserstein Auto-Encoder (WAE) and was introduced by Tolstikhin et al. (2017). Similarly to VAE, the objective of WAE is composed of two terms: the reconstruction cost and the regulariser. It was shown that the minimum reconstruction cost can be defined in the following way:

$$C = \inf \mathbb{E}_{P_X} \mathbb{E}_{Q(Z|X)} [c(X, G(Z))] \quad (3.14)$$

where $G(Z) = P(X|Z)$ is the *Decoder*, which maps from Z to X and $Z \sim Q(Z|X)$. We need to optimise over the *Encoder* $Q(Z|X)$ with the penalty to the objective as a regulariser.

Now let's define the WAE Objective Function in comparison to VAE. Let's recall that both VAE and WAE aim to minimise the difference between the true but unknown distribution $P(X)$ and the generated distribution by the *Encoder* $P(X|Z)$. Let's redefine equation 3.9. We minimise the right-hand side of the equation:

$$D_{VAE}(P(x), P(X|Z)) = \inf \mathbb{E}_{P_X} [KL(Q(Z|X)||P(Z)) - \mathbb{E}[\log P(Z|X)]] \quad (3.15)$$

$$D_{WAE}(P(X), P(X|Z)) = \inf \lambda D(Q(Z|X), P(Z)) + \mathbb{E}_{P_X} \mathbb{E}_{Q(Z|X)} [c(X, G(Z))] \quad (3.16)$$

Where λ is a hyperparameter to our regulariser. We expect WAE to provide a better reconstruction without blurry representation as opposed to VAE, figure 3.4. The blurriness often observed in VAE-generated samples results from the trade-off between achieving high reconstruction accuracy and the KL divergence term's regularisation. In contrast, WAE utilises the Wasserstein distance as their primary training objective, aiming to measure the discrepancy between probability distributions and capture the underlying data distribution more effectively, providing sharper sample generation. The only thing we don't know yet in the equation 3.16 is $D(Q(Z|X), P(Z))$.

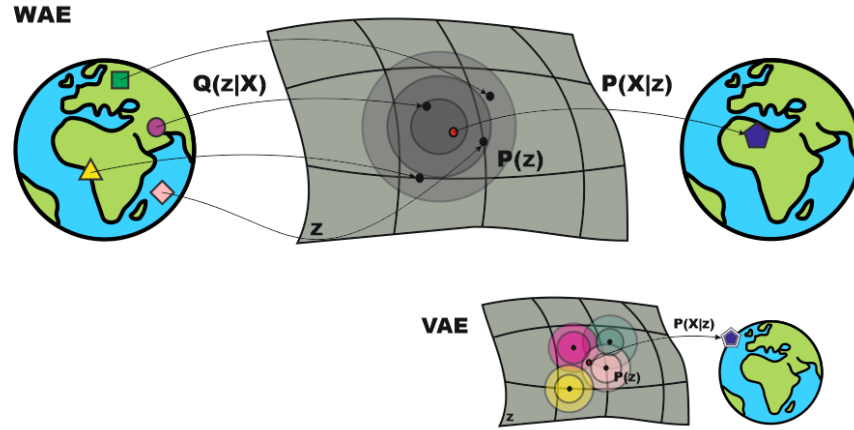


Figure 3.4: WAE forces continuous mixture $Q(z) = \int Q(z|X)dP(z)$ (●) to match $P(z)$ (grey region), which provides better reconstruction (◆), as opposed to VAE, which forces to match $P(z)$ for every single input

3.2.3 MMD regulariser

The regulariser $D(Q(Z|X), P(Z))$ from equation 3.16 is a penalty component in the optimisation process of the right-hand side of the equation. In contrast to the classical formulation of VAE, Tolstikhin et al. (2017) changed KL divergence to Maximum-Mean Discrepancy (MMD) (Gretton et al., 2012) metric, which is a distance on the space of probability measures and has been widely applied in various ML techniques. In the context of Wasserstein Autoencoders, the regulariser penalises the Encoder by imposing constraints on the Wasserstein distance between the encoded distribution and a predefined distribution (e.g., a standard normal distribution). This encourages the model to generate encoded representations that follow a specific distribution, which can have benefits such as better

generalization and improved disentanglement of features.

In general, MMD conveys the idea of estimating distances between probability distributions as the distance between the mean embeddings of features. Suppose we have distributions P and Q over a set \mathcal{Z} . The MMD is defined by a feature map $k : \mathcal{Z} \rightarrow \mathcal{H}$, where \mathcal{H} is a reproducing kernel Hilbert space. It is known that MMD performs well when matching high-dimensional standard normal distributions. If the value of MMD is large, this means that $Q(Z|X)$ and $P(Z)$ are different distributions.

$$MMD_k(Q(Z|X), P(Z)) = \left\| \int_{\mathcal{Z}} k(Z, \cdot) dQ(Z|X) - \int_{\mathcal{Z}} k(Z, \cdot) dP(Z) \right\|_{\mathcal{H}} \quad (3.17)$$

$k(Z, \cdot)$ indicates the kernel has one argument fixed at Z , and the second free. The kernel k is inverse multiquadratic $k(x, y) = C/(C + \|x - y\|_2^2)$ and, $C = 2d_z\sigma_z^2$ which is the expected squared distance between two multivariate Gaussian vectors from $P(Z)$. The quality of the generated objects is expected to depend on the *Encoder* $Q(Z|X)$.

While both KL divergence and MMD are used to measure distributional differences, they have different mathematical formulations, properties, and use cases. KL divergence is asymmetric, has an information-theoretic interpretation, and is sensitive to outliers. In contrast, MMD is symmetric, less sensitive to outliers, and focuses on differences in means in a reproducing kernel Hilbert space.

Wasserstein distance is not new for geoscience and was utilised by Chan et al. (2019b), Corrales et al. (2022), and T. Zhang et al. (2022).

3.3 Importance of Geometric Deep Learning

It has been previously noted that the *Encoder* and *Decoder* are parameterised by deep neural networks, which by default have a “classical” Convolutional Neural Networks (CNN) in architecture. It is worth noting that CNNs were initially developed for regularly-structured (Euclidean) data, such as images, texts, etc., which should be perceived as a limitation.

Works by Feng et al. (2022), Jiang et al. (2021), and Toma et al. (2022) use the architecture based on CNNs and LS as a source for generating geological representations. However, these approaches have limitations and cannot describe the complex structural features of reservoirs or irregularities in flow behaviour. To successfully solve the task of generating three-dimensional geological objects considering structural uncertainty, generative machine learning models need to be modified to work with irregular (non-Euclidean) data. Geometric deep learning is a general term encompassing machine learning methods adapted for domains with structurally non-regular data, such as graphs; these approaches allow convolutional operations on connected topological graph data.

Geological structures often exhibit intricate spatial relationships, connectivity, and dependencies that graph representations can effectively capture, unlike Euclidean data types. Graph-based convolutions are particularly well-suited for geological modelling due to their inherent ability to naturally preserve the topology of geological objects. In a framework based on graph convolution, geological features are represented as nodes, and their relationships are depicted as edges, forming a graph structure. This approach aligns seamlessly with the complex and interconnected nature of geological formations. Graph-based convolutions operate directly on the relationships encoded in the graph, allowing for a more natural representation of geological topology. This is in contrast to grid-based convolutions, which may struggle to handle irregularly shaped geological features and their connectivity. This methodology ensures that convolutional operations adhere to the structure of geological formations, allowing for more precise and realistic modelling of subsurface processes, fault networks, and other complex geological phenomena.

Typically, geological models are represented as grids with a predefined regular lattice. However, if such a structure is transformed into a graph representation, it becomes possible to use more advanced graph convolutions, which enable the processing of irregular geological structures. Each node of such a graph can contain a vector of reservoir properties (such as porosity, permeability, etc.). In contrast, the edges connecting the graph nodes can be represented according to the structural features of the geological object (such as faults, unconformities, etc.).

From a purely technological point of view, if the considered three-dimensional object has a complex structure, additional computational memory needs to be allocated to process it with convolutional networks, as such representations need to be supplemented with “empty cells” necessary only to make the representation Euclidean, but such representation does not play any role in the geological model itself (Bronstein et al., 2021), figure 3.5. To feed the geological model into CNN-based GAN or VAE, it needs to have some regular rectangular shape because the CNN’s filters provide sliding window operations along the grid to perform some aggregations during the training process. So, if the gridded geological model would have a high

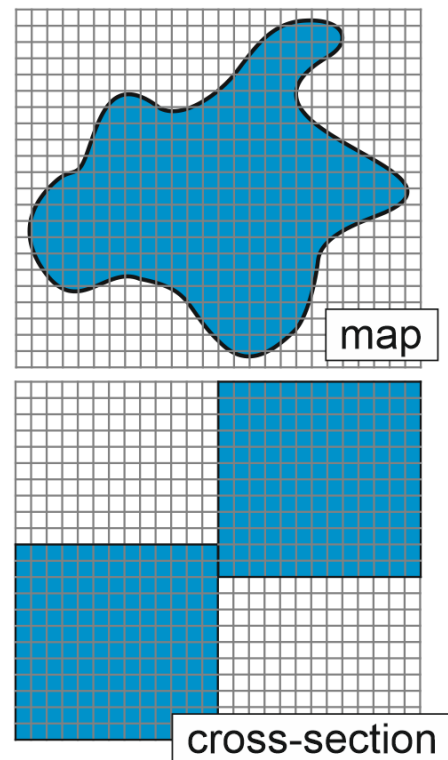


Figure 3.5: Schematic field map. To process a complex geological object with CNN, it is necessary to add empty cells (□) to already existing (■), only in this case the representation will become Euclidean.

proportion of zero-valued cells, it entails the GAN or VAE to train on these cells, which is undesirable.

This section is the transition from the Euclidean CNNs to non-Euclidean.

3.3.1 CNN for Euclidean data types

Before describing and further using generative machine learning models based on graph convolutions, it is necessary first to explain the principles of standard Convolutional Neural Networks. Bronstein et al. (2021) highlight the main principle of deep learning, which consists of two simple algorithmic approaches. The first involves representation learning or feature learning, wherein customized features, often arranged hierarchically, capture the relevant regularities for each task. The second approach is learning through local gradient descent, commonly executed as backpropagation.

CNN is a deep learning algorithm that processes data with a grid-like topology or lattice (Euclidean data), such as images, time series, and texts. CNN assigns importance to different data features through learnable weights and biases (I. Goodfellow et al., 2016; Hastie et al., 2009). Using various filters, CNN successfully captures spatial and temporal dependencies in the data. The efficiency of CNN lies in reducing the dimensionality of the original data to a size more suitable for further processing. At the same time, the quality of predictions is ensured by selecting an optimal number of reduced-dimensional features that describe the original data. In general, any CNN consists of three parts: a convolutional layer, a pooling layer, and an activation function. Let's briefly describe each part.

Convolutional layer

Suppose we have a regularly discretised map of some property (i.e., porosity) $x(n)$, where n is the dimensionality of the map, i.e., the number of cells. Both x and n are real-valued, and we can get a different reading anywhere. If our data is noisy, we want to find some general regularities from our data by applying some smoothing or averaging by the function w and get a representation $s(m)$.

$$s(m) = (x(n) * w) = \sum_n (x * w) \quad (3.18)$$

NOTE: In the terminology of machine learning: the $*$ operator is usually called convolution; $x(n)$ — input; w — kernel or filter; $s(m)$ output or feature map ($m \neq n$ and, usually $m \ll n$).

During the forward pass, the kernel w slides along $x(n)$ and produces a representation $s(m)$ of the kernel at each spatial position, figure 3.6.

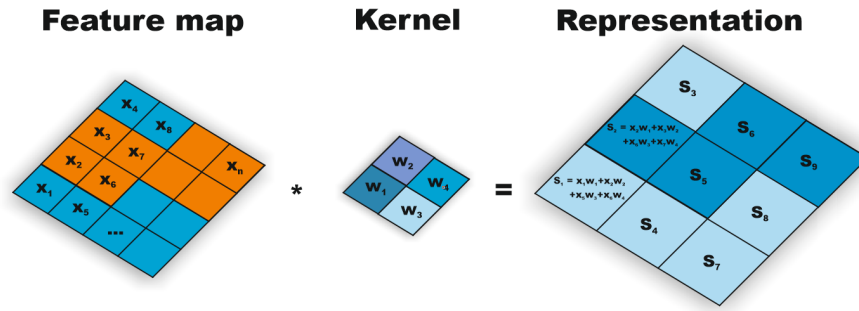


Figure 3.6: The Kernel of size 2×2 is applied to the Feature map to create Representation

From a geoscience perspective, our input data can be represented as a grid of porosity, permeability, lithology, etc., which can already be called a high-dimensional tensor. However, CNNs typically have sparse interactions. This is achieved by setting the kernel size smaller than the input data size. This allows the neural network to capture complex dependencies among multiple features effectively.

Another useful feature of CNNs is equivariance to translation. This means that if the input data changes, the output result will change similarly. For example, if a geological body somehow changes its spatial location due to uncertainty quantification, on the property map $x(n)$, the representation $s(m)$ will reflect this change.

However, CNNs are not naturally equivariant to certain transformations, such as changes in scale or rotations, so more general approaches are needed.

The Pooling Layer

The pooling function replaces the output data with summary statistics, such as maximum or average, figure 3.7. This allows the reduction of the spatial size of the representation, which in turn reduces the required computational power and number of weight coefficients needed for training the neural network model. Moreover, the pooling operation helps make a representation approximately invariant to small perturbations in the input data. If we slightly change the input information, the pooling values will not change.

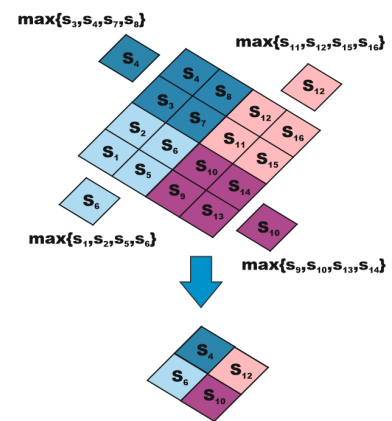


Figure 3.7: Pooling operation with 2×2 filter size that picks the maximum value

Activation functions

The main idea of neural networks is to extract linear combinations of input features and model the target parameter as a nonlinear function of these input character-

istics. Since convolution is a linear transformation and maps can be highly nonlinear, non-linear layers in the neural network (typically denoted by σ) are usually placed immediately after the convolutional layers to provide nonlinearity in the activation map.

$$s(m) = \sigma(x(n) * w) = \sigma \sum_n (x * w) \quad (3.19)$$

Notice that if σ is the identity function, the entire model collapses to a linear model.

General features of CNN and Euclidean Data

Generative models (GAN, VAE, Diffusion) are based initially on “classical” Convolutional Neural Networks, which, as was mentioned earlier, are designed for structured (Euclidean) data such as images, texts, grids, etc., which is a limitation. One of the main assumptions made for this type of data is compositionality, i.e., the ability to break down the problem into different scales to facilitate the step-by-step learning process, figure 3.8.

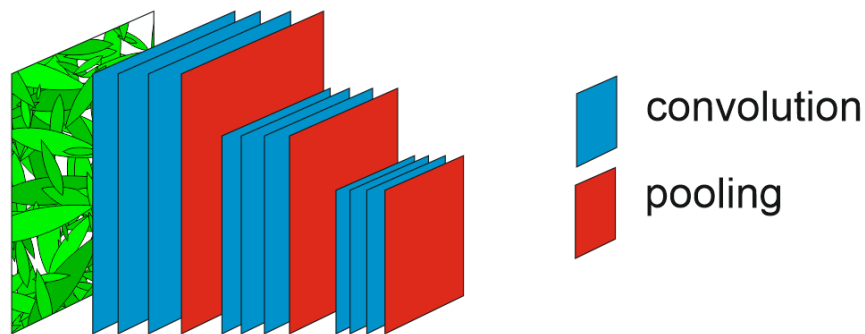


Figure 3.8: Schematic representation of the compositionality property of the data when the task is broken down into a sequence of steps of different scales.

This term means that Euclidean data form patterns, which are (figure 3.9):

- Local – Activation performed if a local feature is present
- Stationary – The quantity to be predicted does not change much if the input is slightly deformed
- Multi-scale (hierarchical) – Combination of simple structures composes more abstract

One of the key reasons for the success of deep neural networks in utilising the statistical properties of data lies in the principles of stationarity and compositionality through local statistics, which are present in videos, images, and sound.

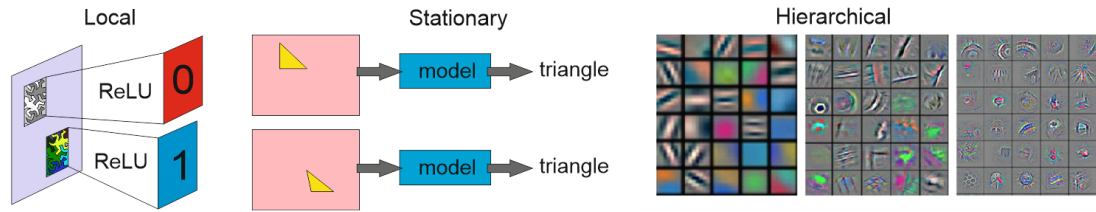


Figure 3.9: Schematic representation of the three main properties of Euclidean data on which the work of CNN is based. From Zeiler et al. (2013).

3.3.2 Convolutions on non-Euclidean data

To effectively perform convolutional operations on non-Euclidean data, it is necessary to formulate requirements similar to standard Euclidean convolutional operations. Therefore, it is necessary to develop the compositionality property for unstructured data, such as graphs. Similarly to standard Convolutional Neural Networks, the key to efficient learning on graphs lies in developing local operations with shared weights that do message passing between nodes and their neighbours. Message passing can be defined as computations performed at each layer l of a graph neural network. During message passing, each graph node’s property is updated to aggregate the properties of neighbouring nodes and update the learning parameters w^l for each layer. The term “message passing” reflects the essence of the signal “travelling” through the network, passing through graph nodes along the path of the neighbourhood between nodes (Papillon et al., 2023).

Compared to classical neural networks that process Euclidean data, the main difference is that non-Euclidean data operations must be permutation-invariant, i.e., independent of the order of passing through neighbour nodes. Additionally, the model should not depend on the size and structure of the graph (Gilmer et al., 2017). All this makes graph-based convolutions a good technique to obtain a reduced-order representation of complex geological objects, keeping the topology. In general, graph convolutions can be divided into two main types: Spectral and Spatial.

Spectral Convolutions

The first method was popularised mainly in graph theory and signal analysis (Defferrard et al., 2017). Graph signal processing allows convolutional operations that consider the graph structure and individual properties of graph components. In general, our task is to decompose a signal, such as a graph, into a combination of simple elements (wavelets) that are usually orthogonal to each other, i.e., linearly independent, and therefore form a basis.

When we talk about the operation of spectral decomposition applied to graph convolutions, we mean the decomposition of the graph Laplacian L . Such a Laplacian can be represented

as a normalised adjacency matrix A . In this case, spectral decomposition allows us to find elementary orthogonal elements (eigenvectors V and eigenvalues Λ) that make up the original graph.

$$L = V\Lambda V^T \tag{3.20}$$

The Laplacian can be interpreted as the difference between the average value of a function in an infinitesimal sphere around a certain point and the value of the same function at that point. From a physical point of view, the Laplacian shows the direction of signal propagation when applying energy to one of the nodes. It describes processes such as heat conduction, quantum mechanics, and wave propagation, figure 3.10.

One limitation of using the spectral approach for graph convolutions is that the graph must be undirected to ensure the symmetry of the adjacency matrix A , which allows us to find the Laplacian.

To perform convolutional operations on a graph, it is sufficient to find several eigenvectors corresponding to the smallest eigenvalues of the matrix. This approach may initially seem counter-intuitive, as the standard PCA approach involves finding eigenvectors corresponding to the largest eigenvalues of the matrix. However, the difference lies only in the fact that the calculation of the Laplacian is performed with the opposite sign, thus making the eigenvalues of PCA inversely proportional to the eigenvalues of the Laplacian.

Thus, the Laplacian of node features X , which allows us to formulate a simple version of spectral convolutions on graphs.

$$X^{(l+1)} = V(V^T X^{(l)} \odot V^T W_{spectral}^{(l)}) \tag{3.21}$$

NOTE: \odot denotes element-wise multiplication, l is the convolutional layer number, and W is the convolutional kernel.

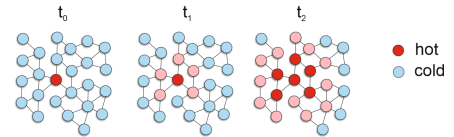


Figure 3.10: From a physical point of view, the Laplacian displays the speed of propagation of an impulse, for example, heat applied to a certain point.

However, this approach imposes several constraints. For example, the dimension of the convolutional kernel W depends on the number of nodes in the graph. Additionally, W depends on the graph structure encoded in the eigenvectors V , and there is no guarantee of spatial localisation of the kernel. Therefore, such an approach is not applicable to large graphs with a non-fixed structure.

Further work proposes ways to overcome these limitations. For example, in the work by Bruna et al. (2014), a method for reducing the size and localising the kernel was proposed. The idea is to represent the filter W in the equation 3.21 as a sum of certain functions,

such as splines, which allows for K coefficients α instead of N values in the W matrix:

$$W_{spectral}^{(l)} \approx \sum_{k=1}^K \alpha_k f_k \quad (3.22)$$

However, the filter W still depends on the eigenvectors V , making this approach inapplicable for datasets with a non-fixed graph structure. Although spectral convolutional operations are used much less frequently than spatial ones, understanding the principles of their operation allows us to understand potential problems that arise when processing non-Euclidean data types.

Spatial convolutions

The idea of spatial graph convolutions is as follows: the neighbourhood of graph nodes determines the computations on the graph, i.e., if we want to calculate an embedding for a target node, the calculations are performed considering the neighbouring connected nodes. Thus, if each graph node is defined by a feature vector h , spatial graph convolutions transform and aggregate features considering the neighbourhood. In other words, the network learns to propagate and aggregate information in the graph.

One example of spatial convolutional networks is presented in the work by Hamilton et al. (2018). Figure 3.11 shows the basic scheme. To calculate the properties of the target node i for a layer $l + 1$, all neighbouring nodes should be considered (a , b , c). The information from the neighbouring nodes is passed to a transformation operation (Q), then aggregated ($\sum(\cdot)$) and transformed (W) concerning the target node i . This sequence of actions is performed for each node in the graph. Thus, matrices Q and W are the training parameters of the neural network.

$$h_i^{(l+1)} = f^l(h_i^{(l)}, h_n^{(l)}) = \sigma(W^{(l)} h_i^{(l)} + Q^{(l)} h_n^{(l)}) \quad (3.23)$$

$$h_n^{(l)} = \frac{1}{(n-1)} \sum_{n \neq i} h_i^{(l)} \quad (3.24)$$

Therefore, such convolutional architecture allows the utilisation of various aggregators, such as mean, LSTM, Max-pooling; loss functions: cross-entropy, hinge loss, ranking loss, etc. The neural network has a fixed set of training parameters and can be applied to any node in the graph.

To effectively aggregate information from complex geological objects with different spatial characteristics, it is necessary to consider the properties of the nearest neighbours and higher-dimensional graph structures. This type of graph convolutional architecture was proposed by Morris et al. (2021) and formed the basis for graph VAE developed within

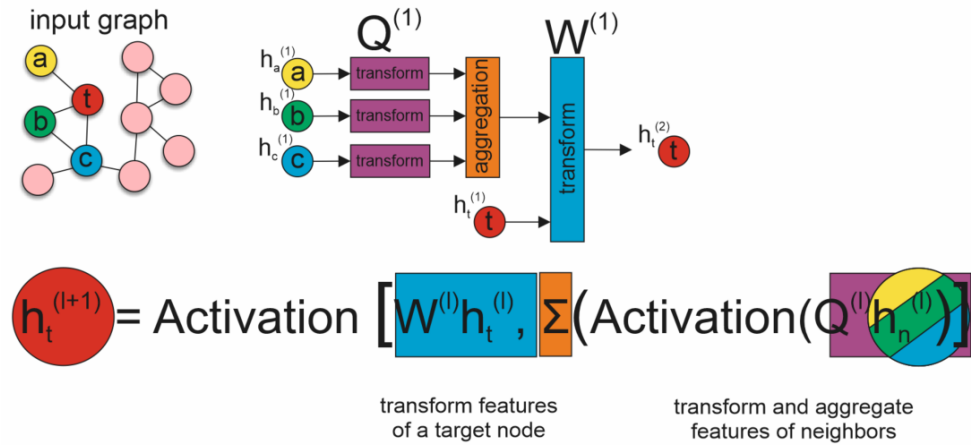


Figure 3.11: Schematic representation of the operation of a convolutional graph neural network for a node t , where h — features of the node on the layer l .

the framework of the presented research.

3.3.3 Graph Wasserstein Variational Autoencoder (GWAE)

The key feature of the proposed approach of performing convolutional operations on graph data types is that message passing occurs on sub-graphs of different dimensions, not just nodes. Thus, it becomes possible to incorporate a hierarchical approach to graph processing, which allows taking into account and combining geological features at the scale of nodes and larger-scale objects in the form of sub-graphs.

NOTE: For a more formal description of this type of convolution operations, let's introduce the following notations: Graph G is a pair (V, E) , where V — finite set of nodes, $E \subseteq \{\{u, v\} \subseteq V | u \neq v\}$ - set of edges. Define (u, v) as a specific edge of G . $N(v)$ — set of neighbours of $v \in V$, i.e. $N(v) = \{u \in V | (v, u) \in E\}$.

The concept of the hierarchy could be linked to the earlier idea of MPS - connected data and higher-order statistics. In geostatistics, particularly in techniques like Training Images/MPS, the approach is lattice-based, where spatial patterns are characterised by transitioning from one state to another in a predefined grid or lattice. Contrastingly, the graph-based approach delves into the topology of data, capturing the relationships and connections between nodes and sub-graphs. This graph-centric methodology enables the elicitation of connected node information at various levels, be it first-order connections, second-order connections, and so forth. Unlike MPS, which primarily involves the collation of statistics and the inference of distributions based on observed patterns, the graph-based model incorporates a learning mechanism. This learning aspect allows the system to adapt and refine its understanding of complex relationships within the data, providing a more dynamic and adaptable framework for capturing and leveraging higher-order

statistical patterns in interconnected datasets.

To introduce the concept of graph-based hierarchy, it is necessary to modify the equation 3.23 to include the neighbourhood of nodes and sub-graphs of different levels. Introduce k -element subset $[V^k] \subset V$, where $s = \{s_1, \dots, s_k\}$ - k -set in $[V^k]$. So the neighbourhood of s is: $N(s) = \{t \in [V^k] \mid |s \cap t| = k - 1\}$. Local neighbourhood $N_L(s)$ consists of $t \in N(s)$, which has edges $(u, v) \in E$ for every $u \in s \setminus t$ and $v \in t \setminus s$. From the other side, global neighbourhood $N_G(s)$ is defined as $N(s) \setminus N_L(s)$. Schematically k -element subset, the local and global neighbourhood is depicted in figure 3.12

Equation 3.23 can be defined as follows:

$$h_t^{(l+1)}(s) = \sigma(W^{(l)}h_t^{(l)}(s) + \sum_{u \in N_L(s)} Q_L^{(l)}h_n^{(l)}(u) + \sum_{v \in N_G(s)} Q_G^{(l)}h_n^{(l)}(v)) \quad (3.25)$$

Using matrices W, Q_L, Q_G makes it possible to train a neural network taking into account local and global neighbourhoods. Additionally, using subsets of nodes (s) as a single element allows for hierarchy in training, figure 3.13.

Thus, the discussed modification of graph CNN allows for the creation of a set of training layers that aggregate information not only for neighbouring nodes but also for subgraphs of different dimensions. Such a hierarchy in graph convolutions is now known as cell complexes and can be considered a multidimensional generalisation of graphs. Unlike graphs, which consist of nodes and edges, cell complexes can also contain structures of higher dimensionality (Battiston et al., 2021; Papillon et al., 2023). This is important when considering geological objects as research objects, as they possess local and global structures.

Combining the above, I present a general version of the Graph Wasserstein Variational Autoencoder (GWAE), which will be used as a generator for geological realisations. In its general form, it consists of two neural networks: an *Encoder* ($Q(Z|X)$) and a *Decoder* ($P(X|Z)$), each of which is a sequence of graph convolutional networks (\bullet). The upper part of figure 3.14 schematically represents the GWAE. The training dataset, a set of unstructured graphs, is passed to the *Encoder*, which performs convolutional operations. The result is the parameters of the distribution $Q(Z|X)$ (\blacksquare), which are the parameters of the latent space. It is worth noting that the dimensionality of the LS is manually specified

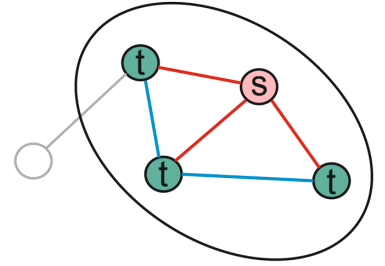


Figure 3.12: The concept of neighborhood $N(s)$ (\circ) of subset of elements s of the graph. Local neighbourhood $N_L(s)$ (\triangle) is determined by the neighbourhood of the nodes of the set $N(s)$ with the elements of a subset s . Global neighbourhood $N_G(s)$ (\triangle) is determined as the difference of sets $N(s)$ and $N_L(s)$

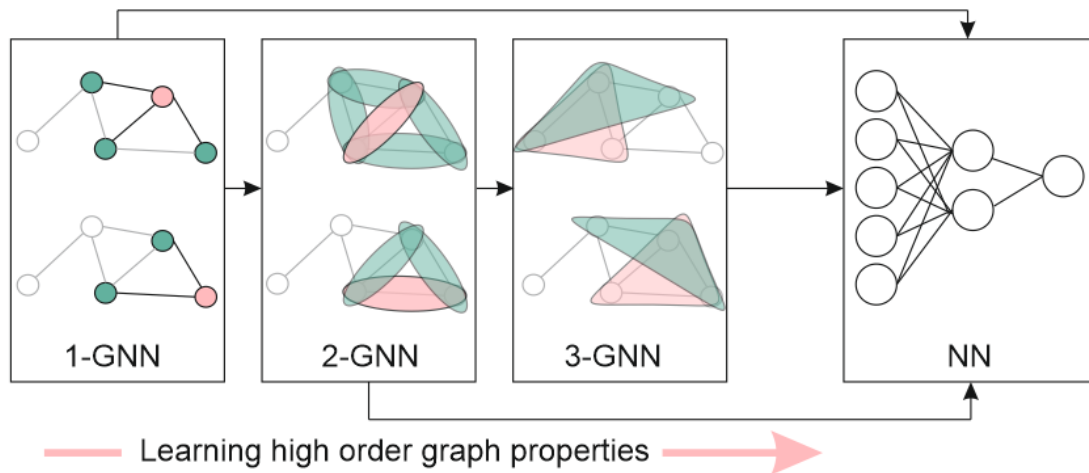


Figure 3.13: Visualisation of the process of hierarchical convolution on a graph. 1-GNN provides message passing on the level of nodes. 2-GNN is the process performed on the sub-graph level, composed of pairs of nodes. 3-GNN – level of sub-graphs with triplets of nodes.

and empirically adjusted until the reconstruction quality of the *Decoder* starts to degrade.

In the next step, a hidden vector z is sampled from the distribution $Q(Z|X)$ (▣). If the GWAE is trained correctly, the vector z contains enough information for the *Decoder* to reconstruct x correctly.

The bottom of figure 3.14 shows the loss function required for training GWAE. As mentioned earlier, the loss function consists of two terms (▣): the reconstruction error, which ensures that the generated model is similar to the original and a regulariser that enforces a Gaussian distribution for $Q(Z|X)$. In the case of GWAE, the reconstruction error is controlled by equation 3.14, and the regulariser is controlled by the Maximum-Mean Discrepancy, equation 3.17.

Thus, this architecture allows minimising one of the main drawbacks of traditional VAEs — blurriness of a generated object and also enables handling objects with complex and irregular topology through graph convolutional networks that consider the spatial complexity of the objects under consideration.

It should be noted that specific parameters of the convolutional layers and the hidden layer are omitted, as they depend on the complexity of the dataset.

3.4 Optimisation theory

When we have trained GWAE on the training dataset, all we need is the hidden space \mathcal{Z} and the *Decoder*, which allows to perform the mapping operation $\mathcal{Z} \rightarrow \mathcal{X}$. However, to solve specific practical problems such as history matching, UQ, etc., conditioned on

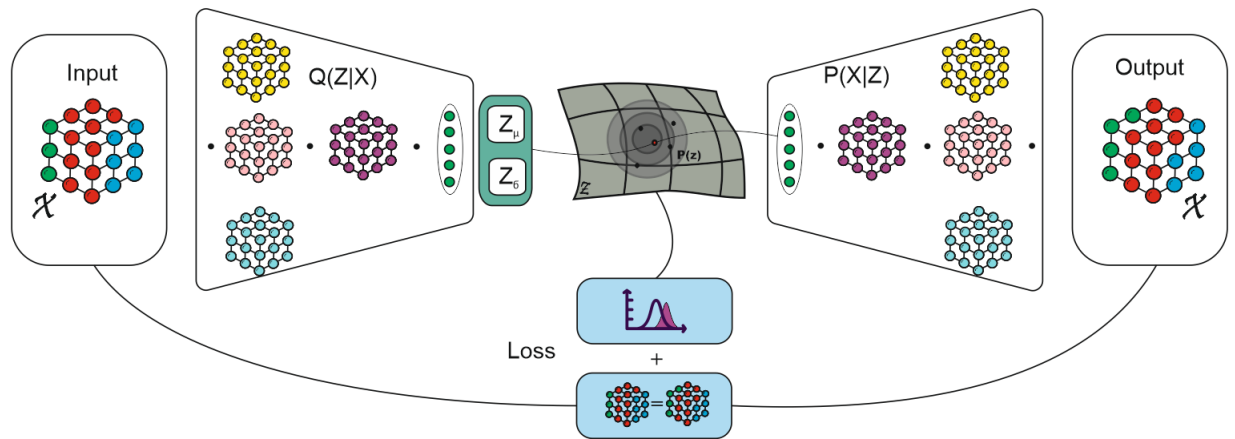


Figure 3.14: Schematic representation of GWAE: a sample $x \in \mathcal{X}$ goes into *Encoder* and converted by a sequence of graph convolutions (\bullet) into a latent vector $z \in \mathcal{Z}$. Then, it could be sampled from \mathcal{Z} and converted back into the initial representation. The quality of GWAE training is controlled by reconstruction loss and regulariser (\square).

data available, it is necessary for the generated models to meet predefined requirements, namely, not contradict static well data in the form of well logs and also satisfy the conditions of matching dynamic data (oil production, water production, water injection, etc.). Therefore, to effectively search such model representations through the hidden space, it is necessary to define the Objective Function (OF) and optimisation algorithm.

The success of deep learning is mainly due to the ability to use the backpropagation algorithm to efficiently compute the gradient of the OF concerning each model parameter. With the help of these gradients, we can effectively search the parameter space to solve the optimisation problem and find neural network parameters.

OpenAI has published an article titled “Evolution Strategies as a Scalable Alternative to Reinforcement Learning” (Salimans et al., 2017), in which they demonstrate that Evolution Strategies (ES) offers several advantages. One of the advantages is that the algorithm does not require gradient calculations for its operation, making it more computationally efficient. Additionally, the Evolution Strategies algorithm’s workload can be distributed across multiple parallel processes. Regarding the variety of solutions obtained, ES also holds a certain advantage because repeated algorithm runs may lead to different optima, unlike standard gradient-based methods.

The article notes that we cannot directly compute gradients even for tasks such as designing a neural network architecture. Meanwhile, evolution strategy or genetic algorithms can be applied to search in the model architecture space. So, for the problem of searching in the multidimensional latent space, it could be very expensive to compute gradients or even impossible because we can get stuck in the regions of low density.

Evolution strategies can be presented with various definitions, but we can characterize

them as algorithms that provide users with a set of candidate solutions for assessing a given problem. The evaluation relies on an OF, which, given a solution, produces a singular fitness value. Utilising the fitness outcomes from the existing solutions, the algorithm generates the subsequent generation of candidates expected to yield superior results. The iterative cycle continues until the optimisation process identifies the best-known solution.

While the population size usually remains constant for each generation, it's not a mandatory condition. ES can produce numerous candidate solutions, as it generates solutions from a distribution with continuously updated parameters in each generation.

3.4.1 Covariance-Matrix Adaptation Evolution Strategy (CMA-ES)

One limitation shared by both ES strategies and basic genetic algorithms is the constant standard deviation for parameters, limiting the search space exploration. There are processes when we desire greater exploration, prompting an increase in the standard deviation within our search space. Conversely, there are times when we are certain about proximity to an optimal solution and aim to refine the existing solution. So, we need an adaptive approach with varying search space exploration.

CMA-ES (Hansen, 2023) is an optimisation algorithm that can take the results of each generation and adaptively adjust the search space for the subsequent generation. It adapts the mean value parameters μ with standard deviation σ and computes the entire covariance matrix of the parameter space. In every generational step, CMA-ES computes parameters for sampling solutions from a multivariate normal distribution (Kriest et al., 2020; Miyagi et al., 2018; Sayyafzadeh et al., 2018).

Let's consider the process of covariance matrix estimation because it is important for understanding the CMA-ES methodology. We should use the equations below to estimate the 2×2 covariance C matrix of a whole 2-dimensional generation $g(x, y)$ of size n .

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)^2 \quad (3.26)$$

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \mu_y)^2 \quad (3.27)$$

$$\sigma_{xy}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y) \quad (3.28)$$

CMA-ES modifies the covariance matrix calculation in a smart way to adapt it to the optimisation task. Firstly, the algorithm picks the top n_{best} solutions by fitness for each generation g . After sorting the solutions, the mean of the next generation μ^{g+1} is estimated

as n_{best} solutions in the current population g .

$$\mu_x^{g+1} = \frac{1}{n_{best}} \sum_{i=1}^{n_{best}} x_i \quad (3.29)$$

$$\mu_y^{g+1} = \frac{1}{n_{best}} \sum_{i=1}^{n_{best}} y_i \quad (3.30)$$

Then, only n_{best} solutions are used to estimate the covariance matrix C^{g+1} of the next generation, but the smart thing than μ^g is used instead of the updated parameters μ^{g+1} .

$$\sigma_x^{2,g+1} = \frac{1}{n_{best}} \sum_{i=1}^{n_{best}} (x_i - \mu_x^g)^2 \quad (3.31)$$

$$\sigma_y^{2,g+1} = \frac{1}{n_{best}} \sum_{i=1}^{n_{best}} (y_i - \mu_y^g)^2 \quad (3.32)$$

$$\sigma_{xy}^{2,g+1} = \frac{1}{n_{best}} \sum_{i=1}^{n_{best}} (x_i - \mu_x^g)(y_i - \mu_y^g) \quad (3.33)$$

Below is a set of steps on how CMA-ES uses the results from the current generation:

Algorithm 1: CMA-ES optimisation

for $g \leftarrow 1, N$ **do**

- Compute the fitness estimate for each candidate solution in generation g ;
- Select the top 50% of the population in generation g ;
- Using only the top 50% solutions together with the μ^g , compute C^{g+1} ;
- Generate a new set of candidate solutions using the updated μ^{g+1} and C^{g+1} ;

end

Adapting both the μ value and C , based on information from the best solutions, CMA-ES can determine whether to broaden the search when the top solutions are distant or to decrease the search space when the top solutions are proximate, figure 3.15.

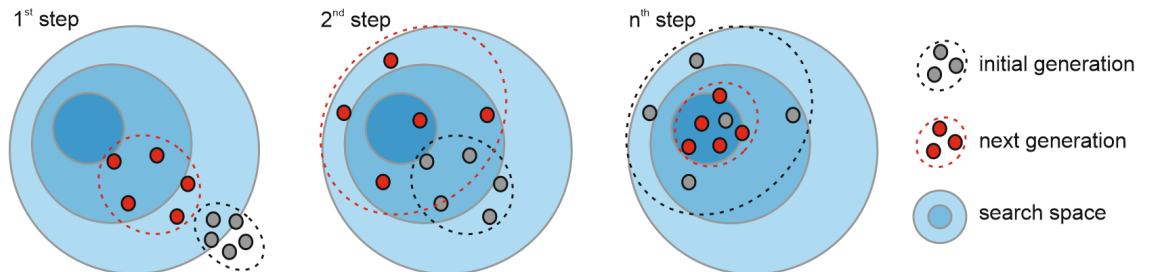


Figure 3.15: Visualisation of an CMA-ES run on a simple 2D problem. Where \bullet - is a search space, \bullet - the generation before estimation of the fitness and \bullet - the generation after estimation of the fitness.

The main limitation lies in its performance when dealing with many model parameters, as

computing the covariance has $O(N^2)$ complexity. However, approximations have recently been proposed to speed up the calculations to $O(N)$. CMA-ES is very efficient if the search space dimension is fewer than a thousand parameters.

3.5 Why it is important to keep LS small enough

Recent research in machine learning has revealed the double descent phenomenon, which is an unexpected finding, that the test error of a model decreases initially, then increases, and then decreases again as the model complexity increases (Belkin et al., 2019; Schaeffer et al., 2023). This phenomenon is intriguing because it goes against the conventional bias-variance trade-off theory, which states that as model complexity grows, the test error should decrease at first and then start to rise continuously. In other words, when the capacity of a model exceeds the interpolation threshold, the model’s accuracy experiences a significant improvement.

The double descent phenomenon has been identified across different machine learning contexts, such as linear regression, neural networks, and other model types. It indicates that there are distinct levels of model complexity where the test error behaves differently, challenging our comprehension of how models generalise to new data. This observation implies that, in the realm of machine learning, the greater the dimensionality of a GWAE’s hidden space, the greater its generative capability will be. However, when aiming to optimise within a latent space in a problem of HM, the notion that "bigger is better" no longer holds true.

Even if we have a rather robust optimisation algorithm, which can handle thousands of parameters, it is still important to keep LS as small as possible.

The “Curse of Dimensionality” concept, first introduced by Bellman et al. (1959), is a critical limiting factor in solving complex optimisation problems. It can be described from various perspectives (Bronstein et al., 2021; L. Chen, 2009; Hastie et al., 2009). For example, let’s assume that all input parameters are uniformly distributed in a d -dimensional unit hypercube. To cover 1% ($n = 0.01$) of all data in a 10-dimensional space, we need to consider 63% of the spread of all input data ($e_d(n) = n^{1/d}$), i.e., the edge of such an embedded hypercube should be 0.63 in length, while the absolute range of information is 1. When increasing the percentage of data description to 10% ($n = 0.1$), it is necessary to consider 80% of the information spread.

If we reduce the dimensionality of the space to $d = 2$, then solving such a problem will be simplified to 10% and 32%, respectively.

Another interpretation of the “Curse of Dimensionality” is described from the perspective

of sample density, which is proportional to $N^{1/d}$, where d is the dimensionality of the input parameters, and N is the sample size. Therefore, if $N_1 = 100$ is the sample size for describing one parameter, then for describing 10 parameters, a sample size of $N_{10} = 100^{10}$ is required.

Moreover, if we return to the data property called Stationary, it can be formulated as follows: $||f(x) - f(x+\epsilon)|| \leq ||x - (x+\epsilon)||$, which means that if we slightly change our input data, the result should not change significantly (Lipschitz continuous function). Thus, the complexity of testing this condition will exponentially increase with the dimensionality of the space.

In the paper “On the Latent Space of Wasserstein Auto-Encoders”, Rubenstein et al. (2018) study the role of dimensionality of the latent spaces. In **Chapter 1: Introduction** I mentioned that through the use of LS \mathcal{Z} , we are trying to model the “real” natural processes that formed a specific geological object. This means that “real” physical processes can also be represented as a set of parameters that shape the geological object. Thus, the question arises: what if the dimensionality of \mathcal{Z} is significantly smaller or larger than the dimensionality of the “real” processes, which Rubenstein calls “intrinsic dimensionality”?

An experiment demonstrated a scenario where a one-dimensional law was encoded into a hidden space of two dimensions. The WAE successfully reconstructed the objects while also attempting to evenly distribute the one-dimensional data manifold within the latent space. To achieve this, the WAE had to curl up the manifold in the LS. In practical terms, the WAE only needed to fill the space enough to deceive the divergence measure employed by the Wasserstein metric. This experiment can be extrapolated into higher-dimensional problems, which will cause latent spaces of complex topology and bring additional difficulties for the optimisation process.

On the contrary, if the dimension of a LS is much smaller than the “intrinsic dimensionality”, it will cause difficulties for the *Decoder* to reconstruct an object x due to insufficient information encoded into the latent vector z .

The dimensionality of the LS is crucial in manifold projection methods. Significant data characteristics are compressed into a single dimension if the dimensionality is too small. At the same time, if the dimensionality is too large, the projections become noisy and may even become unstable in certain cases.

Currently, the problem of automatically determining the intrinsic dimensionality comes to the forefront when it is necessary to decide on the number of independent variables for a complete description of the state of a system (B. Chen et al., 2021; Levina et al., 2004). There are two general groups to estimate the internal dimensionality: Eigenvalue and Geometric methods. The first can be a good tool for exploratory purposes, where one

can plot eigenvalues and look for a clear boundary, but not for obtaining reliable estimates of the internal dimensionality. The second method uses the internal geometry of the dataset and is more effective because it can process non-linear manifolds. However, a more in-depth consideration of the issue goes beyond the scope of this work.

3.6 Inner Geometry of a Latent Space

Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line.

Benoît Mandelbrot

In the final sections of this chapter, I will move on to more advanced technologies necessary for a deeper understanding and analysis of hidden spaces. Therefore, we will elevate the level of abstraction to Riemannian geometry and Topological Data Analysis.

At the beginning of the chapter, I mentioned the manifold concept and promised to provide a more detailed definition. Recall that if the dimensionality of the hidden space \mathcal{Z} is greater than the intrinsic dimensionality, then in this case, GWAE had to curl up the manifold in the latent space. In general, we assume that the data (observations) lie in a space of a nonlinear low-dimensional manifold. In simpler terms, a geological object is generated by some hidden nonlinear process, so the data we observe, due to field exploration, lies in the parameter space of this hidden process – this space can be called a manifold. Thus, ideally, our latent space should fully replicate the manifold, which is impossible.

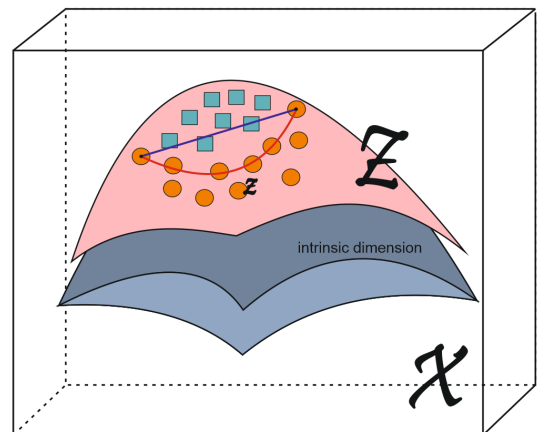


Figure 3.16: The concept of ambient space \mathcal{X} which in 3D (\square); intrinsic space in 2D (\square); LS \mathcal{Z} (\square), which tries to cover intrinsic space; latent vectors z of different classes (\bullet , \square); strait (Euclidean) interpolation ($—$); shortest (geodesic) path (\curvearrowright)

On the other hand, a manifold can be viewed as a curved surface in an ambient space, where the ambient space is the dimensionality of the space of our features that we observe as a result of field exploration. Therefore, when analysing a LS, it is necessary to consider the curvature of space, which leads to errors using Euclidean metrics when working with latent vectors. We must use metrics that consider the non-Euclidean geometry of the hidden space, figure 3.16.

From here arises the question of what methods let us determine the curvature of the latent space and thus define a metric that allows us to work in such a space. Therefore, replacing Euclidean “straight lines” with “shortest paths” (from the perspective of differential geometry, such paths are called geodesics) for working with latent spaces GWAE will allow us to achieve the following practical goals:

- improve the optimisation process. Introducing a metric that corresponds to the density of the regions in the LS will provide the optimisation process with information that we are staying in a high-density space, which means that the generator reproduces the most probable objects, which can be interpreted as the most realistic objects from a geological point of view.
- shortest paths. Allows the selection of vectors from hidden spaces that correspond to similar geological objects more confidently.
- smooth interpolation through the latent space. Provides the ability to smoothly transition from one geological realisation to another through the latent space

To implement this concept, it is necessary to include the notion of Riemannian spaces in the analysis.

The works by Arvanitidis et al. (2021), Lawrence (2005), Shao et al. (2017), and Tosi et al. (2014) were among the first to focus on the analysis of latent spaces generated by generative neural networks in terms of their internal nonlinearity. They showed that taking into account the internal nonlinearity improves the quality of sampling latent vectors z of the LS \mathcal{Z} , which leads to an increase in the quality of generated objects x .

The present section introduces a metric for assessing the curvature of the hidden space, which was adapted from the work of Arvanitidis (2019).

According to the beginning of this chapter, our dataset $\{x\} \in \mathcal{X} = \mathbb{R}^n$. Similarly, let us assume that \mathcal{X} "lives" in some nonlinear hidden space \mathcal{Z} , which ideally is our intrinsic manifold.

NOTE:

S_+^m — positive definite matrix of size m .

\mathbb{J} - Jacobian, which reflects the rate of change.

Denote the Riemannian space as $\mathcal{Z} = \mathbb{R}^m$, since the hidden space of GWAE should ideally be as close to it as possible. Therefore, to take into account the curvature of the Riemannian space, we need to introduce a metric tensor $\mathbf{Z} : \mathcal{Z} \rightarrow S_+^m$. Ideally, the values of the metric tensor should be small in regions of \mathcal{Z} confirmed by observations x , otherwise \mathbf{Z} should be large, indicating that in such regions of \mathcal{Z} there was no data from the training set x .

By definition, the *Decoder* is a function of the form: $f(Z) = P(X|Z) = \mu_\theta + \sigma_\theta \odot \epsilon$, where $\epsilon \sim \mathcal{N}(0, I_n)$. According to Arvanitidis (2019), f should generate a pull-back metric (Khlevniuk et al., 2017) in the space \mathcal{Z} , taking into account its internal geometry (Arvanitidis et al., 2021).

$$\mathbf{Z}_f(z) = \mathbb{J}_f(z)^\top \mathbb{J}_f(z) = \mathbb{J}_\mu(z)^\top \mathbb{J}_\mu(z) + \mathbb{J}_\sigma(z)^\top \mathbb{J}_\sigma(z). \quad (3.34)$$

This metric \mathbf{Z} has several desirable properties. The first term considers the curvature of the hidden space \mathcal{Z} , while the second term increases with uncertainty. Therefore, regions of the hidden space without training data should have high metric \mathbf{Z} values.

However, one of the critical features of neural networks is that they have high interpolation capability but low extrapolation capability. Therefore, we cannot guarantee that regions of the latent space without training data will have high uncertainty, i.e., high values of σ_θ .

To avoid this negative impact, we need an inverted estimate of σ_θ , with a neural network that extrapolates the metric values to zero in uncertain regions of the hidden space. According to Que et al. (2016), it is sufficient to train the precision $\beta_\psi(z) = \frac{1}{\sigma_\psi(z)}$ using a neural network with Radial Basis Function.

$$\beta_\psi(z) = W_+ v_k(z) + \xi \quad (3.35)$$

$$v_k(z) = \exp(-\lambda_k \|z - c_k\|_2^2) \quad (3.36)$$

where ψ - parameters of a neural network, W_+ - positive definite matrix of weights, required for a positive estimate, $c_k \in \mathcal{Z}$, ξ - prevents division by zero. λ_k — centres and bandwidths of $k = 1, \dots, K$ Radial Basis Functions. This estimate is interpreted as follows: as the distance from the centroid increases, the uncertainty of the metric increases.

To calculate the metric \mathbf{Z} , it is necessary to employ the method of Jacobian calculation. This is based on the calculation of partial derivatives of the Decoder. Given that the Decoder is comprised of a sequence of linear transformations of graph convolutions, it is essential to perform calculations over the weight coefficients of the hidden layers in order to find the Jacobian. It should be noted that Jacobian estimation is a computationally demanding process, so it's not always possible to calculate it in practice.

To improve the efficiency of the optimisation algorithm (CMA-ES) presented in **Subsection 3.4.1: Covariance-Matrix Adaptation Evolution Strategy (CMA-ES)**, it is necessary to use geodesic distances between points instead of Euclidean distances, as is currently done. However, on the other hand, this may not be efficient in terms of computational time for finding optima, as calculating Jacobians is a computationally intensive task. To address this, it seems necessary to resort to approaches for approximating Ja-

cobians (Gatilov, 2014; G. Wang et al., 2023) only then the optimisation in Riemannian space be feasible. In any case, improving and speeding up the optimiser’s performance in hidden spaces is not the objective of this study but rather one of the directions for further development.

In summary, we have described a metric \mathbf{Z} that provides an estimate of the curvature of the hidden space and increases the uncertainty estimate as we move away from dense regions of the hidden space covered by training data.

3.7 Objective function in HM process

Assisted history matching (AHM) methods have become widely used to provide an automated adjustment of reservoir model parameters to achieve the best resemblance between modelled and historical dynamic indicators such as production rates, cumulative production and pressure measurements. The essential components of AHM process are an optimisation algorithm and an Objective Function, which estimates the agreement of a model under consideration with the data at hand. The OF is a function of multiple variables that must be minimised using an optimisation algorithm. Serving as a measure of discrepancy between calculation and history, the OF, with the help of an optimisation algorithm, provides an iterative improvement, allowing for refining reservoir parameters.

With the development of generative neural networks as a method of matching geological models, the OF must control the quality of matching based on production data and static indicators such as well logs. Furthermore, even good matching to static well information does not guarantee the geological realism of the final model, as the spatial distribution of properties may be disrupted. Therefore, similar to Mosser et al. (2019), the OF should consist of three main components: the error in reproducing dynamic production data ($Loss_{flow}$), the error in reproducing static information such as well logs ($Loss_{static}$), and the control of geological realism of the model in general ($Loss_{realism}$), which is responsible for property propagation in the inter-well space, providing geologically reliable representations.

$$Loss = w_f Loss_{flow} + w_s Loss_{static} + w_r Loss_{realism} \quad (3.37)$$

Where w_f, w_s, w_r are weights to provide additional flexibility to the overall loss function.

The first method for estimating the degree of error can be traced back to the early development of statistics and probability theory. The concept of Mean Squared Error (MSE) can be attributed to the work of mathematician Carl Friedrich Gauss (Gauss et al., 1957) in the early 19th century.

Gauss introduced the concept of least squares estimation, which involves minimising the sum of squared errors between observed data points and their corresponding estimated values. This approach was a significant advancement in statistical estimation, providing a systematic way to find the best-fitting line or curve through a set of data points.

Squared errors in Gauss's method were primarily motivated by mathematical convenience, as squaring the errors allowed for more straightforward calculations and mathematical derivations. Additionally, squaring the errors ensured they were always positive, avoiding issues with cancelling out positive and negative errors.

The MSE as a formal measure of discrepancy was further developed and popularised in the 20th century, particularly through the works of Fisher (1920) and Splawa-Neyman (1990). Fisher emphasised the importance of using unbiased estimators and introduced the concept of mean squared deviation as a measure of precision.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.38)$$

Neyman, another influential statistician, expanded on Fisher's work and formalised the concept of Mean Squared Error. He recognised that minimising the MSE was equivalent to maximising the efficiency of an estimator, which measures how much information about the population is captured by the sample.

Since then, the MSE has become a widely accepted criterion for assessing the quality of estimators in various fields, including statistics and machine learning. Its simplicity and mathematical properties make it a convenient measure for comparing different estimation methods and evaluating their performance.

However, the MSE has some limitations. It gives equal weight to all errors, regardless of their magnitude or importance. This means that large errors have the same impact on the MSE as small error values, which may not accurately reflect the true performance of an estimator.

Alternative error measures have been proposed to address this issue. One such measure is the Mean Absolute Error (MAE), which takes the absolute value of errors instead of squaring them. This gives more weight to larger errors and can provide a more robust performance measure in certain situations.

Another commonly used error measure is the Root Mean Squared Error (RMSE), the square root of the MSE. The RMSE has the advantage of being in the same unit as the original data, making it easier to interpret. However, it still has the same limitations as the MSE in giving equal weight to all errors.

Overall, the choice of error measure depends on the specific requirements and characteristics of the problem. It is essential to consider the implications and limitations of different measures when evaluating and comparing estimators. Most of the HM publications utilise MSE or RMSE error measure (Hutahaean et al., 2015; 2017; Purshouse et al., 2003; Schulze-Riegert et al., 2017). In the context of uncertainty quantification, MSE and RMSE are particularly valuable as they facilitate the evaluation of posterior probabilities within a Bayesian framework (Christie et al., 2006; O’Sullivan, 2004). This assumes that the errors in the likelihood model follow a Gaussian distribution. In further analysis, we will refer to MSE error measure because data considered in experiments is semisynthetic without any special importance differentiation.

Let’s consider equation’s 3.37 components in more detail.

3.7.1 Flow loss function

One of the main criteria for evaluating the quality of HM of a simulation model is the assessment of the degree of reproduction of production data from each specific well of a field. The parameters involved in model matching are selected so that the simulation model reproduces the data obtained from the field as accurately as possible.

Thus, the component of the loss function responsible for dynamic error estimation is represented by the following expression 3.39:

$$Loss_{flow} = w_f \sum_{c=1}^C \sum_{i=1}^M \frac{1}{N} \sum_{n=1}^N \left(y_{c,i,n}^{obs} - y_{c,i,n}^{sim} \right)^2 \quad (3.39)$$

Where, y^{obs} - production data observed, y^{sim} - production data simulated, N - number of time steps, M - number of wells, C - number of production components (oil, water, gas, injection).

It is recommended (Yeremian, 2022) use current flow rates for all types of fluids. Using current indicators helps to avoid accumulated errors that are inherent in accumulated well performance indicators. Accumulated well performance indicators are integral quantities. As a result, when using them in the loss function, the following problems may arise. For example, suppose there is an outlier in the flow rate measurements of a well at the beginning of its operation. In that case, this will affect the values of the accumulated oil production, which will worsen the overall matching result after the outlier, as the optimisation algorithm will be based on the distorted historical curve of accumulated oil production. When using current flow rates in the loss function, such a problem does not

exist.

The second reason is that in some sections, the model values for the oil flow rate may exceed historical values, while in others, they may be lower. Integral matching based on accumulated values will be good, but the behaviour of the well in terms of flow rate will be reproduced inaccurately.

The third disadvantage of using accumulated production occurs when there is a sharp change in the oil flow rate at the end of the well's operation period. A typical example is some intensification events, such as hydraulic fracturing, to enhance fluid inflow to the well. If the loss function uses accumulated production, then against the large value of this indicator in the late period of well operation, local changes in flow rate will not significantly contribute to the OF, leading to the disregard of matching to this sharp local change in flow rate.

When only two fluids are defined mathematically, one component should ensure convergence to the second, as control of the model is based on liquid rate (oil + water). However, this is not always achieved for the following reasons. When it is impossible to provide fluid flows at the current reservoir productivity and minimum bottomhole pressures, control based on fluid flow switches to control based on the specified bottomhole pressure.

Including watercut in the OF is undesirable because this parameter is derived and contains combined errors from flow rate measurements. Additionally, the option with flow rates is more flexible than watercut, allowing for different measurement errors for oil and water. Including the watercut component in the OF when flow rate components are present will double-account for flow rate discrepancies, shifting the focus away from the numerical contribution of other components, such as bottomhole pressure. This, in turn, will negatively affect the overall criteria adaptation process.

In general, during the HM process, more attention is paid to the convergence of oil production rates than bottomhole pressures.

3.7.2 Static loss function

Petrophysical properties, such as porosity and permeability, are crucial in determining a rock's ability to store hydrocarbons and facilitate fluid flow. These properties are typically measured in laboratories using rock samples obtained from cores or outcrops. After that, these measurements are propagated along the wellbore by correlation with well logs, such as density, neutron, sonic, and nuclear magnetic resonance. This data is then interpolated to an inter-well space of an entire volume of a field by the various geostatistical approaches with the seismic data and sedimentological concept (Ringrose et al., 2015).

Since “direct measurements” of properties are conducted at well locations, the obtained data have a lower uncertainty level than the inter-well space. Therefore, the second term of the OF, equation 3.37 ($Loss_{static}$), should control that the generated values of porosity and permeability at well locations do not contradict the results of interpretations of well log measurements.

$$Loss_{static} = w_s \sum_{c=1}^C \sum_{i=1}^M \frac{1}{N} \sum_{n=1}^N \left(y_{c,i,n}^{obs} - y_{c,i,n}^{sim} \right)^2 \quad (3.40)$$

Where, y^{obs} - well logs observed, y^{sim} - well logs simulated, N - number of grid blocks along the wellbore, M - number of wells, C - number of log types (porosity, permeability due to well logs).

3.7.3 Realism loss function

The last component of the loss function, equation 3.37 ($Loss_{realism}$), should be responsible for the quality of the generated grids of porosity and permeability in the inter-well space. In the concept of this thesis, the quality of the generated objects implies that they correspond to the geological concept initially established by the geologist.

It is assumed that during the training of GWAE, the training dataset consists only of geologically realistic scenarios, which means the neural network is trained to reproduce the underlying geology correctly. However, it is worth noting that the LS (\mathcal{Z}), which is the “source” of geological realisations, has a probabilistic structure. Therefore, if we sample from low-probability areas, the reconstructed geological realisations may contain disrupted geology. For us, this means that the optimiser should receive information about the probability of occurrence of a particular realisation. In this work, I replaced the probability estimation with an estimation of the density of the LS $\mathbf{Z}_f(z)$ from equation 3.34, which was presented in the **Section 3.6: Inner Geometry of a Latent Space**.

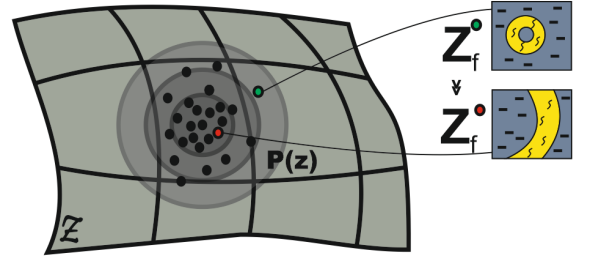


Figure 3.17: If we pick a sample z (●), from a region with high metric value $\mathbf{Z}_f(z)$, GWAE Decoder will reproduce an object of high geological reliability. Otherwise, the sample z (●) from regions of low density, which provide low-quality geological realisation.

Thus, as the third term, the loss function will receive an estimation of density $\mathbf{Z}_f(z)$. If this estimation starts to increase, it means that the optimiser is starting to sample from areas of

the hidden space with low density, i.e., areas with few training samples, indicating a low probability of occurrence of such a realisation, figure 3.17. As the estimation of density decreases, the probability of occurrence of such a geological realisation will also decrease.

It is worth noting that during the description of experiments in the following chapters, various modes of testing the performance of the Objective Function are not considered, such as the prevalence of static information over dynamic information (i.e. a small number of historical measurements), etc., as this is not the subject of the current research. However, the conducted experiments have shown that the static and dynamic components of the OF work coherently, which is logical since well-logging data is one of the most critical indicators for subsequent extraction. Meanwhile, the third component operates more against it as it imposes constraints on the optimization process, restricting the search exclusively to the zone of geologically realistic scenarios.

3.7.4 Optimisation process

Thus, the process of AHM includes an optimiser (CMA-ES) that, based on the three-component loss function, takes the next step to search for the optimum at each iteration. The optimum in this case involves minimising the production error for each well, well logs error, and while maintaining geological realism.

It is worth noting that to evaluate the production error, it is necessary to perform a simulation, which itself is computationally expensive. A simulation is not required to assess well logs error; only the generated by GWAE property graphs are needed, which is fast. However, an implicit dependence exists between production and geology. It can be argued that if we find the optimum for the static component of the loss function, we will approach the optimum for the dynamic component. This results in significant computational resource savings. Furthermore, once we have completed the optimisation process for the static component, we can perform fine-tuning and include the dynamic error component in the loss function, thereby controlling all three components, figure 3.18.

The third component of the loss function should be involved throughout the optimisation process, as it acts as a limiting factor (regulariser) that prevents the optimiser from sliding into the space of unlikely geological realisations, ensuring geological realism.

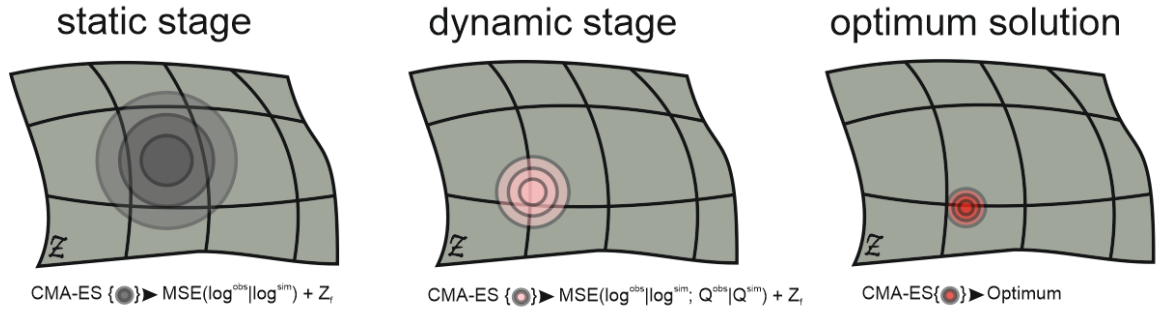


Figure 3.18: Two-stage optimisation process. Left: CMA-ES seeks optimum subspace for static properties by minimising $Loss_{static}$ component. Centre: in the region of static optimum (○) CMA-ES switch on the $Loss_{dynamic}$ component to find the total optimum (●), for static and dynamic components of the loss function. $Loss_{realism}$ is a regulariser for the optimisation process.

3.8 Tools for analysing multidimensional spaces

Топология! Стратосфера человеческой мысли! В двадцать четвёртом столетии она, может быть, и понадобится кому-нибудь, а пока... А пока...^a

^aTopology belonged to the stratosphere of human thought. It might conceivably turn out to be of some use in the twenty-fourth century, but for the time being...

In the first cycle
A. Solzhenitsyn

In the case of GWAE latent space \mathcal{Z} is always too large to conduct visual analysis and draw any conclusions about its inner structure. That is why we need some dimensionality reduction techniques, which are used to reduce the number of dimensions in a dataset while preserving its important structure or relationships.

Different dimensionality reduction methods allow for extracting information of various kinds. There is no universal method that would comprehensively analyse multidimensional data. Therefore, a researcher should have a set of such algorithms in their arsenal. There are quite a few dimensionality reduction methods (Holmström, 2008; Keim, 2002; Tang et al., 2016; Tierny, 2018; Y. Wang et al., 2021), but I will be using some of the most widely used ones in practice.

In this work, I am using algorithms such as Principal Components Analysis (Jolliffe, 2002), t-Distributed Stochastic Neighbour Embedding (t-SNE) (Maaten et al., 2008), as well as Topological Data Analysis (TDA) (Tierny, 2018).

These techniques are commonly used for exploratory data analysis, visualisation, and feature extraction. They can help to understand the underlying structure of complex datasets and reduce the computational complexity of subsequent analyses. However, it's important

to note that while PCA is a deterministic technique, t-SNE is stochastic and can produce different results with different random seeds.

The basic approach for analysing multivariate data is orthogonal component analysis such as Principal Components Analysis and Multidimensional Scaling (Céline Scheidt, 2018). PCA and MDS are both powerful techniques in the field of multivariate analysis that are commonly used for dimensionality reduction and data visualisation. PCA is a linear technique that projects the data onto a lower-dimensional space by finding the principal components, which are new uncorrelated variables that capture the maximum variance in the data. It helps to identify the most important features or patterns in the data. MDS is a technique used to visualise the similarity or dissimilarity between objects or cases based on a distance or dissimilarity matrix. MDS aims to represent the relationships between objects in a lower-dimensional space, typically two or three dimensions, while preserving the original pairwise distances as much as possible.

On the other hand, t-SNE is a non-linear technique that focuses on preserving the local structure of the data. It converts similarities between data points into probabilities and minimises the divergence between them in the high and low-dimensional space. This allows for better visualisation of clusters or groups within the data.

Topological Data Analysis is a mathematical framework that aims to analyse and understand complex datasets' underlying shape and structure. It combines concepts from algebraic topology, computational geometry, and machine learning to extract topological features from data. TDA focuses on capturing the global and local geometric properties of data, such as clusters, holes, loops, and connectivity, which may not be easily captured by traditional statistical methods or linear dimensionality reduction techniques like PCA. It provides a way to analyse the shape and structure of data in a robust and mathematically rigorous manner.

3.8.1 Principal Components Analysis

The Principal Components Analysis is a statistical technique used to reduce the complexity of multidimensional data by identifying patterns and relationships among variables. PCA transforms the original variables into a new set of variables called principal components, which are linear combinations of the original variables. These principal components capture the maximum amount of variation in the data with minimum redundancy. It was proposed by Pearson (1901) and then further developed by the American economist and statistician Hotelling (1933).

From a mathematical point of view, PCA is an orthogonal linear transformation that projects data from its original feature space to a new space with reduced dimensionality. During

this procedure, the primary axis of the newly established coordinate system is identified to maximize the variance of the data along that axis. The second axis is then constructed orthogonally to the first, aiming to maximize the variance among the remaining possibilities, and this process continues. The initial axis is referred to as the first principal component, followed by the second principal component, and so on.

Figure 3.19 shows the reduction of dimensionality of the original 2-dimensional space (X_1, X_2) using PCA to a 1-dimensional space. The first principal component, PC_1 , aligns with the direction of the greatest variance observed in the scatter plot of points derived from the original dataset.

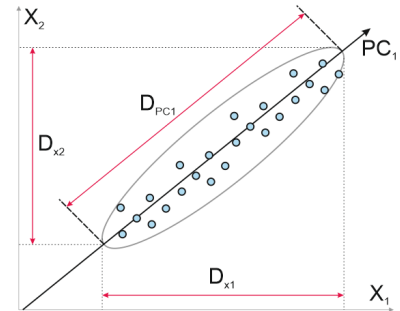


Figure 3.19: The concept of dimensionality reduction by PCA

In figure 3.19, the projection of the data variance onto the axis of the first principal component D_{PC1} is greater than its projections onto the original axes D_{X1} and D_{X2} , but less than the total variance. In other words, the entire range of data variance could not be captured solely through the first principal component. Therefore, in the case of multidimensional data, the second, third, and so on principal components are constructed until they collectively capture all the variance.

Hence, the concept of PCA is that each principal component corresponds to a specific fraction of the overall variance in the original dataset. Consequently, variance, as an indicator of data variability, can signify the degree of informativeness within the data.

Variability can be substantial, minimal, or even non-existent within specific axes of the initial feature space. The underlying assumption is that a lower variance along an axis corresponds to a less impactful contribution from the variable associated with that axis. Consequently, eliminating such an axis (i.e., excluding the variable from the model) makes it feasible to diminish the problem's dimensionality without sacrificing a significant portion of the data's informativeness.

Therefore, the task of PCA is to construct a new feature space of lower dimensionality, where the variance between the axes is redistributed in such a way as to maximise the variance along each of them. To do this, the following sequence of actions is performed (I. Goodfellow et al., 2016):

- The overall variance of the initial feature space is computed, and this task involves more than just adding up individual variances because, in many cases, the variables are not independent. Consequently, it is essential to sum up the covariances among variables derived from the covariance matrix.

- Computation of eigenvectors and eigenvalues of the covariance matrix. These components define the directions of the principal components and quantify the respective amounts of variance.
- Dimensionality reduction is performed. The covariance matrix's diagonal elements represent the variance in the original coordinate system, while its eigenvalues indicate the variance in the new system. To determine the proportion of variance for each principal component, one can divide the variance linked to each component by the total variance across all components. After that, the least informative principal components are discarded.

It's important to note that direct selection of the number of components may not consistently produce favourable outcomes. This is because some variance in the data might be due to noise rather than informative components. Hence, practical approaches often rely on specific criteria, like the Kaiser criterion (Kaiser, 1960) or the broken stick criterion, and so on, to determine the appropriate number of components (Saccenti et al., 2015).

The main limitations of the principal component analysis method are (Céline Scheidt, 2018):

- The inability to interpret the components in a meaningful way as they combine the variance from multiple original variables.
- The method does not account for categorical variables, as it is designed for continuous data only. Categorical variables need to be preprocessed or transformed before applying PCA, which can introduce additional complexity and potential loss of information.
- PCA assumes linearity between variables, which may not hold true in all cases. Nonlinear relationships between variables can lead to inaccurate results.
- The method assumes that the data is normally distributed, which may not be the case in real-world datasets. Non-normal data can lead to biased results and inaccurate interpretations of the components.
- PCA is sensitive to outliers in the data. A single outlier can significantly influence the results and distort the interpretation of the components.
- The method requires a large sample size to capture the underlying structure of the data accurately. With a small sample size, PCA may be unable to reduce dimensionality or provide reliable results effectively.
- The method assumes no missing values in the dataset. Missing values can introduce bias and affect the accuracy of the results obtained from PCA.

3.8.2 t-Distributed Stochastic Neighbour Embedding

The classic Stochastic Neighbour Embedding (SNE) was proposed by Hinton et al. (2002). The 2002 article describes several “tricks” that allowed for simplifying the process of finding global minima and improving visualisation quality. The t-SNE algorithm, also considered a feature learning method, was published by Dutch researcher Maaten et al. (2008). One replaced the normal distribution with the Student’s t-distribution for low-dimensional data.

First, let’s briefly describe the classical SNE algorithm. We have a dataset z with points defined by a multidimensional variable with a dimensionality significantly larger than three. We need to obtain a new variable that exists in a two-dimensional or three-dimensional space, preserving the original data’s structure and patterns as much as possible. SNE starts by transforming the multidimensional Euclidean distance between points into conditional probabilities that reflect the similarity between points. Mathematically, this can be represented as follows:

$$p_{j|i} = \frac{\exp(-\|z_i - z_j\|^2/2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|z_i - z_k\|^2/2\sigma_i^2)} \quad (3.41)$$

This equation shows how close a point z_j to z_i under a Gaussian distribution around z_i with a given deviation σ . σ will be different for each z . It is chosen so that points in denser regions have smaller variances. To achieve this, an estimation of perplexity is used, equation 3.42:

$$Perp(P_i) = 2^{H(P_i)} \quad (3.42)$$

$$H(P_i) = - \sum_j p_{j|i} \log_2 p_{j|i} \quad (3.43)$$

Where $H(P_i)$ - Shannon entropy in bites. Perplexity can be interpreted as a smoothed estimate of the effective number of “neighbours” for a point z_i . It is used as a hyperparameter in the method. The authors recommend using a value in the range of 5 to 50. σ is determined for each pair of z_i and z_j using a binary search algorithm (Williams, 1976).

Let define \hat{z} as a 2 or 3-dimensional representation of z .

$$q_{j|i} = \frac{\exp(-\|\hat{z}_i - \hat{z}_j\|^2)}{\sum_{k \neq i} \exp(-\|\hat{z}_i - \hat{z}_k\|^2)} \quad (3.44)$$

$$\sigma = 1/\sqrt{2} \quad (3.45)$$

If the points \hat{z}_i and \hat{z}_j accurately model the similarity between high-dimensional source

points z_i and z_j , then the corresponding conditional probabilities $p_{j|i}$ and $q_{j|i}$ will be equivalent. As an obvious measure of the quality with which $q_{j|i}$ reflects $p_{j|i}$, the authors use the KL divergence or distance. SNE minimises the sum of such distances for all mapping points using gradient descent, equation 3.47. The loss function for this method is determined by the equation 3.46.

$$Dist_{SNE} = \sum_i KL(P_i||Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}} \quad (3.46)$$

$$\frac{\partial Dist_{SNE}}{\partial \hat{z}_i} = 2 \sum_j (p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j})(\hat{z}_i - \hat{z}_j) \quad (3.47)$$

The authors propose the following physical analogy for optimisation: All mapping points are connected by springs. The stiffness of the spring connecting points i and j is determined by the similarity of two points in the multidimensional space \mathcal{Z} and points in the mapping space $\hat{\mathcal{Z}}$. In this analogy, the gradient is the resulting force acting on a point in the mapping space $\hat{\mathcal{Z}}$. If the system is “released”, it will eventually reach equilibrium, which will be the desired distribution. Algorithmically, finding equilibrium is suggested by taking into account moments.

The loss function is the main difference of t-SNE. First, t-SNE has a symmetric form of similarity in the high-dimensional space and a simpler gradient variant. Second, instead of using a Gaussian distribution for points in the mapping space, a Student’s t-distribution with heavy tails facilitates optimisation and solves the sparsity problem.

As an alternative to minimising the sum of Kullback-Leibler divergences between conditional probabilities $p_{j|i}$ and $q_{j|i}$, it is proposed to minimise a single divergence between the joint probability P in the high-dimensional space and the joint probability Q in the mapping space.

$$Dist_{t-SNE} = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (3.48)$$

$$p_{ij} = p_{ji} = \frac{p_{j|i} + p_{i|j}}{2n} \quad (3.49)$$

$$q_{ij} = q_{ji} = \frac{(1 + \|\hat{z}_i - \hat{z}_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|\hat{z}_k - \hat{z}_i\|^2)^{-1}} \quad (3.50)$$

$$p_{ii} = q_{ii} = 0 \quad (3.51)$$

And the gradient:

$$\frac{\partial Dist_{t-SNE}}{\partial \hat{z}_i} = 4 \sum_j (p_{ij} - q_{ij})(\hat{z}_i - \hat{z}_j)(1 + \|\hat{z}_i - \hat{z}_j\|^2)^{-1} \quad (3.52)$$

Returning to physical analogues, the resulting force, defined by equation 3.52, will significantly contract the points \hat{z} for close points z of multidimensional space and repulse remote ones.

The search for equilibrium is proposed to be done taking into account the moments, equation 3.53:

$$\hat{Z}^t = \hat{Z}^{t-1} + \eta \frac{\partial Dist_{t-SNE}}{\partial \hat{Z}} + \alpha(t)(\hat{Z}^{t-1} - \hat{Z}^{t-2}) \quad (3.53)$$

t-SNE has three potential weaknesses (Maaten et al., 2008):

- Regarding dimensionality reduction, it is unclear how t-SNE will perform when reducing data to higher dimensions ($d > 3$) as it is primarily evaluated for data visualisation. The behaviour of t-SNE in two or three dimensions cannot be directly extrapolated due to the heavy tails of the Student-t distribution to higher dimensions.
- The curse of intrinsic dimensionality (Bengio, 2009) poses a challenge for t-SNE as it relies on the local properties of the data. The assumption of local linearity may be violated in datasets with high intrinsic dimensionality and a varying underlying manifold.
- Unlike other dimensionality reduction techniques with convex cost functions, t-SNE's cost function is non-convex. This leads to the need to choose optimisation parameters, and the resulting solutions may vary based on these choices and initial random configurations.

Below is a set of steps on how t-SNE works:

Algorithm 2: t-SNE dimensionality reduction

Data: $\{z\}$

Input: $n_{iter}, Perp, \eta, \alpha(t)$

Result: $\{\hat{z}\}$

estimation of $p_{j|i}$ eq.3.41;

set $p_{i,j}$ with eq.3.49;

initialise $\hat{Z}^0 = \{\hat{z}\} = \mathcal{N}(\mu = 0, \sigma = 1e - 4)$;

for $t \leftarrow 1, T$ **do**

estimation of q_{ij} with eq. 3.50 ;

estimation of the gradient by 3.52 ;

set \hat{Z}^t with 3.53 ;

end

3.8.3 Topological Data Analysis

Topology is the science of data shape. Studying the topology of an object means understanding its geometry and structure at a certain level of abstraction without dimensionality reduction itself.

Topological Data Analysis allows for the exploration of complex, multidimensional, noisy data and extracting important properties that provide insight into the shape of the object being studied by Carlsson (2009) and Edelsbrunner et al. (2000). TDA could effectively capture the intrinsic geometry of complex data sets. In the case of geological representations, latent spaces encode high-dimensional relationships between geological representations. TDA can reveal the underlying topological structures that may not be apparent through traditional methods.

The tool used for TDA, which computes topological features of objects of different nature, is called Persistent Homology (Biasotti et al., 2008). Simplicial Complexes are commonly used to represent multidimensional data clouds. A Simplex can represent different elements, such as a single data point (0-simplex), a line connecting two data points (1-simplex), a triangle (2-simplex), and so on. Generally, a subset of $(k + 1)$ data points is called a k -simplex. A simplicial complex is a collection of these simplices. Figure 3.20 illustrates some examples of k -simplices and a simplicial complex.

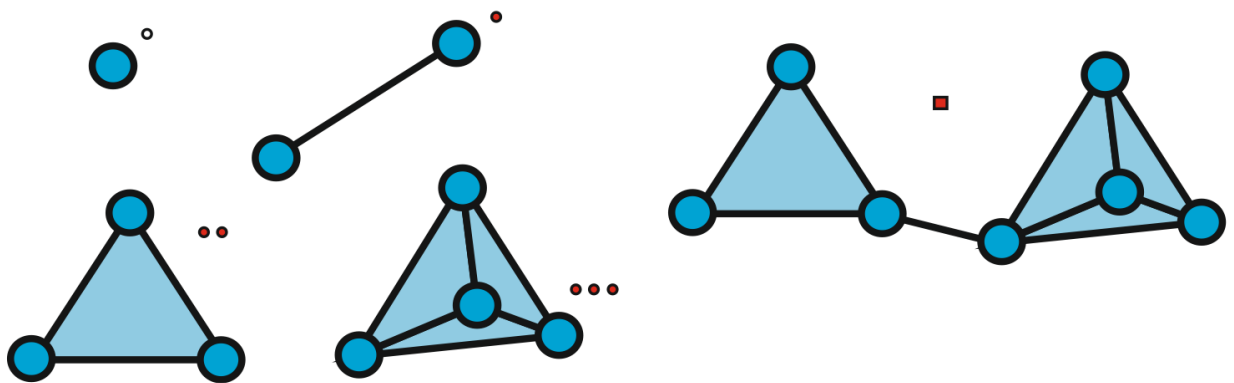


Figure 3.20: Example of 0-simplex (○), 1-simplex (●), 2-simplex (●●), 3-simplex (●●●), and simplicial complex (■)

Betti numbers are used to describe topological properties, where β_0 represents the number of connected components, β_1 represents the number of 1-D holes, β_2 represents the number of 2-D voids, and so on, figure 3.21.

One of the most common topological representations of data through the connected components is the Persistence Homology Diagram (Fabio et al., 2015). Persistent homology is a tool that captures the topological signature of a data cloud by gradually increasing spa-

tial resolution and observing the appearance and disappearance of loops or holes in each dimension. The topological characteristics of the studied space are represented as a set of points on the diagram, where the moments of appearance of topological “holes” are plotted on the x-axis, and the moments of their disappearance are plotted on the y-axis. With such a diagram, we can understand which characteristics are irrelevant to the original space, as their stability or lifespan (the difference between the moment of disappearance and the moment of appearance) is small compared to the stability of other topological characteristics. We can also identify which characteristics have high stability and are fundamental in shaping the space.

A persistent diagram can be considered for different dimensions (H_0 , H_1 , etc.). For H_0 , we track the time of appearance and disappearance of connected components, which tells us about the stability in dimension 0. For H_1 , we follow the stability of the appearance and disappearance of loops. H of dimension n can be described as the appearance and disappearance of a hypersphere that encloses multidimensional voids, figure 3.22. There are many examples of using this approach to analyse multidimensional data. In Gholizadeh et al. (2018), it is shown how to extract topological signatures in the text using persistent homology to utilise these signatures for text classification. Pereira et al. (2015) presented the usefulness of the TDA in the clustering problem of time series and spatial data in biology, medicine, and ecology by the Betty numbers.





				
β_0	1	1	1	1
β_1	0	1	0	2
β_2	0	0	1	1
...

Figure 3.21: Betti numbers for a single point, a circle, a sphere, and a torus

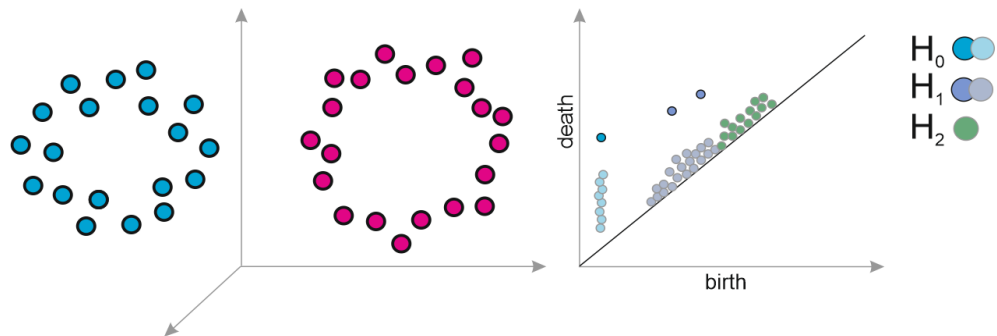


Figure 3.22: Example of persistence homology diagram. Left: 3-dimensional data $\{z\} \in \mathcal{Z}$ (● ●). Right: Persistence homology diagram for H_0 (●), H_1 (●) and H_2 dimensions (●). Noise or short-living components are in faint colours, and long-living components are bright. For H_0 , there are two distinct bodies because we can see one point on the persistence homology diagram distant from the birth-death diagonal (●). This is the moment when two sets of points form a single one (● ●). For H_1 , we can see two bright points (●), this indicates two stable circle bodies. For H_2 , we can see only short living spheres, which indicates no distinct bodies (noise)

Thus, visualising multidimensional spaces using persistent diagrams combined with Betti

numbers allows us to create a particular story of what is happening in the multidimensional space. In further experiments, persistent diagrams will be utilised to identify hidden patterns that help understand the relative structure of geological representations within the latent space (Moon et al., 2019). The primary constraint in topology data analysis is the significant amount of computing resources needed (Cheng, Liang, 2020), so to speed up the process of TDA metrics computation, we will randomly pick points from $\{z\}$ to describe and analyse the whole space \mathcal{Z} .

3.9 Conclusion

In this chapter, I have introduced the fundamental technologies employed throughout the applications in the further chapters. I have described the essence of *Variational Autoencoders*, justified why their basic configuration is sub-optimal for geological modelling under conditions of uncertainty, and transitioned to the *Wasserstein Metric* and *Graph Architecture* for autoencoders. Next, I presented the optimiser (CMA-ES) along with three components Objective Function that will be used to solve the history matching problem. Then, issues related to determining the dimensionality of the latent space were discussed, which can be crucial in the context of HM through the LS search. The final sections of this chapter were dedicated to more abstract concepts, such as interpolation in nonlinear spaces and topological analysis of multidimensional spaces. These approaches provide a deep understanding of how to work with geological representations in hidden space. They also enable the analysis of the structure of hidden space, which is crucial when dealing with generative networks as relatively new tools in geoscience. It can already be asserted that the concept of distance, and therefore similarity, of geological representations in hidden space is non-trivial. This, in turn, imposes certain limitations on optimisation processes, directly influencing the geological modelling and history matching processes using generative networks. Introducing these techniques is necessary for a more in-depth analysis of the performance of GWAE.

The key principle concerns traditional Convolutional Neural Networks, which have limitations when applied to non-Euclidean geological data. Geological models often have complex, non-regular internal structures that don't fit well with CNNs. To address this, we discussed the transition from Euclidean CNNs to non-Euclidean data representations, specifically using *graphs*. This approach enables more advanced *graph convolutions*, making it possible to describe irregular geological structures and their variation by representing geological models as graphs with nodes containing reservoir properties and edges representing structural features of the geological object.

Chapter 4

Modelling uncertainty and datasets construction process

The initial step in geological modelling using generative neural networks involves creating a set of prior subsurface models based on existing information. The prior ensemble is used to train the model and generate the corresponding data and prediction variables. The process of generating the training set with the target answers involves forward modelling, e.g. reservoir flow simulation, geophysical simulation, or simpler static operators.

The next common step is pre-processing, where the prediction and data variables are transformed via Encoders into lower dimensional latent space or appropriate distributions that meet the necessary assumptions, e.g. Gaussianity or independence of features.

In a typical subsurface system, the model variables are parameterised in a specific manner, such as using a grid or a set of objects with defined characteristics. The prior distribution of these model variables is often referred to as a “geological scenario” (Céline Scheidt, 2018). The question then becomes whether the collection of information and extraction of parameter values accurately produce realistic priors.

To describe the prior geological scenario uncertainty explicitly, it is necessary to understand the sedimentary systems based on physical principles rather than solely relying on information principles. Ideally, this involves stating all possible configurations of system architectures and their frequencies (represented as a probability density). But in reality, it is impossible to cover, so it is important to prioritise and focus on the most relevant and significant configurations for further decisions. Sensitivity analyses and validation against newly observed data can further help in assessing the robustness and reliability of the chosen prior. This probability density does not necessarily follow a Gaussian or uniform distribution. Therefore, the question arises: *What is the probability density for geological systems, and how can it be represented?*

There are a lot of benchmark datasets that exist and are used for data assimilation or optimisation, e.g. Stanford VI (Castro et al., 2005), COSTA (Costa Gomes et al., 2022), PUNQ S3 (Cutts, 1991) etc. However, the internal architecture of the dataset should allow to test a specific hypothesis and get an answer about the applicability of a particular approach. This is why the present work required the creation of a set of in-house datasets.

This section describes the process of building training datasets and prior ensembles, based on which further training of GWAE and testing of the generative technologies described in the previous chapter will be performed. To conduct the experiments, I created three training datasets that will be used throughout this thesis.

- i The first dataset (*FLUVIAL*) consists of a set of three-dimensional models of porosity and permeability properties of a simplified channelised synthetic cube of properties. The main uncertainty of this dataset is the uncertainty in the location of channel bodies, their quantity, and channel parameters.
- ii The second dataset (*STRUCTURAL*) represents structural uncertainty in a simplified synthetic reservoir setting, namely the location of the fault and block displacements (throw) along it.
- iii The third dataset (*BRUGGE*) consists of multiple realisations of the semi-synthetic full-field Brugge benchmark reservoir model, which consists of four stratigraphic zones with different geological scenarios and depositional conditions.

The two first datasets represent relatively simple models of deposits that allow testing specific generative capabilities of GWAE under geological and structural uncertainty. Additionally, all these datasets have uncertainty in petrophysical parameters such as porosity, permeability, and their dependence. A simplified fluvial setting was chosen as the base geological concept, allowing for visual quality control of generated realisations. To test the generative capabilities of GWAE in more realistic conditions, the third dataset was created, which is closer in complexity to a real deposit.

Each of these datasets aims to answer the questions:

- i Is it possible to create a generative model that could combine geological uncertainty of different nature?
- ii Which configurations of GWAE are the most promising?

4.1 Set up for various geological scenarios case study (*FLUVIAL*)

A prior set of models for training GWAE should include the following features: various geological scenarios and the spatial distribution of petrophysical properties. These scenarios should be explicit enough to allow for visual verification of geological realism during result analysis but challenging due to geological bodies' complex connectivity and geometry. Therefore, a fluvial geological setting was chosen.

The concept of the training dataset is presented in figure 4.1. It is worth noting that the different geological scenarios consist of a scenario with a single channel and a scenario with two channels. The channels' parameters for each scenario vary within different uncertainty ranges, table 4.1.

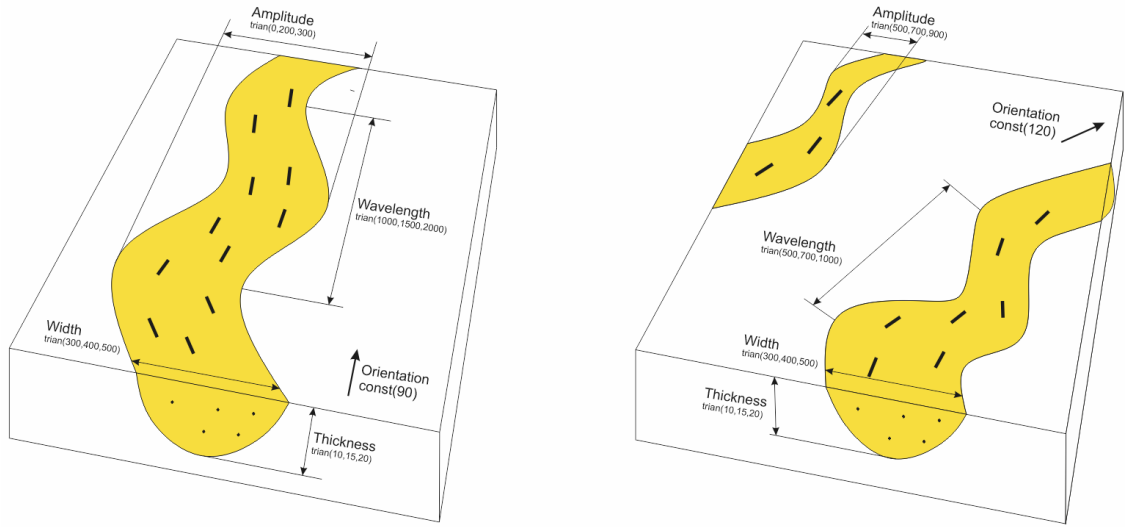


Figure 4.1: Settings of the channelised synthetic dataset. Two distinct geological scenarios: single and double channels are considered

The goal of creating such a dataset is to test the ability of the generative algorithm to work with geological representations in a wide enough space of geological scenario uncertainty.

Table 4.1: Table of parameters for *FLUVIAL* dataset construction

Scenario	Single	Double
Number of channels	one	two
Width	[300; 500]	[300; 500]
Thickness	[10; 20]	[10; 20]
Wavelength	[1000; 2000]	[500; 1000]
Amplitude	[0; 300]	[500; 900]
Orientation	90	120

It is worth noting that the absolute values of the ranges that determine the geometry of the channels are not as important for the validation of GWAE generative capability as the fact

that a dataset describing a different number of channels can be perceived as prior information about the reservoir. The full-field *BRUGGE* case will consider more geologically realistic reservoir parameters.

The petrophysical relationship between porosity and permeability was defined by a deterministic function with some slope variation and bias to provide some uncertainty. This is one of the critical parameters to impact the flow.

The following figure 4.2 shows examples of implementing a porosity dataset for a single (left column) and double channels (right column). It is important to note the imbalance in the volumes of channel and non-channel facies. There is also a clear spatial correlation in the direction of the channels. Additionally, implementations marked with ● reproduce only one channel. This can be explained by the stochastic nature of the object modelling process, which essentially merged two channels into one.

Upon closer examination, it can be observed that each implementation lacks the first cell. The purpose of this irregular structure is to disrupt the grid's regularity. As mentioned earlier, standard convolutional networks cannot work with such data. On the other hand, graph neural networks should be able to process the data since regularity is not required. This approach could be used as a geological simulator capable of generating only "useful" fields for property modelling. Currently, the ACTNUM parameter is used for this purpose, providing information to the flow simulator about which cells require calculations ($ACTNUM = 1$) and which do not ($ACTNUM = 0$), thereby reducing computation time. Therefore, in the case of geological modelling using a graph approach, there is no need to create property fields with inactive cells.

The distributions of porosity and permeability parameters are shown in figure 4.3.

The petrophysical relationship between porosity and permeability follows a "standard" exponential relationship presented in figure 4.4. It should be noted that the uncertainty range for the entire dataset is quite broad, which can be attributed to the inherent uncertainty in the petrophysical relationship of the parameters. However, each individual dataset realisation has a much smaller variability, as illustrated in figure 4.5.

Therefore, a properly trained GWAE should adhere to the following conditions: preserving the statistical distributions of porosity and permeability properties, maintaining the petrophysical relationship between the properties, and preserving the correct spatial geometry for the channel facies regardless of the number of channels.

A commercial Petrel subsurface software (*SLBTM*) was used to create the set of models. The process of creating models for the training dataset was structured as follows. Stochastic Object Modelling was chosen as the geological modelling process, as it is the

4.1. SET UP FOR VARIOUS GEOLOGICAL SCENARIOS CASE STUDY (*FLUVIAL*)

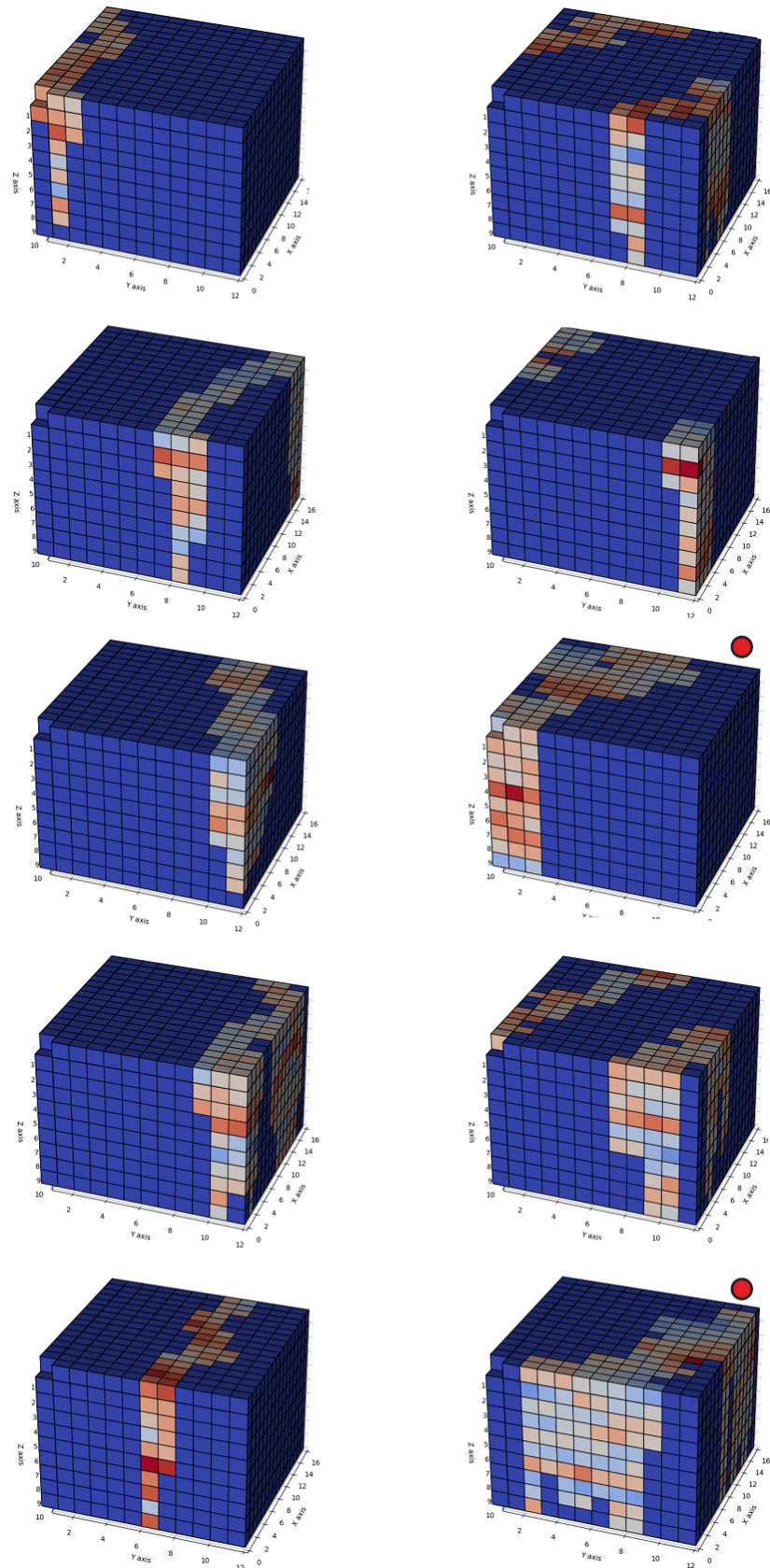


Figure 4.2: Examples of *FLUVIAL* synthetic dataset: a single channel scenario (left) vs the double channel scenario (right) are considered. ● highlight realisations in which only one channel is visually observed, but in reality, two channels merged into one due to the stochastic component of the object modelling algorithm

4.1. SET UP FOR VARIOUS GEOLOGICAL SCENARIOS CASE STUDY (FLUVIAL)

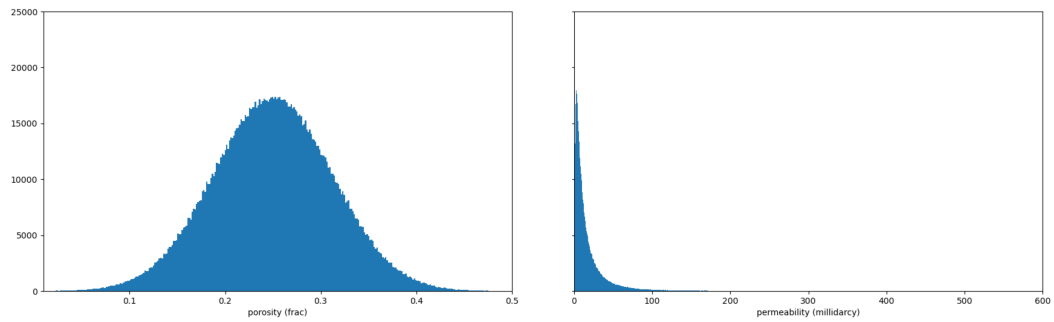


Figure 4.3: Porosity and permeability distributions for the entire dataset. Right: porosity, left: permeability.

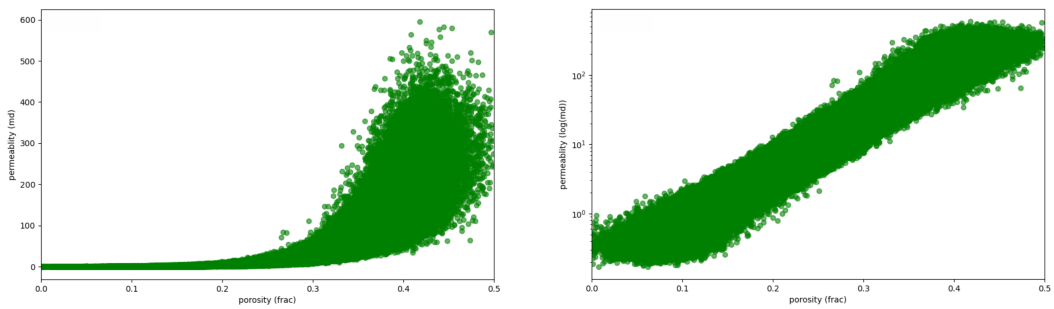


Figure 4.4: Porosity/Permeability petrophysical relationship for the channel facie for the entire ensemble of models. Right: y-axis in millidarcy, left: the same relationship, but the y-axis in $\log(\text{millidarcy})$.

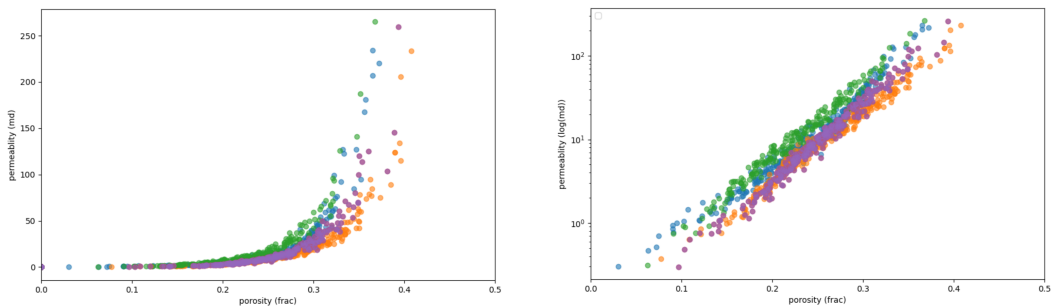


Figure 4.5: Porosity/Permeability petrophysical relationship for the channel facie for some random model samples from prior. Right: y-axis in millidarcy, left: the same relationship, but the y-axis in $\log(\text{millidarcy})$.

most preferable for modelling fluvial systems. This algorithm's operation principle was described in **Chapter 2: Review of recent advances in generative deep learning and problem statement**. Two workflows were implemented to create the dataset.

The workflow simulates the propagation of channel/non-channel facies in 3D volume. In addition to the spatial distribution of channels, their parameters were also varied, table 4.1.

The workflow results in porosity and permeability grids (*.GRDECL format) to be used further in flow simulation. Five thousand realisations were created, 2,500 for each scenario with one or two channels. The model dimension is $x = 16$, $y = 12$, $z = 10$, where $dx = dy = 50m$ and $dz = 1m$, which means that the dimension of the search space for optimisation is $x \cdot y \cdot z \cdot NumberOfProperties = 3840$, as the optimisation algorithm needs to find values for porosity and permeability for each cell.

In addition to the prior dataset of static porosity and permeability properties, a flow simulation model has been prepared for conducting simulation experiments. It is expected that the main structure of the flow model will remain unchanged while the property cubes will be substituted, thereby altering the dynamic response. The model consists of two phases: oil/water. It was created using a line drive scheme, with one line of producers in the middle and two lines of injectors, figure 4.6. This well placement was chosen due to the uncertain position of a channel. The well performance calculation is set for 60 time steps, each time step representing a month. The main parameters of the flow simulation model are presented in the table 4.2.

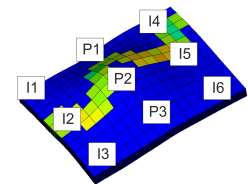


Figure 4.6: Well placement scheme

Table 4.2: Flow simulation parameters

Parameter	Value	Unit
P_{init}	240	bar at OWC
P_{prod}	45	bar
P_{inj}	330	bar
N_{wells}	9	–
OWC	2460	<i>meter</i>
Water viscosity (μ_w)	0.40	<i>cP</i>
Oil viscosity (μ_o)	3.3	<i>cP</i>
Water density (ρ_w)	1020	<i>kg/m³</i>
Oil density (ρ_o)	875	<i>kg/m³</i>
Connate Water Saturation (SWL)	$-0.048 \cdot \ln(perm) + 0.5$	<i>frac</i>
Critical Water Saturation (SWCR)	$1.1 \cdot SWL$	<i>frac</i>
Initial Water Saturations (SWATINIT)	$-0.18 \cdot \ln(J) + 0.57$	<i>frac</i>
Scaled Critical Oil-in-Water Saturations (SOWCR)	if(SWATINIT=1, 0, 0.25)	<i>frac</i>
Maximum Water Saturation (SWU)	if(SWATINIT=1, 1, 1-SOWCR)	<i>frac</i>

4.2 Set up for structural uncertainty case study (*STRUCTURAL*)

One of the key features of graph neural networks is their ability to work with unstructured data. In the case of geological modelling, this can be used to simulate structural uncertainty. The literature review has shown that dealing with structural uncertainty is quite a complex process and generally does not allow for the simultaneous consideration of both property and structural uncertainty. However, structural uncertainty is often crucial for accurate flow modelling, as structural unconformity directly affects the dynamic response, disrupting reservoir connectivity.

Thus, the second dataset should help test the hypothesis that GWAE can generate geological models considering structural and geological uncertainty as a unified process.

Figure 4.7 illustrates the concept of a *STRUCTURAL* dataset. As before, the geological foundation is represented by simplified implementations of fluvial deposits with the addition of a fault that divides the entire model into two blocks. The dataset has the following key features: the shape and spatial location of the channel (always one), the fault's spatial position and the blocks' vertical displacement relative to each other along the fault. The complete set of uncertainty embedded in the dataset is presented in the following table 4.3.

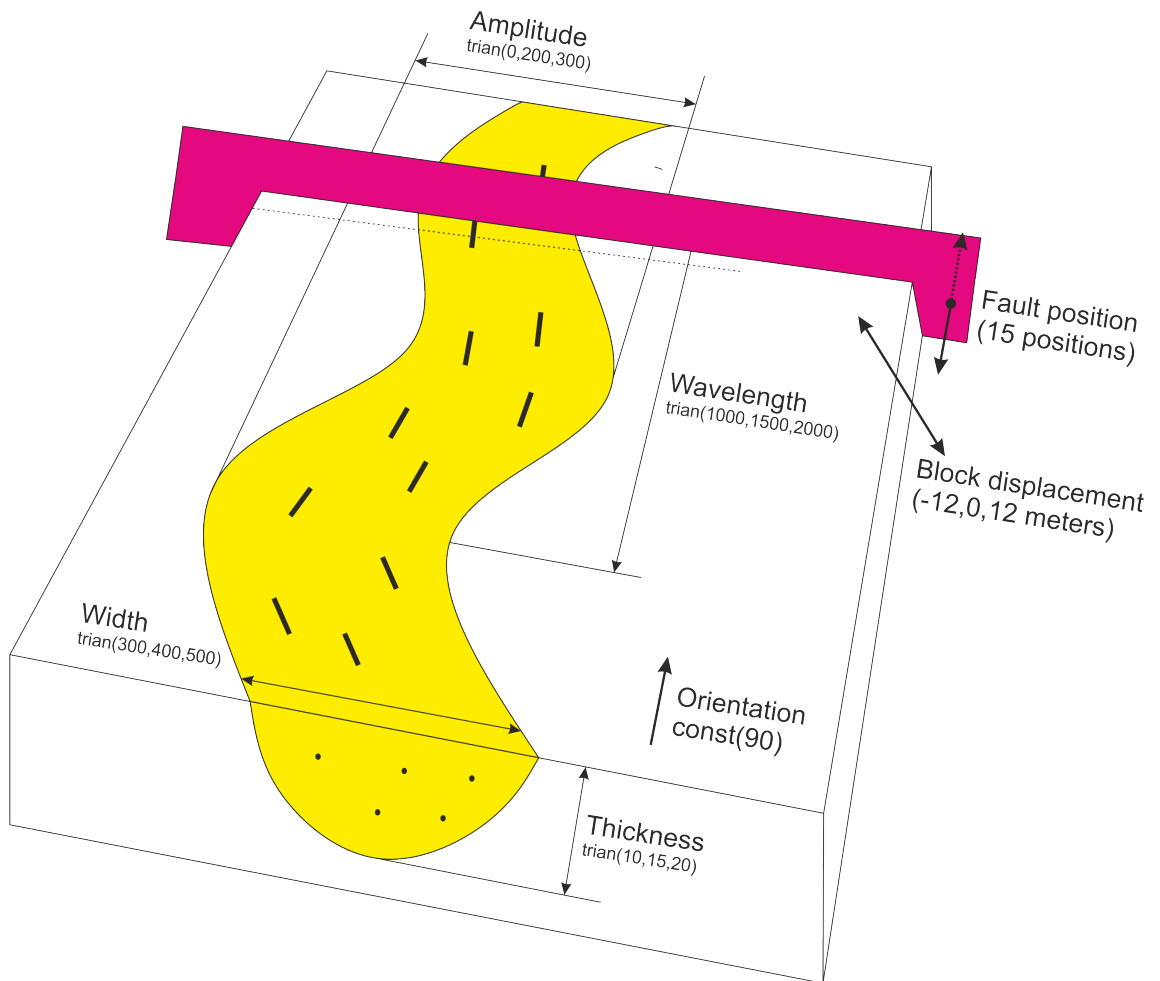


Figure 4.7: Settings of the *STRUCTURAL* synthetic dataset. The Structural uncertainty of the fault and block displacement are considered

Table 4.3: Table of parameters for *STRUCTURAL* dataset construction

Scenario	Structural
Number of channels	one
Width	[300; 500]
Thickness	[10; 20]
Wavelength	[1000; 2000]
Amplitude	[0; 300]
Orientation	90
Fault location	15 discrete positions
Block displacement	[-12; 12]

The variability in the position of blocks relative to each other adds up to 24 meters in total, making blocks partly or entirely isolated. This provides a significant variation in the dynamic response. The following figure 4.8 shows random realisations of the *STRUCTURAL* dataset. It can be observed that the fault location, blocks displacement relative to each other and the position of the channel vary.

The statistical characteristics of porosity and permeability (figure 4.9), as well as their interdependence (figure 4.10) with uncertainty taken into account, have been preserved, just like in the *FLUVIAL* dataset. Figure 4.11 presents the petrophysical dependency of some random examples from the *STRUCTURAL* dataset.

As in the previous case with the *FLUVIAL* dataset, the Petrel software workflow (*SLBTM*), based on the object modelling method, was used for data generation. However, to generate grids in *.GRDECL format with an undefined fault location and block displacement relative to the faults, a set of fifteen workflows had to be created since this is the maximum number of positions along the x-axis that the fault can occupy. The dimension of the x-axis is 16. Each workflow has a fixed fault position, while the block displacement along the fault continuously varies. The results of workflows were exported as porosity and permeability, and cells coordinate grids in *.GRDECL format. Ten thousand such realisations were created, 665 for each scenario with fault location. The model dimension is $x = 16$, $y = 12$, $z = 10$, where $dx = dy = 50m$ and $dz = 1m$.

The flow calculations were performed using a model from the *FLUVIAL* dataset with a line drive well placement scheme, and the calculations were conducted over 60 time intervals.

In summary, this subsection aims to create a dataset for training a generative network that considers uncertainty in the spatial distribution of geological bodies, petrophysical dependencies and structural uncertainty, such as fault location and vertical block displacement.

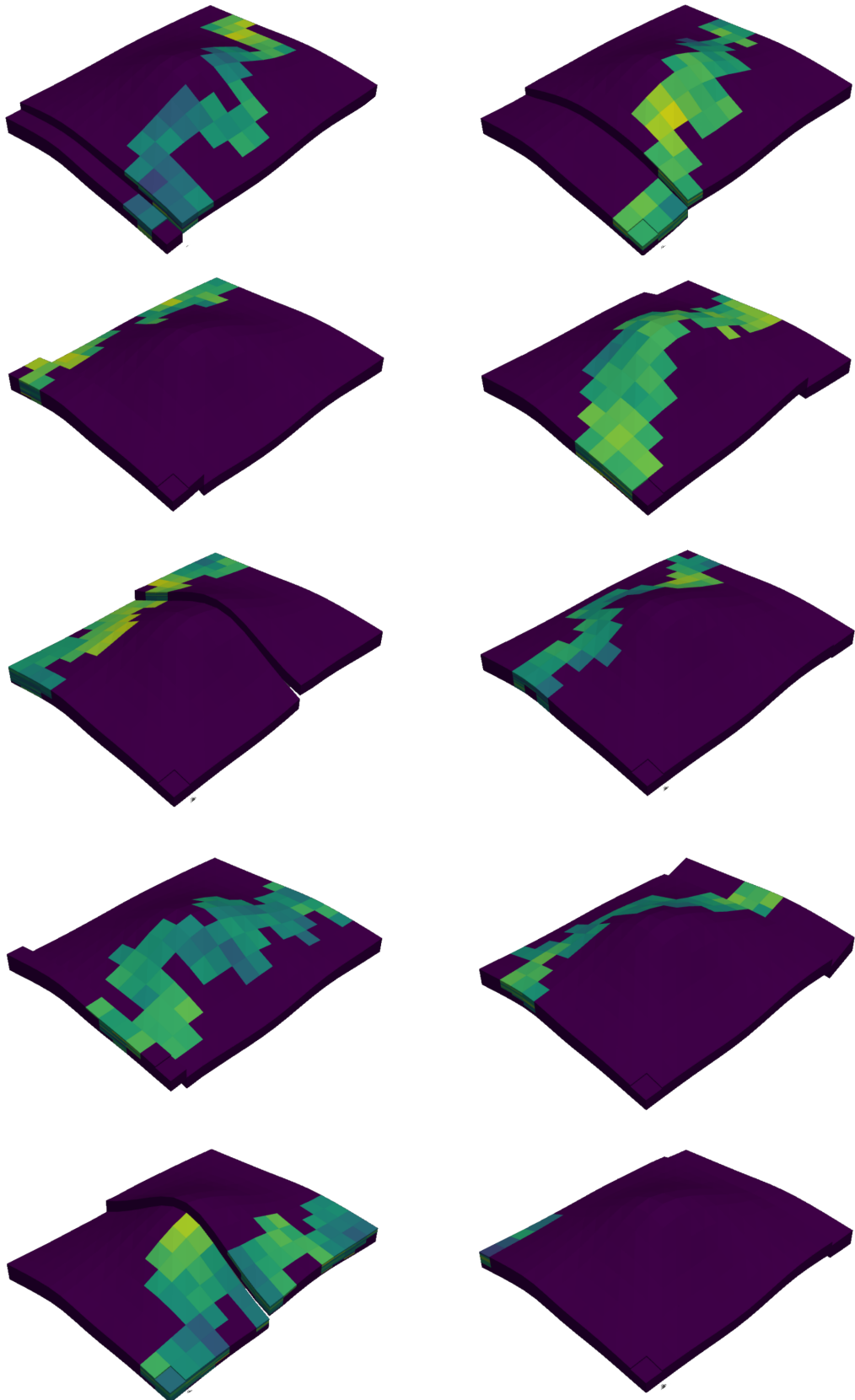


Figure 4.8: Examples of *STRUCTURAL* synthetic dataset.

4.2. SET UP FOR STRUCTURAL UNCERTAINTY CASE STUDY (*STRUCTURAL*)

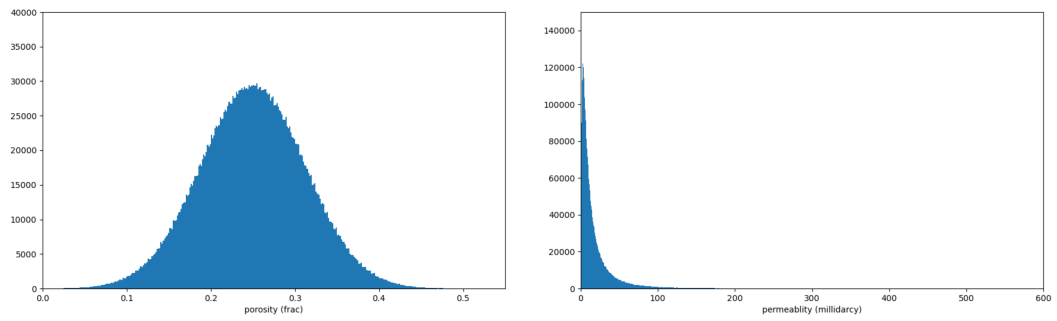


Figure 4.9: Porosity and permeability distributions for the entire dataset. Right: porosity, left: permeability.

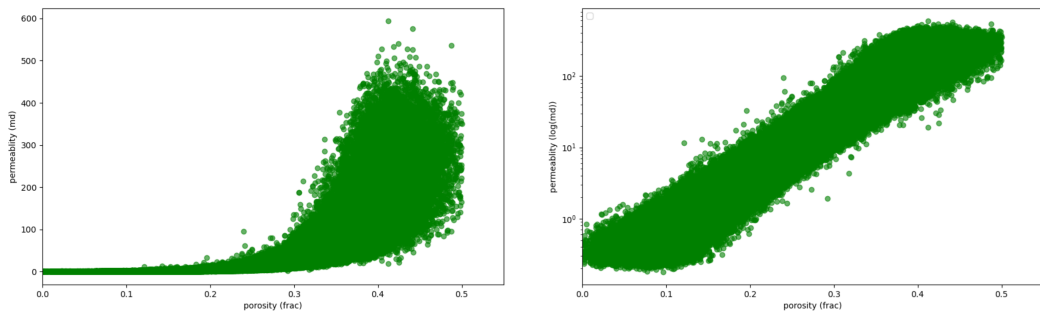


Figure 4.10: Porosity/Permeability petrophysical relationship for the channel facie for the entire dataset. Right: y-axis in millidarcy, left: the same relationship, but the y-axis in log(millidarcy).

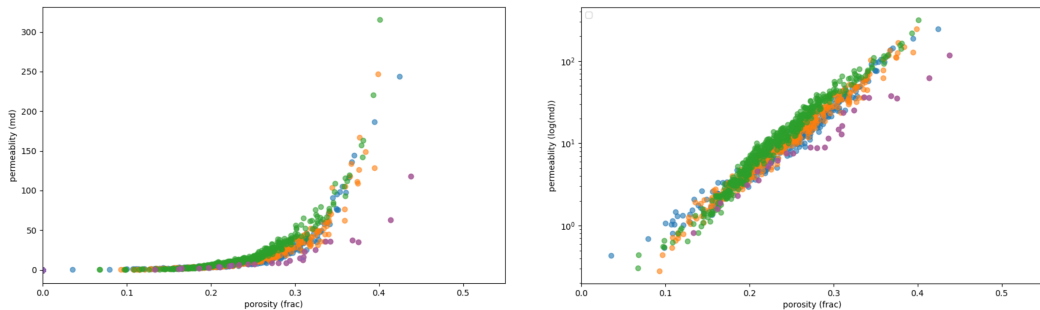


Figure 4.11: Porosity/Permeability petrophysical relationship for the channel facie for some random samples. Right: y-axis in millidarcy, left: the same relationship, but the y-axis in log(millidarcy).

4.3 Set up for Brugge field case study

4.3.1 Brugge benchmark

The Brugge field is a semi-synthetic oil reservoir model used for benchmarking and research in reservoir engineering and geoscience. The model is used to develop and test optimisation and history matching methods for oil reservoir management. It was chosen as the basis for creating a training dataset for GWAE. The three-dimensional model of the semi-synthetic field was presented as a set of 104 realisations (TNO) based on well-log curves with fixed locations at the SPE-ATW workshop in Brugge by Peters et al. (2010). The field consists

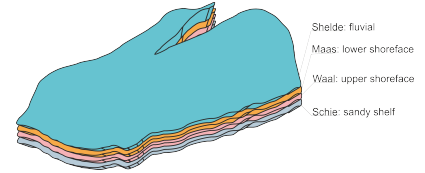


Figure 4.12: Structure of the Brugge field. Each colour represents a zone with distinct geological settings. Schelde ●, Maas ●, Waal ●, Schie ●

of 30 vertical wells, with 20 functioning as producers and 10 as injectors. From a geological perspective, the field is a half-dome structure elongated in a west-east direction. A significant fault is observed in the northern part of the structure. The size of the field is approximately $10 \cdot 3$ km, figure 4.12.

The Brugge benchmark is a platform for testing approaches to HM and development optimisation. One such example is the study by Y. Chen et al. (2010). The paper applies ensemble-based closed-loop optimisation to the Brugge field to maximise the net present value. In a later article Torrado et al. (2015) proposed a methodology which provides an efficient and accurate means of ranking greenfield portfolios in the presence of geological uncertainty. By characterising uncertainty in sedimentary variability and flow behaviour through representative geological realisations, decision-makers can rapidly evaluate multiple field-development plans. Examples of using the Brugge benchmark for testing approaches of HM include various works, such as Chalub et al. (2023), Cruz et al. (2022), Demyanov et al. (2015), and Mohamed et al. (2010), since the complexity of the field opens up opportunities for various research and comparisons.

From a stratigraphic point of view, Brugge consists of four zones (Schelde, Maas, Waal, Schie) similar to Brent-type North Sea deposits. However, it should be noted that the stratigraphic sequence of zones was intentionally disrupted to increase the model's complexity. The conceptual model of the deposit is presented in figure 4.13. More detailed characteristics of each zone are presented in the table 4.4

Based on the petrophysical properties of each zone, it can be concluded that a greater volume of production is expected from the Waal zone as it is the thickest layer with good flow properties. The Schelde zone is less favourable for production than Waal, as it consists of high-permeability channel bodies included in shale deposits with low permeability, mak-

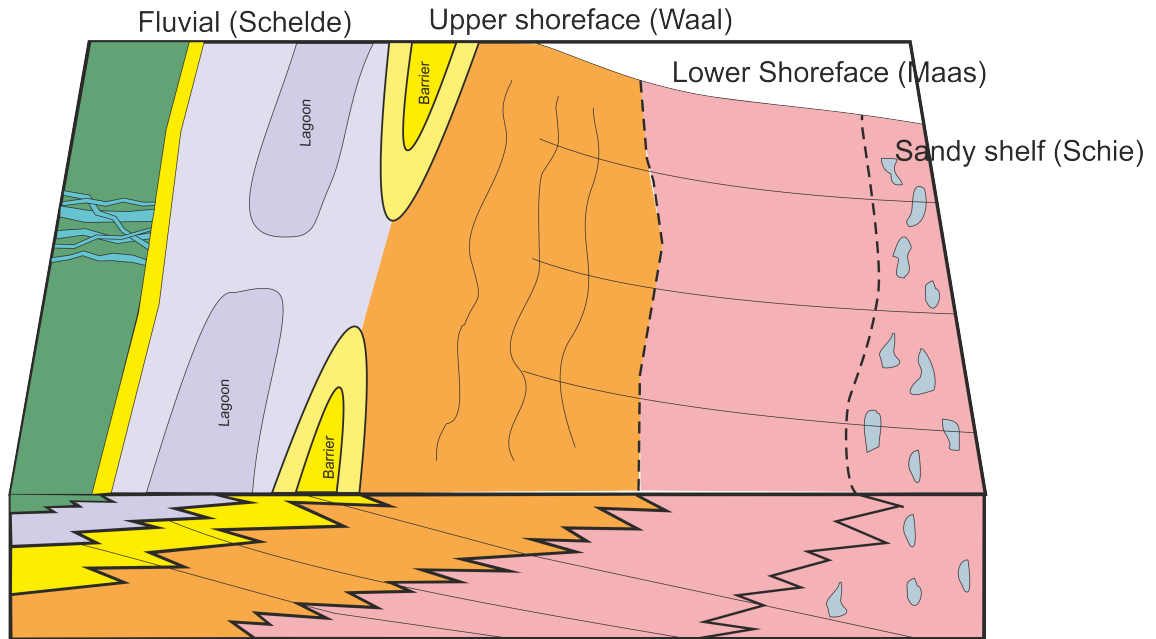


Figure 4.13: Conceptual stratigraphic representation of the Brugge field. Note that the true sequence of stratigraphic zones should be Fluvial (Schelde ●), Upper shoreface (Waal ●), Lower shoreface (Maas ●) and Sandy shelf (Schie ●), but it was intentionally swapped to Fluvial (Schelde ●), Lower shoreface (Maas ●), Upper shoreface (Waal ●) and Sandy shelf (Schie ●).

Table 4.4: Main characteristics of each zone

Zone	Average thickness, m	Average porosity, %	Average permeability, mD	Depositional environment
Schelde	10	21	1100	Fluvial
Maas	20	19	90	Lower Shoreface
Waal	26	24	814	Upper Shoreface
Schie	5	19	36	Sandy Shelf

ing this zone more heterogeneous. Maas permeability is orders of magnitude lower than Waal and contains carbonate concretions.

The Schie zone is the least favourable for production and is of little interest as it has a small thickness and poor flow properties due to irregular carbonate patches. Based on the preliminary flow calculations on the TNO and *BRUGGE* datasets, the distribution of total fluid production proportions is as follows: Schelde — 18%, Maas — 6%, Waal — 78%, Schie — less than 1%.

When compiling the dataset of 104 realisations, the authors (Peters et al., 2010) followed the following principles, which become a base for the algorithm of generating a dataset for training GWAE:

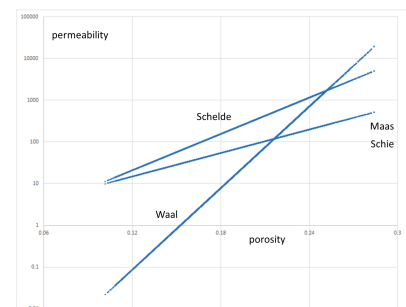


Figure 4.14: Relation of porosity and permeability for each zone of the field

- The facies distribution within each zone with corresponding porosity and permeability characteristics. It is worth noting that for the Schelde zone, representing a fluvial environment, two geological modelling algorithms were used: Object-based modelling and Sequential Indicator Simulation. The other zones were modelled exclusively using the SIS method with different settings.
- The porosity parameter was stochastically modelled using SGS for each zone.
- The permeability parameter was determined with different scenarios: deterministically depending on the porosity (figure 4.14), deterministically for every facie or stochastically with co-Kriging on porosity.

Overall, these principles were followed to ensure that each realisation in the dataset captures the spatial variability and heterogeneity of the Brugge deposit, allowing for a realistic representation of the reservoir conditions.

4.3.2 *BRUGGE* dataset construction for GWAE

Training generative neural networks requires large datasets, so 104 implementations are insufficient. Following the principles outlined in the previous subsection, a dataset of 10,000 realisations was created. Each of the four zones was modelled separately, taking into account the inherent uncertainty of each zone's parameters. The simulation model was created in accordance with the parameters mentioned in the Peters et al. (2010) and has not been modified.

Schelde

The Schelde Zone has a fluvial depositional environment characterised by channel facie with favourable flow properties and shale impermeable facie. According to Peters et al. (2010), Object Modelling and SIS algorithms of geological modelling were used to generate facies of training dataset and SGS for porosity-permeability. The input data for this process were well data. Figure 4.15 shows histograms of porosity distributions for the well data, TNO dataset, and GWAE training dataset. It can be observed that the statistical distribution of porosity in the GWAE dataset is in good agreement with the well data. In contrast, the TNO dataset has a narrower range of distribution. The right part of figure 4.15 compares porosity and permeability dependency. As can be seen, the dependency differs slightly because the initial porosity distributions are different. However, the differences are insignificant, so no major changes are expected in flow modelling.

The table 4.5 displays uncertainty parameters for the Schelde zone when using the Object Modelling and SIS algorithms.

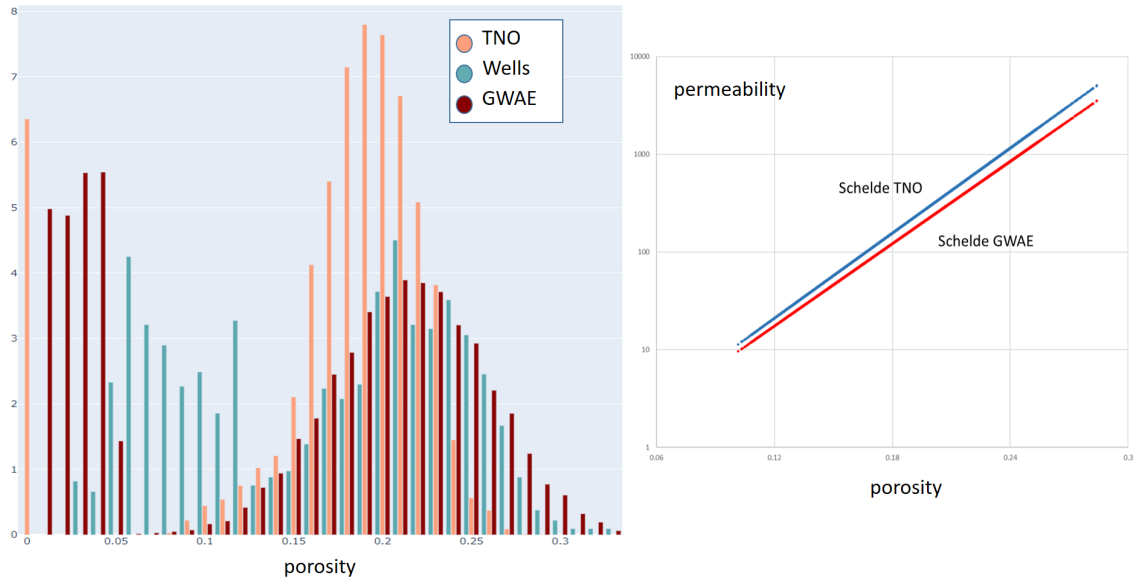


Figure 4.15: Schelde zone. Left: comparison of porosity distributions for initial well data \circ , TNO (104 realisations dataset) \bullet , and GWAE training dataset \bullet . Right: comparison of porosity-permeability dependency of TNO dataset and GWAE training dataset.

Table 4.5: Ranges of uncertainty for Schelde

Modelling Approach	Object Modelling
Channel Width	[50; 550]
Channel Thickness	[5; 20]
Orientation	[200; 300]
Amplitude	[200; 1600]
Wavelength	[250; 2500]
Variogram Major for porosity	[1200; 1700]
Variogram Minor for porosity	[600; 1200]
Variogram Vertical for porosity	[1; 5]
Modelling Approach	SIS
Variogram Major for channel facie	[350; 550]
Variogram Minor for channel facie	[50; 250]
Variogram Vertical for channel facie	[5; 20]
Variogram Major for porosity	[1200; 1700]
Variogram Minor for porosity	[600; 1200]
Variogram Vertical for porosity	[1; 5]

The following figure 4.16 shows random realisations from the training dataset for GWAE of the Schelde zone for the Object Modelling method (left) and SIS (right).

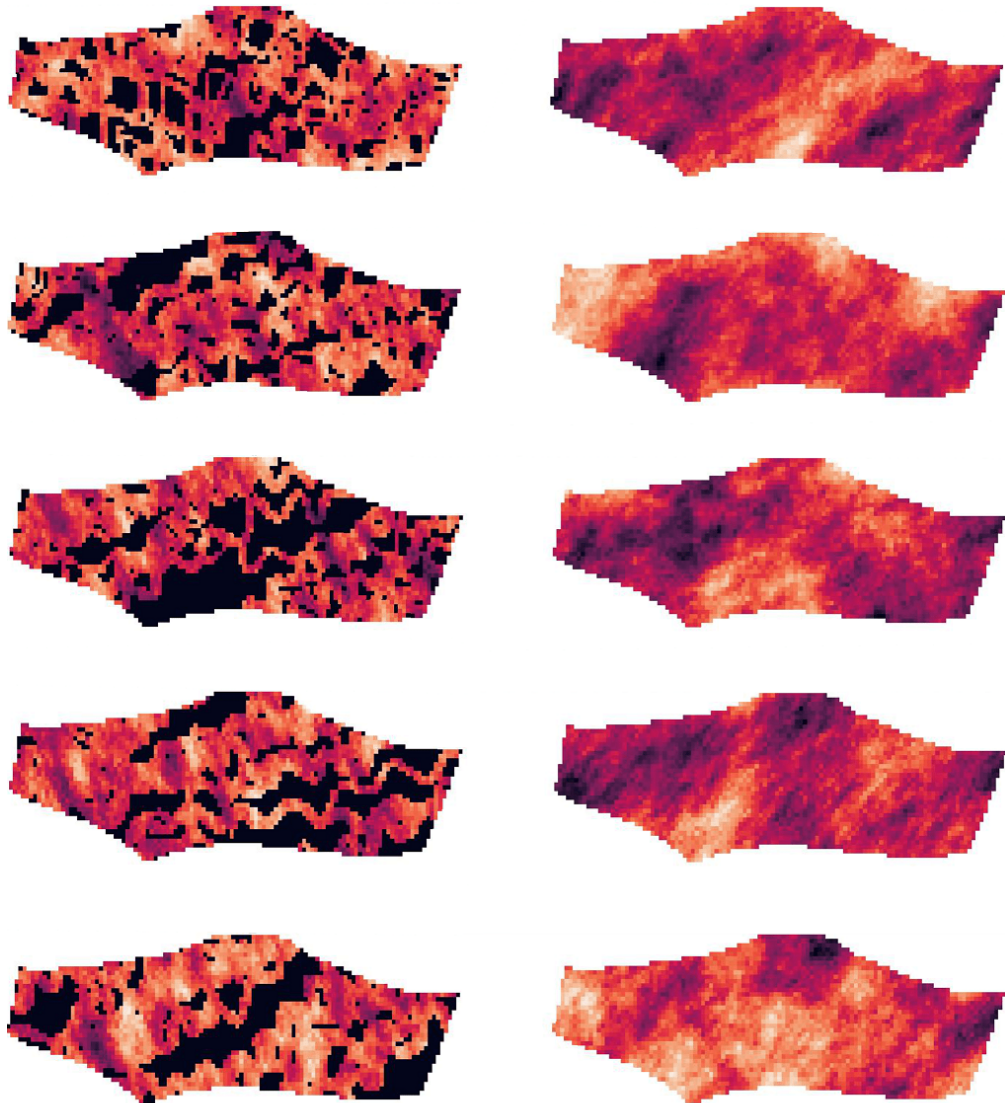


Figure 4.16: Examples of Schelde zone realisations for training GWAE. Left: Object Modelling, Right: SIS

Maas

The next stratigraphic unit of the Brugge field is the Maas zone, which represents the Lower Shoreface depositional setting. The zone consists of sandstone facies with good reservoir properties and carbonate concretions. The most suitable algorithm for modelling such a depositional environment is SIS using well data. The left side of figure 4.17 compares porosity distributions for well data, the TNO dataset, and the training dataset for GWAE. Similar to the Schelde case, the range of porosity parameter for TNO is tighter than that of well data. The right side indicates the dependency of porosity and permeability for TNO and GWAE datasets, which are slightly different. The table 4.6 displays

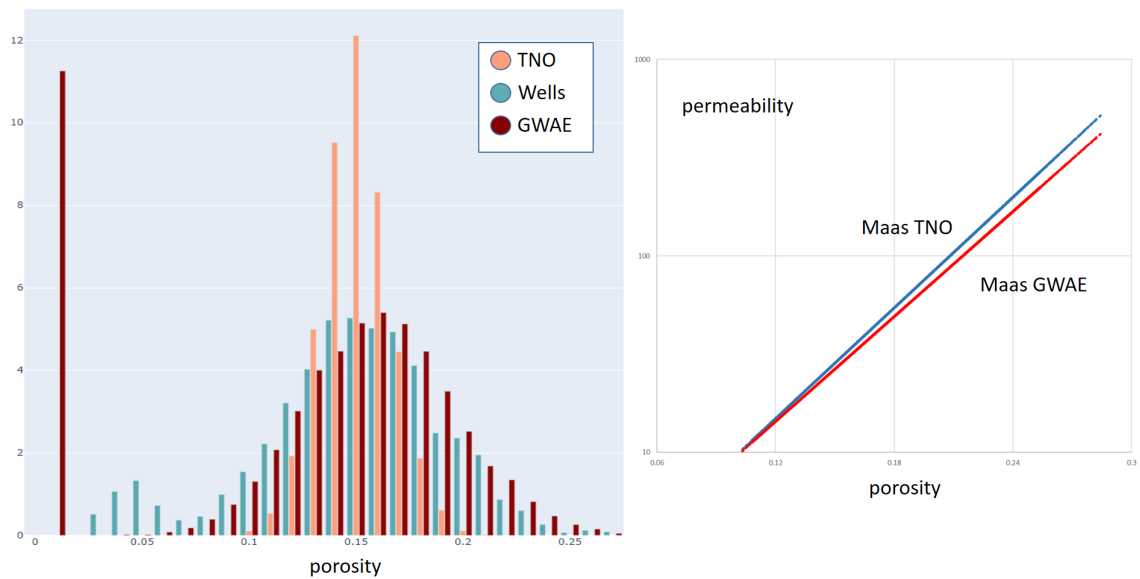


Figure 4.17: Maas zone. Left: comparison of porosity distributions for initial well data ●, TNO ● (104 realisations dataset), and GWAE training dataset ●. Right: comparison of porosity-permeability dependency of TNO dataset and GWAE training dataset.

parameters of uncertainty for the Maas zone when using the SIS algorithms

Table 4.6: Ranges of uncertainty for Maas

Modelling Approach	SIS
Variogram Major for sand facie	[500; 1500]
Variogram Minor for sand facie	[400; 1050]
Variogram Vertical for sand facie	[1; 6]
Variogram Major for porosity	[1500; 2600]
Variogram Minor for porosity	[500; 2000]
Variogram Vertical for porosity	[1; 4]

Next figure 4.18 shows random realisations from the training dataset for GWAE of the Maas zone for the SIS method.

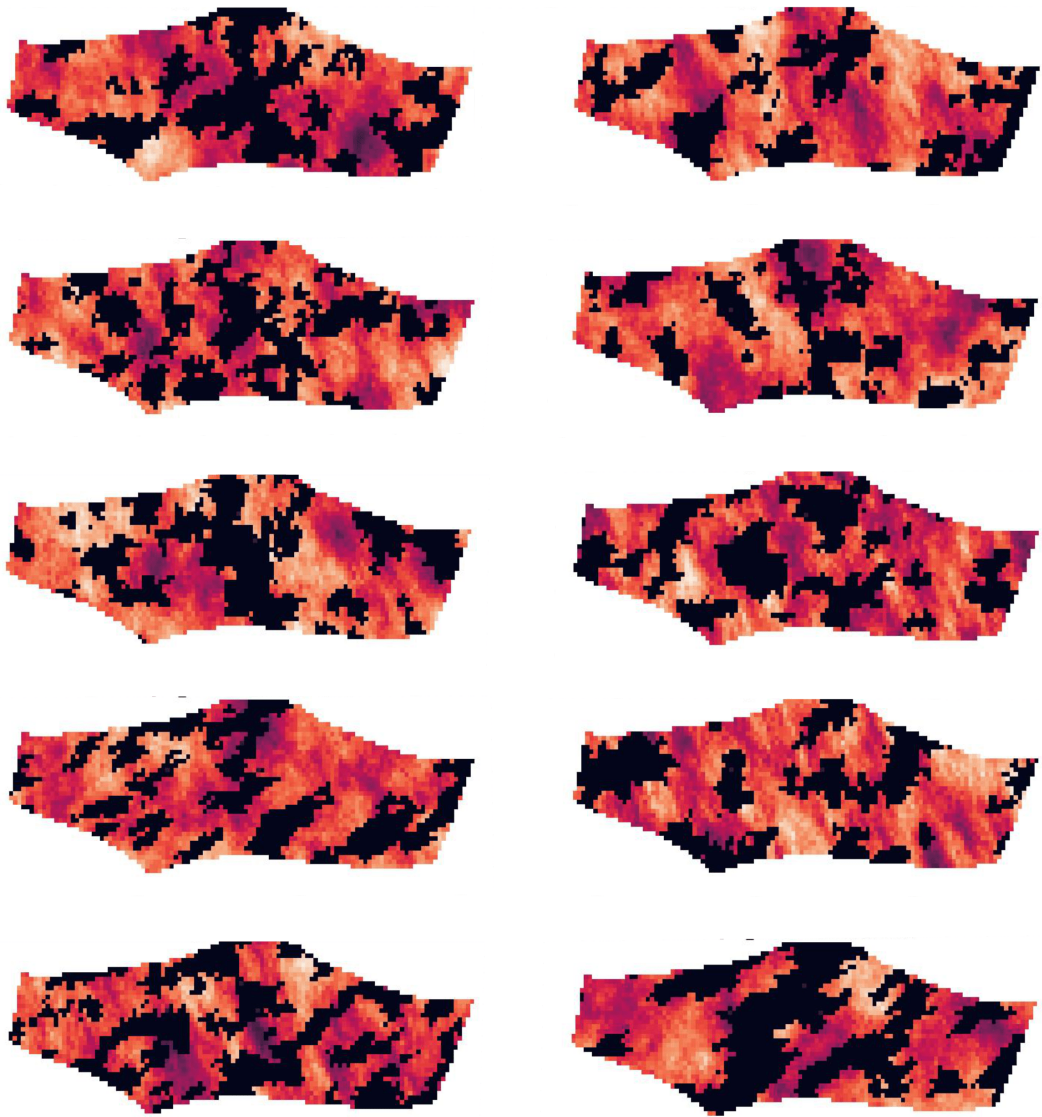


Figure 4.18: Examples of Maas zone realisations for training GWAE.

Waal

The third stratigraphic unit has an Upper Shoreface depositional environment. It is characterised by good filtration and storage properties and is the best zone for development. This reservoir has a high thickness with an NTG up to 97 per cent, and the average permeability is about 800mD. When modelling such a reservoir, the most suitable algorithm is SIS. Figure 4.19 demonstrates the main petrophysical characteristics of the Waal zone.

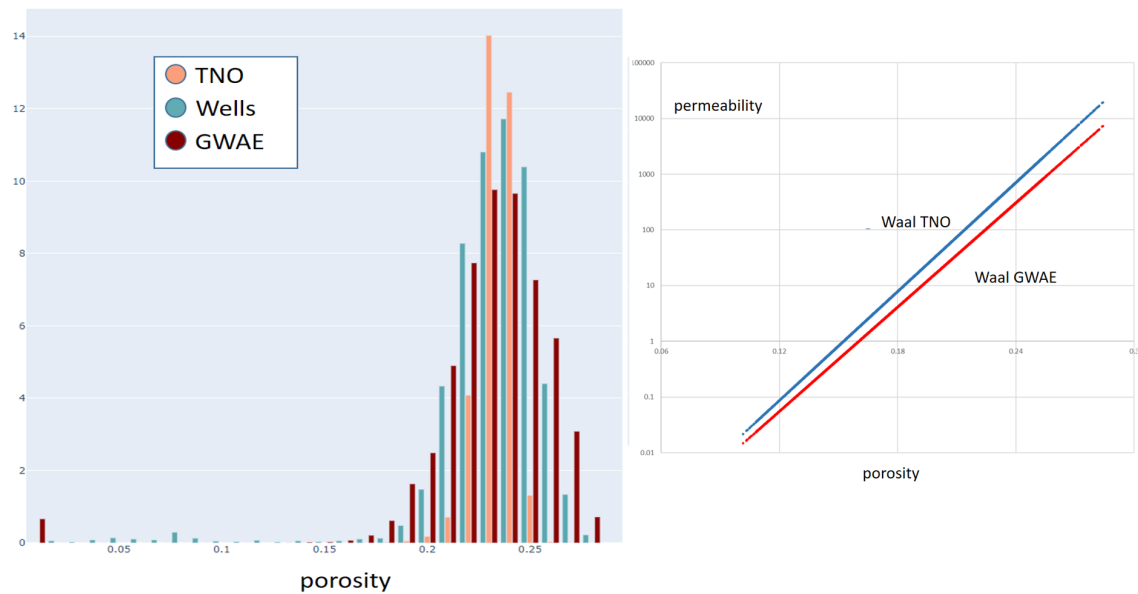


Figure 4.19: Waal zone. Left: comparison of porosity distributions for initial well data \bullet , TNO (104 realisations dataset) \bullet , and GWAE training dataset \bullet . Right: comparison of porosity-permeability dependency of TNO dataset and GWAE training dataset.

The next table 4.7 provides uncertainty, mostly for variograms of facies and petrophysics. Figure 4.20 shows random realisations from the training dataset.

Table 4.7: Ranges of uncertainty for Waal

Modelling Approach	SIS
Variogram Major for sand facie	[1000; 1700]
Variogram Minor for sand facie	[700; 900]
Variogram Vertical for sand facie	[20; 45]
Variogram Major for porosity	[1000; 3000]
Variogram Minor for porosity	[500; 1500]
Variogram Vertical for porosity	[1; 6]

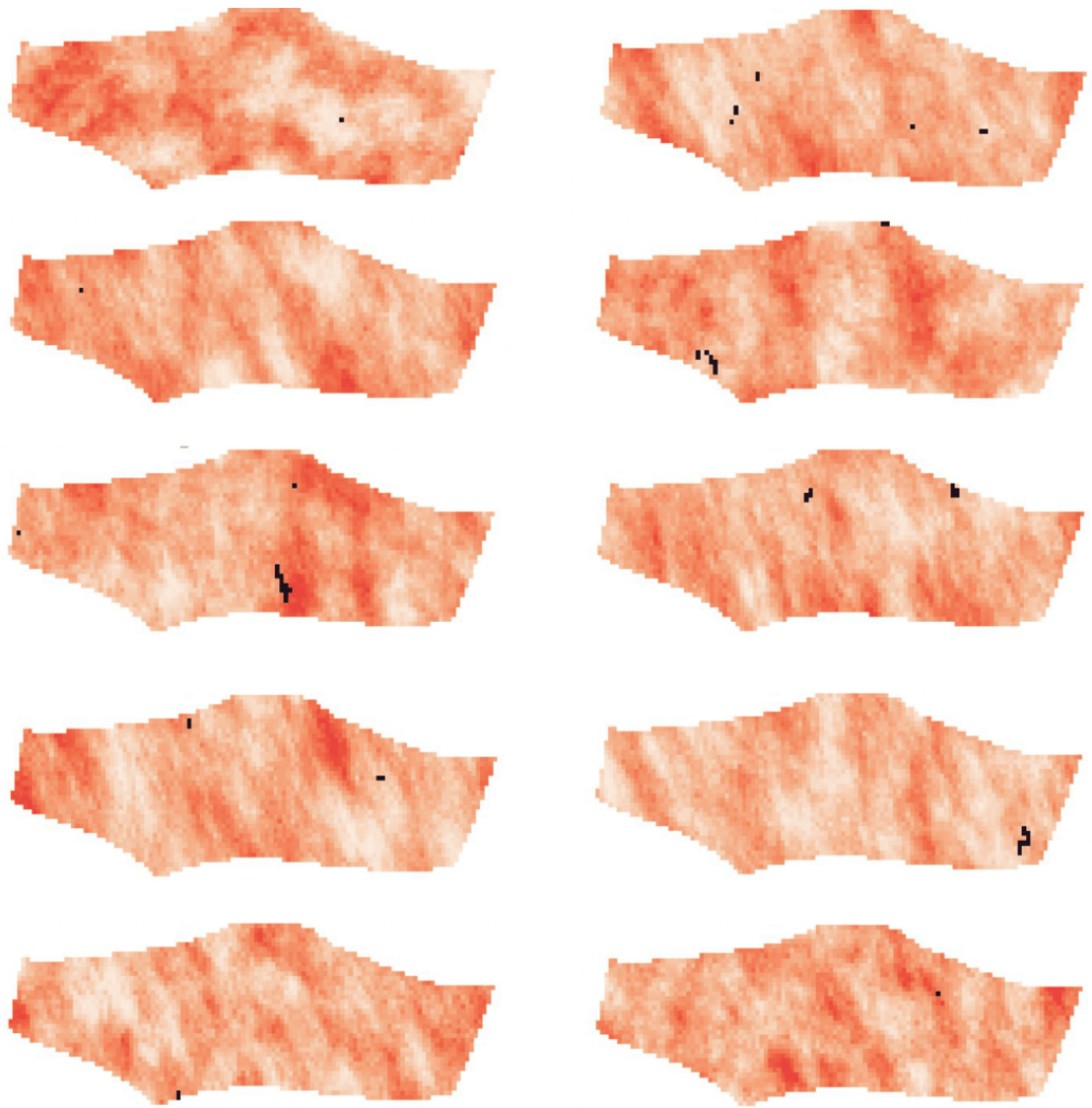


Figure 4.20: Examples of Waal zone realisations for training GWAE.

Schie

The last lower zone of the Brugge field is composed of the Sandy Shelf sedimentary accumulation. This is a relatively thin layer with poor development parameters. The main feature of this zone is the irregular carbonate patches. Figure 4.21 presents the primary petrophysical characteristics. Table 4.8 lists the uncertainty parameters that were included in the creation of the training dataset. Figure 4.21 represents random realisations of the training dataset.

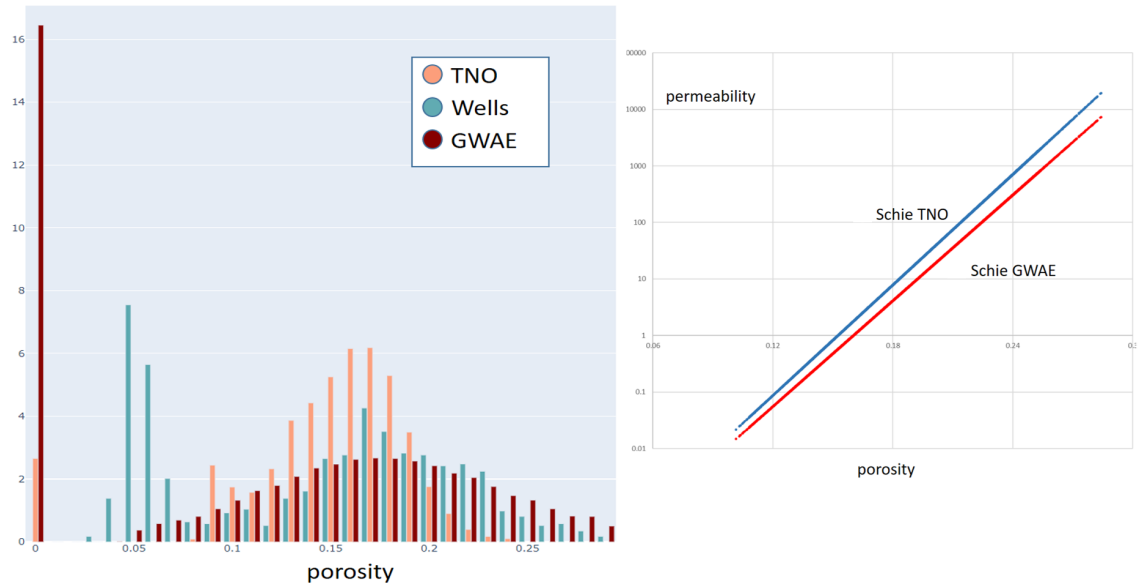


Figure 4.21: Schie zone. Left: comparison of porosity distributions for initial well data ●, TNO (104 realisations dataset) ●, and GWAE training dataset ●. Right: comparison of porosity-permeability dependency of TNO dataset and GWAE training dataset.

Table 4.8: Ranges of uncertainty for Schie

Modelling Approach	SIS
Variogram Major for sand facie	[400; 1500]
Variogram Minor for sand facie	[200; 800]
Variogram Vertical for sand facie	[1; 6]
Variogram Major for porosity	[600; 1400]
Variogram Minor for porosity	[600; 1200]
Variogram Vertical for porosity	[1; 6]

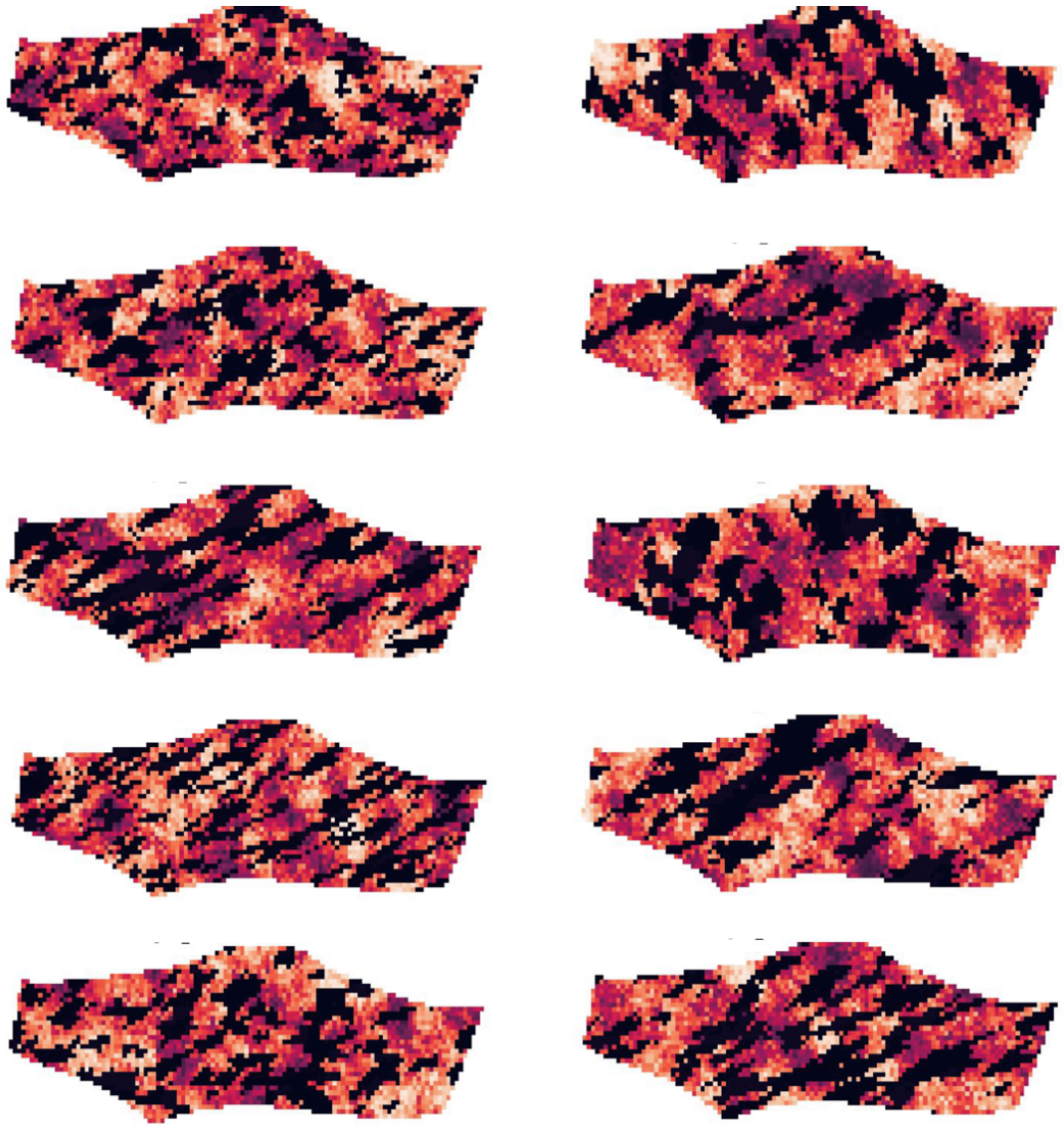


Figure 4.22: Examples of Schie zone realisations for training GWAE.

Summarising this subsection, it's worth noting that the prior dataset of 10,000 Brugge field realisations was created to train a graph generator. Each zone within the field has distinctive features, allowing the evaluation of the generator's capabilities not only on simple datasets like *FLUVIAL* and *STRUCTURAL* but also on a larger dataset with more diverse geological characteristics. It's important to mention that the Schelde zone was modelled using two different algorithms (Object Modelling and SIS).

4.3.3 The process of converting corner point grid into a graph representation

All the datasets described above were created with a corner point (CP) grid representation. Therefore, the geological realisations are initially presented in this format, which needs to be converted into a graph data type for training a graph neural network. This task is not trivial because, in addition to translating the properties of each grid cell into graph nodes, it is necessary to preserve the structural information. This subsection presents the algorithm of converting CP data type into a graph representation.

The algorithm ensures the conversion of the CP data format into a graph type while preserving the structural information. It involves general steps:

- Reading the CP file. The algorithm starts by reading the CP file containing the geological realisation data.
- Creating graph nodes. The algorithm creates a corresponding node in the graph for each cell in the grid. The properties of each cell in the CP file are transferred to the attributes of the graph node. If a cell has the property $ACTNUM = 0$, it indicates that it is irrelevant to the model. In the context of converting data into a graph representation, cells with this property are excluded or removed from the graph, indicating that they are not considered in the conversion process.
- Defining graph edges. The algorithm establishes edges between the graph nodes based on the connectivity of the cells in the grid. Thus, if a face of one cell is adjacent to two or more faces of other cells, then a graph node will contain the corresponding number of edges. The structural information is preserved by connecting neighbouring nodes according to the grid topology.
- Mapping property values: The algorithm maps the property values from the CP file to the attributes of the graph nodes. This step ensures that the information from each cell in the grid is properly transferred to the corresponding node in the graph.

By following these steps, the algorithm converts the CP data format into a graph representation that can be used to train a graph neural network. The most complex and time-consuming step in the transformation process is determining the edges of the graph based on the adjacency of cells in the initial grid. This enables the utilisation of structural information and cell properties for more accurate predictions and analysis.

Each cell is a hexagon (📐) and has eight vertices, each characterised by three coordinates (x, y, z) . The vertices' coordinates are determined through the z coordinate of the vertex from the $zcorn$ section of the CP file and guide pillars (J), passing through these vertices defined by two points $(x_0, y_0, z_0); (x_2, y_2, z_2)$ in the $coord$ section (see figure 4.23). The

formula for recalculating the $x_1^?, y_1^?, z_1$ coordinates of vertices as follows:

$$x_1^? = x_2 + (x_0 - x_2) \cdot \frac{z_1 - z_2}{z_0 - z_2} \quad (4.1)$$

$$y_1^? = y_2 + (y_0 - y_2) \cdot \frac{z_1 - z_2}{z_0 - z_2} \quad (4.2)$$

We convert each active cell into a set of six labelled faces (□). As a result, we obtain a dictionary where the key is the sequential index of the cell, and the values are a list of faces.

After that, it is necessary to find cells whose faces have a common intersection area. For a target t cell:

- 0 face of the t cell is checked for intersection with the 1 face of the neighbours below. In this case, the neighbours below are cells with a $k = k_{init} - 1$ coordinate;
- 1 face of the t cell is checked for intersection with the 0 face of the neighbours above. In this case, the neighbours above are cells with a $k = k_{init} + 1$ coordinate;
- 2 face of the t cell is checked for intersection with the 3 face of the neighbours on the left. In this case, the neighbours on the left are cells with a $i = i_{init} - 1$ coordinate;
- 3 face of the t cell is checked for intersection with the 2 face of the neighbours on the right. In this case, the neighbours on the right are cells with a $i = i_{init} + 1$ coordinate;
- 4 face of the t cell is checked for intersection with the 5 face of the neighbours in front. In this case, the neighbours in front are cells with a $j = j_{init} - 1$ coordinate;
- 5 face of the t cell is checked for intersection with the 4 face of the neighbours behind. In this case, the neighbours behind are cells with a $j = j_{init} + 1$ coordinate.

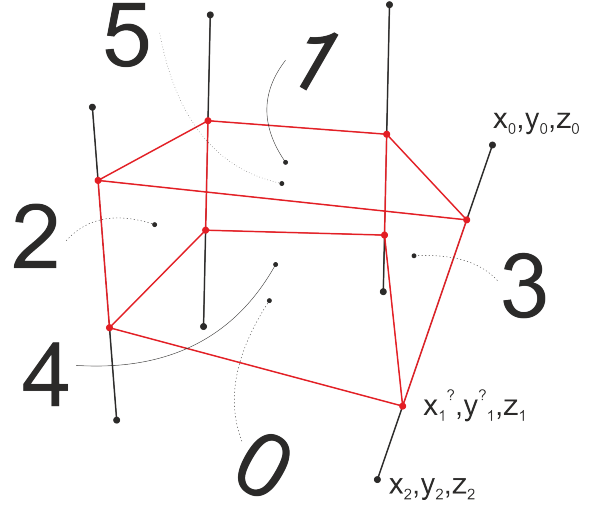


Figure 4.23: CP format cell geometry notation, where cell (□) is characterised by x, y, z coordinates. z is usually set in the $zcorn$ section and x, y can be calculated from pillars (l) and z coordinate for every vertice.

If the intersection is a plane, the cells are considered neighbours, and an edge appears between the corresponding nodes in the graph. To speed up the calculation of the adjacency search algorithm, it is assumed that if the ratio of intersection areas of neighbouring

cells is close to one, then for such cells, there cannot be other neighbours concerning the considered faces. Thus, such faces are excluded from further search.

Thus, after traversing all active cells ($ACTNUM = 1$) of the grid, we obtain data in a graph representation that contains information about the internal topology of the object and the spatial distribution of reservoir properties, such as porosity and permeability, figure 4.24. This is more suitable for geological modelling because it avoids non-necessary cells in lattice and only needs to make a lattice of proper structure. Moreover, unstructured grids are suitable for modelling under structural uncertainty. So, if the gridded geological model had a high proportion of zero-valued cells, it would entail the neural network to train on these cells, which is undesirable.

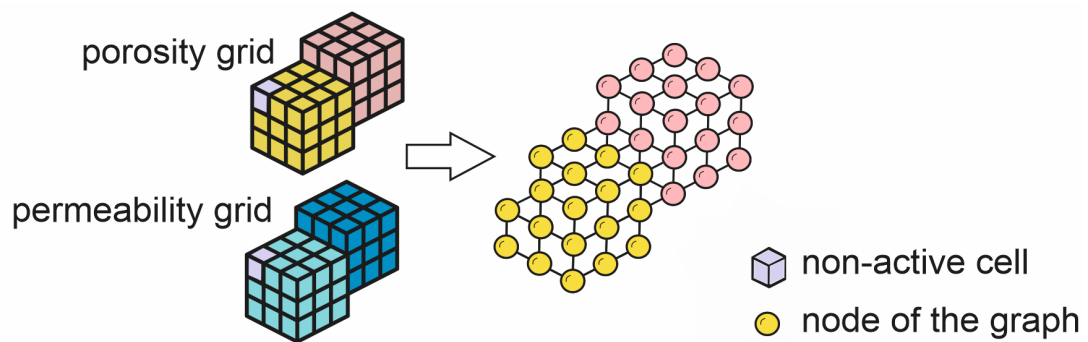


Figure 4.24: Visual representation of translating a grid model into a graph format. The properties of porosity (■, ■) and permeability (■, ■) of each cell of the grid are translated into a property vector that characterises each node of the graph (●, ●). The presence of an edge between nodes of the graph indicates the adjacency of cells in the grid. Inactive cells (■) of the grid are excluded from the set of nodes of the graph.

It should be noted that this subsection gives a scenario where the original geological representation is implemented as a structured grid. However, there are more suitable representations that avoid unnecessary conversion from grid to graph. One such representation is the connection list (Khait et al., 2019). Connection list allows to discretise a space by representing connectivity between control volumes for structured and unstructured representations, only for active cells.

4.4 Conclusion

The process of geological modelling using generative neural networks begins with the creation of a prior ensemble of subsurface models for training based on existing information. These models are the foundation for generating essential data and prediction variables through various techniques.

This section describes the construction principles of *three* training prior datasets, which

will facilitate the subsequent training of GWAE for geological modelling and testing the previously discussed technologies. The three distinct training datasets were created for this purposes:

- i - contains three-dimensional models of synthetic porosity and permeability property cubes of the fluvial geological environment (*FLUVIAL*), with the primary uncertainty centred around channel locations and their parameters;
- ii - focuses on structural uncertainty in reservoir setting, particularly the placement of a fault and block displacements combined with model (i) setting (*STRUCTURAL*);
- iii - encompasses multiple realisations of the synthetic full-field Brugge benchmark reservoir, comprising four zones with different depositional conditions and the variation of geological parameters (*BRUGGE*)

These datasets will be crucial in advancing our understanding and application of generative neural networks in geological modelling.

During further experiments involving training GWAE and testing the network's generative capabilities, we will refer to each of the provided datasets to test various hypotheses and draw conclusions about the applicability of the technology described in the **Chapter 3: Methodology**.

Chapter 5

Graph Wasserstein Autoencoders and channelized synthetic case

In this chapter, I will present an approach to generate reservoir models conditioned on production data and provide uncertainty estimation using the Graph Wasserstein Variational Autoencoder. This approach implicitly parameterises geological representations into a lower-dimensional latent space and allows for uncertainty quantification and production profiling across multiple geological concepts. By applying this approach to the synthetic *FLUVIAL* dataset, I will demonstrate that GWAE reliably reproduces a reservoir by comparing the static and dynamic properties of the reference model with generated representations. In addition, I will provide an in-depth analysis of the structure of the LS of the generative model using tools such as PCA, t-SNE, and TDA. In the second part of the chapter, I will compare the generative capabilities of VAEs, WAEs, and GWAEs. The comparison highlights the limitations of CNN-based generative networks in dealing with structural uncertainty and significant reservoir/non-reservoir imbalances. Graph-based approaches, such as GWAEs, are better suited to adapt to structural uncertainty by removing unnecessary non-reservoir nodes and breaking the regularity of the original grid.

I will introduce the VAE-based workflow to map reservoir description into latent space that depicts associated geological uncertainty. To solve the inverse problem, it is necessary to merge measurements with existing knowledge of model parameters and physical correlation patterns between the model and data. The process involves three significant stages:

Stage One: Parametrisation of the physical system by defining the minimum model parameters under prior uncertainty

The first stage involves creating a prior training dataset that describes the entire space of uncertainty as comprehensively as possible. We assume that the prior dataset is represented by the *FLUVIAL*. After that, the dataset is converted into a graph representation and fed into GWAE's training process, the working principle of which was described in the **Chapter 3: Methodology**. The architecture of GWAE implies that the latent space, located at the junction of the *Encoder* and *Decoder*, is the minimal set of model parameters that can describe the object of interest, considering uncertainty and implicit dependencies between parameters, such as petrophysical ones, figure 5.1.

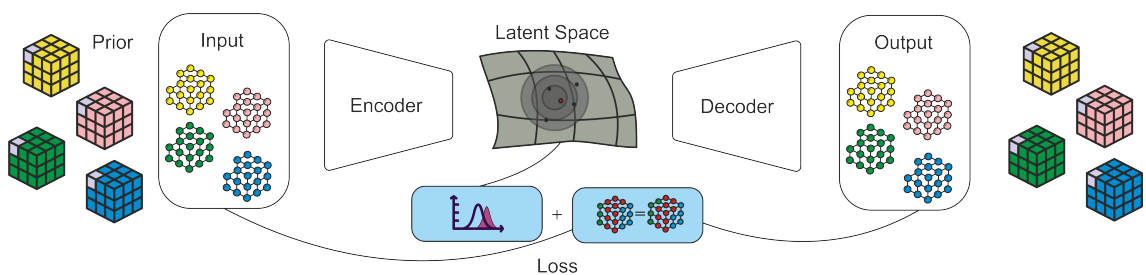


Figure 5.1: The first step of inverse modelling is to parameterise prior dataset into the LS, which are model parameters, by training GWAE.

Stage Two: Conducting forward modelling by unveiling physical laws that can help forecast measurement outcomes

After the neural network is trained and capable of reproducing instances of the training dataset and generating new instances reliably, we can exclude the *Encoder* from further consideration. It should be noted that a new instance is considered reliable if it does not contradict the geological concepts incorporated in the creation of the prior dataset, taking into account uncertainty.

For further work, we need the LS, representing the parameter space characterising the geological object under consideration. Parameter space is the compressed and abstract representation of features or properties of geological formations or structures. The *Decoder* essentially serves as a forward model, allowing us to translate the parameters into a physical representation of the object — a geological model, figure 5.2.

Stage Three: Inverse modelling to deduce the true values of the model

The final step involves searching for the location in the latent space that matches the data within the latent space. The provided decoded geological model should not contradict the available information about the field. Thus, the third stage involves an optimisation process aimed at minimising the OF error described in the **Chapter 3: Methodology**.

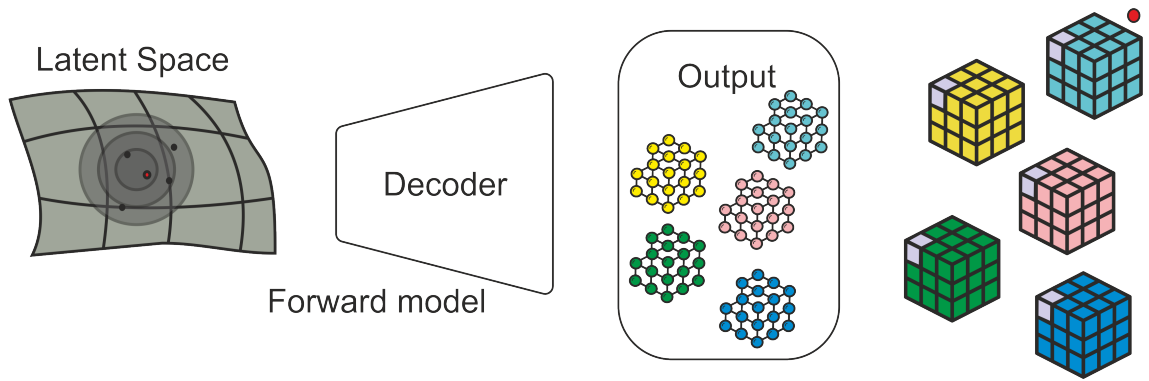


Figure 5.2: The second step of inverse modelling is to provide forward modelling through the LS and *Encoder*. Note that *Encoder* can generate samples from the initial prior dataset and new reliable examples ●

To recap, the OF consists of three components: Static loss, Flow loss, and Realism loss (**Section 3.7: Objective function in HM process**). In the case of minimising the Static loss component, the optimisation process seeks a set of parameters within the LS where the well log information of the generated reservoir closely matches the data from real well logs for each well. The Flow loss accounts for the well’s dynamic response similarity to the real measurements, meaning the well production indicators, after simulation modelling, should closely match the actual measurements. The third component controls the optimisation process by restricting the search space to dense regions within the LS, thereby enhancing the geological reliability of the generated model, figure 5.3.

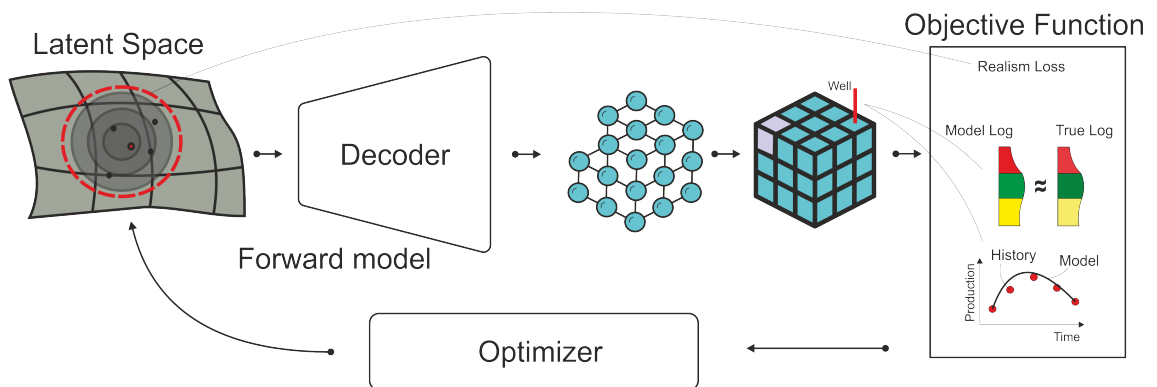


Figure 5.3: A latent vector is sampled from the LS, which is decoded to the geological representation. Three-component OF then calculated and provides information for the optimiser about the next optimisation step in the LS

This approach of solving the inverse problem through a latent space provides the following opportunities:

- Provide unique generic non-parametric description across multiple geological concepts;
- Effective model update with the reduced dimensionality of the model description in

the LS;

- Implicit link between the model update and the flow response via the LS.

This provides an opportunity to obtain multiple models and quantify uncertainty in reservoir prediction with the HM ensemble. The approach addresses the issue that the high-dimensional data originates from a lower-dimensional source that is not directly observable. This means that each data point, or sample, can be considered an indirect measurement of a lower-dimensional phenomenon responsible for the variability between the samples. This phenomenon, described as a lower-dimensional manifold, represents a geometric structure in a lower-dimensional space that generates high-dimensional data points when mapped back to the original space. The aim of dimension reduction, therefore, is to identify the key dimensions or degrees of freedom that capture the majority of the variance in the data.

5.1 Generative quality of GWAE with *FLUVIAL* dataset

When solving the inverse problem under uncertainty of various geological scenarios, a problem arises in that modelling such scenarios requires constructing a set of geological models, each of which must be described by its unique set of parameters. This, in turn, imposes constraints on the optimisation process, as it must navigate through conditions where the parameter set for each geological scenario differs. Essentially, solving the inverse problem boils down to a collection of processes, each corresponding to its own scenario. Thus, the question arises: can we determine a universal set of parameters that would allow us to consider different geological scenarios as a unified space in which we can solve the inverse problem using optimisation algorithms?

Let's consider the geological scenarios of the *FLUVIAL* dataset, which is conceptually divided into a scenario with one river channel and a scenario with two channels, with some variations in the shape (i.e. wavelength, amplitude, orientation). To conduct optimisation, we need to somehow "jump" from one scenario to another. To solve such a task, both scenarios must be translated into a single universal parametric space: the latent space of the trained GWAE.

Thus, after training GWAE on the *FLUVIAL* dataset, I obtained a unified LS parameter space that includes variants of both geological scenarios. Each sample of the training dataset is a graph containing porosity and permeability properties; every graph has a 1919 number of nodes, so the initial dimensionality is $1919 \cdot 2 = 3838$. The *Encoder* of GWAE decreased the initial dimensionality of 3838 to the dimensionality of a LS, which is 30. It is worth noting that the dimensionality of the space was chosen empirically. Several publications (B. Chen et al., 2022; Guss et al., 2018; Levina et al., 2004) provide a dimen-

sionality selection process, but this research went beyond the scope of the current work. Increasing the dimensionality of the space did not improve the quality of the generated geological realisations, which represented three-dimensional fields of porosity and permeability properties. If the dimensionality of the space is reduced below 30, the performance of the *Encoder* starts to deteriorate, indicating that the features become insufficiently informative for the correct generation of properties.

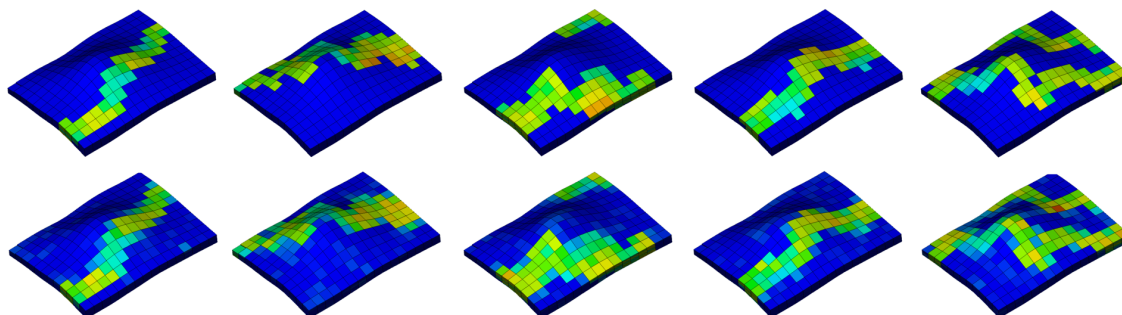


Figure 5.4: Upper row: Examples of the initial porosity grid from the training dataset. Lower row: GWAE reconstructed

Each node of the graph is characterised by porosity and permeability properties. The following figure 5.4 shows random realisations from the training dataset (top row) and those reconstructed using the *GWAE Decoder* (bottom row).

It can be concluded that the vectors of the latent space are informative enough to reconstruct the spatial dependencies of property distribution accurately. It was mentioned earlier that one of the drawbacks of generative models based on variational autoencoders is the blurriness of generated objects, which is partially mitigated by incorporating the Wasserstein metric into the loss function. This blurriness can also be observed in the reconstructed geological representations. It is worth noting that three-dimensional VAEs based on standard CNN perform worse in reconstructing property cubes, which will be presented further.

The following hypothesis regarding the generative capabilities of GWAE was that petrophysical parameter dependencies, such as porosity and permeability, could be implicitly reconstructed. This is a significant departure from standard petrophysical modelling approaches, where porosity-permeability dependencies are determined deterministically. It can be observed that the channel shapes and petrophysical properties are well preserved, figure 5.5. However, it should be noted that the permeability ranges predicted by the GWAE are more condensed compared to the true model permeability ranges concerning porosity.

Therefore, the preliminary analysis confirms the generative capability of GWAE under geological scenario uncertainty.

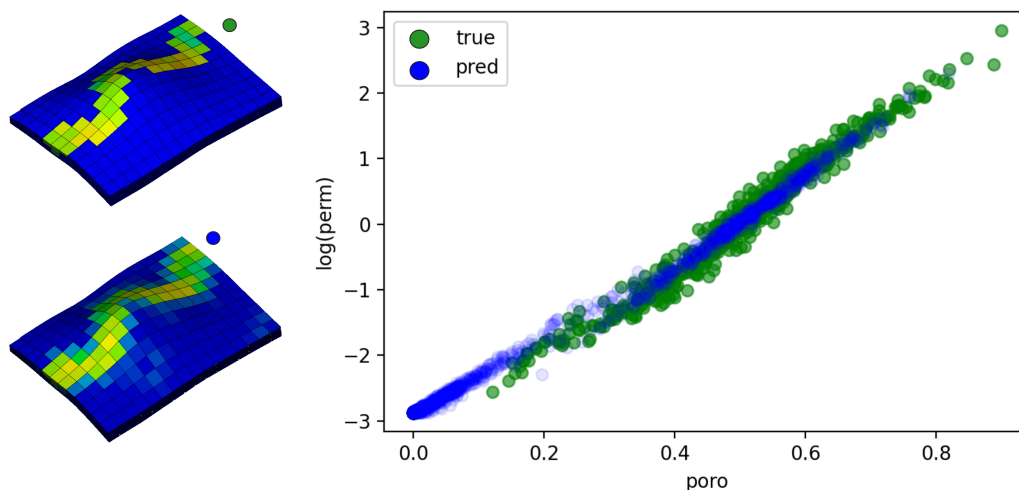


Figure 5.5: GWAE preserves petrophysical relations and spatial propagation of properties under uncertainty of geological scenarios

5.1.1 Latent Space inner structure

The next stage in the exploration of the generative capabilities of GWAE is a more in-depth analysis of the LS. Understanding its structure will help understand the technology’s advantages and limitations. But first, we need to answer the initial question – is there any internal structure in the LS? Is it organised randomly, and we cannot use it for a more efficient search for geological realisations under given conditions?

To answer this questions, it is necessary to employ dimensionality reduction techniques. Initially, the dimensionality of the latent space is set to 30, which makes the analysis quite challenging. I attempted a pairwise exploration of each dimension to identify any internal structure visually, but it yielded no meaningful results. In the **Section 3.8: Tools for analysing multidimensional spaces**, I described dimensionality reduction techniques, based on which I developed a tool for processing and visualising the LS of GWAE. It is worth noting that there are many more methods available, but for my work, I chose PCA, t-SNE, and TDA.

The following figure 5.6 represents a visualisation of the 30-dimensional LS. Undoubtedly, any dimensionality reduction tool loses some information, but even such visualisation allows for certain conclusions. In case of *FLUVIAL* dataset the first three main components of the PCA projection holds only 30 per cent of overall information. Let’s first examine the global properties of the space under each tool and then move on to more local characteristics.

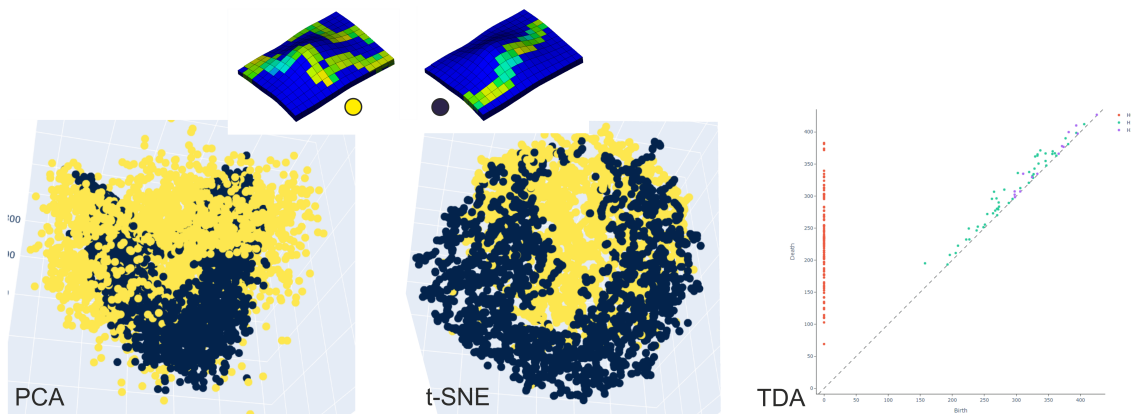


Figure 5.6: Visualisation of the 30-dimensional LS in the PCA, t-SNE, and TDA spaces. Each point in the PCA and t-SNE space represents a 30-dimensional vector characterising a geological realisation from the training dataset. ● depicts a geological scenario with one channel while ● represents scenarios with two channels. In the TDA space, the appearance and disappearance of connected components are displayed.

PCA and t-SNE

I intentionally merged the analysis of the LS visualisation in PCA and t-SNE, as essentially, we see the same thing in both images. Globally, the space consists of a relatively dense cloud of points comprising two large clusters. Each point represents a specific 30-dimensional coordinate of model location in the LS for each geological realisation from the training dataset, with the colour indicating it belongs to a particular scenario. ● depicts a geological scenario with one channel while ● represents scenarios with two channels. It is evident that scenarios are grouped but have areas of overlap. From this, we can infer that if sampling is performed in a specific region of the latent space, we can expect the generated geological realisations to belong to a certain scenario. This could significantly improve sampling efficiency in the LS. In the work of Céline Scheidt (2018), this approach is referred to as *falsification*, where the latent space is considered as the entire prior space of possible geological scenarios. As additional information becomes available, we can falsify specific regions of this space and exclude them from further consideration, rejecting models that do not match the data.

To understand the local structure of the latent space, it is necessary to sample and reconstruct geological realisations manually. Figure 5.7 shows geological realisa-

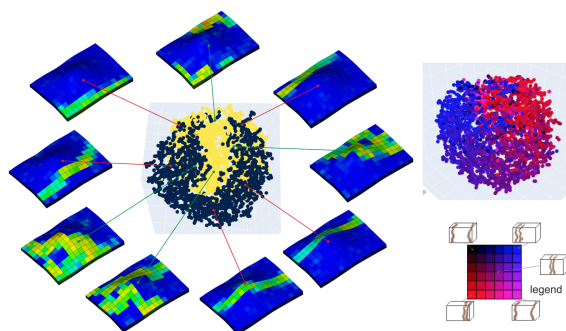


Figure 5.7: 3-d t-SNE visualisation of the 30-d LS. Every point in this cloud is a sample from the training dataset. left: the LS in the scenario dimension, where / represents a scenario with one channel, and / represents two channels. On the right: the same LS in the channel spatial arrangement metric.

tions corresponding to specific regions of space. The latent space possesses an internal structure reflected in the spatial arrangement of channels. If we consider the U-shaped ● cloud of points, it can be observed that the channel “shifts” from the southeast to the northwest. ● cloud of points exhibits a similar dependency. The northern part of the LS produces realisations where the channels are spread across opposite sides. In contrast, the central part of the LS allows for reconstructions with channel placements closer to the central area of the grids. Also, the LS can be represented in the channel arrangement metric. This time, the colour scale corresponds not to geological scenarios but to the spatial placement of channels.

If we randomly select a model coordinate from the latent space and progressively add Gaussian noise with a mean value of 0 and a variance ranging from 0 to 1, we obtain the following figure 5.8. As the variance increases, the geological representation deviates more from the original, but visually, geological realism is not critically lost.

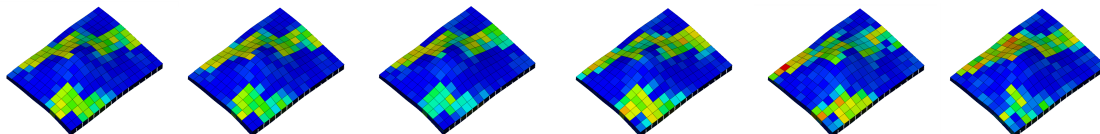


Figure 5.8: Visualisation of the geological representation reconstructed by GWAE, adding Gaussian noise to the original latent vector. Variance step from 0 (left) to 1 (right).

TDA

Topology is the study of the shape of data. Analysing the topology of an object involves comprehending its geometry and structure without necessarily reducing its dimensionality. Topological Data Analysis enables the examination of complex, multidimensional, noisy data and the extraction of essential characteristics that offer insights into the object’s shape under consideration. In our case, we will analyse the entire LS as a single object. To do this, let’s refer to figure 5.9. We can see that in space H_0 , the points are arranged along the vertical axis, indicating that the hidden space vectors are separated from each other by different distances, with some distinct objects marked by ○.

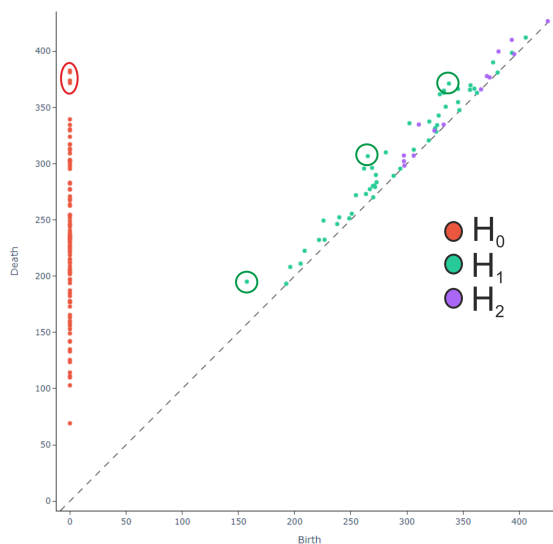


Figure 5.9: The LS in the context of TDA. ○ denote points that are furthest from the rest on H_0 , and ○ most stable rings (H_1). However, it cannot be concluded that these objects are global topological properties of the space.

In H_1 , however, there are no pronounced objects, as the points are stretched along the diagonal more or less equally. Minor outliers are denoted by \circ , suggesting that there may be several rings in this dimension that do not represent global objects. In general, it can be concluded that the 30-dimensional latent space lacks any distinct topological characteristics and is more akin to a simple cloud of points. In the context of TDA, one cannot conclude that the geological scenarios are somehow distinguishable. Instead, both scenarios appear to form a single cloud.

NOTE: It is worth noting that the remote objects marked with a red circle can be either large clusters of points already merged into a massive body or individual remote points. It is likely that in the 30-dimensional space, a situation similar to figure 3.22 when we have two remote accumulations of geological concepts. We will return to this hypothesis in later subsections of this chapter.

5.1.2 Variation of models along the shortest path trajectories in the LS

In the previous sections of this chapter, we examined the generative ability of GWAE solely in reconstructing instances from the training dataset. However, a question arises: what are the generative capabilities – to reconstruct from the dataset or to generate something new? Of course, it is more likely the latter. To definitively assess the generative capabilities of the neural network, it is necessary to evaluate how well it can generate new objects that are not present in the training dataset. This assessment will use a variation of models along trajectories through the LS as an example. The idea is as follows: select two random points/coordinates in the LS, construct the shortest path between them, and then choose hidden model coordinates in the LS along this path at regular intervals. The GWAE will be used to reconstruct the geological realisations from these hidden vectors. Ideally, we should observe a smooth transition from the initial realisation to the final one.

It is important to note that in the **Subsection 3.6: Inner Geometry of a Latent Space**, it was mentioned that the LS is not linear inside. Therefore, the shortest path will not be defined by Euclidean distance but rather by Geodesic distance. One should not expect the distance between ensemble members to be measured as Euclidean. This distance should traverse areas of high density in the LS to allow the generator to produce high-quality realisations. Realisations in less dense regions of the LS become less probable and may eventually become highly improbable, indicating that they are geologically unrealistic.

I selected two random geological realisations for the experiment and identified their positions in the LS. Then, I calculated the shortest paths through the latent space using both Euclidean and Geodesic metrics. Ten realisations were chosen along these paths and recon-

structured back into geological representations using the *Decoder*. Figure 5.10 shows three such experiments. It can be observed that Geodesic interpolation provides higher-quality realisations, while Euclidean interpolation becomes less confident at intermediate steps. This indicates that Euclidean interpolation passes through regions of the 30-dimensional LS with lower density than Geodesic interpolation.

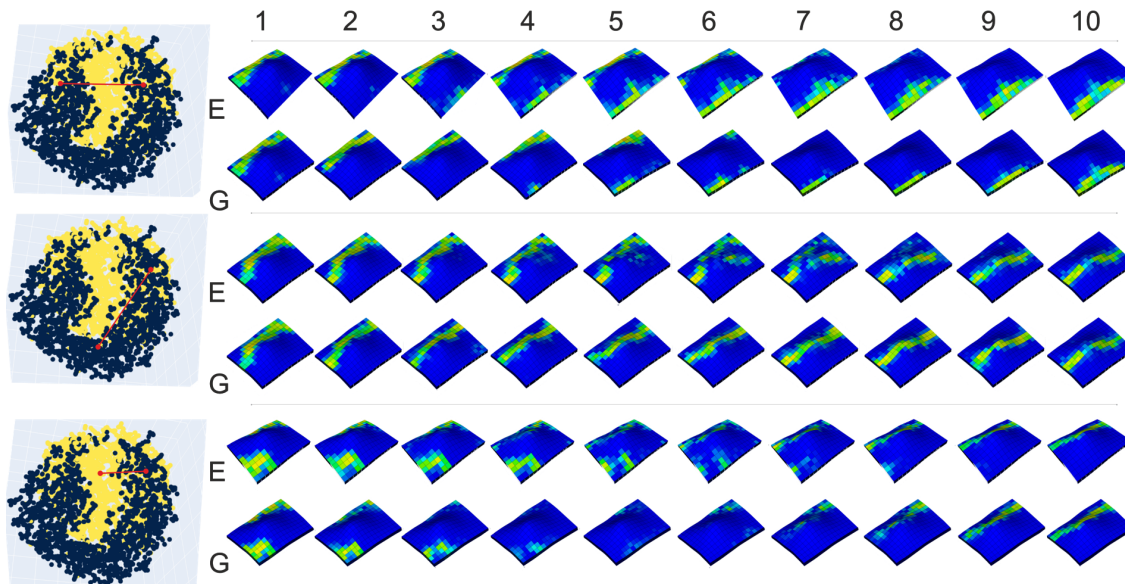


Figure 5.10: Variation of ensemble members along the shortest path trajectories in the LS. Left side: starting and final points in the LS. Right side: Euclidean (E) and Geodesic (G) interpolation options in 10 steps. It’s noticeable that Geodesic interpolation enables the reconstruction of higher-quality geological realisations.

In the figure, special attention should be paid to the third case – the transition from one concept to another. Here, it can be seen that Euclidean interpolation practically deviates into unrealistic implementations from steps 4 to 8, while Geodesic interpolation only shows non-quality implementations at steps 7 and 8. This behaviour can be explained as follows: according to the TDA analysis, objects in the space are distant from each other in H_0 . These objects can be geological scenarios. Therefore, when interpolating between scenarios, we inevitably enter a low-density space, resulting in unrealistic geological variations. However, the Geodesic metric attempts to “stay” in dense regions of the latent space for as long as possible.

It should be emphasised that the intermediate realisations are not instances from the training dataset. This implies that the quality of the GWAE *Decoder’s* performance depends on the quality of the training process and the hidden vector used for reconstruction. It cannot be expected that a generated object will be of high quality if the latent position was sampled from a region of the LS where no training examples are present. This conclusion can be generalised to a fundamental ability of machine learning methods, which work well as interpolators but not as extrapolators.

5.2 Uncertainty of dynamic response and history matching

This section of the chapter will focus on studying the dynamic response of geological realisations generated by GWAE. To achieve this, I prepared a simulation model where GWAE-generated porosity and permeability grids were automatically loaded. Parameters of the flow simulation models are presented in the table 4.2 of the **Chapter 4: Modelling uncertainty and datasets construction process**. I investigated how the dynamic response varies across the latent space, especially for models that are close concerning latent coordinates. The synthetic reservoir model was created using a line drive scheme with one line of producers in the middle and two lines of injectors, figure 5.11.

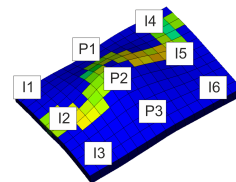


Figure 5.11: Well placement scheme

5.2.1 Uncertainty of dynamic response

To assess the impact of changes in model location in the LS on the dynamic response, I conducted the following experiment, figure 5.12. A reference realisation (●) was selected as a basis for further comparison. Similar to the previous subsection, Gaussian noise with different variances was added to the latent vector of the reference realisation and geological realisations (●) were generated by the *Decoder*. Additionally, the three nearest neighbours (●) were chosen based on the LS vector and were also reconstructed. In the next step, simulations were conducted. The only difference between them was the porosity and permeability grids. For analysis, the producing well P2 and injecting well I2 were selected, as these wells intersected the channel in all realisations and had the strongest flow connection, exerting the greatest influence on each other.

Despite the geological similarity of realisations, some production profiles vary in a wide range, which helps us to estimate production uncertainty. Local variations in property distribution could explain the high variability of dynamic profiles. The channel shape is similar across the models, but the flow dynamic is driven by connectivity that is affected by local features. That is why this intensive dimensionality reduction from 30d to 3d leads to global features reflection of the models and loss of local features information. Moreover, the wide range of production comes from the absence of historical conditioning. In other words, the production history is unknown to us, so we have a wide range of uncertainty in dynamic response despite the high similarity of geological realisations.

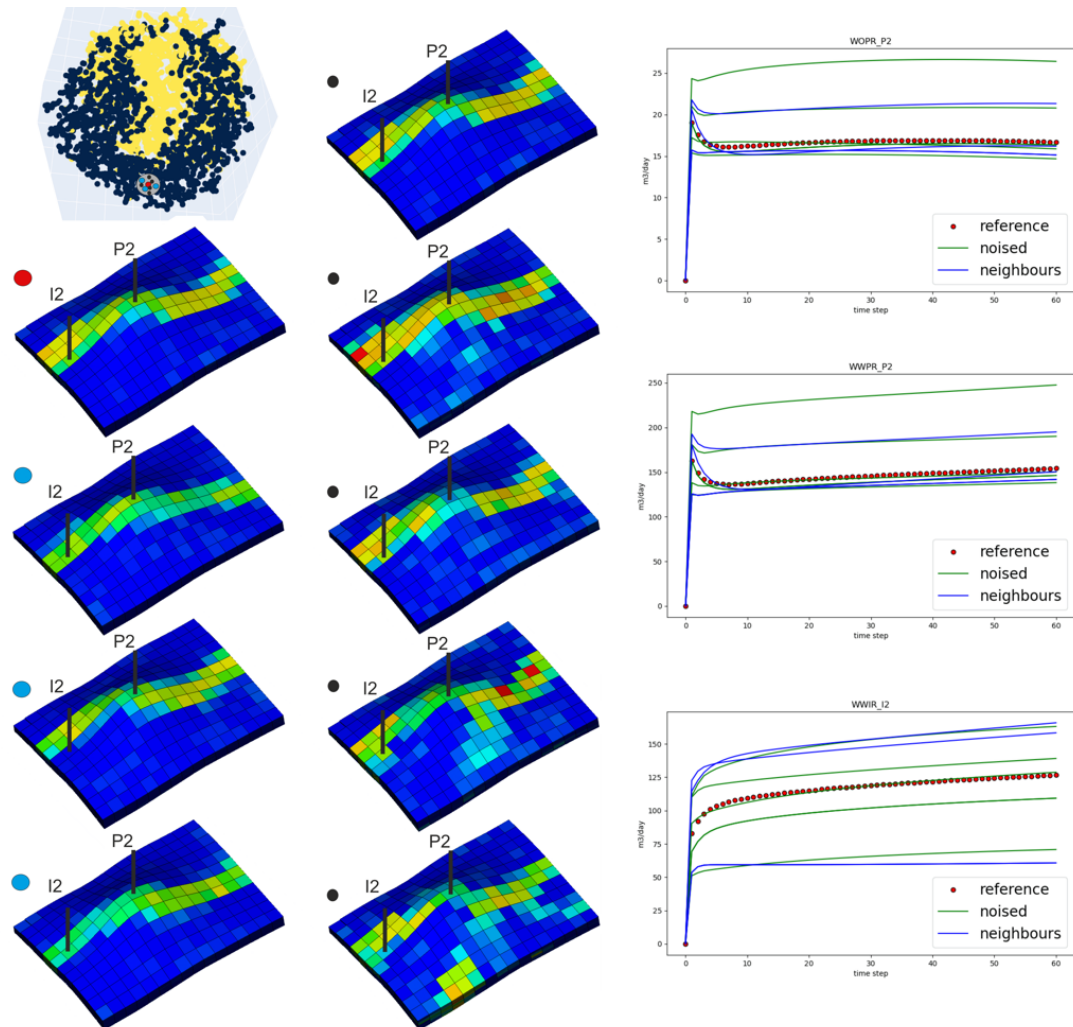


Figure 5.12: Visualisation of the uncertainty of the dynamic response of different geological realisations. I selected a region in the LS (●) where the realisation ● identified as the reference. Different noise levels were added to the latent vector of the reference, and it was reconstructed by the *Decoder* (●). Additionally, three neighbours were selected in the latent space (●). All these realisations were simulated. The right part of the figure indicates the oil and water production rates for well P2 and the water injection rate for well I2.

5.2.2 History matching through the latent space

The optimisation experiment was conducted through the LS conditioned on “hard” static and dynamic well data. The scheme of the experiment was presented in figure 5.3. I picked a reference model which will be unknown to us except for petrophysical and production well data. Then, we start an optimisation process by iteratively generating the population of models and comparing it with the reference well data, consequently minimising a loss function.

The loss function consists of three components: Flow loss, which is the MSE difference in oil /water production and injection rates between the reference and generated models; Static loss, which is the difference in static (petrophysical) properties in wells, i.e. porosity/permeability values; Realism Loss is the third parameter that implicitly measures the realism of every GWAE generated realisation.

The LS has various densities, so it is considered that the density reflects the probability of a particular geological setting. I estimated the density of latent space in different regions by the parameter metric \mathbf{Z} (was introduced in the **Section 3.6: Inner Geometry of a Latent Space**) that provides an estimate of the curvature of the LS and increases the uncertainty estimate as we move away from dense regions of the LS covered by training data. If the value of Realism loss is high, the optimisation algorithm sampled from the subspace is of low density, i.e., highly unlikely realisation or geology of low reliability. Otherwise, the sampling process is conducted in regions of high density, i.e., the realisation is likely.

The Covariance-Matrix Adaptation Evolution Strategy (CMA-ES) optimisation algorithm (was introduced in the **Section 3.4: Optimisation theory**) was chosen to search the LS. CMA-ES is an algorithm that can take the results of each generation and adaptively increase or decrease the search space for the next generation. I launched four optimisation processes, each composed of 100 iterations, with a population size of 51 samples. The optimisation process has found various realisations that all meet the following requirements: similarity in static and dynamic wells data along with geological realism of every found solution. Results of the optimisation process are presented in figure 5.13. Due to high uncertainty in regions without well data, the geological realisations have not been completely similar to our reference model, so the uncertainty of geology could be considered.

In figure 5.14, oil and water production profiles for well P2 and injection profiles for wells I3 and I5 are presented. In grey lines there are dynamic responses of the initial population of generated models, which shows great variability due to the randomness of the initial population to spread across the latent space evenly. Final optimised realisations are shown in solid black lines, in proximity to history data. Only three wells penetrated the channel, so the rest were performed without production. Values of production comparing reference

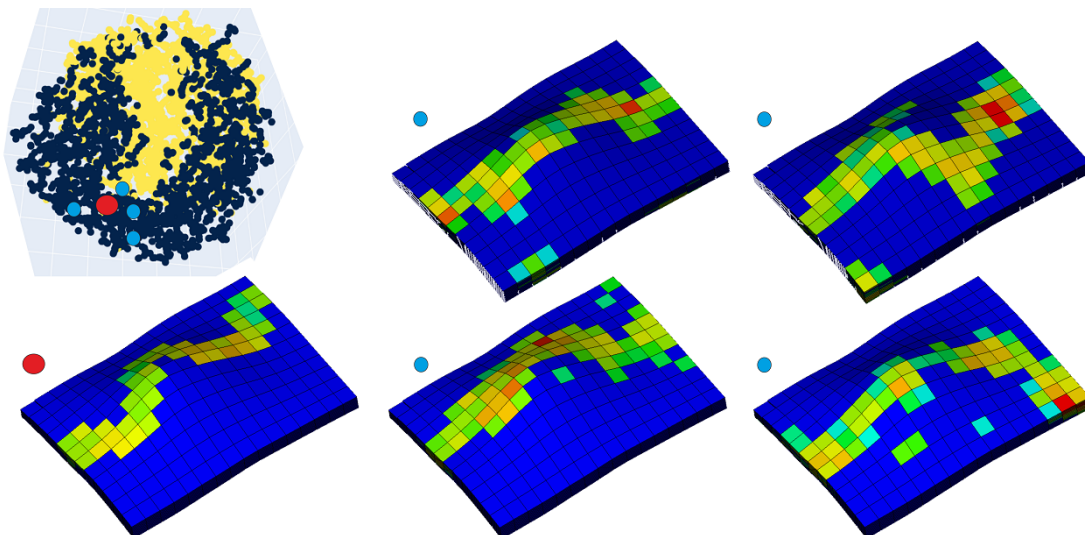


Figure 5.13: Visualisation of the reference porosity property (●) and the cubes obtained through the optimisation process (●). Due to high uncertainty in regions without well data, the geological realisations have not been completely similar to our reference model

and optimised models are quite close to each other, so the similarity in dynamic response can be considered acceptable.

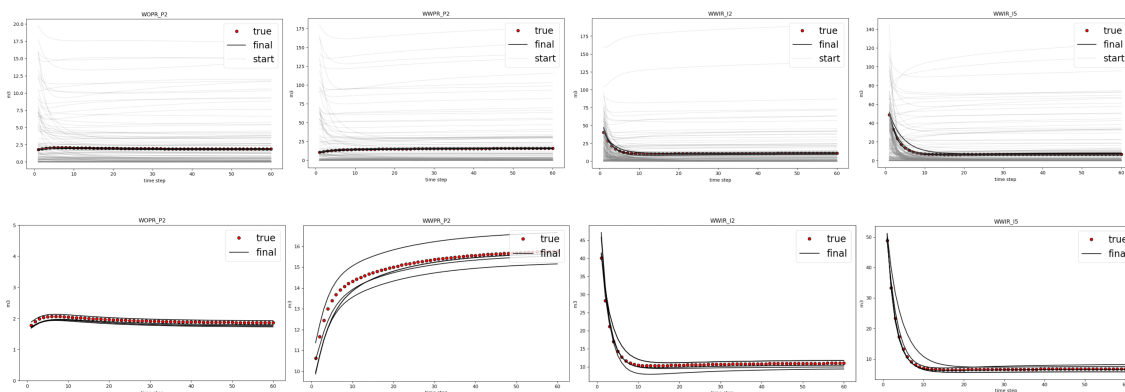


Figure 5.14: Production dynamics of injectors and producers for the initial model ensemble (top row) in comparison to the final optimised ensemble (bottom row) vs the reference dynamic response

Earlier, I considered the static well properties known to us, so the optimisation process should have found solutions similar to the reference model. GWAE along with the optimisation process, managed to find models with appropriate static properties, figure 5.15. It can be stated that the GWAE can be used to generate an ensemble of reliable realisations conditioned to “hard” well data.

5.2.3 Objective Function efficiency study

To test the effectiveness of the OF as a tool for controlling realism of GWAE generated realisations, the following experiment was conducted: the third component of the Objective Function (Realism Loss), which implicitly measures the realism of GWAE generated

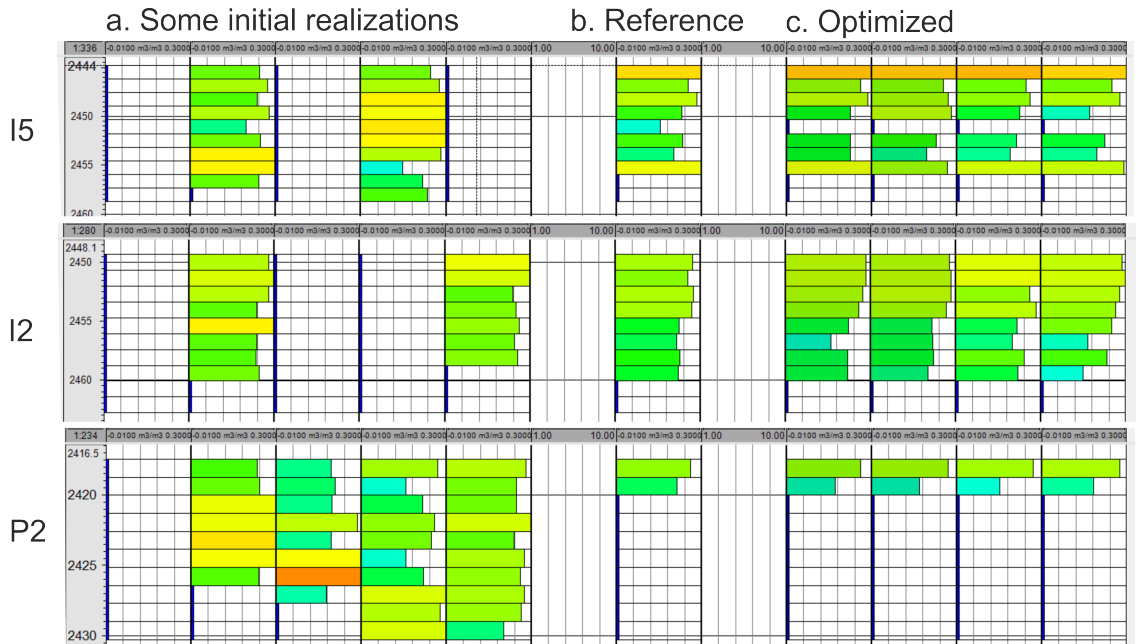


Figure 5.15: Static (porosity) property in wells P2, I2, I5: a. examples of initial (start) realisations, b. reference (true) realisation, c. optimised (final) realisations

realisation, was disabled. Then an optimisation process was run with the settings that are described in the previous subsection. Thus, we simulate a HM process in which the optimiser does not receive information about the spatial location of the geological representations in the LS, and the direction of the optimisation with respect to the location in the LS. As a result of HM, we expect the optimiser to find a geological representation that provides similar static properties at well points as well as dynamic response. However, in general, the geological realism of the model is not expected, as no information has been obtained by the optimiser.

The following figures show the results of the experiment. Hereinafter in this subsection, the model obtained as a result of optimisation with a truncated Objective Function will be called *short_OF*, and the models obtained using the full Objective Function - *full_OF*. The left part of figure 5.16 shows the spatial location of the *short_OF* geological representation in the PCA representation. It can be seen that the location of *short_OF* (●) is significantly distant from the main cluster of points. This suggests that the reconstructed geological realisation will be significantly different from the prior set (●), which is a violation of geological realism. The right part of the figure confirms this hypothesis, as the reconstructed geological representation of *short_OF* (●) differs from the reference model (●) and the reconstructed realisations of *full_OF* (●).

The following figures 5.17, 5.18 show that the optimisation process was able to find implementations that provide reliable static and dynamic characteristics of the geological realisation.

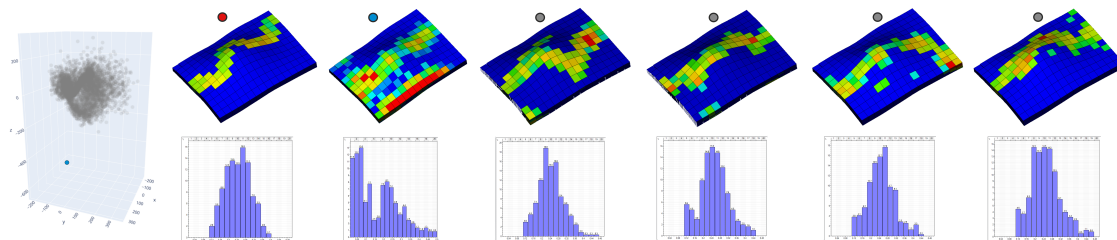


Figure 5.16: Left: spatial location of the *short_OF* geological representation (●) in the PCA representation. Right: the reconstructed geological representation of *short_OF* (●) differs from the reference model (●) and the reconstructed realisations of *full_OF* (●)

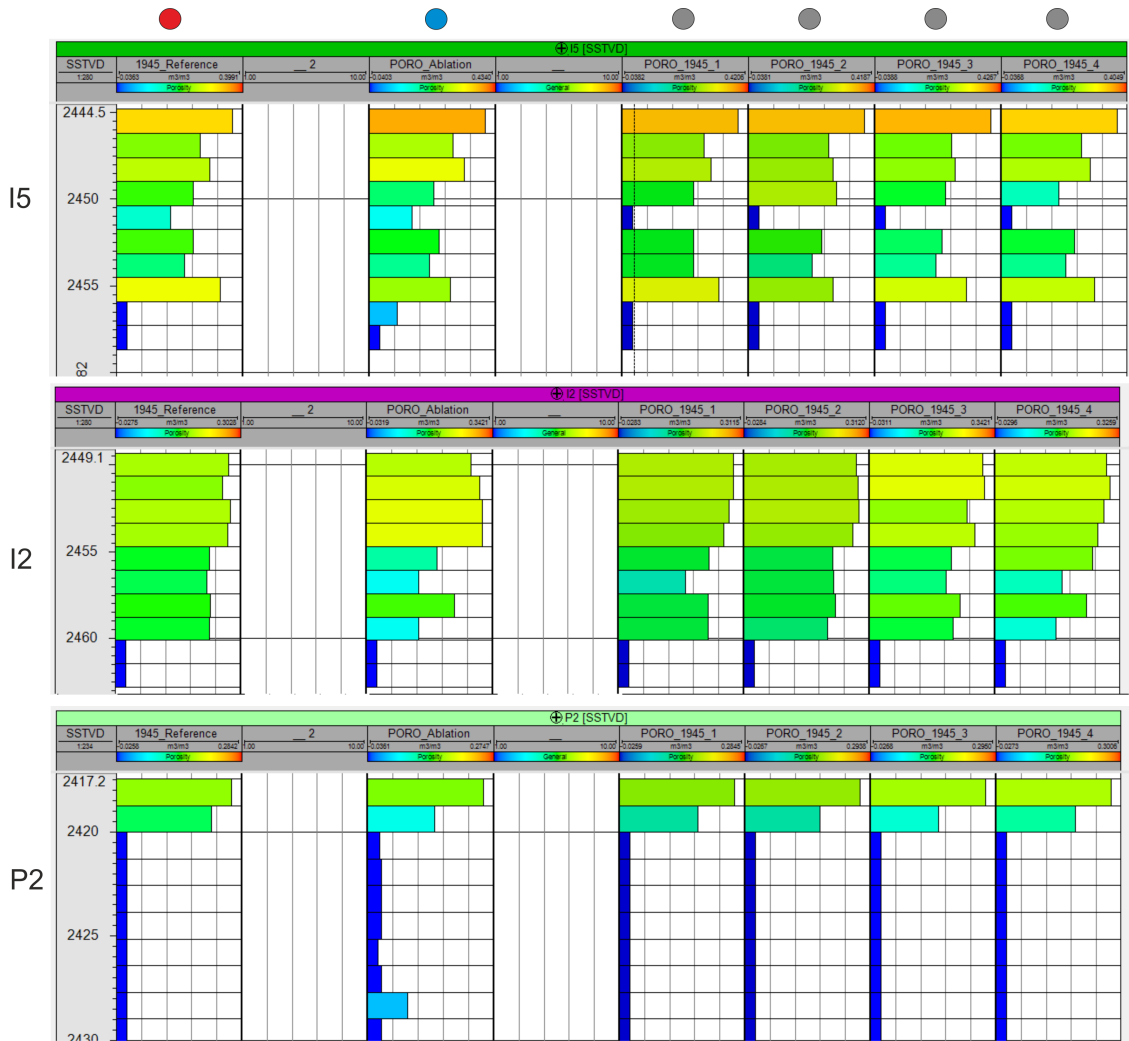


Figure 5.17: Static (porosity) property in wells P2, I2, I5. Reference model (●), *short_OF* (●) and *full_OF* (●). The overall similarity of well logs can be seen

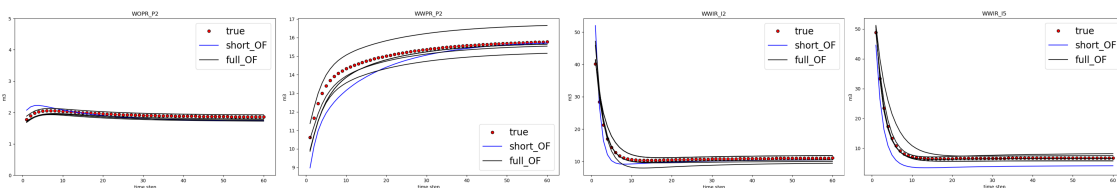


Figure 5.18: Production dynamics of injectors (I2, I5) and producer (P2) for the reference, *short_OF* and *full_OF*

5.3 Variational Autoencoders and 3D reservoir representations

In **Chapter 3: Methodology**, a transition from standard Variational Autoencoder to Wasserstein Auto-Encoder was presented, followed by a shift from CNNs to graph convolutions. In this section, I will provide a comparative analysis of the generative capabilities of various modifications of the variational autoencoders and a comparison of computational costs using the *FLUVIAL* dataset as an example.

Let's start with a brief recap of VAE, WAE and GWAE. VAE is an unsupervised generative model comprising an encoder network, a decoder network, and a probabilistic latent space. The encoder network takes in input data and maps it to a LS distribution. The decoder network then reconstructs the input data from samples drawn from the LS distribution. VAEs aim to learn a low-dimensional latent space representation that captures the underlying structure of the data and allows for generation of new samples from the assumed distribution that constitutes the space of reservoir descriptions. The main drawback of VAE is the blurriness of generated objects, which can be addressed using the Wasserstein distance.

WAE is another unsupervised generative model that aims to learn a latent space representation of reservoir descriptions. WAEs use the Wasserstein distance, or Earth Mover's distance, to measure the discrepancy between the LS distribution and a prior distribution. This encourages the LS to be close to a prior distribution, which can improve the generation quality and robustness of the model.

GWAE is an extension of WAE specifically designed for graph-structured data. It leverages graph neural networks to encode and decode graph representations. GWAEs consider the graph's structural information when learning the latent space representation. This makes them suitable for tasks such as graph generation, graph classification, and graph embedding.

VAEs, WAEs, and GWAEs are all unsupervised generative models that learn LS representations of data. VAEs use the probabilistic approach, WAEs optimise the Wasserstein distance, and GWAEs are tailored for graph-structured data.

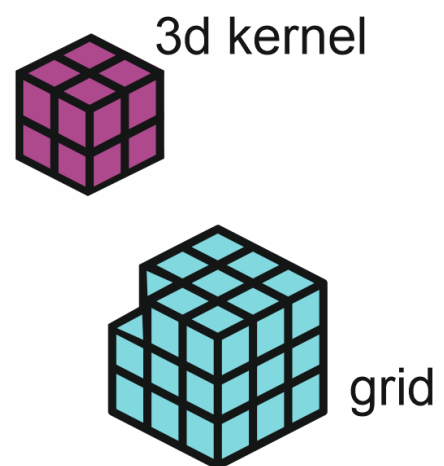


Figure 5.19: Visualisation of an example where a three-dimensional convolutional kernel cannot process a grid with an irregular structure when one of its cells is missing.

It was noted that to disrupt the regularity of the original grid, the first cell was “removed” from each instance of the training dataset. As a result, the size of the original grid was $x \cdot y \cdot z \cdot \text{NumberOfProperties} = 3840$, while the graph used for training was 3838. This small but significant difference means that CNN-based generative networks cannot learn from data with such a structure, figure 5.19. This small change in data structure is purely technical, aiming to demonstrate that even such simple cases unrelated to geological modelling, could be a limiting factor.

5.4 Architecture of Generative Models

A pre-built solution described in the paper by Tolstikhin et al. (2017) was used for the implementation of CNN-based VAE and WAE. The dimensionality of the latent space for each network was set to be the same for all networks – 30.

I use batches of size 100 with the number of epochs to train 10,000 for VAE and WAE. 3DConv is 3D convolution, which is a type of convolutional operation applied to three-dimensional data. Batch Normalisation (BN) is a technique to normalise the mean and variance of the inputs within a batch of training examples. Rectified Linear Unit (ReLU) is a non-linear activation function which is defined as $f(x) = \max(0, x)$, which was first mentioned by Fukushima (1975). ConvTranspose3D is the inverse operation of 3D convolution.

Table 5.1: Architecture of VAE

VAE Encoder $x \in \mathbf{R}^{16 \cdot 12 \cdot 10}$	VAE Decoder $z \in \mathbf{Z}^{30}$
$\rightarrow 3DConv(2; 16) \rightarrow BN \rightarrow ReLU$	$\rightarrow Linear(30; 1024; bias = True) \rightarrow ReLU$
$\rightarrow 3DConv(16; 32) \rightarrow BN \rightarrow ReLU$	$\rightarrow ConvTranspose3D(16; 10) \rightarrow BN \rightarrow ReLU$
$\rightarrow 3DConv(32; 64) \rightarrow BN \rightarrow ReLU$	$\rightarrow ConvTranspose3D(10; 8) \rightarrow BN \rightarrow ReLU$
$\rightarrow 3DConv(64; 128) \rightarrow BN \rightarrow ReLU$	$\rightarrow ConvTranspose3D(8; 4) \rightarrow BN \rightarrow ReLU$
$\rightarrow Linear(128; 128; bias = True)$	$\rightarrow ConvTranspose3D(4; 2) \rightarrow BN \rightarrow ReLU$
$\rightarrow Linear(128; 30; bias = True)$	$\rightarrow Sigmoid()$

For the GWAE, the batch size was set to 10 and the number of epochs 1000 because the training process is much slower due to graph convolutions. GraphConv is a type of neural network designed to operate on graph-structured data. ELU stands for Exponential Linear Unit (Clevert et al., 2016), which is an improvement over other activation functions like ReLU. For $x \geq 0$, $ELU(x) = x$ (same as ReLU). For $x < 0$, $ELU(x) = \alpha \cdot (\exp(x) - 1)$, where α is a hyperparameter that controls the output for negative values. It has a smoother gradient than ReLU, which can help improve the convergence of the optimisation process during training. It does not suffer from the “dying ReLU” problem, where neurons can

Table 5.2: Architecture of WAE

WAE Encoder	WAE Decoder
$x \in \mathbf{R}^{16 \cdot 12 \cdot 10}$	$z \in \mathbf{Z}^{30}$
$\rightarrow 3DConv(2; 16) \rightarrow BN \rightarrow ReLU$	$\rightarrow Linear(30; 1024; bias = True) \rightarrow ReLU$
$\rightarrow 3DConv(16; 32) \rightarrow BN \rightarrow ReLU$	$\rightarrow ConvTranspose3D(16; 10) \rightarrow BN \rightarrow ReLU$
$\rightarrow 3DConv(32; 64) \rightarrow BN \rightarrow ReLU$	$\rightarrow ConvTranspose3D(10; 8) \rightarrow BN \rightarrow ReLU$
$\rightarrow 3DConv(64; 128) \rightarrow BN \rightarrow ReLU$	$\rightarrow ConvTranspose3D(8; 4) \rightarrow BN \rightarrow ReLU$
$\rightarrow Linear(128; 30; bias = True)$	$\rightarrow ConvTranspose3D(4; 2) \rightarrow BN \rightarrow ReLU$
	$\rightarrow Sigmoid()$

become permanently inactive, leading to dead pathways in the network.

Table 5.3: Architecture of GWAE

GWAE Encoder	GWAE Decoder
$x \in \mathbf{G}^{1919}$	$z \in \mathbf{Z}^{30}$
$\rightarrow GraphConv(2; 2) \rightarrow ELU$	$\rightarrow Linear(30; 500; bias = True) \rightarrow ELU$
$\rightarrow GraphConv(2; 1) \rightarrow ELU$	$\rightarrow Linear(500; 1919; bias = True) \rightarrow ELU$
$\rightarrow Linear(1919; 30; bias = True)$	$\rightarrow GraphConv(1; 8) \rightarrow ELU$
	$\rightarrow GraphConv(8; 2) \rightarrow ELU$

The models were trained using a 12th Gen Intel(R) Core(TM) i7-12700 2.10 GHz processor and an NVIDIA GeForce RTX 3070 graphics card. The training times of epochs per second and total training in minutes are presented in the table 5.4.

Table 5.4: Time for training each model

Model	ep./sec.	Total (min.)
VAE	4.3	37
WAE	2.6	62
GWAE	0.4	40

5.5 Results of training Generative Models

In this section, the results of training the three variants of generative models will be presented. As mentioned, the training was conducted based on the *FLUVIAL* training dataset. This choice was made because it is the simplest dataset among the three datasets described in **Chapter 4: Modelling uncertainty and datasets construction process**. Additionally, *FLUVIAL* has a fixed structure, which allows for training CNN-based neural networks.

Before analysing the quality of the generative models, let's compare the internal structure of the latent space for the three generative models: VAE, WAE and GWAE. The

FLUVIAL dataset consists of two main geological scenarios (one/two channels). In figure 5.20, a three-dimensional t-SNE projection of the 30-dimensional hidden spaces for VAE, WAE, GWAE are shown respectively. The colouring shows the affiliation of the model representation in the LS to one of the two geological scenarios (● - single channel; ● - two channel).

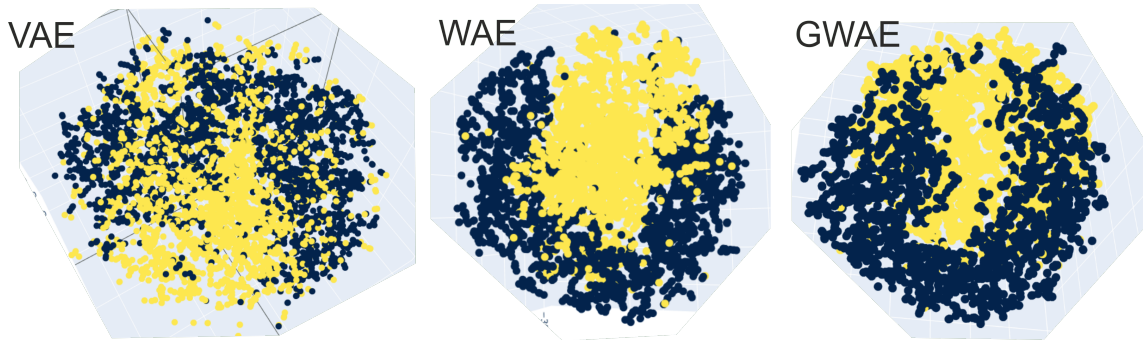


Figure 5.20: Visualisation of the 30-dimensional LS in the t-SNE projection for VAE, WAE, GWAE. Each point represents a 30-dimensional vector characterising a geological realisation from the training dataset. ● depicts a geological scenario with one channel, while ● represents scenarios with two channels

Noticeably, point cloud distribution trends are similar between the three generative models. Still, the LS of VAE is significantly more scattered compared to WAE and GWAE, which can lead to difficulties when navigating in latent space. This can be explained by the replacement of the KL divergence metric with the Wasserstein metric in the architecture of the generative networks (**Subsection 3.2.2: From VAE to WAE**).

To assess the quality of the *Decoder*, each generative network selected and reconstructed a random realisation. In figure 5.21 (top), the original/reference porosity cube is displayed, as well as the reconstructed outputs by the VAE, WAE, and GWAE *Decoders*. The middle row compares histograms generated for porosity property relative to the reference distribution. It should be noted that, for visualisation, histograms do not display porosity values equal to zero, as their quantity is very large. A cut-off at porosity = 10% is adopted, so anything below this threshold is considered non-reservoir. The percentage of non-reservoir in all presented grids is 85%. The reservoir distribution (porosity > 10%) generated by GWAE is the most concurrent with the reference. It can be concluded that VAE and WAE do not reproduce the reference histogram and are skewed towards non-reservoir, as this is the predominant property. Also, it should be noted that the right part of the histograms for VAE and WAE has inflated values compared to the reference, which is reflected in volumes. The lower part of figure shows the reconstructed porosity-permeability relationships respectively. The VAE has the lowest generation quality, as the *Decoder* failed to reproduce the spatial connectivity of the channel and only partially captured the petrophysical dependencies. This indicates that replacing the KL with the Wasserstein metric has worked correctly.

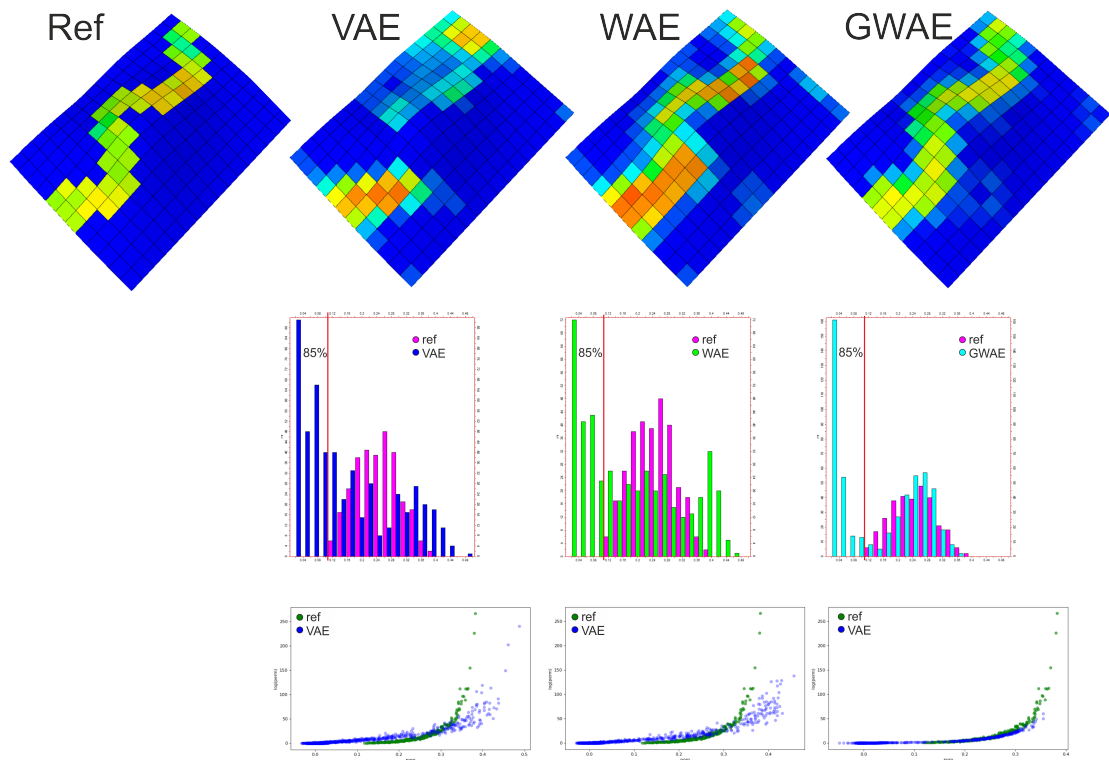


Figure 5.21: Top row: Visualisation of the reference porosity grid and decoded by VAE, WAE and GWAE respectively. Middle row: Histograms of porosity distribution of generative models concerning reference. For visualisation purposes, porosity is presented within the $(0; 0.5]$ range. 85% corresponds to porosity parameter values $< 10\%$, considered as non-reservoir. The percentage of non-reservoir cells amounts to 85% for all presented property cubes. Bottom row: Porosity-permeability dependency of generative models concerning reference.

WAE provides a good qualitative spatial reproduction of the channel position. However, the propagation of the porosity property was not accurately reproduced. A significant increase in porosity is observed at the southern end of the channel compared to the reference model. Like the VAE case, the petrophysical dependency was not reproduced accurately.

Figure 5.22 shows a cross-section of the porosity grids for a reference model and reconstructed by VAE, WAE and GWAE. VAE failed to recover the property in the central part of the model (I) because the values inherent to the channel facies are completely absent. WAE shows inflated values in the southern and northern parts of the model (I, J). As for GWAE, the cross-section shows a high degree of similarity to the reference model in space and absolute values of porosity property.

The quality of reconstruction by the GWAE *Decoder* was thoroughly examined in the previous chapter and significantly outperformed the previous models. However, it should be noted that the VAE and WAE were taken as it was introduced by Tolstikhin et al. (2017), without additional tuning. In the case of VAE, the generative ability is significantly inferior due to its internal architecture, while WAE can improve with additional tuning. It is worth mentioning that the graph-based variant will significantly outperform WAE the graph-based variant in the following cases:

- Presence of structural unconformity: If the geological object is characterised by structural uncertainty, CNN-based generative networks are inherently unable to account for them. The structural unconformity case will be presented in the following chapter.
- Significant reservoir/non-reservoir imbalance. If there is a large imbalance towards the non-collector in the property cube, the CNN during training will be biased towards the dominant properties. Therefore, additional engineering tricks will be required to mitigate the data imbalance. As for the graph-based approach, we can “win” against data imbalance simply by removing unnecessary non-reservoir nodes, thereby breaking the regularity of the original grid

The next question to consider is how the quality of the generated geological objects can

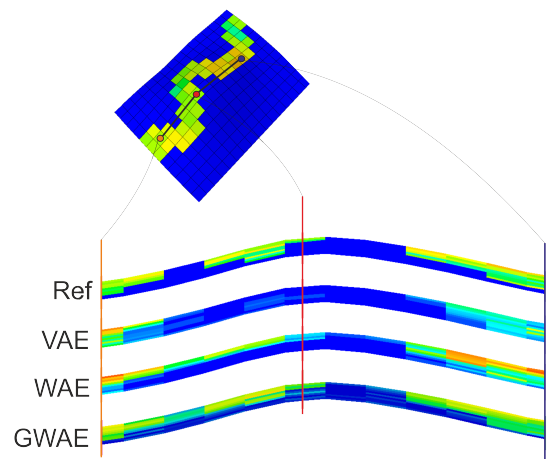


Figure 5.22: Cross-section of the porosity grids for a reference model and reconstructed by VAE, WAE and GWAE

impact the corresponding dynamic responses. To conduct such an analysis, flow simulations were performed for the reference porosity and permeability grids, as well as for the grids generated by the VAE, WAE, and GWAE *Decoders*. The following figure 5.23 (top) displays the average pressure maps at the final simulation step, as well as the rates of oil production, water production, and injection for the entire reservoir (bottom).

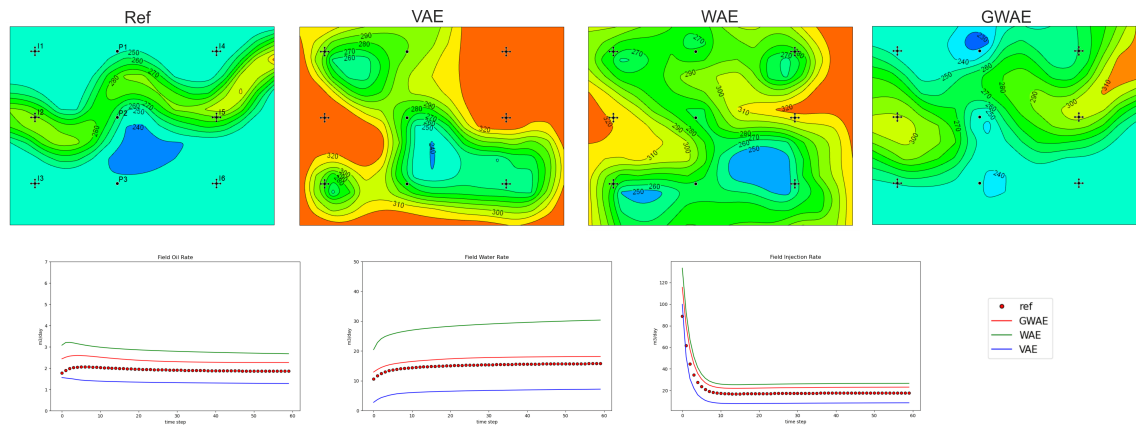


Figure 5.23: Top row: Average pressure maps at the final step for reference model, VAE, WAE and GWAE. Bottom row: Field oil, water and injection rates

It is noticeable that the dynamic response of the VAE significantly differs from the reference model. Due to the incorrect spatial reconstruction of the channel, the pressure map displays significant pressure changes in regions that should not be involved in reservoir development. Due to the channel's fragmentation, the production and injection rates are lower than the reference. Despite the correct spatial positioning of the channel facies, the WAE realisation also shows significant deviations from the reference model. Two reasons can explain this. Firstly, there is an overestimation of the porosity distribution, which results in an overestimation of permeability and, therefore, filtration properties of the rock. Secondly, the inherent predisposition of variational generative models to generate blurry representations leads to an increased drainage volume from the wells. The GWAE implementation shows the closest approximation in reproducing the dynamic response. This is primarily due to the accurate spatial reconstruction of the channel and correct filtration properties. However, the inherent blurriness in the reproduction also slightly increases the filtration volume from the wells, resulting in higher production rates compared to the reference case.

5.6 Conclusion

In this chapter, I have presented an approach to generating reservoir models conditioned on production data and estimating uncertainty with GWAE. This approach can implicitly parameterise geological representations into a latent space of reduced dimensionality and provides ways to uncertainty quantification and production profiling across multiple geological concepts. GWAE links the static model and its dynamic response via the LS, a reduced-order representation of geological model complexity. On the synthetic *FLUVIAL* dataset, I showed that GWAE reliably reproduces geology by comparing the reference model's static and dynamic properties with generated representations. GWAE utilises the notion of a LS, depicting the variation of geological concepts and other geological properties. Navigation and optimisation through the latent space provide a model update and an ensemble of history-matched models.

The comparison of VAE, WAE, and GWAE models for generating geological objects and simulating dynamic responses reveals several key findings.

Firstly, VAEs suffer from blurriness in the generated objects (i.e. channel location is blurred) and cannot accurately reproduce the spatial positioning of channels and petrophysical dependencies. This is due to the limitations of their internal architecture.

WAEs show improvements over VAEs, with better reproduction of spatial positioning of the channel. However, they still struggle to reproduce the propagation of porosity properties accurately, resulting in discrepancies compared to the reference model.

GWAEs, specifically designed for graph-structured data, demonstrate the highest quality in reproducing dynamic responses. The accurate spatial reconstruction of channels and correct filtration properties contribute to their superior performance.

The comparison also highlights the limitations of CNN-based generative networks when dealing with structural uncertainty and significant reservoir/non-reservoir imbalances. CNNs are inherently unable to account for structural unconformities (**Chapter 6: Graph Wasserstein Autoencoders and fault synthetic case**), and data imbalances require additional engineering tricks to mitigate biases. On the other hand, graph-based approaches, such as GWAEs, can easily adapt to structural uncertainty by removing unnecessary non-reservoir nodes and breaking the regularity of the original grid.

Overall, the results suggest that GWAEs outperform VAEs and WAEs in terms of generating geological objects and simulating dynamic responses. However, further research and tuning are needed to fully optimise the performance of all three models and address their limitations.

Chapter 6

Graph Wasserstein Autoencoders and fault synthetic case

6.1 Structural Uncertainty

This chapter will describe an approach of structural uncertainty parametrisation along with uncertainty of reservoir properties, utilising a latent feature space. I will present an approach to solve an inverse problem through the LS, allowing for the generation of new realisations in the joint uncertainty space of both types. To achieve this, two distinct generative networks were trained independently: one focused on the reservoir's structural characteristics, and the other filled the generated model with properties. The presented experiments will reveal the importance of the prior and limitations associated with the internal structure of the hidden space of the generative model, which could potentially negatively impact the optimisation process.

Faults, horizons, and other structural elements have a critical impact on the dynamic response of reservoirs, making the consideration of structural uncertainty a crucial factor in modelling. This can also impact the effectiveness of reservoir management and development (Bardy et al., 2019; Lescoffit et al., 2005; Manzocchi et al., 2008). The primary cause of structural uncertainty is the low quality of seismic data utilised to create reservoir models. This results in imprecise fault/horizon positions due to the low resolution of seismic images. Moreover, the interpretation of the structural data depends on the personal judgment of an expert. In such cases, conventional HM methods that rely on a single interpretation of reservoir geometry may not match past production.

The difficulties of modelling under structural uncertainty lie in defining and parameterising the process of constructing structural surfaces. In the work of Suzuki et al. (2008a), the uncertainty in structural interpretation, including fault identification and fault throw, is

addressed by creating multiple structural models based on expert knowledge. To account for fault positioning uncertainty, horizons and faults are stochastically perturbed from the initial interpretation. Jackson et al. (2013) proposed that surfaces can be utilised to represent facies boundaries along with structural features, even in cases where they cannot be correlated or identified in seismic data. Traditionally, these boundaries are modelled using grid-based stochastic algorithms, but the author suggests using surfaces instead. This approach allows for the integrated modelling of structural, stratigraphic, and sedimentological heterogeneities by creating a hierarchy of surfaces to represent the boundaries between facies types rather than relying on geostatistics algorithms on a predefined grid. He suggests several ways to generate these surfaces, such as deterministic or stochastic interpolation between well data points, purely stochastic, or processed based with geomechanics. Sabatino et al. (2014) performed a HM process with structural uncertainty by elastic gridding, which saves reservoir model topology.

The solution to the inverse problem should consider both structural uncertainty and uncertainty in property distribution. IC fault model is a classical case where even a very basic, easily parameterised structure significantly impacts flow and remains a stiff inverse problem to infer (Demyanov et al., 2010). Another example is presented in the work of Suzuki et al. (2008b), where an ensemble of realisations was generated as the prior set of models to capture the full-field uncertainty. During the solution of the inverse problem, the search is conducted within this pre-existing ensemble.

To automate the process of structural uncertainty quantification for AHM or UQ, there must be some tools for controlling the physicality of the modelled structure, particularly in the case of multi-variant modelling. There are several methods for constructing and verifying the realism of structural models, such as manual inspection, developable meshes (Thibert et al., 2005), elastic gridding (Sabatino et al., 2014) or more advanced topology (Schaaf et al., 2021) and graph (Hillier et al., 2021a) based structural modelling.

6.2 Generative quality of GWAE with *STRUCTURAL* dataset

I decided to train two generative models to examine the generative capabilities of GWAE in the presence of both structural uncertainty and uncertainty in property distribution. The first generative model will be responsible for generating a graph-type structural carcass, where the output of the Decoder will be a graph that incorporates the fault location and block displacement, figure 6.1 (left). The second GWAE fills the generated graph with porosity and permeability properties, ensuring that there are no geological inconsistencies, figure 6.1 (right). For example, a channel should be present on both sides of the fault, accounting for block throw relative to each other. This two-decoder architecture is justified because structural parameters (i.e. fault, block displacement) of geological ob-

jects typically do not depend on parameters related to sediment accumulation settings (i.e. porosity, permeability). Therefore, such an architecture does not contradict the physics of the natural system; the structural changes of a geological object occur after sediment accumulation, but not at the same time. A set of geological models called *STRUCTURAL* is used as the training dataset for conducting further experiments. The main features are the presence of uncertainty in the fault position (discrete parameter) and the displacement of blocks along the fault (continuous parameter). The properties distribution is performed by only one river channel variant, where the position and properties of the channel serve as a source of uncertainty (**Chapter 4: Modelling uncertainty and datasets construction process**).

During training, the GWAE responsible for the generation of the structural carcass is isolated from depositional properties, thereby “cleaning out” the generative model from uninformative parameters (porosity, permeability). The two-decoder approach allows for a more detailed analysis of the quality of each neural network, explicitly separating the latent space into two sets of parameters, facilitating better orientation in the LS. However, it is worth noting that the combined LS, i.e. combination of LS of structural GWAE and LS of depositional GWAE, can be considered as a unified parameter field that describes the uncertainty of the geological object.

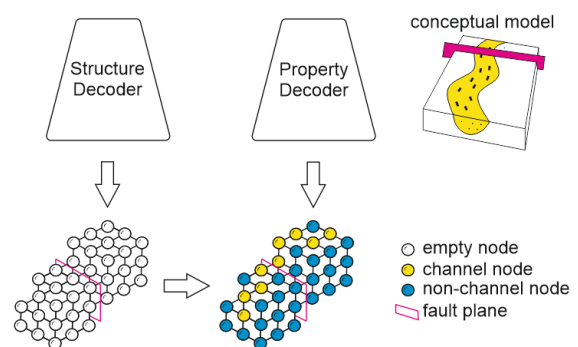


Figure 6.1: The combination of Structural and Property Decoders allows for the generation of a graph-type structural carcass (○), which is then filled with petrophysical properties (●, ●). This combination enables the generation and filling of the structural graph so that the distribution of petrophysical properties does not contradict depositional processes.

Figure 6.2 depicts an example of a random sample from the training dataset *STRUCTURAL* and its reconstructed analogue by the *Decoder*. On the left, the propagation of porosity properties and the presence of a block structure are shown in the familiar format of grids. The middle figure shows a graph representation, where each graph node corresponds to a grid’s cell, and the presence of an edge represents adjacency. This graph representation allows for preserving the topological features of a geological object since the nodes of the graph, which model the properties of the rock, have a common edge only if they are physically adjacent. On the right, a map is presented in graph format. It is worth noting that the LS is informative enough to allow the *Decoder* to reconstruct both structural features and geological aspects correctly.

While solving the problem of edges prediction or generation of graphs, many literature

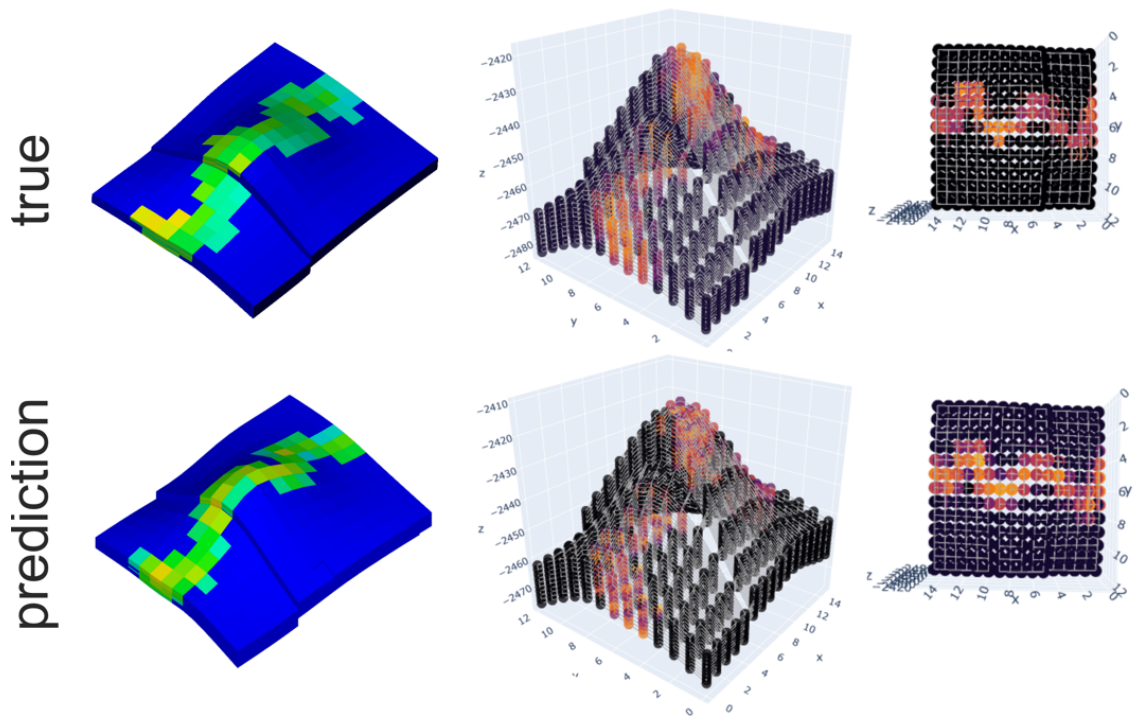


Figure 6.2: Upper row: Examples of the initial porosity grid from the training dataset. Lower row: GWAE reconstructed. The left column displays porosity grids and the arrangement of blocks along the fault in the form of a grid of properties. The central column shows the same properties in the form of a graph, where nodes correspond to grid cells and graph edges represent cell adjacency. The right column depicts a porosity map in the form of a graph.

sources point out a problem concerning incorrect recovery of edges connecting nodes (Agrawal et al., 2019; Pandey et al., 2019; Rahmati, 2023). The experiments conducted in this work were not an exception, and despite the overall high quality of the graph generation, sometimes inaccurate edge recovery was observed. Figure 6.3 illustrates one such example. The grid-like geological representation is shown on the left. The \blacksquare indicates grid cells that are on different sides of the fault, but are in contact with each other and therefore connected. The centre shows a similar geological representation, but only in the form of a graph. It can be seen that the graph nodes (\bullet) have three edges. Such edges fully reproduce the neighbourhood of cells of the original representation. However, when reconstructing the geological representation using GWAE, the neighbourhood of nodes (\bullet) was broken, instead of the expected three edges two were generated. This means that the connectivity of neighbouring nodes will be reproduced with an error. It is worth noting that the overall quality of edge reproduction during training was of the order of 99 percent, so there was no need to invent any tools for automatic quality control of edge recovery other than the latent space density metric.

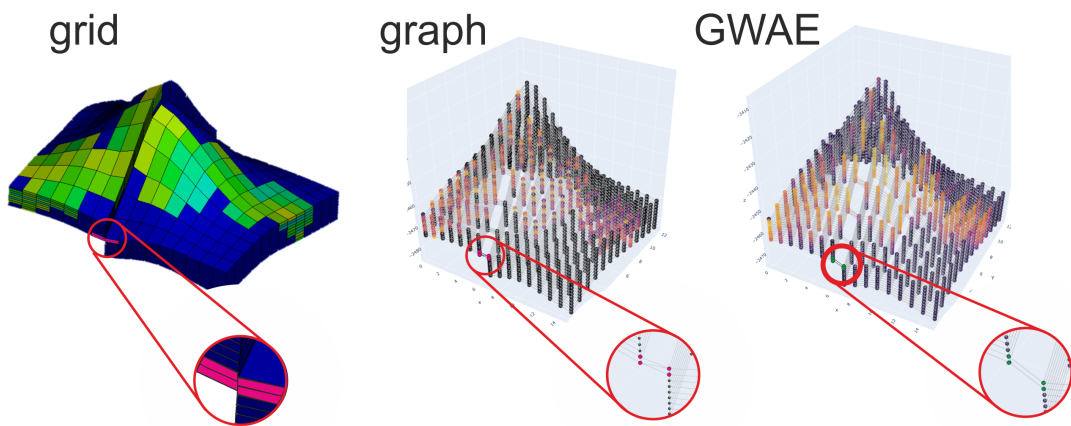


Figure 6.3: Left: Grid-like representation with nodes indicated as neighbours (\blacksquare). Centre: the same geological realisation in the graph representation. Connected nodes represented by \bullet . Note three edges connecting highlighted nodes. Right: GWAE reconstructed geological representation with broken node (\bullet) connections. Only two edges were reconstructed.

6.2.1 Latent Space inner structure with structural uncertainty

Since the generative model consists of two GWAEs, the first responsible for generating the structure and the second for propagating properties, we can divide the hidden space into two parts. The first part parametrises structural uncertainty, and the second parametrises petrophysical properties distribution uncertainty. The dimensionality of the hidden space was empirically chosen to 48, while the dimensionality of the hidden space for the network responsible for properties remained unchanged at 30. So the entire dimensionality of the LS is 78. It's important to note that in the case of training a single universal GWAE, such

analysis (without the use of special techniques, (Higgins et al., 2017)) would be impossible.

PCA and t-SNE

Figure 6.4 provides 3d visualisation of model grids in 30-dimensional LS for property distribution (top), described by graph nodes with attached property values, and 48-dimensional LS for structural features (bottom), described as a connected graph. Each point presented in the 3D projection of the hidden space represents an instance of the training dataset (i.e. a model grid), with its structural and depositional characteristics encoded as 48-dimensional and 30-dimensional model representations, respectively. The first three principal components of PCA for depositional and structural LSs describe 24 and 46 per cent of the total variation respectively.

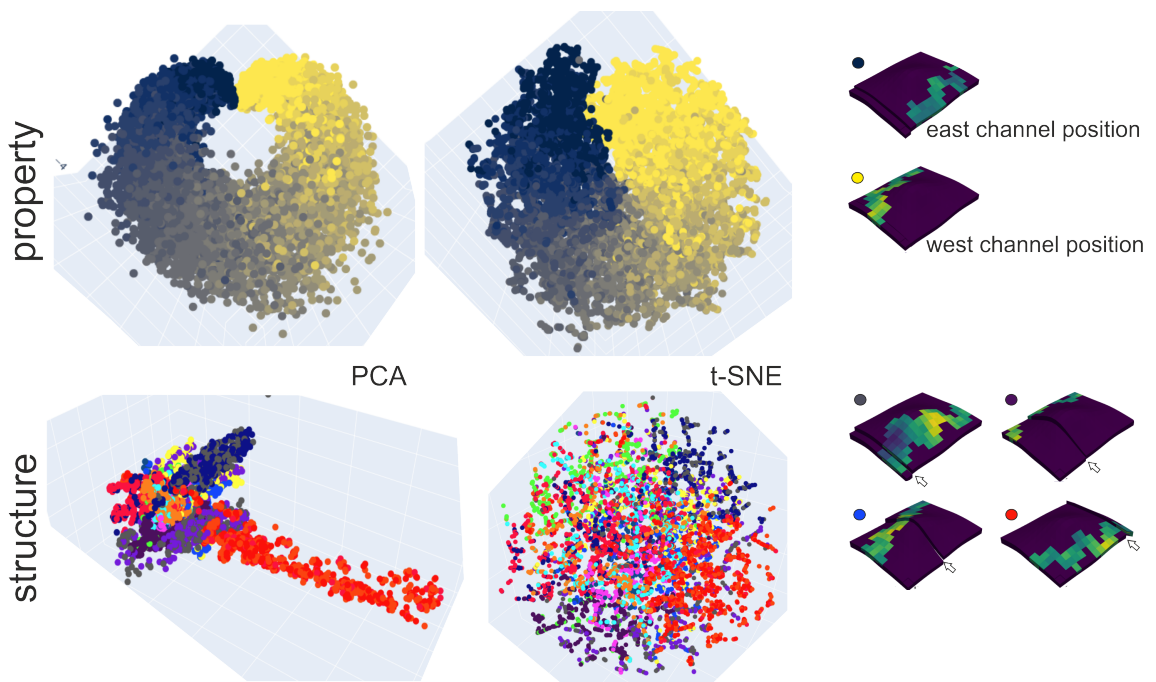


Figure 6.4: Upper row: visualisation depositional LS in PCA, t-SNE projection in channel position settings. Lower row: the same for structural LS in fault position settings

Let's visually analyse both LSs representation in PCA and t-SNE projections. The hidden space responsible for properties (i.e. porosity, permeability) distribution in the reservoir, figure 6.4 (top), has a well-separated structure similar to the *FLUVIAL* dataset. In the case of a *FLUVIAL* dataset, the central part of the hidden space was filled with model grids corresponding to the scenario of two channels. Such a scenario is absent in the *STRUCTURAL* dataset. The colour scale corresponds to the spatial position of the channel, indicating a distinct internal structure of the LS. If we traverse from the ● region to the ● region, the channel will smoothly transition from one part of the property cube to the opposite. It's worth noting that t-SNE projection tends to make the representation denser, thereby limiting the understanding of the actual space structure. PCA suggests that the

central part of the space may have low density, thereby constraining the search space for a potential optimisation process during AHM.

The structural hidden space, figure 6.4 (bottom), is displayed in a discrete colour scale representing the fault's position towards the boundary of the reservoir block. PCA visualisation of the space indicates significant sparsity, which largely depends on the fault's position in the cube. It is possible to trace certain patches of points with similar colour, which means that the spatial arrangement of faults affects the mutual positioning of compressed representations of geological models in latent space. However, it is quite challenging to make a definitive conclusion about any internal structure; besides the spatial arrangement of the fault, there is another type of uncertainty in the form of the mutual arrangement of the blocks of the reservoir model that the fault separates. The combination of structural uncertainty makes the compressed representation of the structural carcass of the reservoir model challenging to interpret. As for the t-SNE projection, like in the previous case, it tightens the point cloud, making it random and leaving no clear understanding of the actual density in the hidden space. In the following section, I will expand on the interpolation capability in the LS.

TDA

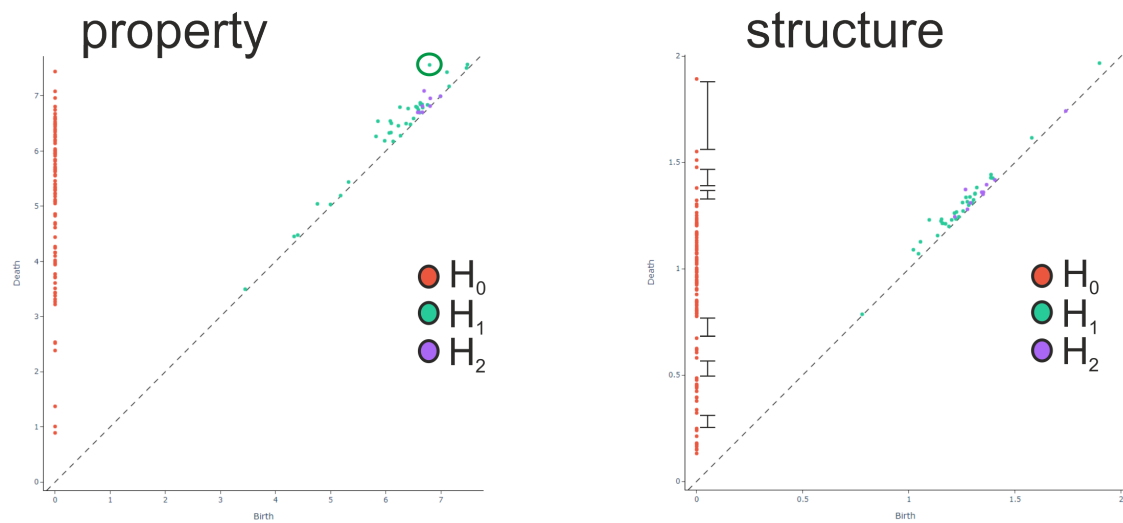


Figure 6.5: TDA projection of depositional (left) and structural (right) LS. Special attention should be paid to the point highlighted by \circ in H_1 dimension of property TDA projection. Multiple hollow spaces (\square) are observed in H_0 dimension of the structural TDA

Moving on to the analysis in the TDA projection for the depositional LS, figure 6.5, attention should be paid to the point highlighted by \circ in H_1 dimension. As a reminder, according to the theory of persistent diagrams, objects located far from the diagonal are the most stable (they live the longest) and can characterise the parameter space from a geometric point of view. Here, we see one standout point that may indicate that points in the original 30-dimensional space form a ring, which is somewhat confirmed by the PCA

visualisation. This is an important indicator because, in further interpolation or solving the inverse problem, one should “avoid” the ring’s central region.

Considering the persistent diagram for the structural LS, the internal structure’s primary indicator is the H_0 dimension. In this dimension, we observe multiple hollow spaces (\sqcup), indicating substantial heterogeneity in the point cloud in the 48-dimensional space. This fact suggests that it will likely be quite challenging to “navigate” through such a discrete space during interpolation or solve the inverse problem without a significant loss in the quality of structural constructions.

6.2.2 Interpolation in the LS

To test the hypothesis of the sparsity of the hidden space and the robustness of the reconstruction to the different noise levels, I selected a random instance from the training dataset and sequentially add Gaussian noise with a mean of 0 and variance of 0.2, 0.4, 0.6, 0.8, and 1 to the LS coordinate, figure 6.6.

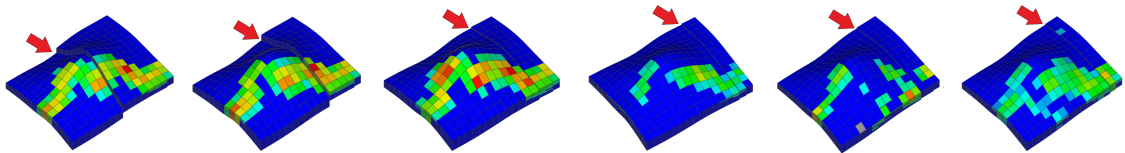


Figure 6.6: Geological realisation reconstructed using the *Decoder* (first instance) with the sequential addition of Gaussian noise to the latent code. The fault location is depicted with \Rightarrow . Note the migration of the fault as the noise increases

It is evident that as the noise increases, the quality of spatial property reconstruction begins to degrade. Meanwhile, the presence of structural features appears to be more random, as the positions of faults and block displacements occur chaotic. From this, we can conclude that the hidden space’s internal structure is a complex object with regions of varying density. This undoubtedly will affect sampling quality during the optimisation process, as navigation in LS becomes less controllable for structural uncertainty; this affects the value of the Objective Function during optimisation. Structural uncertainty influences compartmentalisation, which reflects on the system’s dynamic response.

In **Section 5.2.2: Variation of models along the shortest path trajectories in the LS**, we experimentally demonstrated the advantages of Geodesic interpolation over Euclidean. However, it was noted that such calculations can be computationally expensive (**Section 3.6: Inner Geometry of a Latent Space**). In this subsection, I will present examples of Geodesic interpolation in a joint (hybrid) 78d hidden space with complex and heterogeneous structure. Based on the conducted analysis, it can be inferred that there is a high probability that sampling during interpolation will result in varying levels of reconstruction quality. The discrete nature of the “fault position” property is likely to yield a difficult-

to-interpret result, as the hidden vector probably does not contain information about the similarity of fault positions relative to each other. At least, the analysis using PCA, t-SNE, and TDA did not provide us with information in this regard.

Figure 6.7 illustrates three scenarios of Geodesic interpolation through the hybrid LS with a discretisation of 10 steps. Let's illustrate interpolation capacity with three considered distinct scenarios, depicted as trajectories of different colours. The first row represents the paths ($/, /, /$) of interpolation for the spatial (left) and structural (right) hidden spaces.

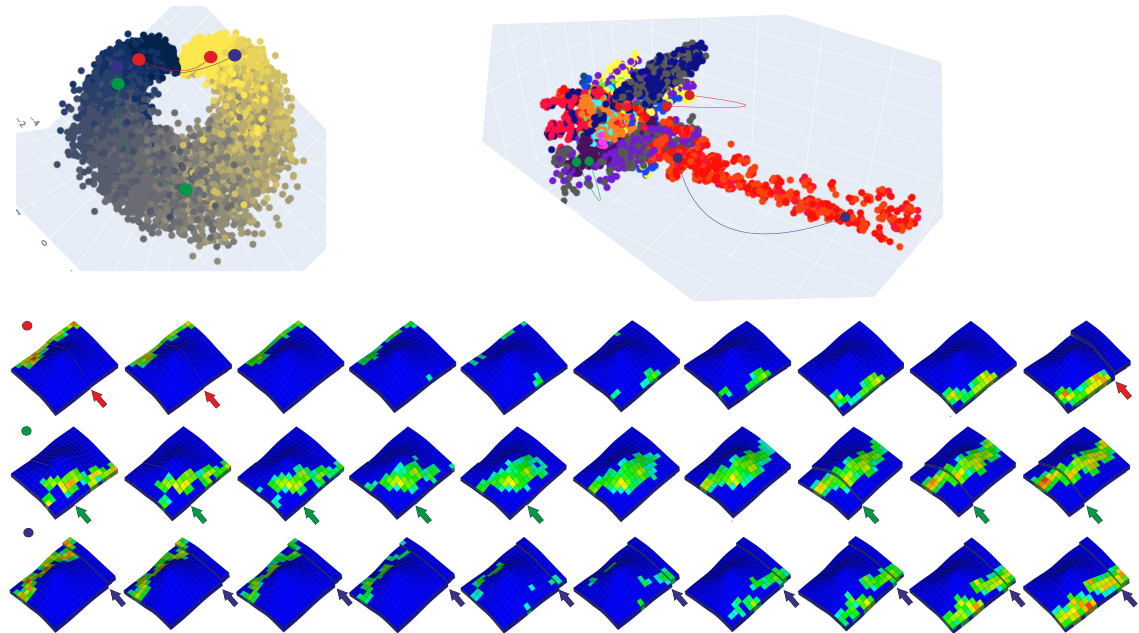


Figure 6.7: Examples of interpolation ($/, /, /$) through a hybrid LS. Upper row: LS for spatial uncertainty (left), LS for structural uncertainty (right). Geological realisations (\bullet , \bullet , \bullet) reconstructed using the *Decoder* according to the interpolation path. Arrows depict the fault location (\Rightarrow , \Rightarrow , \Rightarrow)

The first row (\bullet) illustrates that the volume of the channel facies gradually decreases on one side and increases on the other. Moreover, the transition of the channel occurs not through the cube but seemingly outside of it. This can be explained by the fact that according to TDA (and visually confirmed by PCA), the 30-dimensional hidden space is somewhat looped, meaning there is a “bridge” that makes the transition continuous. This behaviour is partly since the training dataset contains instances where the proportion of the “channel” facies is very small (the entire channel is seemingly outside the grid). In general, this property of hidden space should have a positive impact on the AHM process, as the optimiser will have the ability to transition from one “tail” of the hidden space to another, even in the case of an unsuccessful initialisation of the optimisation process. This once again confirms the importance of constructing a prior.

As for the changes in structural features (\Rightarrow), the fault almost immediately disappears, and the grid becomes free of structural disruptions. Only in the last step does a well-defined

fault and structural inconsistencies appear. This can be explained by the interpolation path exiting the discrete class of “presence of a fault”; all intermediate steps remained in the low-density region and then returned to another discrete cloud with a fault in a different location. In the case of AHM, it means the optimisation process needs to stay within the validity domain of the LS point cloud to stick to the prior — the presence of a fault.

To track how the channel “travels” through the cube volume, a second interpolation experiment was conducted (●). It is evident that the channel facie smoothly moves within the volume. The combination of these two experiments (●, ●) confirms the non-Euclidean nature of the LS, in this case of depositional properties LS, we observe a “ring”.

The structural features (⇒) during interpolation become more predictable. The position of the fault remains unchanged during the first and last few steps of interpolation, with only two intermediate steps showing no structural unconformity. However, it should be noted that the fault does not move across the cube during interpolation but rather disappears and reappears. This can be explained by the discreteness of this property, which is more categorical.

The third interpolation example (●) shows a variation moving through the red “tail” of the structural latent space. The experiment illustrates a case of falsification, where we limit the search in the LS. Based on, for example, seismic data, we assume the presence of a fault but do not know the spatial location of the channel. Therefore, correct interpolation should preserve the fault in place since the densest area for movement is precisely there. We see that structural features (⇒) remain unchanged throughout the entire path, while spatial features change. This fact aligns with our expectations that for proper movement in the LS, a tool for controlling the realism of property reconstruction is necessary, which in our case is a density estimation metric, as well as the proper construction of a prior dataset to support the training of the generative network. To conclude, the trajectories along the high-density regions of the LS represent the expected change in model representation.

6.3 Uncertainty of dynamic response and history matching under structural uncertainty

To test how the dynamic response depends on structural and spatial uncertainty parameterised through the hybrid LS, sensitivity analysis was conducted, similar to **Section 5.3: Uncertainty of dynamic response and history matching**. A simulation model was prepared for the experiments with a *line drive* well scheme, figure 5.11. The fault was defined as a conducting one to emphasise the importance of displacement. Simulations were performed using tNavigator software by the Rock Flow Dynamics with parameters, presented in the table 4.2 of the **Chapter 4: Modelling uncertainty and datasets construction pro-**

cess. It is important to note that, in addition to the GWAE generated grids of porosity and permeability properties, GWAE generated structural properties are also incorporated into the flow simulation model. The *Decoder* generates an irregular graph on the output, where nodes only contain active grid cells. Therefore, the reverse operation of adding inactive cells was necessary before conducting simulations to make the model regular. Currently, this is a limitation of the approach that requires the implementation of graph-based simulators (Sanchez-Gonzalez et al., 2020) or simulators capable of working with irregular grids (Y. Wang et al., 2020).

6.3.1 Uncertainty of dynamic response

To conduct experiments with uncertainty estimation of flow response based on changes of the hidden coordinate in the region of the hybrid LS (●), a point in the LS was selected as the reference (●) against which the variability of the response was measured. Multiple alternatives were generated for this ● by adding Gaussian noise (●) and the nearest options from the training dataset (●). All the hidden model representations were reconstructed by the *Decoder* into geological realisations, and simulations were performed, figure 6.8.

For each geological realisation, a map of the average pressures at the last time step of the calculation is presented, as well as the production profile for production well P3 and injection well I2 since they almost always intersected with the channel, the rest of the wells are not presented here because production is zero or negligible due to poor reservoir properties. It can be seen that factors such as the spatial location of the channel, its continuity, and structural unconformity in the form of fault and block displacement significantly impact the flow. Therefore, further experiments with inverse problem solutions are impossible without considering depositional and structural uncertainty.

6.3.2 History matching through the hybrid latent space

The next step is the conditioned search through the hybrid latent space to solve the inverse problem. Just like in **Chapter 5: Graph Wasserstein Autoencoders and channelized synthetic case**, the conditioned search is based on two types of data: well log curves and well production data. Information about the presence of a fault and its position is not provided. Therefore, the dynamic response can be influenced by both the spatial location of the channel and structural features. Moreover, the search will be conducted in the hybrid LS, meaning a 78-dimensional latent model representation consists of two parts simultaneously. As mentioned earlier, this approach combines uncertainty in two geological domains (depositional and structural), enabling optimisation in a unified parametric space.

The key point for optimisation is the homogeneity and density of the LS. If the part re-

6.3. UNCERTAINTY OF DYNAMIC RESPONSE AND HISTORY MATCHING UNDER STRUCTURAL UNCERTAINTY

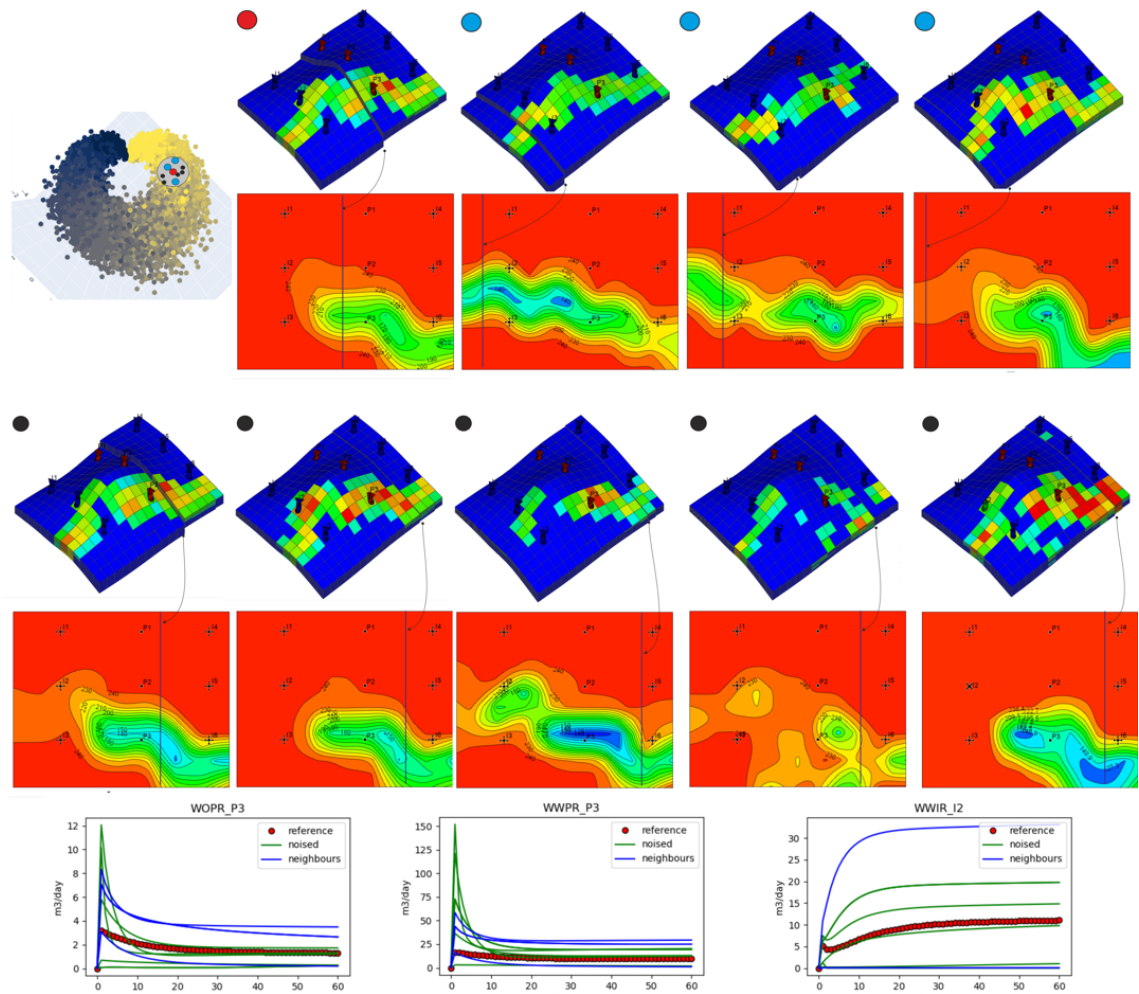


Figure 6.8: Visualisation of the uncertainty of the dynamic response of different geological realisations. A region in the hybrid LS (●) was selected where the realisation ● identified as the reference. Different noise levels were added to the latent vector of the reference, and it was reconstructed by the *Decoder* (●). Additionally, three neighbours were selected in the LS (●). All these realisations were simulated; pressure maps and flow rate for wells P3 and I2 are presented.

6.3. UNCERTAINTY OF DYNAMIC RESPONSE AND HISTORY MATCHING UNDER STRUCTURAL UNCERTAINTY

responsible for the uncertainty of the spatial distribution of porous matrix properties appears homogeneous, then the structural part is likely not, which can be a source of problems. It's important to reiterate that the flow response in this case depends on two types of uncertainty - depositional and structural. Hence, even if you manage to find a spatial distribution of properties to match production history without considering structural features, such a model won't have predictive power. In other words, it's crucial to find a balance between depositional and structural uncertainty in the optimisation process to adapt the model to the observed data successfully.

To conduct an experiment of a model history matching in the presence of structural uncertainty, a reference model with structural unconformity dividing the geological model into two isolated volumes was selected. I simulate the following scenario: no high-quality seismic information is available for the studied geological object, resulting in significant structure uncertainty. Well-log curves indicate the dominance of the channel facies in some wells. Additionally, production data and tracer studies are also available, which gives us information about well flow connectivity. One of the examples is presented by Arnold et al. (2012), where tracer studies data (Produced Water Chemistry) reveals quality improvement HM and UQ.

NOTE: Tracer studies in wells involve using chemical tracers to track the movement of fluids within the subsurface. These studies are conducted by injecting a known quantity of tracer material into a well and monitoring its movement over time. The tracers can be dyes, radioactive isotopes, or stable isotopes that can be easily detected and quantified. Tracer studies in wells aim to gather information about fluid flow pathways, connectivity between different reservoir zones, and the efficiency of various recovery techniques. By analysing the concentration and arrival times of the tracer at different monitoring points, valuable insights can be gained regarding the behaviour of fluids in the subsurface.

Before describing the results, it is worth reminding the HM process. I selected an instance from the prior dataset and extracted static well data in the form of well logs and dynamic data in the form of production profiles. During the HM process, a family of 78-dimensional latent geological model representations is sampled, which are then transformed into full-scale geological realisations by the *Decoder*. Static and Flow losses are calculated, and at this stage, an estimation of the density of the hidden space sampling area is known (Realism loss). According to the **Section 3.7: Objective function in HM process**, the HM process performed in two stages, the Static loss is minimised first with realism control, without evaluating the Flow loss. This saves computational resources since calculating the Flow loss requires simulations, table 6.1. Generating a geological model is nearly instantaneous compared to flow calculations. Therefore, until we find geologically satisfactory

6.3. UNCERTAINTY OF DYNAMIC RESPONSE AND HISTORY MATCHING UNDER STRUCTURAL UNCERTAINTY

realisations, there is no need to calculate the flow. Only in the second optimisation stage is a complete calculation of the OF performed, which includes evaluating the Flow loss, Static loss, and Realism. Calculating the Static loss and Realism is necessary throughout the entire process of solving the inverse problem to maintain geological realism.

Table 6.1: CPU resources needed for the first optimisation step with and without Flow loss calculation

CPU time for:	with Flow loss	without Flow loss
One model generation (model/sec)	0.07	7
Static opt. stage (min)	616	6

Figure 6.9 illustrates an example of a two-stage inverse problem solution. The realisation, marked with a ● is taken as the reference. Four wells (I1, P2, I5, I6) intersected the facies of the river channel with good reservoir properties. It's important to note that a fault and displacement separate the central part of the model, making two blocks isolated. This is confirmed by the pressure map and streamline scheme. There is a connectivity between the producing well P2 and the injection wells I5 and I6, while well I1 is isolated.

In the first stage of HM, Static error is minimised with Realism control. Examples of such realisations are represented by ●. It's worth noting that the spatial distribution of the channel facies is maintained; however, even at this stage, geological realisations do not look very promising due to the high degree of noise in petrophysical distribution. Structural features are not represented at all, i.e., no block displacement, so fault can't be identified. The heterogeneity of the 78-dimensional space explains this generation quality (it will be shown in the next chapters that 78 dimensions are not many for the optimisation algorithm; homogeneity of the LS is much more important). During optimisation, the algorithm searches for a set of hidden model representations that would minimise the error. Hence, a trade-off exists between the Static and Realism components of the OF. The Realism component tries to keep the sampling in dense areas of the LS, preventing the optimiser from "jumping" from one cloud to another. Moreover, the categorical nature of the hidden structural representations responsible for the fault position complicates the search.

When it comes to the flow response of the reservoir, it's essential to note the connection between well I1 and P3, which should not exist due to block displacement in the reference model.

● represents realisations after the second stage of HM, taking into account the dynamic response of the wells. It is noticeable that the quality of geological realisations has deteriorated – the channel appears more fragmented. However, the quality of the dynamic response has improved – the connection between I1 and P3 has disappeared, and the connection between P2 and I5 and I6 has strengthened. This is explained by the fact that the

6.3. UNCERTAINTY OF DYNAMIC RESPONSE AND HISTORY MATCHING UNDER STRUCTURAL UNCERTAINTY

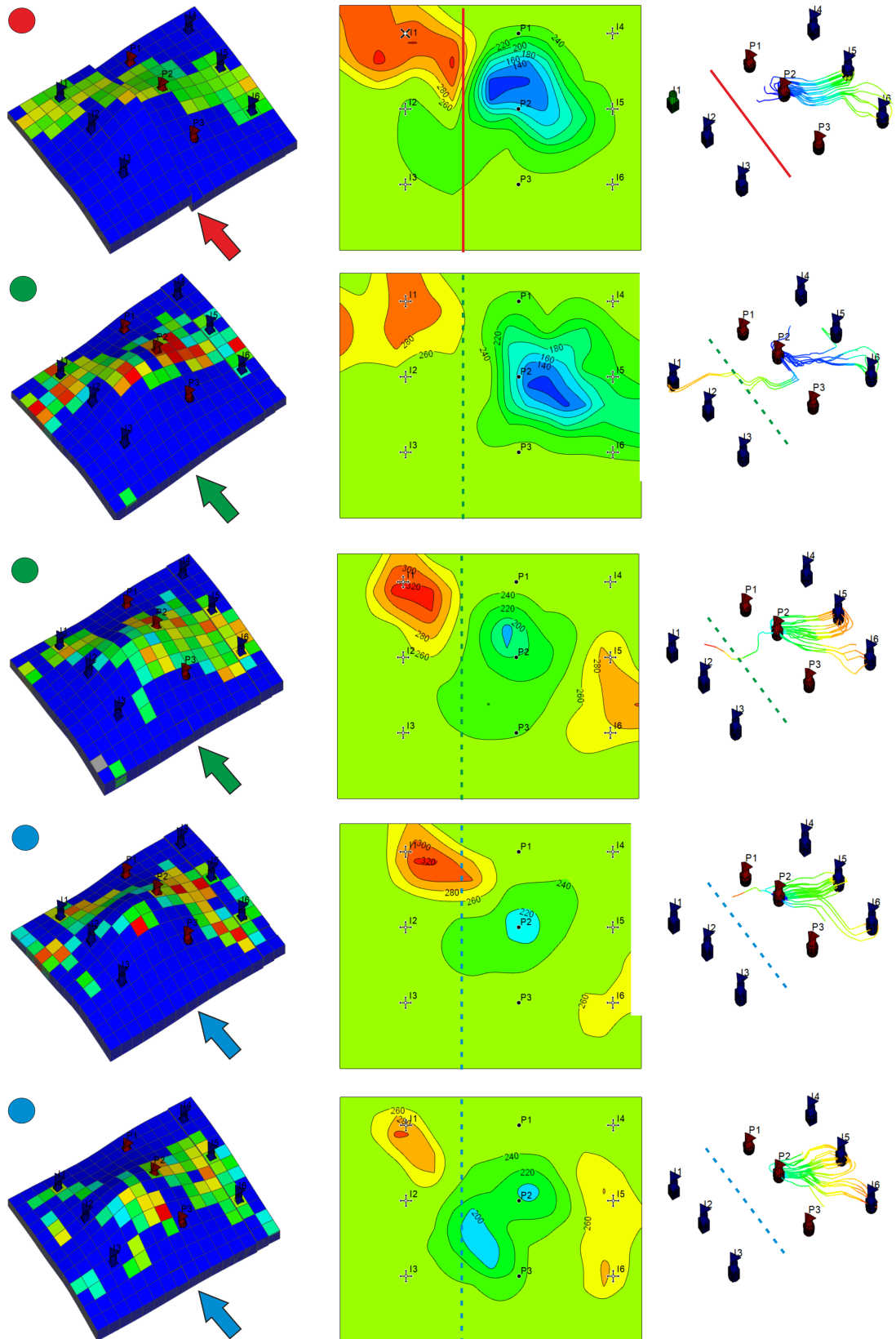


Figure 6.9: Results of the two-stage history matching. The left column shows porosity cubes. The centre column displays pressure maps at the final calculation step. The right column presents streamline visualisations, reflecting the dynamic connectivity between wells. ● - reference model. ● - examples after the static optimisation stage. ● - examples after the static and dynamic optimisation stage.

6.3. UNCERTAINTY OF DYNAMIC RESPONSE AND HISTORY MATCHING UNDER STRUCTURAL UNCERTAINTY

optimiser found realisations that helped improve the dynamic response. Still, it did so by isolating wells I1 and P2 and removing the channel facies, thus creating an impermeable barrier, which was achieved through structural unconformity in the reference model. This result again highlights the importance of constructing a homogeneous latent space that accounts for structural uncertainty.

Figure 6.10 shows a more detailed HM with the example of wells P2, I5. It is worth noting that the optimiser could not perfectly match the static distribution of porosity along the wellbore, as well as the production and injection rates. However, the static matching stage captured the trends in property distribution — it is noticeable that along the wellbore of well P5, high porosity values prevail in the upper part of the reservoir, while porosity is almost absent in the lower part, which is consistent with the reference model data. Since the dynamic error is not calculated at this stage, the production data differs significantly from the reference model.

During the dynamic HM stage, production data match was significantly improved without critical loss in the quality of static property distribution along the wellbore. However, it is worth noting that the overall geological realism decreased due to increased Realism error.

To summarise the model HM process under structural uncertainty, the optimisation process was able to reproduce the main depositional features of the reference geological model. The trend of channel facies distribution and compartmentalisation of blocks was achieved. However, unlike the reference model, compartmentalisation was achieved by interrupting channel propagation rather than through structural inconsistencies. As for local properties in wells, there is also convergence with the overall trends, but a detailed match was not achieved.

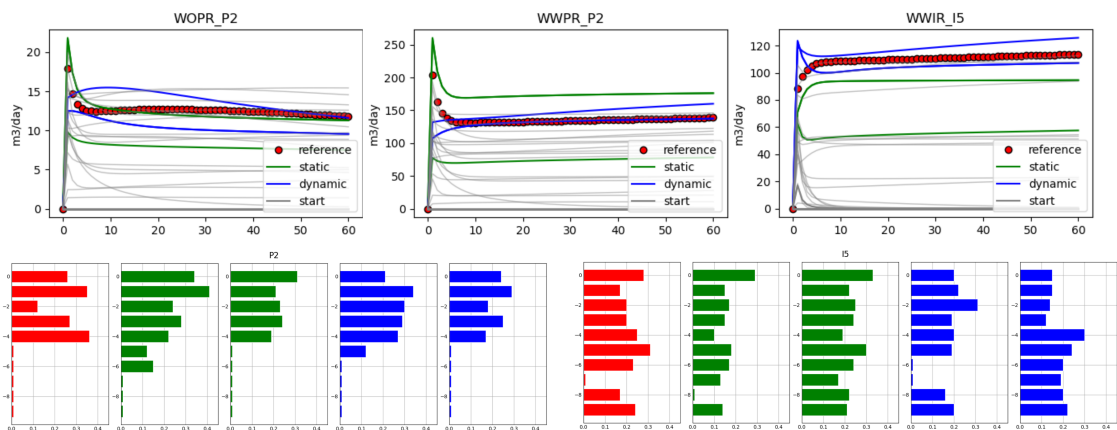


Figure 6.10: Examples of wells P2 and I5 match are marked in green for optimization after the static stage and in blue for HM after the entire optimisation process. Reference is depicted by the red colour, gray - initial optimisation step. Top: dynamic mode response. Bottom: static well data match for each model layer.

6.4 Conclusion

In this chapter, the generative capabilities of GWAE were tested using geological settings that include structural uncertainty and uncertainty in porous property distribution. For this purpose, two generative networks were trained independently, one responsible for the structural features of the reservoir and the other for filling the generated model with properties.

The analysis of the hybrid LS, consisting of structural and depositional components, showed that the depositional part of the latent space forms a ring. Similarly to what *FLUVIAL* example demonstrated in **Chapter 5: Graph Wasserstein Autoencoders and channelized synthetic case**, by interpolating along this ring, instances with predictable channel locations can be generated. The second part of the hybrid LS, responsible for the structural unconformities, showed high heterogeneity due to the categorical nature of the features related to fault positions. Interpolation between regions of the LS is possible but usually leads to the disappearance of structural features (i.e. presence of fault and block displacement) in the generated realisations at intermediate steps. Overall, the generative capabilities of the hybrid GWAE allow us to conclude that by staying in dense regions of the LS, we can expect high-quality realisations that reflect both structural and spatial uncertainty.

The experiments on matching the model confirmed conclusions that geological realism control is a key factor in the *Decoder's* performance quality. Global trends in property distribution and flow connectivity between wells were reproduced, but this was achieved by partitioning the channel into parts rather than through structural features. In many ways, this was achieved through a trade-off between minimising the error in reproducing static properties along the well, dynamic well response, and overall realism control. However, finding an implementation close to the reference was hindered by the heterogeneity of the hidden space. In general, it is worth noting that a key factor in ensuring the generative capabilities of GWAE is not only the architecture of the network itself and the choice of the dimensionality of the hidden space but, above all, the internal structure of the prior. If the prior has a heterogeneous structure, this will affect the generative capabilities of the network, as regions poorly covered by training examples will produce low-quality and unpredictable results.

Chapter 7

Graph Wasserstein Autoencoders for AHM Brugge case

Chapters **Graph Wasserstein Autoencoders and channelized synthetic case** and **Graph Wasserstein Autoencoders and fault synthetic case** demonstrated the generative capabilities of graph neural networks using simple geological settings. Experiments were conducted to test the effectiveness of the approach applied to cases that represent uncertainty of geomodelling parameters across different geological scenarios and structural uncertainty.

This chapter aims to examine the generative capabilities of GWAE in a more geologically realistic model description of heterogeneity of porous properties. A semi-synthetic benchmark case of the Brugge field was chosen for history matching. The field data is presented in **Section 4.3: Brugge benchmark**.

7.1 GWAE under uncertainty of modelling approach

Brugge reservoir comprises four stratigraphic zones (Schelde, Maas, Waal, Schie), each characterised by geological uncertainty. In this work, I primarily focused on uncertainty related to geological reservoir characteristics. I built a prior training dataset based on our understanding of the reservoir and key uncertainty that may later influence any decision-making process. However, in constructing the prior, I utilised well-known geostatistical algorithms, each with advantages and limitations. This raises the question of how we would know which specific algorithm is suitable for modelling our prior. To what extent do the modelling algorithm assumptions constrain our prior dataset? Can we use a combination of geostatistical modelling algorithms to construct the prior, and how will it impact the generative capabilities of GWAE? While the first two questions have a more general

theoretical nature, we will attempt to answer the third question through experimentation.

Brugge TNO prior ensemble utilises two geostatistical algorithmic scenarios: SIS/SGS and Object Modelling/SGS for modelling fluvial depositional environment in the upper zone Schelde. Examples of training dataset instances are shown in figure 4.16. Therefore, I trained the generative network using the training dataset composed of a combination of 2500 instances of Object Modelling and 2500 instances of SIS for the Schelde layer.

7.1.1 Combined LS analysis

The visualisation of the LS generated as a result of training the GWAE model is presented in figure 7.1. The dimensionality of the hidden space is 100.

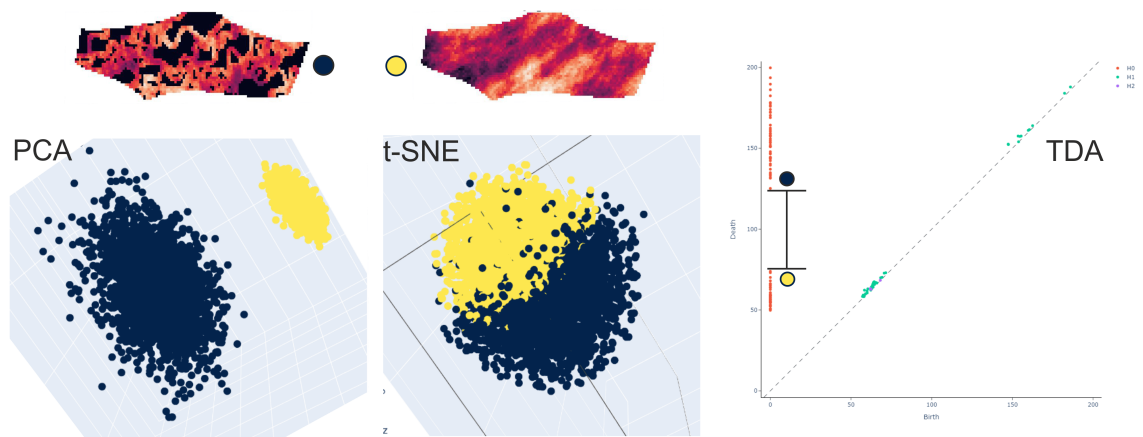


Figure 7.1: from left to right: PCA, t-SNE and TDA visualisation of the combined LS for Schelde zone. ● - region of SIS samples; ● - the region of Object Modelling samples.

In the PCA projection, two distinct clusters of points can be observed, which are significantly separated from each other. The colouring of these clusters reflects the affiliation of the hidden vector to a specific type of geostatistical modelling algorithm — SIS or Object Modelling. The t-SNE projection is presented as a densely packed cloud composed of two hemispheres. However, it is worth noting that without additional labelling, it would be difficult to determine that the space is divided into hemispheres. In other words, if we did not know the structure of the training dataset in advance, determining such a structure through t-SNE projection would be practically impossible. TDA analysis also provides a reasonably clear answer about the internal structure of the latent space. It is divided into two large regions, each with a different density. When looking at the H_0 space, the lower cluster of points corresponds to SIS realisations, which are tightly packed together, likely representing quite similar objects. The second cluster of points is more sparse and pertains to the part of the space associated with Object Modelling.

Based on the analogy with **Chapter 6: Graph Wasserstein Autoencoders and fault synthetic case**, it can be concluded that such a discrete hidden space presents a challenge

for further tasks related to HM. There is a need for a tool to “teleport” from one cloud to another or a tool that can make the LS a unified dense object (Drozdova et al., 2021; Giovanis et al., 2024).

7.1.2 Generative quality of GWAE with combined dataset

To assess the generative ability of the GWAE trained on the combined dataset, I selected an instance of a training sample in the hidden space, which was initially generated using the Object Modelling method. Then I reconstructed it using the *Decoder*.

Figure 7.2 shows an example of a training sample of porosity and permeability cubes generated using the Object Modelling method. This sample was then encoded into a 100-dimensional model representation in the LS and decoded back to the original representation using the trained GWAE. As can be seen, the spatial features were completely lost and resembled the reproduction of the SIS method, which is based on variograms. However, it is worth noting that the porosity-permeability relationship was reproduced in the reservoir values after decoding. It is important to note that Object Modelling utilises two facies: river channel (sand) and shale. For the shale facies, which do not act as a reservoir, all porosity and permeability values were taken as constants, as seen on the dependency graph. In SIS modelling, facies were not explicitly used, and properties were distributed according to variograms. The GWAE captured a more global and continuous poro/perm dependency during training. It should be noted that, in practice, before flow modelling, a cut-off is often applied to eliminate cells with poor reservoir properties that have minimal influence on flow. This cut-off replaces all cells below a certain threshold of porosity or permeability with a constant value. Therefore, the green tail in the low porosity values would still be cut off as a non-reservoir.

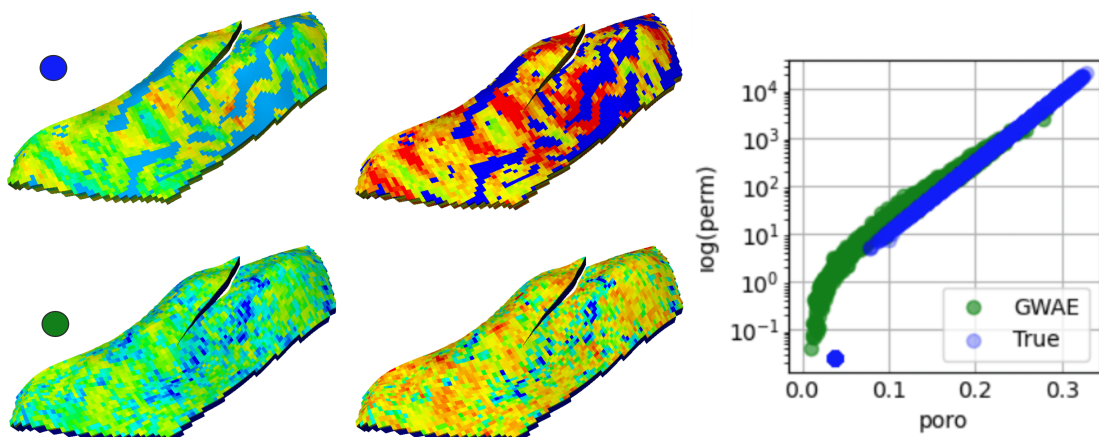


Figure 7.2: ● - example of porosity and permeability grids from the training dataset, modelled by Object Modelling approach; ● - GWAE generated sample, which should be close to the training example. Right: porosity/ log(permeability) dependency for true and GWAE generated samples

Based on the conducted experiment, it can be concluded that the GWAE model could not effectively learn to reproduce the training samples with the desired level of geological realism. This was primarily due to the inherent structure of the training dataset, which consisted of two distinct groups of scenarios that had no overlap. During training, the LS was divided into two clouds, but the hidden vectors were not informative enough for the *Decoder* to perform accurate generation. This experiment demonstrated that if the internal structure of the prior training dataset is discrete, it is better to train separate generative networks for each discrete group rather than attempting to train a single universal generator. Another approach would be creating continuous prior datasets where one class smoothly transitions into another, as shown in the **Chapter 5: Graph Wasserstein Autoencoders and channelized synthetic case**.

7.2 Generative quality of GWAE with Brugge dataset

Céline Scheidt (2018) discusses the concepts of epistemic uncertainty, fuzzy logic, and vagueness. Epistemic uncertainty refers to the uncertainty that arises in the process of forming scientific knowledge and research. They demonstrate that scientific knowledge always contains a degree of uncertainty, which is associated with the limitations of our cognitive abilities and the insufficiency of information or incomplete observability. They also address fuzzy logic, a theory that formally describes fuzzy concepts and fuzzy sets. Fuzzy logic allows for considering uncertainty and fuzziness in problem-solving, where logic and traditional methods are applied. Vagueness is related to uncertainty in language and concepts. Vagueness describes a situation where the boundaries of concepts are blurred or vaguely defined, which can create difficulties in precisely defining and understanding these concepts. These concepts enable us to build more realistic and adaptive models and research methods.

In geoscience, classification systems like depositional environments (deltaic deposits, fluvial deposits and so on) are extensively utilised. However, it is crucial to acknowledge that these classifications are human-made constructs, which introduces vagueness into the equation. The difficulty lies not just in establishing precise boundaries but also in the sequence and variants. As a result, inherent ambiguity arises when determining the point where the delta ends and the river channel begins. Furthermore, geologists may provide interpretations based on their personal experiences and levels of expertise, leading to a range of perspectives on the nature of “deltaic” phenomena. Fuzzy logic, unlike a binary approach of simply voting “for” or “against”, enables a more nuanced evaluation from each participant, allowing for a smoother transition between the two classifications.

In my work, the prior is constructed from the perspective of precise classification: this zone corresponds to a fluvial setting (Schelde), this one to the upper shoreface (Waal), and this

one to the lower shoreface (Maas) with defined boundaries. In reality, these boundaries do not exist in nature, so such an approach to constructing the prior initially restricts our understanding of the reservoir and subsequent modelling. It is worth noting that constructing the prior using fuzzy logic approaches is a complex task that requires separate research.

Thus, to overcome the limitations imposed by the categorical nature of the training dataset, four separate GWAEs were trained for each zone (Schelde, Maas, Waal, Schie), with LS dimensions of 30, 50, 30, and 50, respectively. The overall dimensionality of the hybrid LS is 160.

Figure 7.3 demonstrates the quality of generating geological realisations for each of the four generative neural networks using the example of porosity property. Each row of the figure represents a specific zone, with the original instance from the training dataset (●) and the reconstructed realisation using GWAE (●). Only the first layer of each zone is clearly visible in the figure, so to show that the properties are correctly reproduced throughout the volume, I subtracted the reference instance from the generated one (●). The histogram shows that the error of reproducing the porosity property is normally distributed with an average of zero. The porosity-permeability relationship is also accurately reproduced with a high level of quality and was implicitly identified by the neural network, hence it is reproduced with noise. To confirm the correctness of the property distribution spatially and in depth, the following figure 7.4 shows a cross-section of porosity property between wells BR-P-15 and BR-P-16 of the same original instance and a reconstructed realisation using GWAE. As can be seen, the wells are located on different sides of the fault. Overall, it can be seen that regardless of the depositional environment and reservoir volume, each GWAE was able to learn to generate objects. The variogram analysis shows a resemblance of the initial and reconstructed spatial distribution of porosity properties for each zone, figure 7.5. In addition, an increase in the nugget parameter for the variogram of the difference between the porosity cubes of the original and reconstructed models indicates a high level of variation at a small scale. Combined with a quick approach to sill, this indicates a weak spatial dependence of error propagation.

Therefore, such an ensemble of GWAEs can be further used to solve the inverse problem. It is important to note that this ensemble approach is applicable only in the case of zone independence, i.e., the realisation of Schelde should not affect Maas and so on. When zones depend on each other, it is necessary to train GWAE crosswise or train one extensive network (the limitations of such an approach have already been highlighted).

7.2.1 LS analysis

The analysis of the internal structure of generative networks' LSs for each depositional environment layer did not reveal any significant features. Due to the prior dataset being

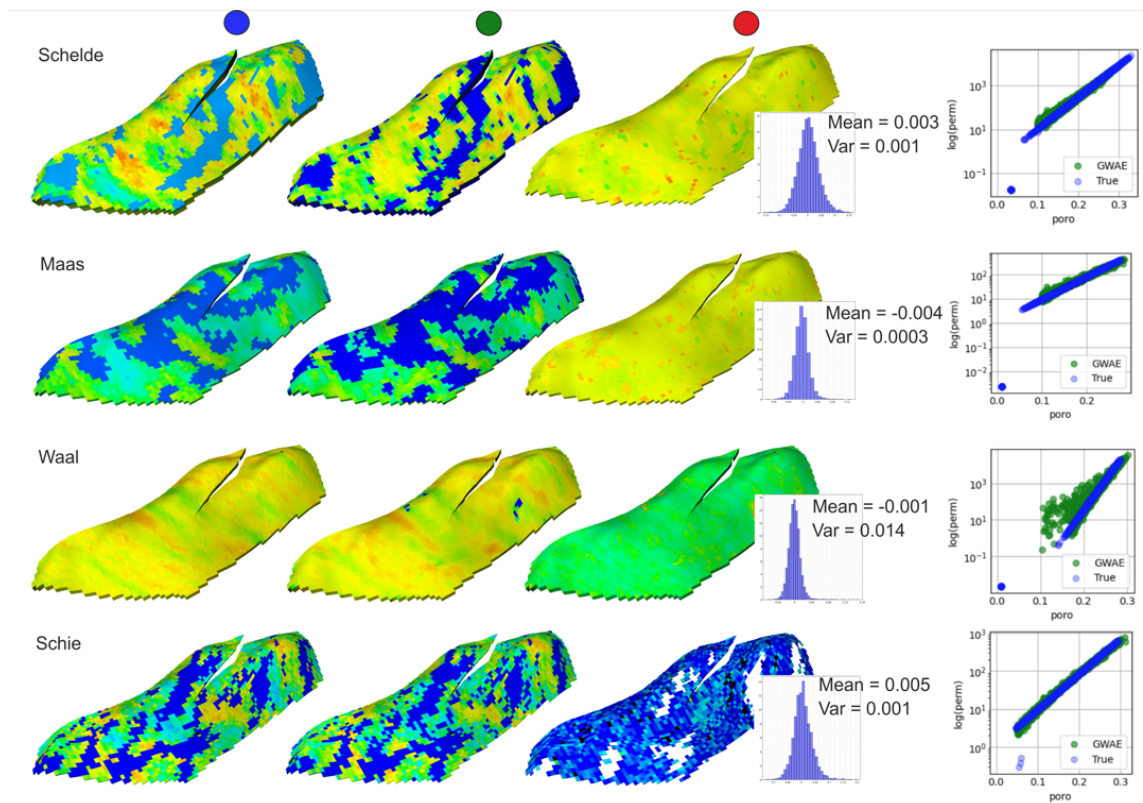


Figure 7.3: Examples of the initial ●, decoded reconstruction ● of each horizon. Their difference ●. Porosity/Log(Permeability) dependency

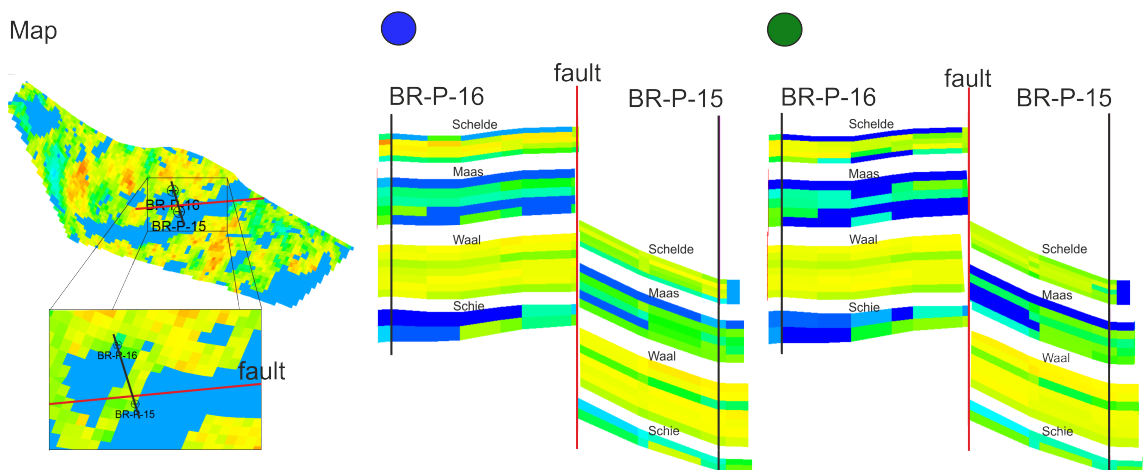


Figure 7.4: Left: the map of Brugge field to locate the cross-section position between wells BR-P-15 and BR-P-16. Centre: Cross-section of the initial realization ●. Right: decoded reconstruction ● of each horizon

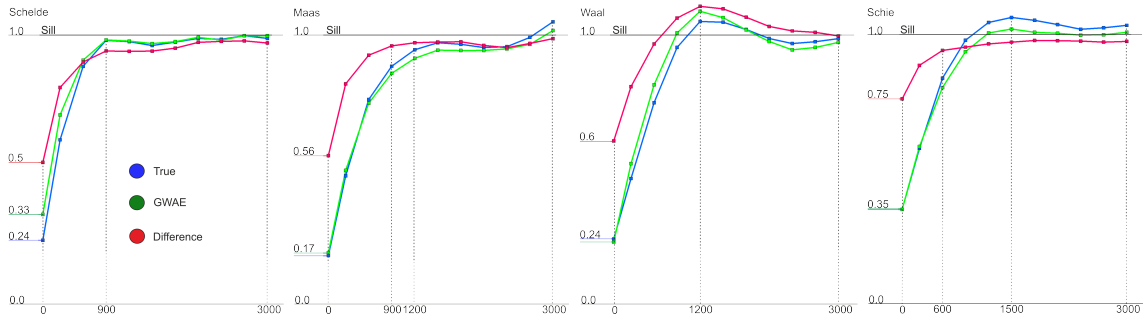


Figure 7.5: Variograms of initial porosity property - ● and decoded reconstruction - ● for each horizon. Variogram of the error of reconstruction - ●

generated as a unified distribution, it becomes challenging to visually identify any patterns or regularities in the structure of the LS. The complexity of geological objects has made it difficult for three-dimensional mapping of the multidimensional set of hidden model locations in the LS to reveal any specific visual correlations. Figure 7.6 (left column) shows an example of the LS in the PCA and t-SNE projections for the Schelde zone. It is worth noting that the three principal components of the PCA projection contain 18, 12, 12, 7 per cent of the total variation for Schelde, Maas, Waal and Schie respectively. Providing visualisations for other zones is not meaningful because they look similar. The colouring reflects the variation of one of the uncertainty parameters, and this colouring does not change globally when the uncertainty parameter is changed. Thus, it can be concluded that each hidden space of the generative network has a homogeneous structure, which is a good sign for solving the inverse problem. As for the internal structure, it was not possible to identify it.

It is worth noting the visualisation of TDA. In the H_0 projection, it is noticeable that points accumulate in certain intervals for each hidden space, which differ significantly. Two factors can explain this: firstly, each hidden space has its density, i.e., connected components start to appear and disappear when reaching the radii of multidimensional spheres of different sizes. Secondly, the dimensions of the hidden spaces for Schelde and Waal are 30, while Maas and Schie are 50. Therefore, the original dimensionality makes densities different. Additionally, for the H_1 projection of Schelde and Schie, we can see a clear indication of a global structure in the form of a ring, as there are points noticeably deviating from the diagonal.

7.3 HM of Brugge field with GWAE

The Brugge is a valuable benchmark for testing and comparing techniques in history matching, production optimisation, and closed-loop optimisation. Its widespread use in literature underscores its significance. The benchmark's key attributes that enhance its

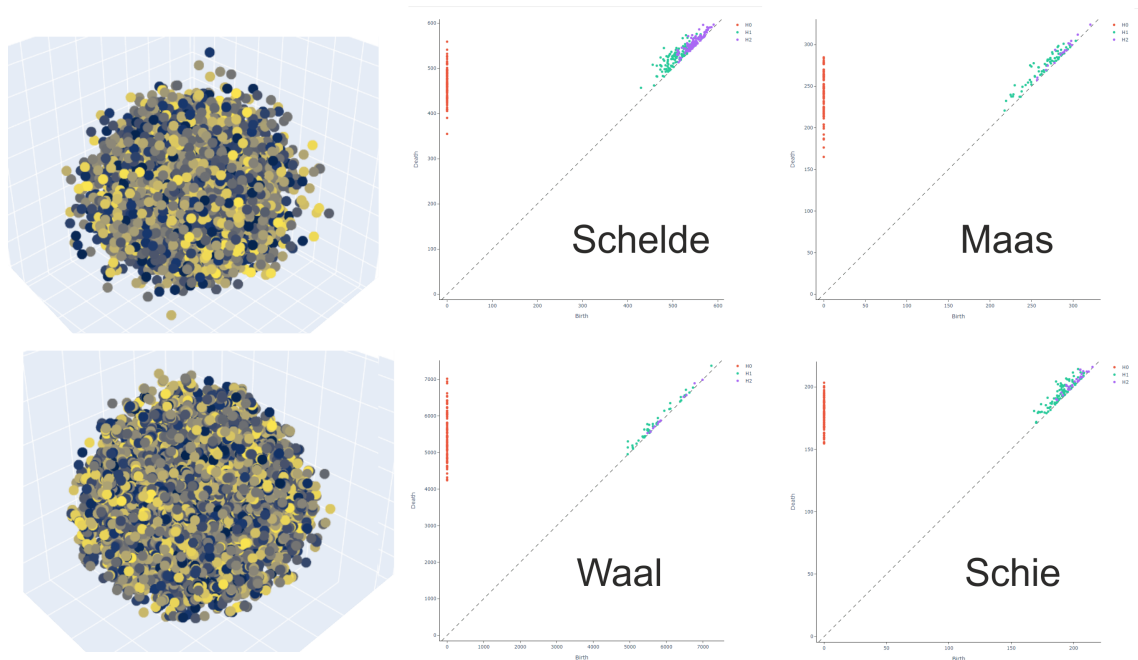


Figure 7.6: The left column: the visualisation of the LS using PCA (top) and t-SNE (bottom) projections for the Schelde zone. The colouring represents the variation of one uncertainty parameter. The left part shows the visualisation of the LS using TDA projection for all four zones. It is worth noting the homogeneity of the density of each LS and the difference in densities between the LS.

utility for history matching include the substantial scale of the problem, the abundance of data, and the intricate nature of the detailed truth model, leading to realistic challenges. Regarding optimisation, the effectiveness of the Brugge field case is due to the complexity and uncertainty of the reservoir model, the need to optimise multiple control parameters, and the presence of numerous constraints governing these variables.

Schulze-Riegert et al. (2017) investigated the Ensemble Kalman Filter (EnKF) (Aanonsen et al., 2009) approach for field HM. The approach consists of a prediction phase and an assimilation phase, where the variables characterising the state of the system are adjusted according to observations. The geological model is represented as a vector containing porosity, permeability and NTG data, which are used to predict the dynamic response. The predicted data may include bottomhole pressure, water recovery, gas-oil ratio, etc.

Mohamed et al. (2010) explores the utilisation of Particle Swarm Optimization (PSO) algorithm in history matching the Brugge reservoir. PSO is employed to discover a variety of history-matched models for the Brugge reservoir, assessing uncertainty in production forecasts. The paper delves into the parameters and workflow for history matching, incorporating the use of PCA for parameterisation. By comparing with EnKF, the paper demonstrates PSO's capability to address large history matching challenges and achieve comparable outcomes. It concludes that both PSO and EnKF have their own strengths and weaknesses, with each algorithm being suitable for specific problems. However, both

methods are noted to suffer from a lack of geological realism during the matching process.

Demyanov et al. (2015) proposes an approach to Brugge HM using a Multiple Kernel Learning (MKL) approach with PSO that integrates different types of data (such as seismic, well data and geological concepts). The approach has an important advantage compared to EnKF and PCA-PSO, as it allows to obtain interpretable parameters, and to determine the importance of each of them. As in previous papers, the authors use the algorithm to model the porosity and permeability fields, and PSO was used as the optimiser. As a result, the quality of the HM was similar or superior to previous works.

Chalub et al. (2023) and Cruz et al. (2022) have combined production, tracer, and 4D seismic data to calibrate reservoir models, improving forecasting accuracy. Their flexible workflow can incorporate various data types for further enhancements. This research highlights the benefits of using multiple datasets in history matching, showing that the integrated approach is valuable for field studies. The study employed the Iterative Ensemble Smoother based on a Regularized Levenberg-Marquardt Algorithm (IES-RLM) as a history matching method. This method aims to generate a set of reservoir models that closely minimise nonlinear least squares without requiring simulator restarts and involving fewer model variables updates during history matching compared to EnKF. Practical implementation of IES-RLM is considered simpler than EnKF for history matching tasks, and empirical evidence suggests that IES-RLM outperforms EnKF and ES. A key focus and contribution of this study is to showcase the advantages of integrating tracer and 4D seismic data alongside production data in an ensemble-based history matching workflow through a 3D field-scale case study.

Before starting the description and analysis of the results of the HM process of the Brugge model through the hybrid hidden space, let's briefly describe the principle scheme, figure 7.7. During the training process, every grid from the dataset was converted into graph representation and fed into the GWAE model. GWAE was trained for each horizon, resulting in a set of 4 generative networks. I plug the trained composition of GWAE model generator into the history matching framework where the optimisation is done by sampling from the

lower dimensional latent space, then generating new models through decoding. The optimiser looks to generate new models (porosity and permeability properties) in the LS that

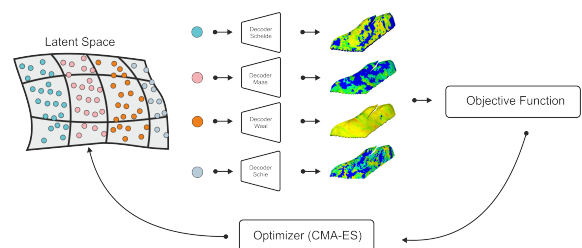


Figure 7.7: General scheme of inverse modelling through the hybrid LS. The hybrid LS consists of four regions (●○●○) for each stratigraphic zone. Latent vectors are sampled and fed into a stack of *Decoders* to generate geological realisations. After OF assessment, the optimiser made the next step in the LS.

comprise a graph for every horizon and comes up with the match solution by stacking graphs into the stratigraphic succession. Usually, there is no guarantee that the AHM will provide an ensemble of solutions that would infer the response of the unknown. Ideally, our prior should be constructed in such a way as to cover the entire space of uncertainty and include an unknown geological representation which we are trying to find. If this does not happen, the “true” geology is outside the hidden space, and the generator will not reproduce it.

I picked the reference model, which will be considered unknown except for petrophysical and production well data. The Covariance-Matrix Adaptation Evolution Strategy optimisation algorithm was chosen to search the latent space. After the CMA-ES step, the algorithm generates a new set of latent vectors for decoding until convergence.

I implemented two-stage optimisations. In the first stage, the optimiser minimises the MSE error between static well log data of the reference model and GWAE generated representations, which is relatively fast because it goes without any simulation process. In the second stage, I optimise in the vicinity of the found solution by minimising the MSE of both the static log data and dynamic production of water and oil fluid, figure 7.8.

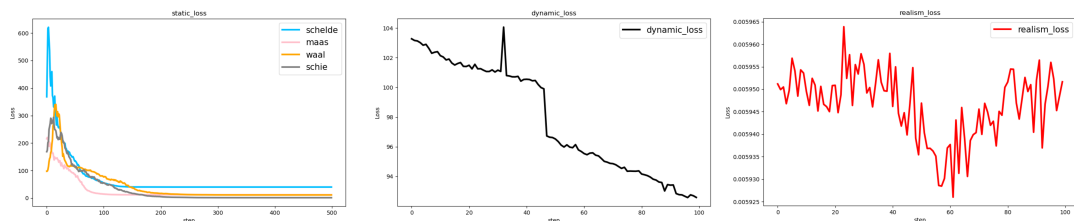


Figure 7.8: Visualisation of the dynamics of error change by components during the HM process. Left: Static error; Center: Dynamic error; Right: Realism error

Three optimisation processes were launched to obtain a set of models that allow assessing the uncertainty of property distribution and dynamic response. If necessary, more such processes can be conducted. The matching results, i.e. the search through the hybrid hidden space of the reference model, are shown in figure 7.9 (left column). The right part of the figure represents the visualisation of the distribution of recovered porosity properties for each zone. It should be noted that the obtained property distribution for the Schelde and Schie zones turned out to be quite noisy, as the training datasets for these zones exhibit a high degree of heterogeneity in property distribution, making spatial dependencies challenging to trace. In addition, the low reservoir quality provides a small portion of production from these zones. Therefore, the variation of hidden positions has little influence on the overall assessment of the loss function. The realisations of the Maas zone turned out to be practically identical to the reference model. The recovered realisations of the Waal zone are characterised by similarity in property distribution trends across the

reservoir, with minor local differences.

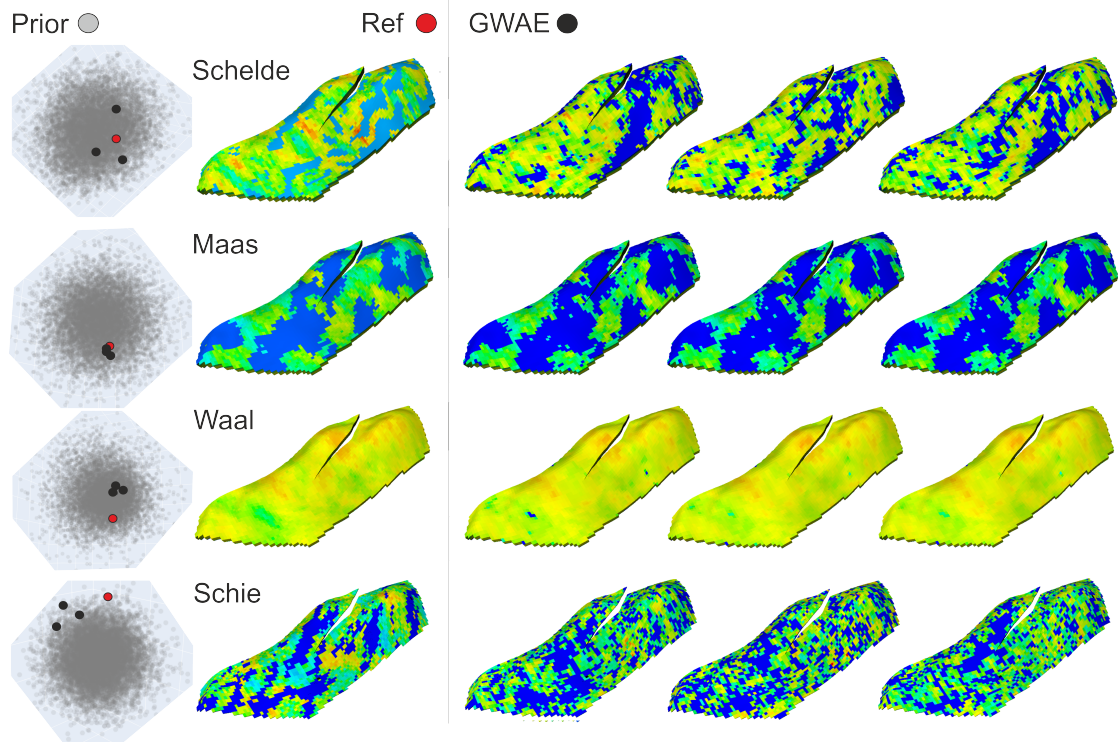


Figure 7.9: Left: A prior (●) of geological realisations represented in the LS (PCA representation). ● - location of the reference model in the latent space. ● - locations of GWAE realisations. Centre: The visualisation of porosity properties for the reference model (●). Right: variants found during the optimisation process (●).

Variogram analysis showed that despite the wide range of variograms in the prior distribution (●) for each zone, the optimisation process achieved very close convergence of the variograms generated by GWAE (●) with the reference model (●), figure 7.10. Let's focus on some specific points. The nugget parameter for the Schelde zone is around 0.4 – 0.5, indicating high variability in properties at small scales, i.e. the absence of short-scale continuity, which indicates abrupt changes in rock properties. In the case of the Schelde zone, this is due to fluvial sedimentation. The variogram for the Maas zone does not reach the sill and has a wavy character. The first sign indicates zonal anisotropy, which is an indicator of cyclicity. The waviness confirms this conclusion. A similar conclusion about cyclicity at large scales can be drawn for the Waal zone, although zonal anisotropy is not observed. The nugget parameter for the Schie zone in the reference model is around 0.45, but the same parameter for GWAE implementations is significantly higher — 0.6 – 0.7. This indicates that the optimised implementations have become much “noisier” than the reference. This confirms some loss of geological realism for the Schie formation, as I mentioned earlier. Variogram analysis was conducted for all zones with the following settings: standardisation (Sill = 1), isotropic, number of lags = 10, search radius = 3000. Type of variogram — Classical, defined as half of the average squared difference between

the paired data values.

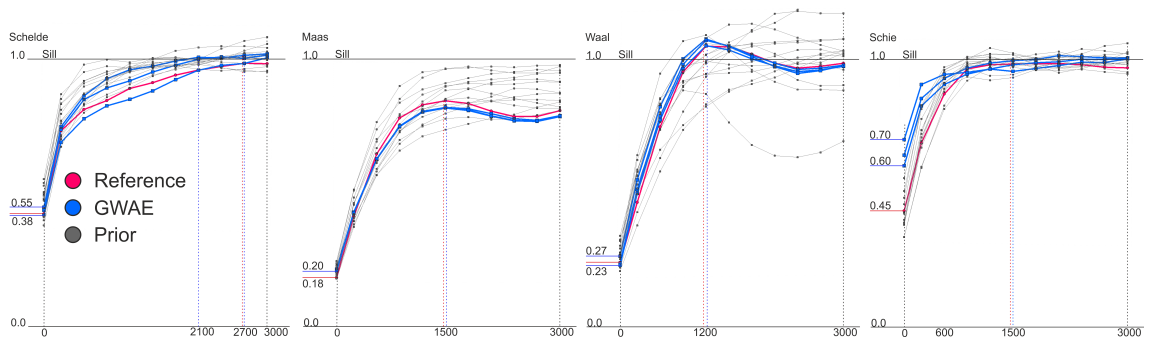


Figure 7.10: Variograms of porosity distribution for four zones of the Bruges field, where ● - prior/initial samples for the optimisation process, ● - reference model variogram, ● - GWAE optimised solutions

Table 7.1 presents the statistical parameters of porosity for each zone and the percentage of total fluid production for each zone. Here, a high level of similarity in the parameters of mean and standard deviation is observed for all zones. At the same time, pore volume and production may vary due to uncertainty in property distribution in the inter-well space and can be considered as uncertainty.

Table 7.1: Main characteristics of porosity property

Realisation	Av Poro* (frac)	Std Poro	Pore Vol ($m^3 10^6$)	Abs Err (%)	% of Total Fluid Prod
Schelde					
Reference	0.21	0.04	53	—	17
GWAE optimised	0.20	0.04	48	9	13
GWAE optimised	0.20	0.04	49	8	13
GWAE optimised	0.20	0.04	47	11	15
Maas					
Reference	0.16	0.04	80	—	7
GWAE optimised	0.16	0.03	77	4	8
GWAE optimised	0.16	0.03	77	4	8
GWAE optimised	0.16	0.03	77	4	8
Waal					
Reference	0.23	0.03	195	—	76
GWAE optimised	0.24	0.02	197	1	79
GWAE optimised	0.24	0.02	198	2	80
GWAE optimised	0.24	0.02	198	2	77
Schie					
Reference	0.18	0.06	25	—	< 1
GWAE optimised	0.18	0.04	23	8	< 1
GWAE optimised	0.19	0.05	27	8	< 1
GWAE optimised	0.18	0.04	24	4	< 1

*Average values for sand only,

The following figure 7.11 illustrates the matching of the dynamic response for the entire reservoir and the wells PR-P-2, PR-P-15, and PR-P-18 as examples. The gray colour represents examples from the prior ensemble of models from which the optimisation process started. There is a slight variation in the dynamic response for oil and water production rates across the field. This can be explained by the fact that the main production zone in the Brugge field is the Waal zone, which exhibits consistent properties throughout the reservoir. However, the initial set of models for individual wells has a much wider range of variation.

It is worth mentioning that the entire HM process is provided by adjusting the distributions of porosity and permeability properties while implicitly preserving their dependencies. Parameters such as relative permeability, which influences the flow ratio of phases (oil, water), remained unchanged. Therefore, there is an imbalance in the matching of oil and water production observed in all the depicted wells.

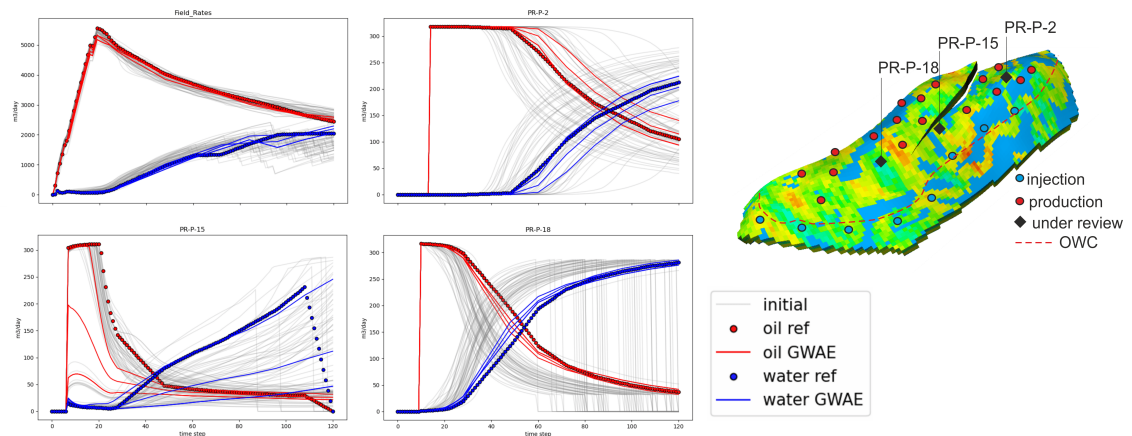


Figure 7.11: Profiles of field and wells PR-P-18, PR-P-15, PR-P-2 production of oil and water in comparison to reference dynamic response)

Figure 7.12 shows the porosity property profiles along the PR-P-2, PR-P-15, and PR-P-18 well bores. There is a very high similarity between the models found by the optimisation process and the reference for all zones. Here, it is worth referring to the work of Ringrose et al. (2015), where he distinguishes between the concepts of *hard* and *soft* conditioning using seismic data as an example. Hard conditioning refers to a method where high-quality seismic data, which is sufficiently resolved at the scale of interest, is directly used to define the architecture of the reservoir model. This approach is applicable when seismic data can accurately represent important elements of the model, such as seismic geobodies. On the other hand, soft conditioning involves using seismic information as a general guide for probabilistic algorithms. Instead of directly defining the reservoir model, the seismic data provides a general understanding or correlation coefficient that guides the probabilistic algorithms in creating the model.

Regarding well-log data, something must be between *hard* and *soft* conditioning. Since the vertical and horizontal resolution of well-logs is only a few tens of centimetres, and they are assigned to a single value per cell, which is much larger than the resolution, it is not accurate to say that well-log can reproduce the properties of the entire rock volume of a cell. On the other hand, using *soft* conditioning only to track property trends is not the optimal use of well-logging data either. Therefore, it is worth allowing some deviations in the generated properties along the wellbore and interpreting them as uncertainty.

It is worth noting that a distinguishing feature of machine learning methods, compared to standard geostatistics, is that ML approaches allow for this intermediate conditioning. In contrast, geostatistics treats well-logs as *hard* data and rigidly constrains the model to the available well data without allowing for deviations.

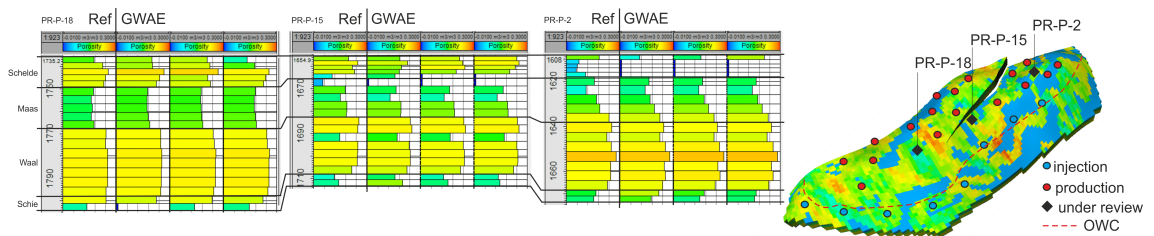


Figure 7.12: Profiles of field and wells PR-P-18, PR-P-15, PR-P-2 production of oil and water in comparison to reference dynamic response)

It can be concluded that the HM framework using the adjustment of porosity and permeability properties does not allow for achieving perfect convergence since it does not provide variation in phase production (water, oil). Additionally, reservoirs such as Schelde and Schie were generated with a noticeable level of noise, indicating that the weight on the geological realism component of the OF could be increased, possibly at the expense of the other two components. However, the search for HM models using a generative model family and a combined (hybrid) latent space of dimensionality 160 allowed finding realisations that meet both static requirements (similarity of well log curves) and dynamic response (production rates of oil and water).

Demyanov et al. (2015) presented a comparative analysis of the quality of HM of the Brugge field using MKL with PSO, PCA with PSO and EnKF methods. Unfortunately, Chalub et al. (2023) and Cruz et al. (2022) do not provide information about the quality of HM for all wells (only the quality of prediction), so these papers are not given in this study. Following figure 7.13 complements the analysis with the results of the GWAE. It should be noted that the comparison is qualitative rather than quantitative. It is noticeable that in general the GWAE approach outperforms previous solutions. At the same time, it allows to preserve and control the geological realism of the spatial distribution of properties.

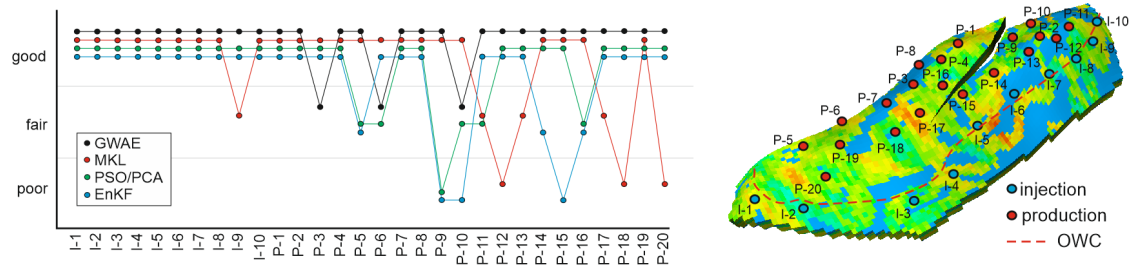


Figure 7.13: A cross-well comparison of the HM quality (good, fair and poor match) of the four methods: ● - GWAE, ● - MKL, ● - PSO/PCA, ● - EnKF. It can be seen from the comparison that in general, the proposed approach outperforms the previous methods.

7.4 Uncertainty of Prediction

Each geological model is created to make investment decisions based on it, such as locating remaining reserves, well drilling planning, or evaluating the effectiveness of reservoir pressure maintenance systems. Multivariate flow simulations allow for uncertainty to be considered that may impact the project's economic feasibility. Flow response predictions were performed on the HM model to conduct such an analysis.

To evaluate the calibrated models under different production modes and to assess the uncertainty of the forecast I considered three different scenarios. The first scenario assumes fixing the bottom hole pressure of production and injection wells at the last step of the historical calculation. The second scenario assumes an increase of reservoir draw-down pressure. This was achieved by decreasing the bottom hole pressure of the producing wells by 10 per cent relative to the last step of the historical calculation and increasing the bottom hole pressure of the injection wells by 10 per cent. The third scenario simulates the opposite mode, where the draw-down pressure is reduced. The bottom hole pressure of the production wells is increased and for injection wells is reduced by 10 per cent.

Then, all wells were put under pressure control and run for a 5-year calculation. As a result, oil and water production rates (figure 7.14), as well as average saturation (S_o) (figure 7.15) and reservoir pressures (P_{res}) maps (figure 7.16) at the last calculation step, were obtained for the reference model and three GWAE models for every scenario. It is important to note that all maps were put in the same scales to provide visual comparison.

Based on figure 7.14, it is evident that the oil and water production profiles for the forecast period are very similar between the reference model and the generated GWAE models. This suggests that the expected production can be anticipated with high confidence. However, some local differences are observed in the pressure and saturation maps, particularly in the reservoir dome and in the water-saturated region to the south. These variations indicate that despite the overall similarity in the maps, there are differences in the reservoir's

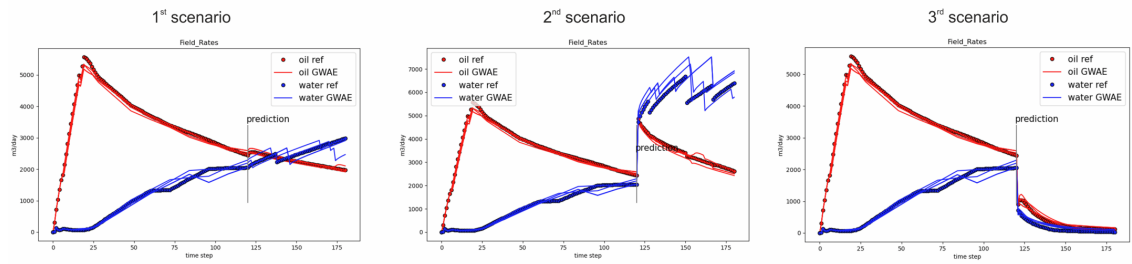


Figure 7.14: Field oil and water production profiles considering the 5-year prediction for every production mode (scenario).

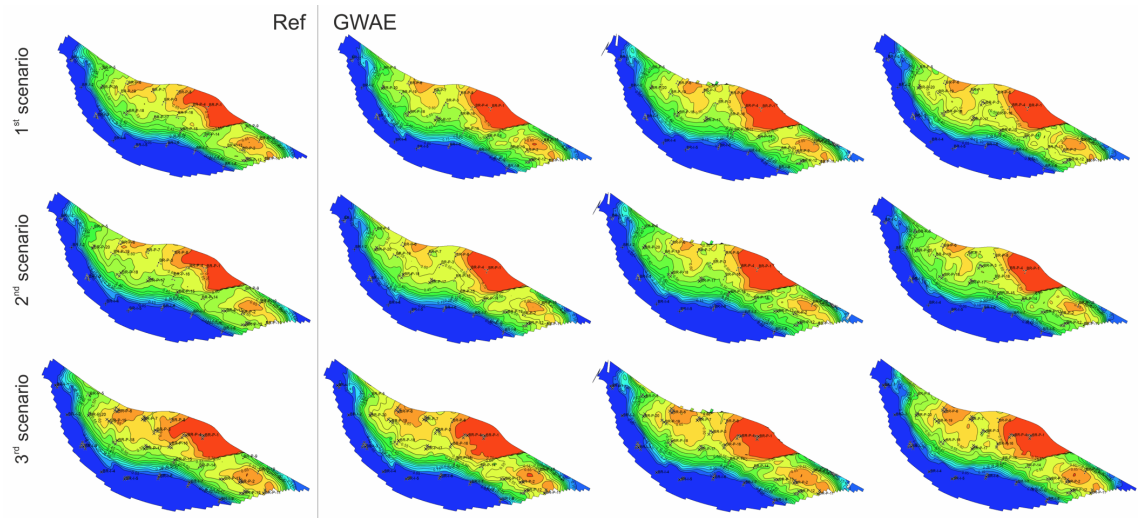


Figure 7.15: Maps of average oil saturation (S_o) at the last time step for the reference model and GWAE generated models. Note: all maps are in the same scales.

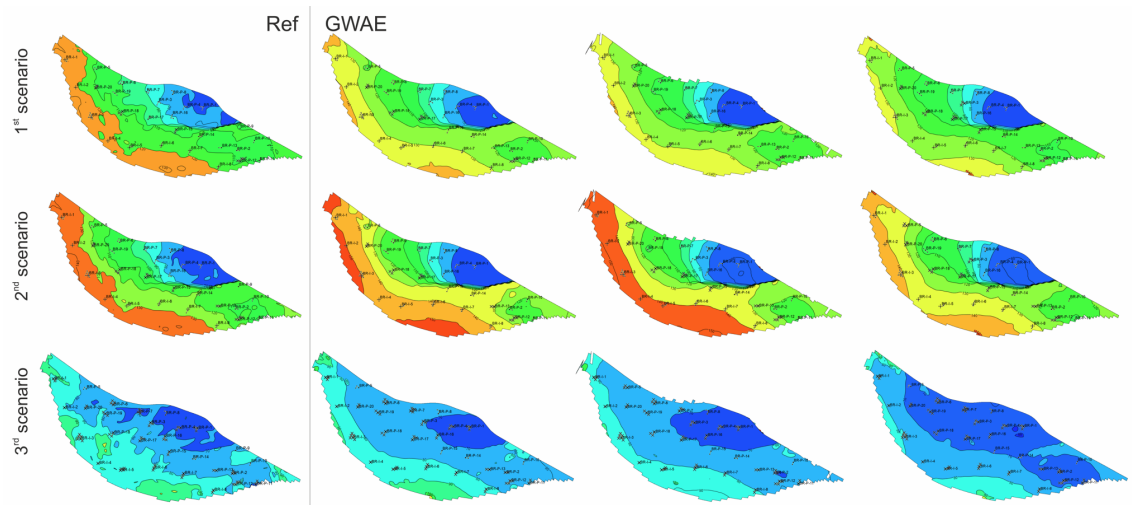


Figure 7.16: Maps of average reservoir pressure (P_{res}) at the last time step for the reference model and GWAE generated models. Note: all maps are in the same scales.

energy that may impact investment decisions.

For instance, different pressure depletion levels in the reservoir dome can be crucial for various technological measures in wells, such as reducing draw-down pressure on the reservoir or deciding to convert a producing well into an injection well. Similarly, variations in the maps within the water-saturated region can serve as a basis for optimising the reservoir pressure maintenance system by adjusting injection levels or shutting down injection wells.

7.5 Conclusion

This chapter conducted experiments to test the generative capability of GWAE using the semi-synthetic Brugge reservoir as an example.

The importance of analysing and understanding the internal structure of the prior dataset was demonstrated. As an example, a hybrid prior training dataset was constructed for the Schelde zone, consisting of 50% realisations generated using the SIS method and 50% created using Object Modelling. The GWAE was trained on this dataset, and it was shown that the structure of the LS split into two unconnected clouds, which is an undesirable effect for solving the inverse problem. Moreover, the generative model showed low-quality results and was unable to reproduce realisations that matched the Object Modelling method.

Since the prior dataset for each Brugge zone represents a separate stratigraphy setting and zones are independent of each other (which is inconsistent with real geology), a particular generative model was trained for each zone. This set of GWAEs can independently generate property distribution for each zone and stack the resulting geological realisations on top of each other to conduct model HM process. The analysis of the latent spaces of each generative model revealed that each LS represents a homogeneous cloud of points with an internal structure that cannot be discerned with current tools. However, the homogeneity of the latent space allows it to be used for solving the inverse problem.

The HM experiment demonstrated the effectiveness of the proposed approach. Overall model and well-level match were achieved as the optimisation algorithm correctly matched production data and porosity profiles along the well bore. Moreover, the implicit dependency of porosity and permeability was correctly identified.

At this stage of development, using a set of generative models seems to be the optimal solution for adapting models with complex stratigraphy. The concepts of epistemic uncertainty, fuzzy logic, and vagueness provide a foundation for the next step for realistic and adaptive models and research methods in geoscience.

Chapter 8

Conclusion and future work

8.1 Conclusion

The aim of the work was to investigate and develop approaches to geological modelling and model conditioning in a single parametric domain. This domain describes uncertain characteristics of a geological object, such as depositional and structural. The accounting of uncertainty in various geological scenarios, structural uncertainty, and uncertainty related to the inter-dependencies of petrophysical parameters is generally resolved through multivariate modelling. This approach is often limited to the unique sets of parameters that correspond to a specific geological scenarios and explicit parameter dependencies. The unique set of parameters to describe a geomodel is subjective and introduces the particular model uncertainty related to the model definition, equations and parametrisation. This brings further constraints to tuning models to data - a given parametrisation may be more or less prone to match data. Fixed parametrisation also may lead to biases and overlooking the aspect of the model not captured by the parameters and therefore under represent uncertainty. To address this challenge, a learning-based approach for a single learning based parameterisation through implicit low-dimensional hidden space latent space was considered.

Based on the literature review, the Variational Autoencoder generative learning approach was chosen as the base tool for parameterising the prior set of geological representations, considering the uncertainty of various kinds. The main idea behind using this approach was to parameterise the ensemble of prior geological realisations with an *Encoder* into a latent space that implicitly describes the prior uncertainty space. Then, the constructed LS, in combination with the *Decoder*, is used as a generator to search for the necessary realisations to condition the model to both static and dynamic data. The considered tasks included uncertainty estimation of dynamic response and HM.

The thesis presents the new GWAE approach, compares its performance with conventional VAE and its improved flavour WAE. The thesis shows how GWAE helps to overcome the limitations of VAE and WAE, such as blurriness and inability to account for structural inconsistencies of complex-shaped objects. Conceptually, such a transition to graph-based generative modelling allowed moving from Euclidean data types to non-Euclidean data, i.e., data that did not have a regular structure. In fact, the GWAE architecture allows taking into account the structural and spatial features of the reservoir by moving away from the classical lattice representation. Compared to standard geostatistical approaches, the GWAE bypasses limitations such as the stationarity assumption, linearity and hard data conditioning.

For a more effective use of generative models, it is necessary to understand the internal structure of the hidden space, as the structure is fundamental to the process of finding the latent representation of a geological model with respect to the conditioning data. I have conducted advanced analytics of the internal structure of hidden spaces for different ensembles of reservoir models with PCA, t-SNE, and TDA. The use of the aforementioned tools has resulted in several conclusions regarding the suitability and existing constraints of utilising the GWAE. It was found that the key property of the latent space is its global continuity, since the optimisation process must take place in "dense" regions to ensure geological realism. If the latent space can be decoupled into two or more continuous parts, the optimization process can be conducted in those segments in parallel. This would reflect the categorical nature of multiple geological concepts. In addition, the structure of the latent space reflected the grouping of latent representations of geological realisations with respect to different geological components, such as the number of channels, channel position, and fault position. TDA revealed the ring structure of the *STRUCTURAL* prior dataset, which partly explains the spatial position of the channel in the Geodesic interpolation experiment. A Geodesic metric was introduced for efficient navigation in the high-dimensional hidden space, because LS has complex non-Euclidean structure (Arvanitidis et al., 2021). This metric was tested on the example of interpolation and showed a more favourable behaviour compared to the standard Euclidean metric. Moreover, the Geodesic metric served as the basis for subsequent control of geological realism in the LS.

For solving the inverse problem, a three-component OF was introduced. The first component was responsible for controlling the static properties along the wellbore, the second component for the dynamic response of the well, and the third component for the overall realism of the geological realisation. This was achieved by calculating the density in the region of the hidden space. In case of its decrease, the optimiser was penalised, as it was assumed that the process begins to generate realisations with reduced geological realism by transitioning into less dense (probable) regions. This type of loss function allows the optimisation process to be carried out in stages, where in the first stage the static error is

minimised with realism control, allowing the first approximation of the model to be set up without expensive flow simulations. In the second stage, the target loss consists of three components, then the optimisation already takes place in a more compact subspace.

To test the generative capabilities of GWAE, three prior datasets of 3D geological objects were developed: *FLUVIAL*, *STRUCTURAL*, *BRUGGE*. The three data sets served different purposes. The comparative study of VAE, WAE and GWAE was performed to reveal limitations and advantages of approaches to generate geologically consistent representations. The *FLUVIAL* dataset models an object whose main uncertainty is the geological depositional scenario of the quantity and shape of fluvial channels. The experiment showed that GWAE parameterised such a prior. The internal structure of the LS featuring the number and spatial placement of the channel as global properties, allowed efficient navigation and solving the inverse problem.

The *STRUCTURAL* dataset combines two different types of geological uncertainty: depositional with respect to channel object and structural associated with a fault in a simple synthetic example. The structural uncertainty includes the position of a fault and the displacement of blocks. Depositional uncertainty includes the geometry and position of a channel. The analysis of the generative model's hidden space revealed discretisation connected with fault position. This is due to the discrete location of the fault in the training dataset, i.e., the fault takes one of 15 discrete positions for each geological realisation in the prior. This discretisation was reflected in the structure of the hidden space and subsequently manifested in navigation and solving the inverse problem. In addition, block displacement relative to the fault was also varied to break connectivity. The optimisation algorithm could not find realisations that replicated the structure of the reference model. However, the dynamics of the reservoir response were accurately reproduced due to the disruption of the channel's integrity. This experiment demonstrated the importance of the continuity of the hidden space, as transitioning to less dense regions limits the generative capabilities of the neural network.

The third dataset represented a semi-synthetic benchmark Brugge field that comprises of four stratigraphic zones. Since GWAE training requires a fairly large amount of data, the initial benchmark was increased to 5000 geological representations. It is worth noting that the resulting prior dataset (*BRUGGE*) maintains geological realism in terms of property distribution and spatial characteristics. It does not statistically contradict the original well data. In the first stage, using GWAE as a geological generator tested the hypothesis of the influence of the geological modelling algorithm to uncertainty when constructing the prior. To remind, the upper stratigraphic object, Schelde, represents a fluvial sedimentary environment, which was modelled by two methods: Object Modelling and SIS. The analysis of the latent space for this prior dataset revealed a limitation as the latent space was

split into two disconnected clusters related to the modelling approach, which is a negative effect for the further optimisation process as the global discreteness of the search space needs to be taken into account. When creating a generative model for the entire Brugge, it was decided to train a set of GWAEs, each of which would be responsible for its stratigraphic zone. While solving the inverse problem, the optimiser used the combined hidden space. This approach proved to be effective, as it allowed the successful identification of a set of matched models.

Graph generative networks have demonstrated potential for modelling the spatial distribution of geological properties, subject to various types of uncertainty. Still, the quality of the generator's performance is directly affected by the construction of the prior dataset. As demonstrated in Chapters **Graph Wasserstein Autoencoders and channelized synthetic case** and **Graph Wasserstein Autoencoders and fault synthetic case**, a geological representations in latent space decompose into discrete clusters, which negatively affects the optimisation process.

In summary, the novelty of this study can be summarised as follows:

- The simultaneous consideration of geological scenario uncertainty, structural uncertainty and petrophysical dependencies through a universal low-dimensional hidden space.
- The use of Graph Wasserstein Autoencoders to handle irregularities of geological structure to maintain the topology and connectivity of 3D geological structures.
- The use of Geodesic metric for efficient navigation in the LS.
- The use of a three-component OF that includes control of geological realism and balance the model soft match to static and dynamic data.
- The use of multidimensional hidden space analysis methods for more efficient work with generative models.
- The use of ensemble generative models for effective parameterisation of geological objects with structural uncertainty of fault position and blocks displacement.
- Solving inverse problems using hybrid hidden spaces that provide non-unique solutions.
- The identification of important limitations in modelling of prior.

8.2 Recommendations

Graph-based machine learning methods are a relatively young direction in data science, but it is already clear that their potential is very high. This work can be considered a first step towards the development of graph-based geological generators that facilitate geological modelling under uncertainty. Recommendations would extend the applicability of the proposed methods beyond the thesis. The logical continuation of this work could involve research in the following directions:

- Development of approaches for constructing prior based on fuzzy transitions between geological environments.
- Development of universal parameterisation approaches in the presence of discrete model parameters (fault positions, number of facies) for continuous representation in the latent space (Giovanis et al., 2024).
- Integration of geodesic-inspired optimisation algorithms for solving the inverse problem and speeding up the calculation of the geodesic metric.
- Integration of graph-based flow simulators (Sanchez-Gonzalez et al., 2020) to eliminate the need for reverse transfer from the graph representation of the geological model to the grid representation.
- Integration of flow diagnostics based on time-of-flight approach (Møyner et al., 2014) or another fast fluid flow approximation techniques (Nakajima et al., 2021) to speed-up simulations.
- Implementation of graph generation approaches to create graphs of different sizes, including graph upscaling and downscaling (Loukas, 2018).
- The implementation of transfer learning approaches to expand the capabilities of the graph generator and its corresponding hidden space by integrating new geological scenarios (eg: number of channels, facies, azimuth and so on).

List of Acronyms

AE Autoencoder. 24, 25

AHM Assisted history matching. vi, 37–41, 43, 44, 73, 78, 141, 146, 148, 149

ANN Artificial Neuron Network. 34

ANN-SAE Artificial Neuron Network — Stacked AutoEncoder. 34

CMA-ES Covariance-Matrix Adaptation Evolution Strategy. vii, 66–68, 72, 128, 166

CNN Convolutional Neural Networks. vii, 54–59, 63, 88, 116, 120, 132–134, 137, 139

CP corner point. 112, 113

DBC Distance-Based Clustering. 34

DSS Direct Sequential Simulation. 14

ELU Exponential Linear Unit. 133

EnKF Ensemble Kalman Filter. 37, 38, 41, 164, 165, 170, 171

ES Evolution Strategies. 65, 66, 165

ES-MDA Ensemble Smoother with Multiple Data Assimilation. 33

GAN Generative Adversarial Network. vi, 21–23, 31, 32, 36, 55, 56, 58

GNN Graph Neural Network. 64

GWAE Graph Wasserstein Variational Autoencoder. vii, viii, 62–65, 68, 70, 71, 77–79, 88, 90–92, 96, 101–111, 115–117, 119–126, 128, 129, 132–139, 141–148, 150, 156–173, 175–177

HM history matching. 7, 11–13, 19, 20, 32, 33, 35, 37, 38, 41, 42, 44, 64, 68, 75, 76, 88, 101, 119, 130, 140, 141, 152, 154, 155, 163, 165, 166, 169–171, 173

- IES-RLM** Iterative Ensemble Smoother based on a Regularized Levenberg-Marquardt Algorithm. 165
- KL** Kullback-Leibler. vi, 25, 48, 50–54, 84, 135
- LS** latent space. 8–12, 26, 28, 32, 36, 41, 42, 49, 50, 54, 63, 65, 68–72, 77, 79, 88, 116–128, 130, 132, 133, 135, 139, 140, 142, 144–151, 153, 155, 156, 158–161, 163–167, 170, 173–176, 178
- MAE** Mean Absolute Error. 74
- MDS** Multidimensional Scaling. 80
- MKL** Multiple Kernel Learning. 165, 170, 171
- ML** machine learning. 8–11, 13, 20–22, 42, 43, 47, 51, 53, 54, 56, 74, 80, 125, 170, 178
- MMD** Maximum-Mean Discrepancy. vii, 53, 54, 64
- MPS** Multi-point statistics. vi, 14, 18–20, 62
- MSE** Mean Squared Error. 25, 73–75, 128
- NF** Normalizing Flow. 27
- NLPCA** Nonlinear Principal Components Analysis. 24
- NTG** Net-to-Gross. 108
- OF** Objective Function. vii, 17, 18, 52, 65, 66, 73, 76–78, 88, 117, 118, 129, 130, 147, 153, 165, 170, 175, 177
- OT** Optimal Transport. vii, 51, 52
- PCA** Principal Components Analysis. vii, 24, 25, 60, 79–82, 116, 121, 122, 130, 131, 145, 146, 148, 158, 163, 164, 167, 170, 171, 175
- PDF** probability distribution function. 47
- PSO** Particle Swarm Optimization. 164, 165, 170, 171
- ReLU** Rectified Linear Unit. 133
- RMSE** Root Mean Squared Error. 74, 75

- RNN** Recurrent Neural Network. 21
- SAE** Stacked AutoEncoder. 34
- SGAN** Spatial Generative Adversarial Network. 32
- SGS** Sequential Gaussian Simulation. 14–16, 40, 103
- SIS** Sequential Indicator Simulation. 14, 16, 17, 103–106, 108, 111, 158, 159, 173, 176
- SNE** Stochastic Neighbour Embedding. 83, 84
- SPADE** Spatially Adaptive Denormalization. 33
- t-SNE** t-Distributed Stochastic Neighbour Embedding. vii, 79, 80, 83–85, 116, 121, 122, 135, 145, 146, 148, 158, 163, 164, 175
- TDA** Topological Data Analysis. vii, 70, 79, 80, 86–88, 116, 121–125, 146, 148, 158, 163, 164, 175
- TGS** Truncated Gaussian Simulation. 14, 16, 17
- UQ** uncertainty quantification. vi, 7, 11, 13, 24, 35, 39–44, 57, 64, 75, 116, 139, 141, 152
- VAE** Variational Autoencoder. vi, vii, 21, 22, 24–28, 32–34, 36, 44, 46–53, 55, 56, 58, 61, 64, 116, 120, 132, 133, 135–139, 174–176
- WAE** Wasserstein Auto-Encoder. vii, 52, 53, 69, 116, 132–139, 175, 176

References

Aanonsen, S.I., Nævdal, G., Oliver, D.S., Reynolds, A.C., and Vallès, B., 2009. The Ensemble Kalman Filter in Reservoir Engineering—a Review. *Spe journal* [Online], 14(03), pp.393–412. eprint: <https://onepetro.org/SJ/article-pdf/14/03/393/2554039/spe-117274-pa.pdf>. Available from: <https://doi.org/10.2118/117274-PA>.

Aggarwal, C., Hinneburg, A., and Keim, D., 2002. On the surprising behavior of distance metric in high-dimensional space. *First publ. in: database theory, icdt 200, 8th international conference, london, uk, january 4 - 6, 2001 / jan van den bussche ... (eds.). berlin: springer, 2001, pp. 420-434 (=lecture notes in computer science ; 1973)*.

Agrawal, R. and Alfaro, L. de, 2019. Learning edge properties in graphs from path aggregations. *The world wide web conference* [Online], Www '19. San Francisco, CA, USA: Association for Computing Machinery, pp.15–25. Available from: <https://doi.org/10.1145/3308558.3313695>.

Ahn, S., Park, C., Kim, J., and Kang, J.M., 2018. Data-driven inverse modeling with a pre-trained neural network at heterogeneous channel reservoirs. *Journal of petroleum science and engineering* [Online], 170, pp.785–796. Available from: <https://doi.org/https://doi.org/10.1016/j.petrol.2018.06.084>.

Arauco, S., Castro, J., Potratz, J., Emerick, A., and Pacheco, M., 2021. Recent developments combining ensemble smoother and deep generative networks for facies history matching. *Computational geosciences* [Online], On-line First. Available from: <https://doi.org/10.1007/s10596-020-10015-0>.

Arauco, S., Emerick, A., and Pacheco, M., 2017. Integration of ensemble data assimilation and deep learning for history matching facies models [Online]. Available from: <https://doi.org/10.4043/28015-MS>.

Arjovsky, M. and Bottou, L., 2017. *Towards principled methods for training generative adversarial networks*. arXiv: 1701.04862 [stat.ML].

Arjovsky, M., Chintala, S., and Bottou, L., 2017. *Wasserstein gan*. arXiv: 1701.07875 [stat.ML].

Use of Water Chemistry Data in History Matching of a Reservoir Model [Online], 2012. Vol. All Days, Spe europec featured at eage conference and exhibition, SPE-154471-MS. eprint: <https://onepetro.org/SPEURO/proceedings-pdf/12EURO/A11-12EURO/SPE-154471-MS/1613161/spe-154471-ms.pdf>. Available from: <https://doi.org/10.2118/154471-MS>.

Arvanitidis, G., 2019. *Geometrical aspects of manifold learning*. PhD thesis. Technical University of Denmark.

Arvanitidis, G., Hansen, L.K., and Hauberg, S., 2021. *Latent space oddity: on the curvature of deep generative models*. arXiv: 1710.11379 [stat.ML].

Bai, T. and Tahmasebi, P., 2022. Sequential gaussian simulation for geosystems modeling: a machine learning approach. *Geoscience frontiers* [Online], 13(1), p.101258. Available from: <https://doi.org/https://doi.org/10.1016/j.gsf.2021.101258>.

Bao, J., Li, L., and Davis, A., 2022. Variational autoencoder or generative adversarial networks? a comparison of two deep learning methods for flow and transport data assimilation. *Mathematical geosciences* [Online], 54. Available from: <https://doi.org/10.1007/s11004-022-10003-3>.

Bardy, G., Biver, P., Caumon, G., and Renard, P., 2019. Oil production uncertainty assessment by predicting reservoir production curves and confidence intervals from arbitrary proxy responses. *Journal of petroleum science and engineering* [Online], 176, pp.116–125. Available from: <https://doi.org/https://doi.org/10.1016/j.petrol.2019.01.035>.

Battiston, F., Amico, E., Barrat, A., Bianconi, G., Arruda, G.F. de, Franceschiello, B., Iacopini, I., Kéfi, S., Latora, V., Moreno, Y., Murray, M.M., Peixoto, T.P., Vaccarino, F., and Petri, G., 2021. The physics of higher-order interactions in complex systems. *Nature physics* [Online], 17(10), pp.1093–1098. Available from: <https://doi.org/10.1038/s41567-021-01371-4>.

Belkin, M., Hsu, D., Ma, S., and Mandal, S., 2019. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the national academy of sciences* [Online], 116(32), pp.15849–15854. Available from: <https://doi.org/10.1073/pnas.1903070116>.

Bellman, R. and Kalaba, R., 1959. On adaptive control processes. *Ire transactions on automatic control*, 4(2), pp.1–9.

- Bengio, Y., 2009. Learning deep architectures for ai. *Foundations* [Online], 2, pp.1–55. Available from: <https://doi.org/10.1561/22000000006>.
- Biasotti, S., De Floriani, L., Falcidieno, B., Frosini, P., Giorgi, D., Landi, C., Papaleo, L., and Spagnuolo, M., 2008. Describing shapes by geometrical-topological properties of real functions. *Acm comput. surv.* [Online], 40(4). Available from: <https://doi.org/10.1145/1391729.1391731>.
- Bronstein, M.M., Bruna, J., Cohen, T., and Veličković, P., 2021. *Geometric deep learning: grids, groups, graphs, geodesics, and gauges* [Online]. arXiv. Available from: <https://doi.org/10.48550/ARXIV.2104.13478>.
- Bruna, J., Zaremba, W., Szlam, A., and LeCun, Y., 2014. *Spectral networks and locally connected networks on graphs*. arXiv: 1312.6203 [cs.LG].
- Caers, J. and Park, K., 2008. A distance-based representation of reservoir uncertainty: the metric enkf [Online], cp-62-00051. Available from: <https://doi.org/https://doi.org/10.3997/2214-4609.20146404>.
- Caers, J., 2011. Modeling uncertainty: concepts and philosophies. In: *Modeling uncertainty in the earth sciences* [Online]. John Wiley and Sons, Ltd. Chap. 3, pp.39–54. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119995920.ch3>. Available from: <https://doi.org/https://doi.org/10.1002/9781119995920.ch3>.
- Caers, J., Mariethoz, G., and Ortiz, J.M., 2023. Multiple point statistics. In: *Encyclopedia of mathematical geosciences* [Online]. Ed. by B.S. Daya Sagar, Q. Cheng, J. McKinley, and F. Agterberg. Cham: Springer International Publishing, pp.960–970. Available from: https://doi.org/10.1007/978-3-030-85040-1_24.
- Carlsson, G., 2009. Topology and data. *Bulletin of the american mathematical society - bull amer math soc* [Online], 46, pp.255–308. Available from: <https://doi.org/10.1090/S0273-0979-09-01249-X>.
- Carneiro, J., Azevedo, L., and Pereira, M., 2018. High-dimensional geostatistical history matching: vectorial multi-objective geostatistical history matching of oil reservoirs and uncertainty assessment. *Computational geosciences* [Online], 22. Available from: <https://doi.org/10.1007/s10596-017-9712-6>.
- Castro, S., Caers, J., and Mukerji, T., 2005. The stanford vi reservoir. *Stanford center for reservoir forecasting (scrf), annual report*, pp.153–154.

- Céline Scheidt Lewis Li, J.C., 2018. Data science for uncertainty quantification. In: *Quantifying uncertainty in subsurface systems* [Online]. American Geophysical Union (AGU). eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/9781119325888.ch3>. Available from: <https://doi.org/https://doi.org/10.1002/9781119325888.ch3>.
- Cetin, I., Stephens, M., Camara, O., and González Ballester, M.A., 2023. Attri-vae: attribute-based interpretable representations of medical images with variational autoencoders. *Computerized medical imaging and graphics* [Online], 104, p.102158. Available from: <https://doi.org/https://doi.org/10.1016/j.compmedimag.2022.102158>.
- Chalub, W., Luo, X., and Petvipusit, K., 2023. Improving history matching performance in a 3d field-scale case study by combing tracers and production data. *Geoenergy science and engineering* [Online], 222, p.211365. Available from: <https://doi.org/10.1016/j.geoen.2022.211365>.
- Chan, S. and Elsheikh, A.H., 2018. *Exemplar-based synthesis of geology using kernel discrepancies and generative neural networks*. arXiv: 1809.07748 [stat.ML].
- Chan, S. and Elsheikh, A.H., 2019a. Parametric generation of conditional geological realizations using generative neural networks. *Computational geosciences* [Online], 23(5), pp.925–952. Available from: <https://doi.org/10.1007/s10596-019-09850-7>.
- Chan, S. and Elsheikh, A.H., 2019b. *Parametrization and generation of geological models with generative adversarial networks*. arXiv: 1708.01810 [stat.ML].
- Chang, Y., Stordal, A., and Valestrand, R., 2016. Integrated work flow of preserving facies realism in history matching: application to the brugge field. *Spe journal* [Online], 21. Available from: <https://doi.org/10.2118/179732-PA>.
- Chen, B., Huang, K., Raghupathi, S., Chandratreya, I., Du, Q., and Lipson, H., 2021. *Discovering state variables hidden in experimental data*. arXiv: 2112.10755 [math.DS].
- Chen, B., Huang, K., Raghupathi, S., Chandratreya, I., Du, Q., and Lipson, H., 2022. Automated discovery of fundamental variables hidden in experimental data. *Nature computational science* [Online], 2, pp.433–442. Available from: <https://doi.org/10.1038/s43588-022-00281-6>.
- Chen, L., 2009. Curse of dimensionality. In: *Encyclopedia of database systems* [Online]. Ed. by L. LIU and M.T. ÖZSU. Boston, MA: Springer US, pp.545–546. Available from: https://doi.org/10.1007/978-0-387-39940-9_133.

- Chen, Y. and Oliver, D.S., 2010. Ensemble-Based Closed-Loop Optimization Applied to Brugge Field. *Spe reservoir evaluation and engineering* [Online], 13(01), pp.56–71. eprint: <https://onepetro.org/REE/article-pdf/13/01/56/2137089/spe-118926-pa.pdf>. Available from: <https://doi.org/10.2118/118926-PA>.
- Cheng, T., 2013. Accelerating universal kriging interpolation algorithm using cuda-enabled gpu. *Computers and geosciences* [Online], 54, pp.178–183. Available from: <https://doi.org/https://doi.org/10.1016/j.cageo.2012.11.013>.
- Cheng, Liang, 2020. The application of topological data analysis in practice and its effectiveness. *E3s web conf.* [Online], 214, p.03034. Available from: <https://doi.org/10.1051/e3sconf/202021403034>.
- Christie, M., Demyanov, V., and Erbas, D., 2006. Uncertainty quantification for porous media flows. *Journal of computational physics* [Online], 217(1). Uncertainty Quantification in Simulation Science, pp.143–158. Available from: <https://doi.org/https://doi.org/10.1016/j.jcp.2006.01.026>.
- Clevert, D.-A., Unterthiner, T., and Hochreiter, S., 2016. *Fast and accurate deep network learning by exponential linear units (elus)*. arXiv: 1511.07289 [cs.LG].
- Cojan, I., Fouché, O., Lopez, S., and Rivoirard, J., 2004. Process-based reservoir modelling in the example of meandering channel [Online]. In: pp.611–619. Available from: https://doi.org/10.1007/978-1-4020-3610-1_62.
- Corrales, M., Izzatullah, M., Hoteit, H., and Ravasi, M., 2022. A wasserstein gan with gradient penalty for 3d porous media generation. [Online], 2022(1), pp.1–5. Available from: <https://doi.org/https://doi.org/10.3997/2214-4609.2022616005>.
- Costa Gomes, J., Geiger, S., and Arnold, D., 2022. The design of an open-source carbonate reservoir model. *Petroleum geoscience* [Online], 28(3), petgeo2021–067. eprint: <https://pubs.geoscienceworld.org/pg/article-pdf/doi/10.1144/petgeo2021-067/5590345/petgeo2021-067.pdf>. Available from: <https://doi.org/10.1144/petgeo2021-067>.
- Cruz, W.C., Luo, X., and Petvipusit, K.R., 2022. Joint history matching of multiple types of field data in a 3d field-scale case study. *Energies* [Online], 15(17). Available from: <https://doi.org/10.3390/en15176372>.
- Cutts, P.L., 1991. The maureen field, block 16/29a, uk north sea. *Geological society, london, memoirs* [Online], 14(1), pp.347–352. eprint: <https://www.lyellcollection.org/doi/pdf/10.1144/GSL.MEM.1991.014.01.43>. Available from: <https://doi.org/10.1144/GSL.MEM.1991.014.01.43>.

- Defferrard, M., Bresson, X., and Vandergheynst, P., 2017. *Convolutional neural networks on graphs with fast localized spectral filtering*. arXiv: 1606.09375 [cs.LG].
- DeMers, D. and Cottrell, G., 1992. Non-linear dimensionality reduction. *Nips* [Online]. Available from: <https://api.semanticscholar.org/CorpusID:2888517>.
- Demyanov, V. and Arnold, D., 2018. *Challenges and solutions in stochastic reservoir modelling: geostatistics, machine learning, uncertainty prediction*, Education tour series. Netherlands: EAGE Publishing BV.
- Demyanov, V., Arnold, D., Rojas, T., and Christie, M., 2018. Uncertainty quantification in reservoir prediction: part 2—handling uncertainty in the geological scenario. *Mathematical geosciences* [Online], 51. Available from: <https://doi.org/10.1007/s11004-018-9755-9>.
- Demyanov, V., Backhouse, L., and Christie, M., 2015. Geological feature selection in reservoir modelling and history matching with multiple kernel learning. *Computers and geosciences* [Online], 85. Available from: <https://doi.org/10.1016/j.cageo.2015.07.014>.
- Demyanov, V., Pozdnoukhov, A., Christie, M., and Kanevski, M., 2010. Detection of optimal models in parameter space with support vector machines [Online]. In: pp.345–358. Available from: https://doi.org/10.1007/978-90-481-2322-3_30.
- Deutsch, C. and Tran, T., 2002. Fluvsim: a program for object-based stochastic modeling of fluvial depositional systems. *Computers and geosciences* [Online], 28(4), pp.525–535. Available from: [https://doi.org/https://doi.org/10.1016/S0098-3004\(01\)00075-9](https://doi.org/https://doi.org/10.1016/S0098-3004(01)00075-9).
- Deutsch, C.V., 2006. A sequential indicator simulation program for categorical variables with point and block data: blocksis. *Computers and geosciences* [Online], 32(10), pp.1669–1681. Available from: <https://doi.org/https://doi.org/10.1016/j.cageo.2006.03.005>.
- Dinh, L., Krueger, D., and Bengio, Y., 2015. *Nice: non-linear independent components estimation*. arXiv: 1410.8516 [cs.LG].
- Dinh, L., Sohl-Dickstein, J., and Bengio, S., 2017. *Density estimation using real nvp*. arXiv: 1605.08803 [cs.LG].
- Doersch, C., 2021. *Tutorial on variational autoencoders*. arXiv: 1606.05908 [stat.ML].

- Drozdova, M., Kinakh, V., Quétant, G., Golling, T., and Voloshynovskiy, S., 2021. *Generation of data on discontinuous manifolds via continuous stochastic non-invertible networks*. arXiv: 2112.09646 [cs.LG].
- Dukalski, M., Rovetta, D., Linde, S. van der, Möller, M., Neumann, N., and Phillipson, F., 2023. Quantum computer-assisted global optimization in geophysics illustrated with stack-power maximization for refraction residual statics estimation. *Geophysics* [Online], 88(2), pp.V75–V91. eprint: <https://doi.org/10.1190/geo2022-0253.1>. Available from: <https://doi.org/10.1190/geo2022-0253.1>.
- Dupont, E., Zhang, T., Tilke, P., Liang, L., and Bailey, W., 2018. *Generating realistic geology conditioned on physical measurements with generative adversarial networks*. arXiv: 1802.03065 [stat.ML].
- Durall, R., Ghanim, A., Fernandez, M.R., Ettrich, N., and Keuper, J., 2023. Deep diffusion models for seismic processing. *Computers and geosciences* [Online], 177, p.105377. Available from: <https://doi.org/10.1016/j.cageo.2023.105377>.
- Edelsbrunner, H., Letscher, D., and Zomorodian, A., 2000. Topological persistence and simplification. *Proceedings 41st annual symposium on foundations of computer science* [Online], pp.454–463. Available from: <https://doi.org/10.1109/SFCS.2000.892133>.
- Emerick, A.A., 2016. Analysis of the performance of ensemble-based assimilation of production and seismic data. *Journal of petroleum science and engineering* [Online], 139, pp.219–239. Available from: <https://doi.org/https://doi.org/10.1016/j.petrol.2016.01.029>.
- Engel, J., Agrawal, K.K., Chen, S., Gulrajani, I., Donahue, C., and Roberts, A., 2019. GANSynth: adversarial neural audio synthesis. *International conference on learning representations* [Online]. Available from: <https://openreview.net/forum?id=H1xQVn09FX>.
- Fabio, B.D. and Ferri, M., 2015. *Comparing persistence diagrams through complex vectors*. arXiv: 1505.01335 [math.AT].
- Fachri, M., Tveranger, J., Braathen, A., and Schueller, S., 2013. Sensitivity of fluid flow to deformation-band damage zone heterogeneities: a study using fault facies and truncated gaussian simulation. *Journal of structural geology* [Online], 52, pp.60–79. Available from: <https://doi.org/https://doi.org/10.1016/j.jsg.2013.04.005>.
- Feng, R., Grana, D., Mukerji, T., and Mosegaard, K., 2022. Application of bayesian generative adversarial networks to geological facies modeling. *Mathematical geosciences* [Online], 54. Available from: <https://doi.org/10.1007/s11004-022-09994-w>.

- Fisher, R.A., 1920. A mathematical examination of the methods of determining the accuracy of observation by the mean error, and by the mean square error. *Monthly notices of the royal astronomical society* [Online], 80, pp.758–770. Available from: <https://api.semanticscholar.org/CorpusID:124194116>.
- Forrester, J.W., 1971. Counterintuitive behavior of social systems. *Technological forecasting and social change* [Online], 3, pp.1–22. Available from: [https://doi.org/https://doi.org/10.1016/S0040-1625\(71\)80001-X](https://doi.org/https://doi.org/10.1016/S0040-1625(71)80001-X).
- Fossum, K., Alyaev, S., and Elsheikh, A.H., 2024. Ensemble history-matching workflow using interpretable spade-gan geomodel. *First break* [Online], 42(2), pp.57–63. Available from: <https://doi.org/https://doi.org/10.3997/1365-2397.fb2024014>.
- Fukushima, K., 1975. Cognitron: a self-organizing multilayered neural network. *Biological cybernetics* [Online], 20, pp.121–136. Available from: <https://api.semanticscholar.org/CorpusID:28586460>.
- Gatilov, S.Y., 2014. Using low-rank approximation of the jacobian matrix in the newton–raphson method to solve certain singular equations. *Journal of computational and applied mathematics* [Online], 272, pp.8–24. Available from: <https://doi.org/https://doi.org/10.1016/j.cam.2014.04.024>.
- Gauss, C., Bertrand, J., and Trotter, H., 1957. *Gauss’s work (1803-1826) on the theory of least squares* [Online]. Statistical Techniques Research Group, Section of Mathematical Statistics, Department of Mathematical [sic], Princeton University. Available from: <https://books.google.ru/books?id=-3ccGQAACAAJ>.
- Gholizadeh, S., Seyeditabari, A., and Zadrozny, W., 2018. Topological signature of 19th century novelists: persistent homology in text mining. *Big data and cognitive computing* [Online], 2(4). Available from: <https://doi.org/10.3390/bdcc2040033>.
- Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., and Dahl, G.E., 2017. *Neural message passing for quantum chemistry*. arXiv: 1704.01212 [cs.LG].
- Giovanis, D.G., Loukrezis, D., Kevrekidis, I.G., and Shields, M.D., 2024. *Polynomial chaos expansions on principal geodesic grassmannian submanifolds for surrogate modeling and uncertainty quantification*. arXiv: 2401.16683 [stat.ML].
- Goodfellow, I., Bengio, Y., and Courville, A., 2016. *Deep learning*. <http://www.deeplearningbook.org>. MIT Press.
- Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., 2014. *Generative adversarial networks*. arXiv: 1406.2661 [stat.ML].

- Gregor, K., Papamakarios, G., Besse, F., Buesing, L., and Weber, T., 2019. *Temporal difference variational auto-encoder*. arXiv: 1806.03107 [cs.LG].
- Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., and Smola, A., 2012. A kernel two-sample test. *Journal of machine learning research* [Online], 13(25), pp.723–773. Available from: <http://jmlr.org/papers/v13/gretton12a.html>.
- Guan, K.M., Anderson, T.I., Creux, P., and Kovscek, A.R., 2021. Reconstructing porous media using generative flow networks. *Computers and geosciences* [Online], 156, p.104905. Available from: <https://doi.org/https://doi.org/10.1016/j.cageo.2021.104905>.
- Guardiano, F.B. and Srivastava, R.M., 1993. Multivariate geostatistics: beyond bivariate moments. In: *Geostatistics tróia'92: volume 1*. Springer, pp.133–144.
- Guss, W.H. and Salakhutdinov, R., 2018. *On characterizing the capacity of neural networks using algebraic topology*. arXiv: 1802.04443 [cs.LG].
- Hamilton, W.L., Ying, R., and Leskovec, J., 2018. *Inductive representation learning on large graphs*. arXiv: 1706.02216 [cs.SI].
- Hansen, N., 2023. *The cma evolution strategy: a tutorial*. arXiv: 1604.00772 [cs.LG].
- Haque, A.K.M., Islam, M.A., and Shalaby, M., 2018. *Three-dimensional facies analysis using object-based geobody modeling: a case study for the farewell formation, maui gas field, taranaki basin, new zealand* [Online]. Available from: <https://doi.org/10.20944/preprints201801.0123.v1>.
- Hastie, T., Tibshirani, R., and Friedman, J., 2009. Neural networks. In: *The elements of statistical learning: data mining, inference, and prediction* [Online]. New York, NY: Springer New York, pp.389–416. Available from: https://doi.org/10.1007/978-0-387-84858-7_11.
- Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., and Lerchner, A., 2017. Beta-VAE: learning basic visual concepts with a constrained variational framework. *International conference on learning representations* [Online]. Available from: <https://openreview.net/forum?id=Sy2fzU9gl>.
- Hillier, M., Wellmann, F., Brodaric, B., de Kemp, E., and Schetselaar, E., 2021a. Three-Dimensional Structural Geological Modeling Using Graph Neural Networks. *Mathematical geosciences* [Online], 53(8), pp.1725–1749. Available from: <https://doi.org/10.1007/s11004-021-09945-x>.

- Hillier, M., Wellmann, F., Brodaric, B., Dekemp, E., and Schetselaar, E., 2021b. Three-dimensional structural geological modeling using graph neural networks. *Mathematical geosciences* [Online], 53, pp.1–25. Available from: <https://doi.org/10.1007/s11004-021-09945-x>.
- Hinton, G.E. and Roweis, S., 2002. Stochastic neighbor embedding. In: S. Becker, S. Thrun, and K. Obermayer, eds. *Advances in neural information processing systems* [Online]. Vol. 15. MIT Press. Available from: https://proceedings.neurips.cc/paper_files/paper/2002/file/6150ccc6069bea6b5716254057a194ef-Paper.pdf.
- Ho, J., Jain, A., and Abbeel, P., 2020. *Denoising diffusion probabilistic models*. arXiv: 2006.11239 [cs.LG].
- Holmström, L., 2008. Nonlinear dimensionality reduction by john a. lee, michel verleysen. *International statistical review* [Online], 76(2), pp.308–309. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1751-5823.2008.00054_10.x. Available from: https://doi.org/https://doi.org/10.1111/j.1751-5823.2008.00054_10.x.
- Hoogeboom, E., Gritsenko, A.A., Bastings, J., Poole, B., Berg, R. van den, and Salimans, T., 2022. *Autoregressive diffusion models*. arXiv: 2110.02037 [cs.LG].
- Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology* [Online], 24, pp.498–520. Available from: <https://api.semanticscholar.org/CorpusID:144828484>.
- Hutahaean, J., Demyanov, V., and Christie, M., 2015. Impact of model parameterisation and objective choices on assisted history matching and reservoir forecasting [Online]. Available from: <https://doi.org/10.2118/176389-MS>.
- Hutahaean, J., Demyanov, V., and Christie, M., 2017. On optimal selection of objective grouping for multiobjective history matching. *Spe journal* [Online], 22. Available from: <https://doi.org/10.2118/185957-PA>.
- Jackson, M., Hampson, G., Saunders, J., El-Sheikh, A., Graham, G., and Massart, B., 2013. Surface-based reservoir modelling for flow simulation. *Geological society, london, special publications* [Online], 387, pp.271–292. Available from: <https://doi.org/10.1144/SP387.2>.
- Japkowicz, N., Hanson, S.J., and Gluck, M.A., 2000. Nonlinear Autoassociation Is Not Equivalent to PCA. *Neural computation* [Online], 12(3), pp.531–545. eprint: <https://direct.mit.edu/neco/article-pdf/12/3/531/814389/089976600300015691.pdf>. Available from: <https://doi.org/10.1162/089976600300015691>.

- Jiang, A. and Jafarpour, B., 2021. Deep convolutional autoencoders for robust flow model calibration under uncertainty in geologic continuity. *Water resources research* [Online], 57. Available from: <https://doi.org/10.1029/2021WR029754>.
- Jolliffe, I.T., 2002. *Principal component analysis*, Springer series in statistics. New York: Springer-Verlag. Available from: <https://doi.org/10.1007/b98835>.
- Journel, A.G., 1983. Nonparametric estimation of spatial distributions. *Journal of the international association for mathematical geology*, 15, pp.445–468.
- Jung, A., Fenwick, D., and Caers, J., 2013. Training image-based scenario modeling of fractured reservoirs for flow uncertainty quantification. *Computational geosciences* [Online], 17. Available from: <https://doi.org/10.1007/s10596-013-9372-0>.
- Kaiser, H.F., 1960. The application of electronic computers to factor analysis. *Educational and psychological measurement* [Online], 20(1), pp.141–151. eprint: <https://doi.org/10.1177/001316446002000116>. Available from: <https://doi.org/10.1177/001316446002000116>.
- Kaplan, W., 2002. *Advanced calculus* [Online], Addison-wesley higher mathematics. Addison-Wesley. Available from: <https://books.google.ru/books?id=wywnAQAAIAAJ>.
- Karras, T., Aila, T., Laine, S., and Lehtinen, J., 2018. *Progressive growing of gans for improved quality, stability, and variation*. arXiv: 1710.10196 [cs.NE].
- Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., and Aila, T., 2021. Alias-free generative adversarial networks. *Proc. neurips*.
- Keim, D., 2002. Information visualization and visual data mining. *iee trans vis comput graph* 7(1). *Visualization and computer graphics, iee transactions on* [Online], 8, pp.1–8. Available from: <https://doi.org/10.1109/2945.981847>.
- Integrated Framework for Modelling of Thermal-Compositional Multiphase Flow in Porous Media* [Online], 2019. Vol. Day 1 Wed, April 10, 2019, Spe reservoir simulation conference, D011S003R004. eprint: <https://onepetro.org/spersc/proceedings-pdf/19RSC/1-19RSC/D011S003R004/1165481/spe-193932-ms.pdf>. Available from: <https://doi.org/10.2118/193932-MS>.
- Khlevniuk, A. and Tymchyshyn, V., 2017. Introduction to pullback metric [Online]. Available from: <https://doi.org/10.13140/RG.2.2.18073.31841>.
- Kingma, D.P. and Welling, M., 2022. *Auto-encoding variational bayes*. arXiv: 1312.6114 [stat.ML].

- Kingma, D.P. and Dhariwal, P., 2018. *Glow: generative flow with invertible 1x1 convolutions*. arXiv: 1807.03039 [stat.ML].
- Kramer, M.A., 1991. Nonlinear principal component analysis using autoassociative neural networks. *Aiche journal* [Online], 37(2), pp.233–243. eprint: <https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1002/aic.690370209>. Available from: <https://doi.org/https://doi.org/10.1002/aic.690370209>.
- Kriest, I., Kähler, P., Koeve, W., Kvale, K., Sauerland, V., and Oschlies, A., 2020. One size fits all? calibrating an ocean biogeochemistry model for different circulations. *Biogeosciences* [Online], 17(12), pp.3057–3082. Available from: <https://doi.org/10.5194/bg-17-3057-2020>.
- Krige, D., 1951. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the southern african institute of mining and metallurgy* [Online], 52(6), pp.119–139. eprint: https://journals.co.za/doi/pdf/10.10520/AJA0038223X_4792. Available from: https://doi.org/10.10520/AJA0038223X_4792.
- Kullback, S. and Leibler, R.A., 1951. On Information and Sufficiency. *The annals of mathematical statistics* [Online], 22(1), pp.79–86. Available from: <https://doi.org/10.1214/aoms/1177729694>.
- Kumar, P., Rao, B., Burman, A., Kumar, S., and Samui, P., 2023. Spatial variation of permeability and consolidation behaviors of soil using ordinary kriging method. *Groundwater for sustainable development* [Online], 20, p.100856. Available from: <https://doi.org/https://doi.org/10.1016/j.gsd.2022.100856>.
- Hierarchical Geological Realism for Reliable Reservoir Predictions* [Online], 2016. Vol. Day 2 Tue, September 27, 2016, Spe annual technical conference and exhibition, D023S099R001. eprint: <https://onepetro.org/SPEATCE/proceedings-pdf/16ATCE/2-16ATCE/D023S099R001/1376090/spe-184495-stu.pdf>. Available from: <https://doi.org/10.2118/184495-STU>.
- Kyriakidis, P.C., Deutsch, C.V., and Grant, M.L., 1999. Calculation of the normal scores variogram used for truncated gaussian lithofacies simulation: theory and fortran code. *Computers and geosciences* [Online], 25(2), pp.161–169. Available from: [https://doi.org/https://doi.org/10.1016/S0098-3004\(98\)00124-1](https://doi.org/https://doi.org/10.1016/S0098-3004(98)00124-1).
- Laloy, E., Héroult, R., Jacques, D., and Linde, N., 2018. Training-image based geostatistical inversion using a spatial generative adversarial neural network. *Water resources research* [Online], 54(1), pp.381–406. eprint: <https://agupubs.onlinelibrary>.

- wiley.com/doi/pdf/10.1002/2017WR022148. Available from: <https://doi.org/https://doi.org/10.1002/2017WR022148>.
- Laloy, E., Héroult, R., Lee, J., Jacques, D., and Linde, N., 2017. Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. *Advances in water resources* [Online], 110, pp.387–405. Available from: <https://doi.org/https://doi.org/10.1016/j.advwatres.2017.09.029>.
- Lawrence, N.D., 2005. Probabilistic non-linear principal component analysis with gaussian process latent variable models. *J. mach. learn. res.* [Online], 6, pp.1783–1816. Available from: <https://api.semanticscholar.org/CorpusID:1969477>.
- Lee, K., Lim, J., Ahn, S., and Kim, J., 2018. Feature extraction using a deep learning algorithm for uncertainty quantification of channelized reservoirs. *Journal of petroleum science and engineering* [Online], 171, pp.1007–1022. Available from: <https://doi.org/https://doi.org/10.1016/j.petrol.2018.07.070>.
- Lee, W., Kim, D., Chae, Y., and Ryu, D., 2011. Probabilistic evaluation of spatial distribution of secondary compression by using kriging estimates of geo-layers. *Engineering geology* [Online], 122(3), pp.239–248. Available from: <https://doi.org/https://doi.org/10.1016/j.enggeo.2011.06.008>.
- Lescoffit, G. and Townsend, C., 2005. Quantifying the Impact of Fault Modeling Parameters on Production Forecasting for Clastic Reservoirs. In: *Evaluating Fault and Cap Rock Seals* [Online]. American Association of Petroleum Geologists. Available from: <https://doi.org/10.1306/1060761H23166>.
- Levina, E. and Bickel, P., 2004. Maximum likelihood estimation of intrinsic dimension. In: L. Saul, Y. Weiss, and L. Bottou, eds. *Advances in neural information processing systems* [Online]. Vol. 17. MIT Press. Available from: https://proceedings.neurips.cc/paper_files/paper/2004/file/74934548253bcab8490ebd74afed7031-Paper.pdf.
- Liu, F., Peng, J., and Zhang, C., 2012. *A non-parametric indicator kriging method for generating coastal sediment type map.* [Online]. Available from: <http://hdl.handle.net/1834/14818>.
- Liu, Y., 2006. Using the snesim program for multiple-point statistical simulation. *Computers and geosciences* [Online], 32(10), pp.1544–1563. Available from: <https://doi.org/https://doi.org/10.1016/j.cageo.2006.02.008>.
- Loukas, A., 2018. *Graph reduction with spectral and cut guarantees.* arXiv: 1808.10650 [cs.DS].

- Maaten, L. van der and Hinton, G., 2008. Visualizing data using t-sne. *Journal of machine learning research* [Online], 9(86), pp.2579–2605. Available from: <http://jmlr.org/papers/v9/vandermaaten08a.html>.
- Madani, N., Maleki, M., and Soltani-Mohammadi, S., 2022. Geostatistical modeling of heterogeneous geo-clusters in a copper deposit integrated with multinomial logistic regression: an exercise on resource estimation. *Ore geology reviews* [Online], 150, p.105132. Available from: <https://doi.org/https://doi.org/10.1016/j.oregeorev.2022.105132>.
- Manzocchi, T., Carter, J., Skorstad, A., Fjellvoll, B., Stephen, K., Howell, J., Matthews, J., Walsh, J.J., Nepveu, M., Bos, C., Cole, J., and Egberts, P., 2008. Sensitivity of the impact of geological uncertainty on production from faulted and unfaulted shallow-marine oil reservoirs: objectives and methods. *Petroleum geoscience - petrol geosci* [Online], 14, pp.3–15. Available from: <https://doi.org/10.1144/1354-079307-790>.
- Mariethoz, G., Renard, P., Cornaton, F., and Jaquet, O., 2009. Truncated plurigaussian simulations to characterize aquifer heterogeneity. *Groundwater* [Online], 47(1), pp.13–24. Available from: <https://doi.org/10.1111/j.1745-6584.2008.00489.x>.
- Maschio, C., Vidal, A., and Schiozer, D., 2008. A framework to integrate history matching and geostatistical modeling using genetic algorithm and direct search methods. *Journal of petroleum science and engineering* [Online], 63, pp.34–42. Available from: <https://doi.org/10.1016/j.petrol.2008.08.001>.
- Matheron, G., 1963. Principles of geostatistics. *Economic geology* [Online], 58(8), pp.1246–1266. eprint: <https://pubs.geoscienceworld.org/segweb/economicgeology/article-pdf/58/8/1246/3481854/1246.pdf>. Available from: <https://doi.org/10.2113/gsecongeo.58.8.1246>.
- Mato, K., Mengoni, R., Ottaviani, D., and Palermo, G., 2021. *Quantum molecular unfolding*. arXiv: 2107.13607 [quant-ph].
- May, D.A. and Gerya, T.V., 2021. Physics-based numerical modeling of geological processes. In: D. Alderton and S.A. Elias, eds. *Encyclopedia of geology (second edition)* [Online]. Second Edition. Oxford: Academic Press, pp.868–883. Available from: <https://doi.org/https://doi.org/10.1016/B978-0-12-409548-9.12520-5>.
- Miyagi, A., akimoto, Y., and Yamamoto, H., 2018. Well placement optimization for carbon dioxide capture and storage via cma-es with mixed integer support. *Proceedings of the genetic and evolutionary computation conference companion* [Online], Gecco '18. Kyoto, Japan: Association for Computing Machinery, pp.1696–1703. Available from: <https://doi.org/10.1145/3205651.3205706>.

- Mohamed, L., Christie, M., Demyanov, V., Robert, E., Kachuma, D., E, T., and Uk, p., 2010. Application of particle swarms for history matching in the brugge reservoir [Online]. Vol. 6. Available from: <https://doi.org/10.2118/135264-MS>.
- Moon, C., Mitchell, S.A., Heath, J.E., and Andrew, M., 2019. Statistical inference over persistent homology predicts fluid flow in porous media. *Water resources research* [Online], 55(11), pp.9592–9603. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019WR025171>. Available from: <https://doi.org/https://doi.org/10.1029/2019WR025171>.
- Morris, C., Ritzert, M., Fey, M., Hamilton, W.L., Lenssen, J.E., Rattan, G., and Grohe, M., 2021. *Weisfeiler and leman go neural: higher-order graph neural networks*. arXiv: 1810.02244 [cs.LG].
- Mosser, L., 2023. Deep diffusion models for facies modeling [Online], 2023(1), pp.1–5. Available from: <https://doi.org/https://doi.org/10.3997/2214-4609.2023101627>.
- Mosser, L., Dubrule, O., and Blunt, M.J., 2017a. Reconstruction of three-dimensional porous media using generative adversarial neural networks. *Phys. rev. e* [Online], 96 (4), p.043309. Available from: <https://doi.org/10.1103/PhysRevE.96.043309>.
- Mosser, L., Dubrule, O., and Blunt, M.J., 2017b. *Stochastic reconstruction of an oolitic limestone by generative adversarial networks*. arXiv: 1712.02854 [cs.CV].
- Mosser, L., Dubrule, O., and Blunt, M.J., 2019. *Deepflow: history matching in the space of deep generative models*. arXiv: 1905.05749 [cs.LG].
- Møyner, O., Krogstad, S., and Lie, K.-A., 2014. The application of flow diagnostics for reservoir management. *Spe journal* [Online], 20. Available from: <https://doi.org/10.2118/171557-PA>.
- Nakajima, K. and King, M., 2021. Development and application of fast simulation based on the pss pressure as a spatial coordinate [Online]. Available from: <https://doi.org/10.2118/206085-MS>.
- Nichol, A. and Dhariwal, P., 2021. *Improved denoising diffusion probabilistic models*. arXiv: 2102.09672 [cs.LG].
- O’Sullivan, A.E., 2004. *Modelling simulation error for improved reservoir prediction*. PhD thesis. Heriot-Watt University.

- Oja, E., 1991. Data compression, feature extraction, and autoassociation in feedforward neural networks. In: T. Kohonen, K. Mäkisara, O. Simula, and J. Kangas, eds. *Artificial neural networks*. Vol. 1. Elsevier Science Publishers B.V., North-Holland, pp.737–745.
- Oliver, D.S. and Chen, Y., 2011. Recent progress on reservoir history matching: a review. *Computational geosciences* [Online], 15, pp.185–221. Available from: <https://api.semanticscholar.org/CorpusID:62772025>.
- Oord, A. van den, Vinyals, O., and Kavukcuoglu, K., 2018. *Neural discrete representation learning*. arXiv: 1711.00937 [cs.LG].
- Pandey, B., Bhanodia, P.K., Khamparia, A., and Pandey, D.K., 2019. A comprehensive survey of edge prediction in social networks: techniques, parameters and challenges. *Expert systems with applications* [Online], 124, pp.164–181. Available from: <https://doi.org/https://doi.org/10.1016/j.eswa.2019.01.040>.
- Papillon, M., Sanborn, S., Hajj, M., and Miolane, N., 2023. *Architectures of topological deep learning: a survey on topological neural networks*. arXiv: 2304.10031 [cs.LG].
- Pearson, K., 1901. Liii. on lines and planes of closest fit to systems of points in space. *The london, edinburgh, and dublin philosophical magazine and journal of science* [Online], 2(11), pp.559–572. eprint: <https://doi.org/10.1080/14786440109462720>. Available from: <https://doi.org/10.1080/14786440109462720>.
- Pereira, C.M. and de Mello, R.F., 2015. Persistent homology for time series and spatial data clustering. *Expert systems with applications* [Online], 42(15), pp.6026–6038. Available from: <https://doi.org/https://doi.org/10.1016/j.eswa.2015.04.010>.
- Peters, E., Arts, R.J., Brouwer, G.K., Geel, C.R., Cullick, S., Lorentzen, R.J., Chen, Y., Dunlop, K.N., Vossepoel, F.C., Xu, R., Sarma, P., Alhutali, A.H., and Reynolds, A.C., 2010. Results of the Brugge Benchmark Study for Flooding Optimization and History Matching. *Spe reservoir evaluation and engineering* [Online], 13(03), pp.391–405. eprint: <https://onepetro.org/REE/article-pdf/13/03/391/2137961/spe-119094-pa.pdf>. Available from: <https://doi.org/10.2118/119094-PA>.
- Peyré, G. and Cuturi, M., 2018. Computational optimal transport [Online]. Available from: <https://doi.org/10.48550/ARXIV.1803.00567>.
- Pham, T.-A., Lee, J.-H., and Park, C.-S., 2022. Mst-vae: multi-scale temporal variational autoencoder for anomaly detection in multivariate time series. *Applied sciences* [Online], 12, p.10078. Available from: <https://doi.org/10.3390/app121910078>.

- Purshouse, R. and Fleming, P., 2003. Conflict, harmony, and independence: relationships in evolutionary multi-criterion optimisation [Online]. Available from: https://doi.org/10.1007/3-540-36970-8_2.
- Pyrcz, M., 2014. *Pyrcz, m.j., and deutsch, c.v., geostatistical reservoir modeling, 2nd edition, oxford university press, new york.*
- Que, Q. and Belkin, M., 2016. Back to the future: radial basis function networks revisited. In: A. Gretton and C.C. Robert, eds. *Proceedings of the 19th international conference on artificial intelligence and statistics* [Online]. Vol. 51, Proceedings of machine learning research. Cadiz, Spain: PMLR, pp.1375–1383. Available from: <https://proceedings.mlr.press/v51/que16.html>.
- Rahmati, Z., 2023. *Edge2node: reducing edge prediction to node classification.* arXiv: 2311.02921 [cs.LG].
- Ren, G., He, J., Wang, Z., Younis, R., and Wen, X.-H., 2019. Implementation of physics-based data-driven models with a commercial simulator [Online]. Available from: <https://doi.org/10.2118/193855-MS>.
- Rezende, D.J. and Mohamed, S., 2016. *Variational inference with normalizing flows.* arXiv: 1505.05770 [stat.ML].
- Ringrose, P. and Bentley, M., 2015. *Reservoir model design: a practitioner's guide* [Online]. Available from: <https://doi.org/10.1007/978-3-030-70163-5>.
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B., 2022. *High-resolution image synthesis with latent diffusion models.* arXiv: 2112.10752 [cs.CV].
- Rubenstein, P.K., Schoelkopf, B., and Tolstikhin, I., 2018. *On the latent space of wasserstein auto-encoders.* arXiv: 1802.03761 [stat.ML].
- Rumelhart, D.E., Hinton, G.E., and Williams, R.J., 1986. Learning internal representations by error propagation [Online]. Available from: <https://api.semanticscholar.org/CorpusID:62245742>.
- Sabatino, R., Viviani, E., Rossa, E., Sala, C., and Mafioli, A., 2014. Structural uncertainty integration within reservoir risk analysis and history matching [Online]. Available from: <https://doi.org/10.2118/170761-MS>.
- Saccenti, E. and Camacho, J., 2015. Determining the number of components in principal components analysis: a comparison of statistical, crossvalidation and approximated methods. *Chemometrics and intelligent laboratory systems* [Online], 149, pp.99–116. Available from: <https://doi.org/https://doi.org/10.1016/j.chemolab.2015.10.006>.

- Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X., 2016. *Improved techniques for training gans*. arXiv: 1606.03498 [cs.LG].
- Salimans, T., Ho, J., Chen, X., Sidor, S., and Sutskever, I., 2017. *Evolution strategies as a scalable alternative to reinforcement learning*. arXiv: 1703.03864 [stat.ML].
- Sanchez-Gonzalez, A., Godwin, J., Pfaff, T., Ying, R., Leskovec, J., and Battaglia, P.W., 2020. *Learning to simulate complex physics with graph networks*. arXiv: 2002.09405 [cs.LG].
- Sayyafzadeh, M., Koochak, R., and Barley, M., 2018. Accelerating cma-es in history matching problems using an ensemble of surrogates with generation-based management [Online], 2018(1), pp.1–15. Available from: <https://doi.org/https://doi.org/10.3997/2214-4609.201802139>.
- Schaaf, A., Varga, M. de la, Wellmann, F., and Bond, C.E., 2021. Constraining stochastic 3-d structural geological models with topology information using approximate bayesian computation in gempy 2.1. *Geoscientific model development* [Online], 14(6), pp.3899–3913. Available from: <https://doi.org/10.5194/gmd-14-3899-2021>.
- Schaeffer, R., Khona, M., Robertson, Z., Boopathy, A., Pistunova, K., Rocks, J.W., Fite, I.R., and Koyejo, O., 2023. *Double descent demystified: identifying, interpreting and ablating the sources of a deep learning puzzle*. arXiv: 2303.14151.
- Brownfield Development Optimization under Uncertainty – A Structured Workflow Design for Complex Case Scenarios* [Online], 2017. Vol. Day 3 Wed, October 18, 2017, Spe russian petroleum technology conference, D033S016R003. eprint: <https://onepetro.org/SPERPTC/proceedings-pdf/17RPTC/3-17RPTC/D033S016R003/1318527/spe-187856-ms.pdf>. Available from: <https://doi.org/10.2118/187856-MS>.
- Sebacher, B., Hanea, R., and Stordal, A.S., 2017. An adaptive pluri-gaussian simulation model for geological uncertainty quantification. *Journal of petroleum science and engineering* [Online], 158, pp.494–508. Available from: <https://doi.org/https://doi.org/10.1016/j.petrol.2017.08.038>.
- An elastic grid approach for fault uncertainty modelling and updating using the Ensemble Kalman filter*. [Online], 2010a. Vol. All Days, Spe europe featured at eage conference and exhibition, SPE-130422–MS. eprint: <https://onepetro.org/SPEEURO/proceedings-pdf/10EURO/All-10EURO/SPE-130422-MS/1772365/spe-130422-ms.pdf>. Available from: <https://doi.org/10.2118/130422-MS>.

- Seiler, A., Aanonsen, S.I., Evensen, G., and Rivenæs, J.C., 2010b. Structural Surface Uncertainty Modeling and Updating Using the Ensemble Kalman Filter. *Spe journal* [Online], 15(04), pp.1062–1076. eprint: <https://onepetro.org/SJ/article-pdf/15/04/1062/2107343/spe-125352-pa.pdf>. Available from: <https://doi.org/10.2118/125352-PA>.
- Shao, H., Kumar, A., and Fletcher, P.T., 2017. *The riemannian geometry of deep generative models*. arXiv: 1711.08014 [cs.LG].
- Shuku, T. and Phoon, K.-K., 2023. Data-driven subsurface modelling using a markov random field model. *Georisk: assessment and management of risk for engineered systems and geohazards* [Online], 17(1), pp.41–63. eprint: <https://doi.org/10.1080/17499518.2023.2181973>. Available from: <https://doi.org/10.1080/17499518.2023.2181973>.
- Silva, D.S.F. and Deutsch, C.V., 2017. Multiple imputation framework for data assignment in truncated pluri-gaussian simulation. *Stochastic environmental research and risk assessment* [Online], 31, pp.2251–2263. Available from: <https://api.semanticscholar.org/CorpusID:125088096>.
- Sleeman, J., Dorband, J., and Halem, M., 2020. *A hybrid quantum enabled rbm advantage: convolutional autoencoders for quantum image compression and generative learning*. arXiv: 2001.11946 [quant-ph].
- Smith, C., Emerick, A.A., and Pacheco, M.A.C., 2019. Towards a robust parameterization for conditioning facies models using deep variational autoencoders and ensemble smoother. *Computers and geosciences* [Online], 128, pp.87–102. Available from: <https://doi.org/https://doi.org/10.1016/j.cageo.2019.04.006>.
- Sohl-Dickstein, J., Weiss, E.A., Maheswaranathan, N., and Ganguli, S., 2015. *Deep unsupervised learning using nonequilibrium thermodynamics*. arXiv: 1503.03585 [cs.LG].
- Song, Y. and Ermon, S., 2020. *Generative modeling by estimating gradients of the data distribution*. arXiv: 1907.05600 [cs.LG].
- Splawa-Neyman, J., 1990. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. sci.* [Online], 5(4), pp.465–472. Available from: [http://links.jstor.org/sici?sici=0883-4237\(199011\)5:4%3C465:OTAOPT%3E2.O.CO;2-M&origin=MSN](http://links.jstor.org/sici?sici=0883-4237(199011)5:4%3C465:OTAOPT%3E2.O.CO;2-M&origin=MSN).

- Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., and Besson, O., 2011. An improved parallel multiple-point algorithm using a list approach. *Mathematical geosciences* [Online], 43, pp.305–328. Available from: <https://doi.org/10.1007/s11004-011-9328-7>.
- Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical geology* [Online], 34, pp.1–21. Available from: <https://doi.org/10.1023/A:1014009426274>.
- Su, J. and Wu, G., 2018. *F-vaes: improve vaes with conditional flows*. arXiv: 1809.05861 [cs.LG].
- Sun, C., Demyanov, V., and Arnold, D., 2023. Geological realism in fluvial facies modelling with gan under variable depositional conditions. *Computational geosciences* [Online], 27. Available from: <https://doi.org/10.1007/s10596-023-10190-w>.
- Suzuki, S., Caumon, G., and Caers, J., 2008a. Dynamic data integration for structural modeling: model screening approach using a distance-based model parameterization. *Computational geosciences* [Online], 12, pp.105–119. Available from: <https://doi.org/10.1007/s10596-007-9063-9>.
- Suzuki, S., Caumon, G., and Caers, J., 2008b. Dynamic data integration for structural modeling: model screening approach using a distance-based model parameterization. *Computational geosciences* [Online], 12, pp.105–119. Available from: <https://doi.org/10.1007/s10596-007-9063-9>.
- Tang, J., Liu, J., Zhang, M., and Mei, Q., 2016. Visualizing large-scale and high-dimensional data. *Proceedings of the 25th international conference on world wide web* [Online]. International World Wide Web Conferences Steering Committee. Available from: <https://doi.org/10.1145/2872427.2883041>.
- Thibert, B., Gratier, J.-P., and Morvan, J.-M., 2005. A direct method for modeling and unfolding developable surfaces and its application to the ventura basin (california). *Journal of structural geology* [Online], 27(2), pp.303–316. Available from: <https://doi.org/https://doi.org/10.1016/j.jsg.2004.08.011>.
- Tierny, J., 2018. *Topological data analysis for scientific visualization* [Online], Mathematics and visualization. Springer International Publishing. Available from: <https://books.google.ru/books?id=M4tHDwAAQBAJ>.
- Tolstikhin, I., Bousquet, O., Gelly, S., and Schoelkopf, B., 2017. Wasserstein auto-encoders [Online]. Available from: <https://doi.org/10.48550/ARXIV.1711.01558>.

- Toma, S.-A. and Sebacher, B., 2022. Bridging deep convolutional autoencoders and ensemble smoothers for improved estimation of channelized reservoirs. *Mathematical geosciences* [Online], 54. Available from: <https://doi.org/10.1007/s11004-022-09997-7>.
- Torrado, R.R., Ciaurri, D.E., Mello, U.T., and Droz, S.E., 2015. Opening New Opportunities With Fast Reservoir-Performance Evaluation Under Uncertainty: Brugge Field Case Study. *Spe economics and management* [Online], 7(03), pp.84–99. eprint: <https://onepetro.org/EM/article-pdf/7/03/84/2100314/spe-166392-pa.pdf>. Available from: <https://doi.org/10.2118/166392-PA>.
- Tosi, A., Hauberg, S., Vellido, A., and Lawrence, N.D., 2014. *Metrics for probabilistic geometries*. arXiv: 1411.7432 [stat.ML].
- Villani, C. and Society, A.M., 2003. *Topics in optimal transportation* [Online], Graduate studies in mathematics. American Mathematical Society. Available from: <https://books.google.ru/books?id=MyPjjgEACAAJ>.
- Wang, G. and Fessler, J.A., 2023. Efficient approximation of jacobian matrices involving a non-uniform fast fourier transform (NUFFT). *IEEE transactions on computational imaging* [Online], 9, pp.43–54. Available from: <https://doi.org/10.1109/tci.2023.3240081>.
- Wang, Y., Voskov, D., Khait, M., and Bruhn, D., 2020. An efficient numerical simulator for geothermal simulation: a benchmark study. *Applied energy* [Online], 264, p.114693. Available from: <https://doi.org/10.1016/j.apenergy.2020.114693>.
- Wang, Y., Huang, H., Rudin, C., and Shaposhnik, Y., 2021. *Understanding how dimension reduction tools work: an empirical approach to deciphering t-sne, umap, trimap, and pacmap for data visualization*. arXiv: 2012.04456 [cs.LG].
- Williams, L.F., 1976. A modification to the half-interval search (binary search) method. *Proceedings of the 14th annual southeast regional conference* [Online], Acm-se 14. Birmingham, Alabama: Association for Computing Machinery, pp.95–101. Available from: <https://doi.org/10.1145/503561.503582>.
- Yarus, J., Chambers, R., Maučec, M., and Shi, G., 2012. Facies simulation in practice: lithotype proportion mapping and plurigaussian simulation, a powerful combination.
- Yeremian, G., 2022. *The choice of the loss function for solving the problem of history matching of the geological and simulation model: thesis for the degree of candidate of technical sciences: spec. 2.3.1*. PhD thesis.

- Zeiler, M.D. and Fergus, R., 2013. *Visualizing and understanding convolutional networks*. arXiv: 1311.2901 [cs.CV].
- Zhang, L., Dou, H., Zhang, K., Huang, R., Lin, X., Wu, S., Zhang, R., Zhang, C., and Zheng, S., 2023. Cnn-lstm model optimized by bayesian optimization for predicting single-well production in water flooding reservoir. *Geofluids (online)* [Online], 2023. Available from: <https://www.proquest.com/scholarly-journals/cnn-lstm-model-optimized-bayesian-optimization/docview/2822120330/se-2>.
- Zhang, T., Liu, Q., Wang, X., Ji, X., and Du, Y., 2022. A 3d reconstruction method of porous media based on improved wgan-gp. *Computers and geosciences* [Online], 165, p.105151. Available from: <https://doi.org/https://doi.org/10.1016/j.cageo.2022.105151>.
- Zhu, Y., Zhang, Y., Yang, H., and Wang, F., 2019. *Gancoder: an automatic natural language-to-programming language translation approach based on gan*. arXiv: 1912.00609 [cs.CL].