



UNIVERSITY OF CALGARY

University of Calgary

PRISM: University of Calgary's Digital Repository

Graduate Studies

The Vault: Electronic Theses and Dissertations

2018-08-31

Visual Analytics Framework for Exploring Uncertainty in Reservoir Models

Sahaf, Zahra

Sahaf, Z. (2018). Visual Analytics Framework for Exploring Uncertainty in Reservoir Models (Unpublished doctoral thesis). University of Calgary, Calgary, AB. doi:10.11575/PRISM/32902
<http://hdl.handle.net/1880/107726>
doctoral thesis

University of Calgary graduate students retain copyright ownership and moral rights for their thesis. You may use this material in any way that is permitted by the Copyright Act or through licensing that has been assigned to the document. For uses that are not allowable under copyright legislation or licensing, you are required to seek permission.

Downloaded from PRISM: <https://prism.ucalgary.ca>

UNIVERSITY OF CALGARY

Visual Analytics Framework for Exploring Uncertainty in Reservoir Models

by

Zahra Sahaf

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN COMPUTER SCIENCE

CALGARY, ALBERTA

AUGUST, 2018

© Zahra Sahaf 2018

Abstract

Uncertainty is related to poor knowledge of a phenomenon. In particular, geological uncertainty is an essential element that affects the prediction of hydrocarbon production. The standard approach to address the geological uncertainty is to generate a large number of random 3D geological models and then perform flow simulations for each of them. Such a brute force approach is not efficient as the flow simulations are computationally costly and as a result, domain experts cannot afford running a large number of simulations. Therefore, it is critically important to be able to address the uncertainty using a few geological models, which can reasonably represent the overall uncertainty of the ensemble. Our goal is to design and develop a visual analytics framework to filter the geological models and to only select models that can potentially cover the uncertain space. In this framework, a new block based approach is proposed using mutual information to calculate pairwise distances between the 3D geological models. The calculated distances are then used within a clustering algorithm to group similar models. Cluster centers are the few representative models of the entire set of models that cover the uncertainty range. The whole framework is complimented by visual interactive tasks to be able to incorporate user's knowledge into the process and make the entire process more understandable. Finally, the framework is applied on many different case studies, and the results are evaluated by comparing with the existent brute force approach. In addition to that, the actual framework is evaluated in formal user study sessions with the domain experts in reservoir engineering and geoscience domain.

To the strong woman who raised me

My Mother

who has been a rock of stability throughout my life

and

whose loving spirit sustains me.

Acknowledgements

It is my pleasure to acknowledge the roles of several individuals who were instrumental for completion of my Ph.D. research. First and foremost I want to thank my loving, supportive, encouraging, and patient husband, Masoud, whose faithful support during all the stages of this Ph.D. is appreciated. His support was not only from the spiritual aspects, but also he helped me generously on the scientific and technical aspects of my research. I could not have done this without the love and support of my family. My parents have been my biggest supporters through thick and thin, and I have blossomed because of that. They gave me strength and allowed me to realize my own potential. I wish to express my gratitude also to my beloved brother and sister who are an important inspiration for me and have shaped me into who I am today, and I love them deeply for that.

I would like to express my gratitude to my supervisors, Dr. Mario Costa Sousa and Dr. Frank Maurer, for their valuable support, advice and feedback on all aspects of my dissertation research. A very special thanks goes out to Dr. Hamidreza Hamdi, without whose motivation and encouragement I would not have accomplished this research. His friendship and collaboration are a few of the many good memories I have of this period of my life. I would also like to thank my committee: Dr. Wesley Willet, Dr. Usman Alim, and Dr. Naser El-Sheimy and Dr. Eric Mackay for their insightful comments. My sincere thanks also goes to Dr. Long Nghiem for offering me the internship opportunity at the CMG company and working on diverse exciting projects. Lastly, thanks to all my great friends for their comfort, support, laughter, and fun times over these years, including both my old friends back home and the new ones that I met here at this beautiful country.

Table of Contents

| | |
|--|------|
| Abstract | ii |
| Acknowledgements | iv |
| Table of Contents | v |
| List of Figures | viii |
| List of Symbols | xii |
| 1 Introduction | 1 |
| 1.1 Background | 1 |
| 1.1.1 Reservoir Modeling | 1 |
| 1.1.2 Uncertainty in Reservoir Modeling | 3 |
| 1.2 Motivation | 8 |
| 1.3 Methodology and Contributions | 10 |
| 1.4 Organization of Thesis | 12 |
| 2 A Visual Analytics Framework for Exploring Uncertainties in Reservoir Models | 18 |
| 2.1 Abstract | 18 |
| 2.2 Introduction | 18 |
| 2.3 Related Work | 21 |
| 2.3.1 Current Approaches for Selecting of Geological Models | 21 |
| 2.3.2 Visual Analytics Techniques for Multirun Data | 22 |
| 2.4 Requirement Analysis | 23 |
| 2.5 Analytical Framework for Descriptive Uncertainty Assessment: Overview | 24 |
| 2.6 Calculation of (Dis)similarity Metric | 25 |
| 2.6.1 Distance Specification | 28 |
| 2.6.2 Block Size Specification | 31 |
| 2.7 Projection with Clustering | 33 |
| 2.8 Evaluation | 36 |
| 2.9 Visual Analytics Application | 40 |
| 2.9.1 Selection Process | 41 |
| 2.9.2 Comparison Analysis | 43 |
| 2.10 Conclusion and Future Works | 44 |
| 2.11 Acknowledgment | 44 |
| 3 Clustering of Geological Models for Reservoir Simulation Studies in a Visual Analytics Framework | 45 |
| 3.1 Introduction | 45 |
| 3.2 Similarity Metric Calculation | 47 |
| 3.3 Clustering | 50 |
| 3.4 Case Study | 50 |
| 3.5 Results | 51 |
| 3.6 Visual Analytics Framework | 53 |
| 3.7 Conclusion | 55 |
| 4 Filtering Geological Realizations for SAGD (Extended Version) | 56 |
| 4.1 Abstract | 56 |
| 4.2 Introduction | 57 |

| | | |
|--------|---|-----|
| 4.3 | Related Work | 59 |
| 4.4 | Proposed Visual Analytics Framework | 61 |
| 4.5 | Distance Calculation | 62 |
| 4.6 | Projection with Clustering | 64 |
| 4.7 | Extended proposed framework for SAGD models | 65 |
| 4.8 | Result | 67 |
| 4.8.1 | Computational Time | 70 |
| 4.9 | Conclusion | 71 |
| 5 | An ROI visual-analytical approach for exploring uncertainty in reservoir models . . | 72 |
| 5.1 | Abstract | 72 |
| 5.2 | Introduction | 73 |
| 5.3 | Related Work | 75 |
| 5.3.1 | Current Approaches for Selection of Geological Realizations | 75 |
| 5.3.2 | Visual Analytics Techniques for Multi-run Data | 75 |
| 5.4 | Analytical Framework | 76 |
| 5.5 | Variance Model and ROI Selection | 78 |
| 5.6 | Distance Calculation | 81 |
| 5.7 | Projection with Clustering | 83 |
| 5.8 | ROI Analysis | 85 |
| 5.9 | Results (with Case Studies) | 86 |
| 5.10 | User Evaluations (Result of User Studies) | 93 |
| 5.10.1 | Participants | 94 |
| 5.10.2 | ROI Selection Task | 95 |
| 5.10.3 | Analysis Task | 97 |
| 5.11 | Conclusion | 97 |
| 5.12 | Appendix - User Interactivity | 98 |
| 6 | VR application for the uncertainty analysis of reservoir ensembles | 102 |
| 6.1 | Introduction | 102 |
| 6.2 | Review of the uncertainty exploration process | 104 |
| 6.3 | Application design | 105 |
| 6.3.1 | Data Import | 106 |
| 6.3.2 | ROI Selection | 109 |
| 6.3.3 | Analytical Calculations | 112 |
| 6.4 | Evaluation | 115 |
| 6.5 | Conclusion | 119 |
| 7 | Conclusion and Future Work | 120 |
| A | Evaluation | 125 |
| A.1 | Evaluation of clustering results with quantitative measurements | 125 |
| B | Clustering algorithm | 132 |
| B.1 | Choice of clustering algorithm | 132 |
| C | User Study Evaluations | 134 |
| C.1 | Consent Form | 134 |
| C.2 | Pre user study questionnaire | 137 |
| C.3 | After user study questionnaire | 138 |
| D | Copyright | 139 |

| | |
|------------------------|-----|
| Bibliography | 142 |
|------------------------|-----|

List of Figures and Illustrations

| | | |
|------|--|----|
| 1.1 | Representation of a sample reservoir model with channels [Frster et al., 2006] (left). Representation of a smple reservoir model with faults and fractures (right). | 2 |
| 1.2 | Representation of a reservoir model with a highlighted cell. | 2 |
| 1.3 | A typical reservoir modeling workflow ([Agada et al., 2017]). | 3 |
| 1.4 | GeoModeling workflow. (a) A structural model showing faults and layering, (b) Example of a 3D Stratigraphic Model, (c) Example of a stochastic model of facies, (d) A geological model showing permeability distribution and values | 5 |
| 1.5 | The workflow of generating geological realizations and assessing uncertainty. (a) Sample sources of information for exploring reservoir models. (b) Observation of the available information and mapping of them to the reservoir models that results into an initial reservoir model with few known and explored areas. (c) Application of geostatistical algorithms for generation of a set of reservoir models to cover the uncertainty for the unknown areas. (d) A set of 3D geological realization models. (e) Running flow simulation for all the generated realizations and observing the uncertainty of a flow related property. (f) Representation of range of uncertainty for an interest dynamic property. | 7 |
| 1.6 | Organization of thesis. “Core” is the first proposed visual analytics framework for exploring uncertainty in reservoir models. Then, that framework is applied to two different reservoir engineering “domains”. In one of the domain applications, the idea of ROI is proposed for that specific domain problem. The concept of ROI generalized to be applicable to different problems, and that creates the “Extended Core”. Finally, “Vis UI” presents the designed and developed visual interactive framework that supports the proposed analytics framework. | 13 |
| 2.1 | Overview of proposed filtering Process. | 25 |
| 2.2 | (a) A sample important 3D structure in the geological models. (b) A sample 3D block | 27 |
| 2.3 | Representation of a sample favorable structure (highlighted in yellow), and move- ment of templates (dark blue frames). | 27 |
| 2.4 | Dataset generated for evaluation of similarity distances. | 30 |
| 2.5 | Comparison of different distance calculation methods including Euclidean, Haus- dorff and Mutual Information. | 31 |
| 2.6 | Mean entropy plot for different block sizes, with highlighted maximum entropy. . . | 33 |
| 2.7 | (a) represents projection with MDS, (b) represents projection with MDS using kernel methods. [Zhang et al., 2010] | 35 |
| 2.8 | Difference between multidimensional scaling with (right) and without (left) kernel transformation on the case study dataset. | 36 |
| 2.9 | Different types of facies property that used for the creation of geological models. . | 37 |
| 2.10 | Simulation results of 15 geological models for Oil Recovery Factor property. . . . | 38 |
| 2.11 | Clustering result for 15 geological models. | 38 |
| 2.12 | Simulation results of 100 geological models for Oil Recovery Factor. | 39 |
| 2.13 | Clustering result for 100 geological models. | 40 |

| | | |
|------|---|----|
| 2.14 | Projection and clustering of loaded modelsscal. | 41 |
| 2.15 | Filtering process for an arbitrary area of interest. | 42 |
| 2.16 | Similarity map for the selected models. | 43 |
| 3.1 | A sample important 3D structure in the geological models. | 48 |
| 3.2 | Representation of block based similarity calculation method for a sample block size 3*3*3. This Figure represents corresponding blocks in two models. | 48 |
| 3.3 | Representation of a sample important structure (highlighted in yellow), and see how similarity between important structures can be captured in a better way by movement of templates in possible directions. (B stand for Block). | 49 |
| 3.4 | Correlation between original mutual information based distance and the projected distance using MDS technique. | 52 |
| 3.5 | Simulation results for 15 realizations, along with expected clustering results. | 52 |
| 3.6 | Actual results of clustering on 15 realizations, projected in 2D space using MDS. | 53 |
| 3.7 | Visualization of spatial contribution of realizations into similarity. | 54 |
| 3.8 | Calculating clustering based on a specific area of models which is selected inter-actively by users. | 54 |
| 4.1 | Workflow of SAGD process [AlbertaEnergy, 2009]. | 58 |
| 4.2 | Proposed visual analytics framework for selecting of reservoir models [Sahaf et al., 2018]. | 61 |
| 4.3 | Block based approach for calculating distance between realizations. | 63 |
| 4.4 | Calculation of distance between two blocks. | 63 |
| 4.5 | Clustering results of projected models with (b) and without (a) kernel transformation. | 65 |
| 4.6 | Simulation result and steam chamber for three different SAGD based reservoirs with varying shale continuity. (Black areas show sample of shale bodies. Smoky areas show the steam chamber and the orange cross shows the steam injection well.) | 66 |
| 4.7 | Proposed custom visual analytics process based on steam chamber area. (a) Original reservoir model, (b) steam chamber area, (c) highlighted steam chamber area for all the realizations (d) distance calculation only for the steam chamber area, (e) projection of models into lower space, (f) transformation of models into new space using kernel functions, (g) clustering of models in the new space. | 67 |
| 4.8 | Illustration of our case study with five reservoir models. First row shows the actual model and the second row shows the simulation results of models. Green areas in the second area show the steam chamber area. | 68 |
| 4.9 | Simulation results for five different realizations for SAGD process. | 69 |
| 4.10 | (a) Representation of a reservoir model without any shale structure. (b) Maximum steam chamber area that is extracted by running simulation for that reservoir model. | 70 |
| 4.11 | (Left) Clustering result with considering only the steam chamber area of models. (Right) Clustering result on the entire area of models. | 70 |
| 5.1 | An overview of our extended visual analytical framework. | 77 |
| 5.2 | Variance calculation between geological realizations. | 78 |
| 5.3 | A sample of calculated variance model. | 79 |
| 5.4 | Interactive filtering of the variance model. | 79 |
| 5.5 | Filtered out variance model. | 79 |

| | | |
|------|---|-----|
| 5.6 | Interactive creation of box for the group selection. | 80 |
| 5.7 | A sample representation of group selection using the interactive box. | 80 |
| 5.8 | Multiple creation of boxes for selection of the ROIs. | 80 |
| 5.9 | Single cell selection around a well. | 81 |
| 5.10 | A sample ROI selection. | 81 |
| 5.11 | Block based approach for calculating distance between realizations. | 82 |
| 5.12 | Calculation of distance between two blocks. | 82 |
| 5.13 | 2D view of clustering results of projected models with (b) and without (a) kernel transformation. | 84 |
| 5.14 | 3D view of clustering results of projected models with (b) and without (a) kernel transformation. | 84 |
| 5.15 | Representation of some outsider models in each cluster that can be considered as potential different models. | 86 |
| 5.16 | ROI analysis with selection of the most outsider model (a) and the second most outsider model (b), and observation of variance changes in each scenario. | 86 |
| 5.17 | Clustering result without the ROI selection (20 models). | 88 |
| 5.18 | Representation of cluster centers on the actual simulation plots (20 models and without the ROI selection). | 88 |
| 5.19 | Clustering result with the ROI Selection (20 models). | 89 |
| 5.20 | Representation of cluster centers on the actual simulation plots (with the ROI selection and 20 models). | 89 |
| 5.21 | ROI selection for the reservoir (20 models). | 90 |
| 5.22 | Clustering result without the ROI selection (100 models). | 91 |
| 5.23 | Representation of cluster centers on the actual simulation plots (100 models and without the ROI selection). | 91 |
| 5.24 | Clustering result with the ROI Selection (100 models). | 92 |
| 5.25 | Representation of cluster centers on the actual simulation plots (with the ROI selection and 100 models). | 92 |
| 5.26 | ROI selection for the reservoir (100 models). | 93 |
| 5.27 | Representation of cluster centers on the actual simulation plots (with the ROI selection and 100 models). | 94 |
| 6.1 | An ensemble of 3D reservoir geological models. | 103 |
| 6.2 | An overview of the proposed analytical framework for the uncertainty exploration of reservoir ensembles. | 105 |
| 6.3 | An overview of the visual analytics application design with the utilized technologies. | 106 |
| 6.4 | Structure of a CMG data file format. | 108 |
| 6.5 | Structure of a VTK data file format. | 108 |
| 6.6 | Representation of cell (structure and types) and point in the reservoir 3D geometry. | 109 |
| 6.7 | Filtering process for removing the low variance cells. (Left) Represents how the controllers are used to perform filtering. (Right) Represents filtering on a sample data. Blue cells are the representative of the low variance cells. For instance, when variance threshold is set to 0.5, such the low variance cells are discarded. | 110 |

| | | |
|------|---|-----|
| 6.8 | Group selection process. (Left) Represents the important controller keys that are used during the group selection process. (Right) Represents the creation of a sample box for selecting a group of cells. | 111 |
| 6.9 | Single cell selection process, which shows that users can look inside the model and select any cells by putting the controller inside the cell and pressing the trigger button. | 112 |
| 6.10 | Representation of cluster nodes for three clusters. (Left) Highlighted models show the center of clusters, which are the representative models. (Right) Process of hovering and selecting outlier models and inspecting the distances. | 113 |
| 6.11 | ROI analysis process for a sample scenario, which shows how selection of outlier models can change the overall variance of the representative models. | 115 |
| A.1 | Quantitative measurements for evaluating clustering results. (a) Actual simulation results for the entire ensemble shown for the oil recovery factor property. (b) Represents the expected clusters in two top and bottom tiers. (c) Illustrates the actual clustering groups, which can be seen that some of the clusters are grouped in the wrong clusters. | 126 |
| A.2 | Different facies distribution that are used to generate different realizations. | 127 |
| A.3 | An example of controlled case study that reservoir models are only different in terms of the controlled parameter. | 128 |
| A.4 | Porosity distribution in a sample real reservoir ensemble that multiple geological properties are changed at the same time. | 129 |
| A.5 | Sample clustering for three different shapes. | 130 |
| A.6 | Representation of expected clusters using the actual flow simulation results. | 131 |
| A.7 | Comparison of expected and actual cluster results and calculation of the correctly clustered items in each cluster. | 131 |
| B.1 | Representation of hierarchical clustering. | 132 |

List of Symbols, Abbreviations and Nomenclature

| Symbol | Definition |
|--------|---------------------------------|
| COP | Cumulative Oil Production |
| KMC | K-Means Clustering |
| OOIP | Original Oil In Place |
| MDS | Multi-Dimensional Scaling |
| MI | Mutual Information |
| NPV | Net Present Value |
| RBF | Radial Basis Functions |
| ROI | Region of Interest |
| SAGD | Steam Assisted Gravity Drainage |
| 3D | Three Dimensional |
| 2D | Two Dimensional |
| U of C | University of Calgary |
| VR | Virtual Reality |

Chapter 1

Introduction

1.1 Background

1.1.1 Reservoir Modeling

A reservoir is a trapped portion of the subsurface earth formations that contains hydrocarbon as a hydraulically connected system. Reservoir engineering concerns with the application of engineering principles for developing and producing oil and gas fluids from reservoirs in such a way to obtain a high economic recovery. In particular, it includes a study of the factors affecting the reservoir recovery [Terry et al., 2013]. Reservoir engineering studies are heavily based on reservoir models, where a high resolution computerized representation of reservoir (a portion of Earth's crust) is created. The outcome model is called geomodel which is represented by an array of discrete cells, delineated by a three-dimensional grid which may be regular or irregular in shape. One of the important features of this model is that it should embed important geological structures of the reservoir such as well paths, faults, fractures, etc (Figure 1.1); which have major impacts on the oil/gas flow performance [Cannon, 2018].

As a result, the main purpose of geomodel (or geological model) is to provide a static representation of the reservoir. The properties that are defined for the geological models are usually called “static” properties. Some of the important properties include porosity, permeability, and facies, whose values are associated with each cell and color-coded to the model display (Figure 1.2). Geological models are usually very large, and their size is defined by a large number of cells (between thousand to million cells).

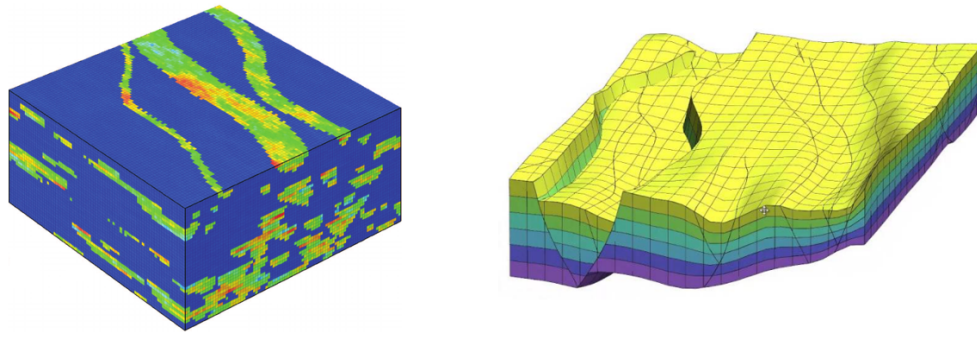


Figure 1.1: Representation of a sample reservoir model with channels [Frster et al., 2006] (left). Representation of a smple reservoir model with faults and fractures (right).

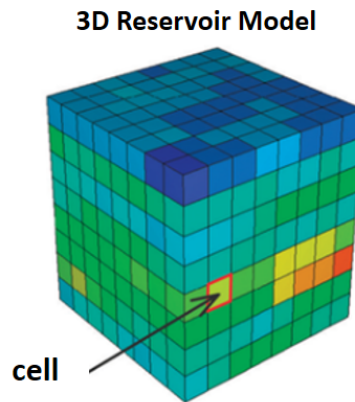


Figure 1.2: Representation of a reservoir model with a highlighted cell.

The next step is to use the geological models to predict the flow of fluids (typically oil, water, and gas) through the reservoir. However, the large number of cells present in the geological model can be problematic for flow simulations as a single flow simulation on the resolution of geological models is very time consuming. Therefore, the geological model should be scaled up to lower resolution models with coarser cells. The models that are created at this stage are called reservoir simulation models. The purpose of reservoir simulation is to predict field performance, ultimate recovery for various development scenarios and perform economic evaluations.

These models help the reservoir development strategies by identifying some factors such as

the number of wells, the present and the future needs for artificial lift and the expected hydrocarbon production. The properties that are resulted from reservoir simulations are usually called “dynamic” properties which include pressure, saturation, temperature, and recovery rates. Similar to the static properties, dynamic properties are also assigned to each cell. However, dynamic properties change over time, in a way that each cell has different values per time step. A typical workflow of reservoir modeling is shown in Figure 1.3, where geological models are first upscaled to coarser model (in order to reduce the computation cost of the reservoir simulation process), and then reservoir simulation model is developed to predict fluid flows over the time using different recovery mechanisms.

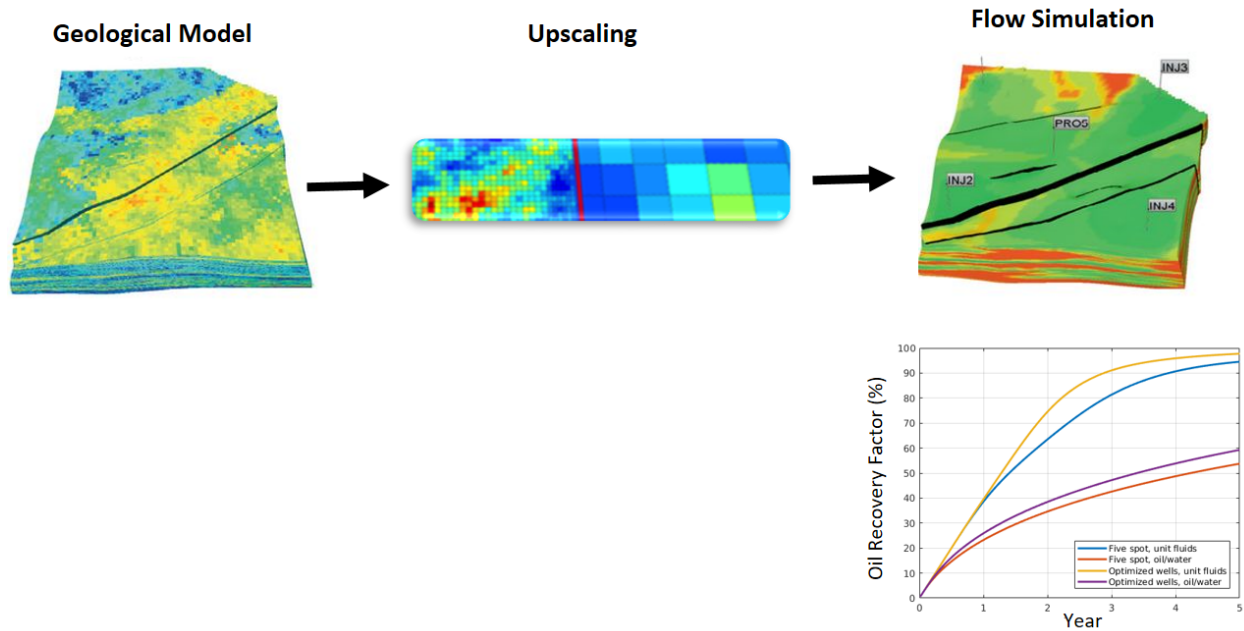


Figure 1.3: A typical reservoir modeling workflow ([Agada et al., 2017]).

1.1.2 Uncertainty in Reservoir Modeling

In the previous section, I explained about the static models (geomodels), where they incorporate all the geological features of an underground volume of rock that can store fluids (hydrocarbons and/or water) and can allow their movement. A geological model is the final product of structural,

stratigraphic and lithological modeling activities, where each of these steps is developed according to its specific workflow. In the structural modeling (Figure 1.4 - (a)), the structural properties of the reservoir are constructed, by defining a map of its structural top and the set of faults. This stage of the work is carried out by integrating interpretations of the geophysical surveys with the available well data. In the stratigraphic modeling (Figure 1.4 - (b)), a stratigraphic scheme of a reservoir is defined using well data, which forms the basis for well to well correlations. The data consist of electrical, acoustic and radioactive wireline logs and core analysis. In the lithological modeling (Figure 1.4 - (c)), basic facies types are defined for the reservoir, which are characterized on the basis of lithology, sedimentology, and petrophysics. Then, a petrophysical modeling stage provides a quantitative interpretation of well logs and core data to determine some of the main petrophysical characteristics of the reservoir rock (e.g. porosity, facies, water saturation, and permeability). Finally, the result of these different stages are integrated in a two (2D) or a three-dimensional (3D) context, to build an integrated geological model of the reservoir (Figure 1.4 - (d)). This model represents the reference frame for calculating the quantity of hydrocarbons in place, and on the other, forms the basis for the initialization of the dynamic model [Dake, 1983].

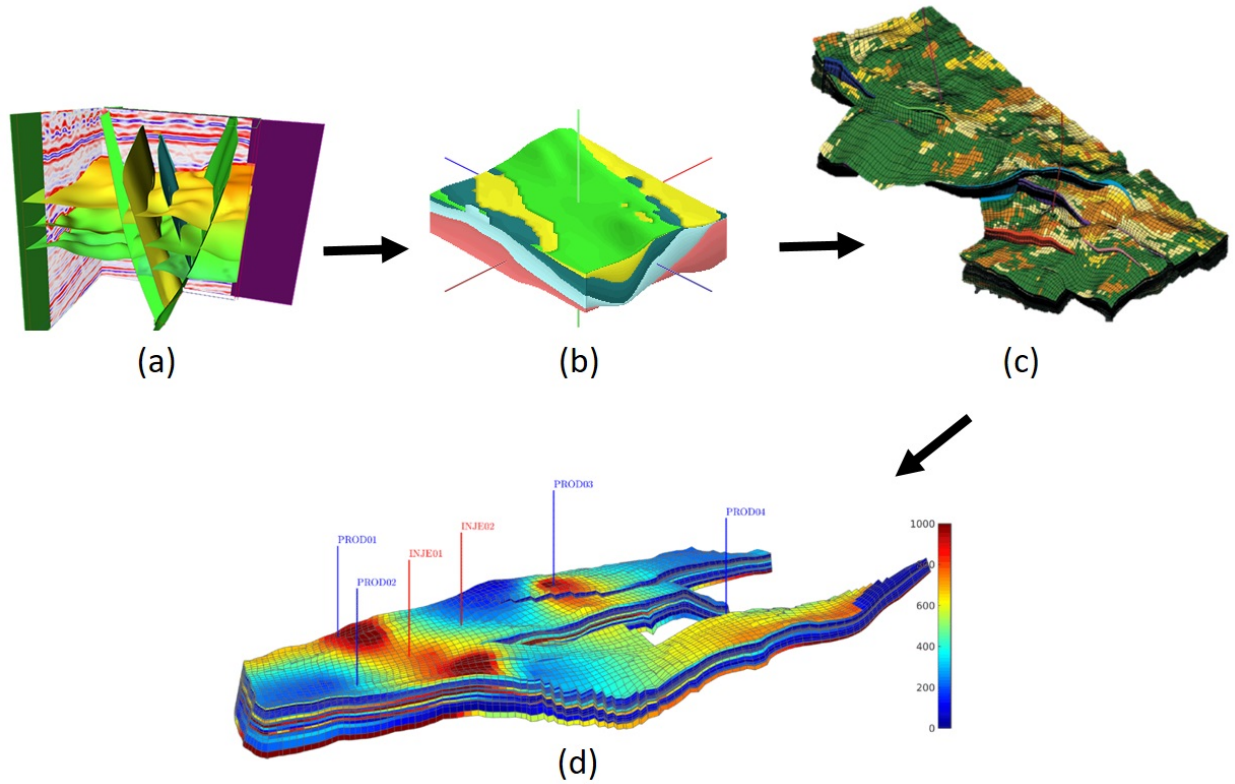


Figure 1.4: GeoModeling workflow. (a) A structural model showing faults and layering, (b) Example of a 3D Stratigraphic Model, (c) Example of a stochastic model of facies, (d) A geological model showing permeability distribution and values

The subsurface complexity and limited data make the reservoir characterization and modeling complex and indeterministic which imports large uncertainty in all the steps of the geomodeling workflow. In general, uncertainty is related to the poor knowledge of a phenomenon and it can be seen in many aspects of technical researches, engineering as well as everyday life [Caers, 2011]. Specific reasons of uncertainty are different in each domain and different terminologies are used to describe, quantify and assess uncertainty. This also occurs in the petroleum engineering applications and there are a variety of sources of uncertainty in the reservoir modeling process and also other areas like reservoir simulation [Wellmann and Regenauer-Lieb, 2012]. The main focus of this research is on the uncertainty assessment of high resolution or up-scaled geological models

(the last product of geomodeling process) and its relationship with the reservoir simulation models, using the visual analytics techniques.

Geological properties are important parameters used in the design and optimization of oil extraction processes from reservoirs. These parameters dictate the ease with which oil can be extracted and also the quantity of oil that can be stored and recovered. Optimal location of wells and the well controls depend on the geological properties of the reservoir. Geological uncertainty exists because it is not possible to know the exact geological properties of every section of a realistic reservoir. Figure 1.5 illustrates a typical workflow for addressing the uncertainty in reservoir models. Techniques such as well logging and coring can give some idea about the geological properties of particular areas of the reservoir (Figure 1.5 - (a)). However, the geological parameters of many areas like the areas between the exploration wells will still be unknown (Figure 1.5 - (b)). As a result, geological uncertainty will always exist in the distribution of properties within the gridded reservoir volume. Therefore, to quantify uncertainty, many equiprobable reservoir models, called geological realizations (also called ‘geo-realizations’ or simply ‘realizations’), are generated by the geostatistical algorithms [Goovaerts, 1998] to reflect the lack of knowledge in representing the geology (Figure 1.5 - (c,d)). Nevertheless, none of these realizations is exactly the same as the real reservoir itself. The difference between several realizations is assessed to obtain a range of uncertainty in the geological properties of the reservoir [Caers, 2011]. This assessment is performed by running the flow simulation for all realizations to determine the reservoir performance that is reflected in terms of some parameters such as Cumulative Oil Production (COP) or Net Present Value (NPV) (Figure 1.5 - (e,f)). Practically, running flow simulations for all the possible realizations (especially at the scale of geological models) can be cumbersome and wasteful, since many of the realizations may exhibit nearly equal flow performance. In addition to that, depending on the type and the size of reservoir models, flow simulations can be very time-consuming which might take hours to even months depending on the size of the geomodels. Therefore, it is very critical to be able to select few models to run in order to make a good decision with a short

turnaround.

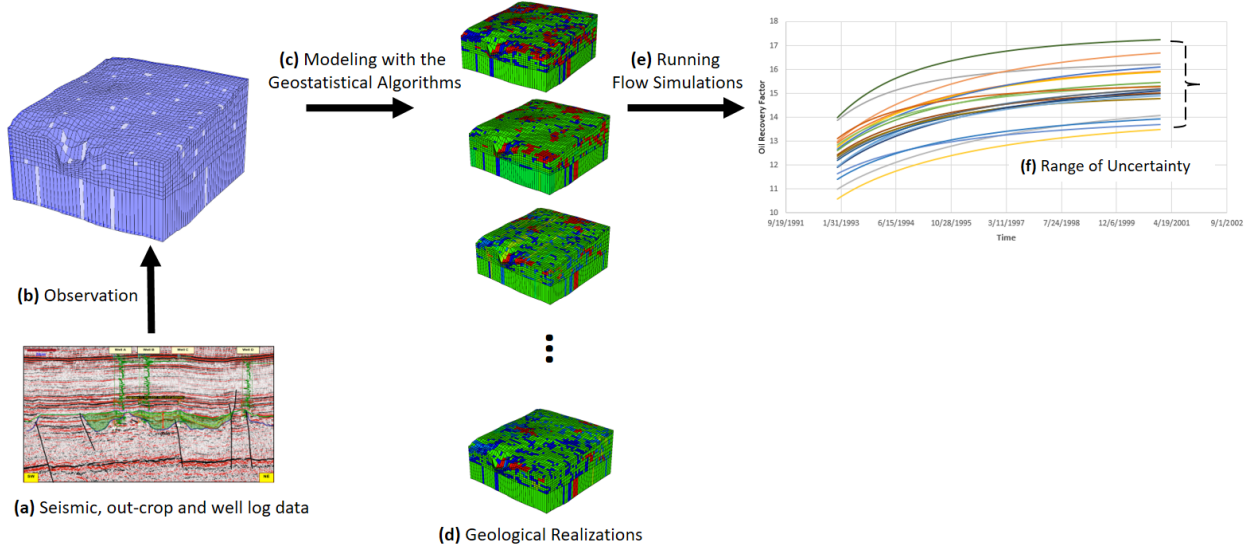


Figure 1.5: The workflow of generating geological realizations and assessing uncertainty. (a) Sample sources of information for exploring reservoir models. (b) Observation of the available information and mapping of them to the reservoir models that results into an initial reservoir model with few known and explored areas. (c) Application of geostatistical algorithms for generation of a set of reservoir models to cover the uncertainty for the unknown areas. (d) A set of 3D geological realization models. (e) Running flow simulation for all the generated realizations and observing the uncertainty of a flow related property. (f) Representation of range of uncertainty for an interest dynamic property.

In this research, the fundamental dataset is that an ensemble of 3D realizations that is generated as part of the workflow presented earlier. In this set of realizations, as it can be seen from the Figure 1.5, all the realizations have the same structure, but they are different in terms of the distribution of geological properties. It is very critical to understand that 3D structure because it embeds important geological structures in the reservoir such as well paths, faults, fractures, etc; that have a major impact on the dynamic flow behaviors.

1.2 Motivation

The geological uncertainty may be described by generating and incorporating a large set of realizations. However, in practice, only a few realizations are selected from a large superset for the assessment of uncertainty range of production behaviors, due to the intensive computational efforts in reservoir flow simulation processes. Various methods are available in the reservoir engineering literature for selecting geological realizations that can be broadly classified as random selection, ranking [Ballin et al., 1992] and distance based techniques [Scheidt et al., 2009] [Rahim and Li, 2015]. Random selection is the easiest approach for implementation where the reservoir engineers randomly select a few realizations, to run the flow simulation. Obviously, with the random selection of models, there is no guarantee to properly describe the certainty. On the other hand, the selected realizations may exhibit very similar flow behaviors, and result in a wrong estimation of uncertainty range. Ranking is one of the well-known approaches for selecting the reservoir models. In this technique, realizations are sorted based on an easily computable measurement, and then models are selected that have low, medium and high values of that defined measurement. Many measures have been defined for different types of reservoirs, and that is one of the main limitations of this approach, i.e. there is not a general approach for selection of models. In addition, ranking of models is accurate, if the measure has a high correlation with the flow related properties. The other recent category of selection techniques is the distance based methods, that calculate pair-wise distances between the models, group the models based on their similarities, and finally select few different models. Although this category relies on the pair-wise distances between realizations, they all suffer from a comprehensive definition of distance that considers 3D attributes within the reservoir models.

Besides the limitations that mentioned for all the current selection methods, they are all considered to be one-time automated processes, that lack any visualization perspectives and more importantly do not incorporate the user's knowledge into the selection process. Additionally, most of the proposed approaches are defined for very specific scenarios and type of reservoirs. There-

fore, a definition of a general approach is very needed for selection geological realizations from a superset of realizations.

The type of dataset that I am studying in this research (a superset of 3D geological realizations) is referred to as multi-run data in computer science domain [Kehrer and Hauser, 2013]. This type of dataset can be found in other domains such as climate research [Hibbard et al., 2002] [Nocke et al., 2007] and engineering [Matkovic et al., 2005], and it is often generated to study the variability of a model and to understand the model sensitivity to a certain control parameter. In general, multi-run data stems from a type of process (like geostatistical algorithms for the generation of realizations) that is repeated multiple times with varied parameter settings, leading to a large number of collocated data volumes given for the same space and time [Wong and Bergeron, 1994] [Wilson and Potter, 2009]. Therefore, a collection of values coexists for the same data attribute at each space-model location [Love et al., 2005]. Since multi-run data consists of a superset of volumetric models, their representation and analysis are found to be very challenging in the computer science literature [Kehrer and Hauser, 2013]. Many researches have been performed on the visualization and analysis of multi-run data. Most of them are proposed either for the 2D datasets such as 2D images and maps or the mapping and aggregation of 3D data into 2D objects. The well-known approaches for visual analysis of uncertainty in multirun data include box-plots [Kao et al., 2002], line charts [Demir et al., 2014], glyphs [Kehrer et al., 2011] and parallel coordinates [Nocke et al., 2007]. Unlike these proposed techniques, in this research, it is tried to more focus on the 3D structure of the 3D models in both visualization and analysis aspects, other than mapping them into the 2D objects.

In conclusion, according to the above mentioned limitations and my continuous discussions with the domain experts, I found out that the need from geoscientists and reservoir engineers to address the geological uncertainty using a limited number of geological realizations necessitates designing an analytical framework that is computationally less expensive, dependent to the static properties of geological models rather than flow simulation results, visual and interactive, and

capable of showing differences and similarities between the models.

1.3 Methodology and Contributions

To address all these requirements as discussed in the previous section, I have designed and developed a visual analytics framework to identify the models that can potentially cover the full uncertain space (that is referred to as "representative models"). The proposed process can resolve the existing issues of previous studies in terms of computation cost, time efficiency, and the user involvement. Moreover, unlike the existent approaches, my proposed approach is modular and can be targeted for different types of reservoirs.

The first step in the proposed process is to establish a representative metric for the calculating distance between a pair of 3D reservoir models. There are limited comprehensive distances defined in the domain for the reservoir models. As such, I designed and introduced a new distance measure based on the mutual information (MI) concept [Lin, 1998] [Goshtasby, 2012]. One critical aspect of the new proposed measure is that it first divides the model into smaller 3D blocks, where each block consists of several cells. Distances are first calculated locally between the blocks. The final distance value is the average of all the block-wise distances. This feature helps to consider the important local variabilities. Thereafter, dimensionality reduction techniques are implemented to use the measured distances to calculate new distances required for mapping the reservoir models into lower spaces (like a point in a 2D or 3D space). The lower dimension helps in obtaining more accurate clustering results, that is performed at the next step. Clustering would be then performed on the lower space. Each cluster contains similar geological realizations, and the cluster center is the representative of the subject cluster. Therefore, consequently, I have the cluster centers that can represent the uncertainty of the entire ensemble of the realizations.

In order to have more accurate results (in terms of distance calculation between the reservoir models and clustering of the models) and better performance (in terms of time), different region of interest (ROI) selection techniques are also introduced to let the users perform the selection

analysis on particular areas of the reservoir model. I have designed and developed the entire framework in a visual interactive process, both in a desktop application and also in a virtual reality environment, which enhances the incorporation of the user's knowledge in the entire process.

In summary, the outcome of the framework is the representative models (or cluster centers) that have potentially different simulation results and can cover the full range of uncertainty. Therefore, instead of running simulation for all the models for the uncertainty assessment, users need only run the flow simulation for the filtered representative models and save considerable time in finding out the uncertainty range. The effectiveness of my proposed framework is evaluated in two phases. Initially, I compared the actual flow simulation results of the selected models (representative models) with the brute-force approach. A particular selection is accurate when the simulation results of all models in one cluster are very similar to each other. In addition, the representative models should cover a similar cumulative hydrocarbon production uncertainty range as the brute force approach. In the second phase of the evaluation, I performed a complete formal user study with a number of engineers and geologists. User studies involve certain tasks that cover all the key aspects of the framework. User feedback suggests that the usability and visual interactivity of our application are the key strengths of our approach. In summary, the main contributions of this thesis are:

- Novel dis(similarity) metric for calculating pairwise distances between the 3D geological models. (Chapter 2)
- Analytical framework for uncertainty assessment of dynamic properties (e.g. cumulative oil production) that utilizes the proposed similarity metric. (Chapter 3, 4)
- Different selection techniques for extracting the region of interest (ROI) for the reservoir 3D models. (Chapter 5 and 6)
- Spatial visual inspection of similarities and differences between the reservoir 3D models. (Chapter 5 and 6)

- A visual analytics desktop application that supports the proposed framework and provides visual and interactive tasks for steering of the uncertainty assessment process. (Chapter 2)
- A virtual reality environment that supports the proposed framework and provides additional interactivity and user engagement in the uncertainty assessment process. (Chapter 6)

1.4 Organization of Thesis

This thesis is organized as a combination of published and submitted peer-reviewed conference and journal papers (Figure 1.6).

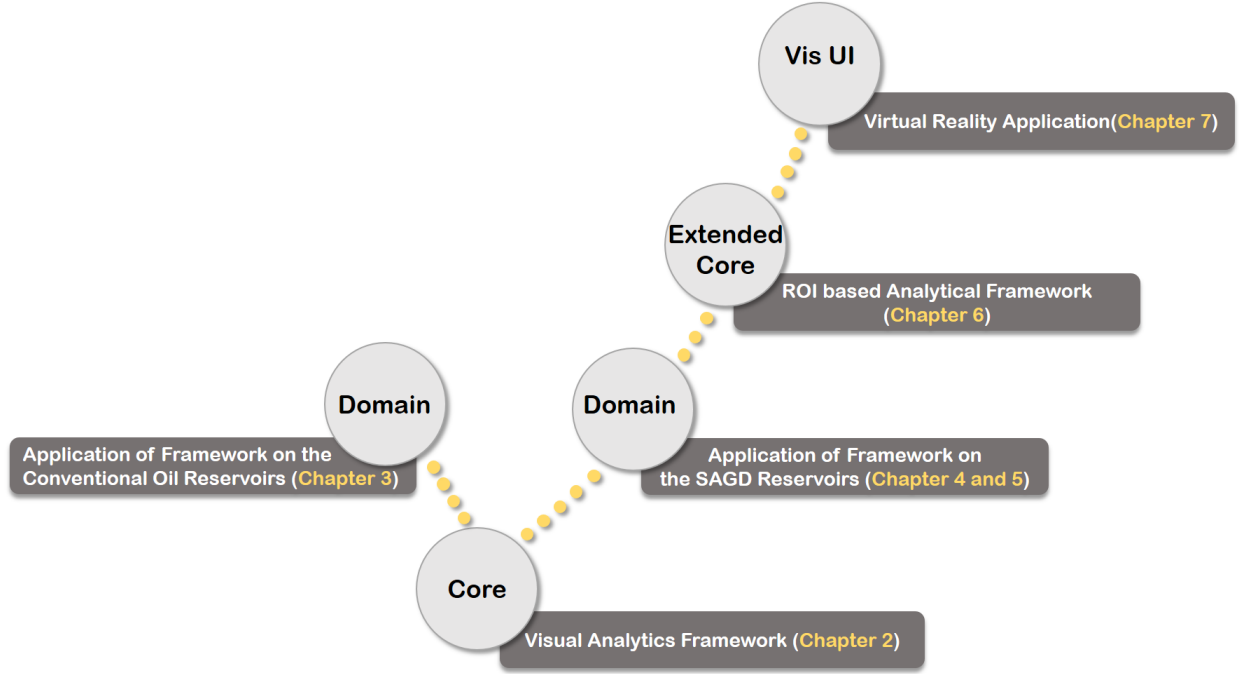


Figure 1.6: Organization of thesis. “Core” is the first proposed visual analytics framework for exploring uncertainty in reservoir models. Then, that framework is applied to two different reservoir engineering “domains”. In one of the domain applications, the idea of ROI is proposed for that specific domain problem. The concept of ROI generalized to be applicable to different problems, and that creates the “Extended Core”. Finally, “Vis UI” presents the designed and developed visual interactive framework that supports the proposed analytics framework.

Chapter 2 presents a published paper that introduces the complete proposed analytical framework for exploring uncertainty in the reservoir model ensembles. In this paper, I introduce the new dissimilarity metric for the calculation of the pair-wise distance between reservoir 3D models. Then, I would explain the main concepts used in defining this distance and elaborate on why this distance is more effective than other common distances. Then, I show how this distance is used with the clustering algorithms - to provide groups of similar models, with the representative models. Lastly, it is been shown how this analytical framework is designed and developed within

a visual interactive desktop application, with multiple coordinated views and several interactive tasks. This paper is published and presented in the IVAPP conference (International Conference on Information Visualization Theory and Applications), which is among the top venues in visual analytics and information visualization domains with the acceptance rate of 15%.

Sahaf, Z., Hamdi, H., Cabral Mota, R., Costa Sousa, M. and Maurer, F. (2018). A visual analytics framework for exploring uncertainties in reservoir models. In 9th International Conference on Information Visualization Theory and Applications (IVAPP 2018).

Chapter 3 provides a published domain paper that applies the proposed analytical framework on a very specific case study for the conventional oil reservoirs. I explain how the proposed distance and clustering technique is effectively applied on a specific case study. It was my first publication in the EAGE conference (European Association of Geoscientists and Engineers), which is one of the largest multidisciplinary venue in the reservoir engineering and geoscience domain and has the acceptance rate of around 35%. The main purpose of this publication was to observe how the framework is evaluated by the domain experts.

Sahaf, Z., Hamdi, H., Maurer, F., Nghiem, L., and Sousa, M. C. (2016). Clustering of Geological Models for Reservoir Simulation Studies in a Visual Analytics Framework. In 78th EAGE Conference and Exhibition 2016, DOI 10.3997/2214-4609.201601143.

Many domain experts very well accepted this application of the proposed analytical framework on the conventional oil reservoir. As such, I received many requests to observe if the proposed framework can be also used for the SAGD process (Steam Assisted Gravity Drainage is a very well-known reservoir technique for extracting heavy oils from the reservoirs located in Alberta). After detailed investigations, I found out that the proposed framework can not be directly applied on a SAGD model without additional consideration. The main reason is that SAGD models consist of many shale structures that affect the oil drainage volume. To extract oil from such reservoirs, depending on the distribution of shales, two wells are drilled in parallel on top of each other. Steam is injected into the top well to warm the in situ oil and to drain the mobilized oil into

the other well (producer). As a result, steam rises with injection time forming a steam chamber within the reservoir. I found out that the steam chamber area is the most important region in such reservoir that controls the flow performance. Therefore, I changed the proposed framework to calculate the distance between the reservoirs only based on a specific area, which in this case is the steam chamber area. This new adapted framework is published in a second paper in the EAGE venue. Due to the space limitation for the prepared papers in this EAGE venue, I prepared a complete extended version of this paper and submitted it to the First Break journal. This extended complete version is available in Chapter 4. First Break journal publishes top-quality submitted and commissioned research articles related to the applied geophysics, petroleum geoscience, and reservoir engineering fields. In addition to that, this journal has a very high visibility from both industry and academia.

Sahaf, Z., Hamdi, H., Maurer, F., Nghiem, L., Chen, Z., and Sousa, M. C. (2017). Filtering Geological Realizations for SAGD. In 79th EAGE Conference and Exhibition 2017, DOI 10.3997/2214-4609.201701313.

Sahaf, Z., Hamdi, H., Maurer, F., Nghiem, L., Chen, Z., and Sousa, M. C. (2017). Filtering Geological Realizations for SAGD. Submitted to First Break Journal (<http://fb.eage.org/>).

For Chapter 4, the proposed analytical framework has modified to only consider a specific area of the reservoir, that is the steam chamber area, to calculate the distance between the models to perform the clustering. Since this idea was performed well on different cases studies, I decided to make this idea more general in a way that depending on the type of reservoir, the users should have this flexibility to interactively select any regions of interest in the reservoir. Subsequently, the entire analytical framework performs on the user selected regions of interest and thereby generates faster (fewer areas are involved in the calculations) and more accurate (in terms of distance calculation and uncertainty range explorations) results. To support this idea, I designed three visual interactive selection techniques and then extended the framework to have a region of interest (ROI) selection mechanism using the selection techniques. The effectiveness of ROI selection mechanism is

elaborated with two comprehensive case studies. Finally, I performed a complete user study with the geologists and reservoir engineers to evaluate the extended framework as part of the designed user study sessions. Chapter 5 provides the paper that I wrote about the materials concerning this extended framework. This paper is submitted to the journal of Communications in Computer and Information Science series published by Springer.

| |
|---|
| Sahaf, Z., Cabral Mota, R., Hamdi, H., Costa Sousa, M. and Maurer, F. (2018). An ROI Visual-analytical Approach for Exploring Uncertainty in Reservoir Models. Submitted to the Journal of Communications in Computer and Information Science (Springer), https://www.springer.com/series/7899 . |
|---|

In order to let the users have a natural engagement feeling while performing the ROI selections and distance inspections between the models (according to the mentioned extended framework), the extended framework is designed and developed in a virtual reality (VR) environment. Many visual and interactive tasks have been designed for the users in this environment, that helps them easily and visually explore the uncertainty within the ensemble of reservoirs. Chapter 6 describes all the details of the designed and developed VR tool that supports my proposed analytical framework. Finally, Chapter 7 presents the conclusion and remarks on future work.

In all the papers, I was the primary contributor in terms of the idea and also writing. Dr. Hamidreza Hamdi helped me a lot with the validity and evaluation of techniques in the reservoir engineering domain since his background is in the geoscience and reservoir engineering domains. He also helped with reviewing and proof-read the papers. My lab colleague, Ms. Roberta Cabral Mota, collaborated a lot on the designing and implementing an immersive visualization and visual analytics version (Chapter 6) of my desktop-based research application that I originally developed (Chapter 2). She also helped me with the designing and conducting the user studies and system evaluation (Chapter 5 and 6). Dr. Costa Sousa and Dr. Frank Maurer (supervisors), helped with the review of papers and the overall structure and content in relation to the key research goals of my Ph.D. research. Dr. Long Nghiem, vice president of research and development at the Computer

Modeling Group (CMG) company, helped me with his constructive comments about the reservoir engineering aspects of the thesis research.

Chapter 2

A Visual Analytics Framework for Exploring Uncertainties in Reservoir Models ¹

2.1 Abstract

Geological uncertainty is an essential element that affects the prediction of hydrocarbon production. The standard approach to address the geological uncertainty is to generate a large number of random 3D geological models and then perform flow simulations for each of them. Such a brute-force approach is not efficient as the flow simulations are computationally costly and as a result, domain experts cannot afford running a large number of simulations. Therefore, it is critically important to be able to address the uncertainty using a few geological models, which can reasonably represent the overall uncertainty of the ensemble. Our goal is to design and develop a visual analytics framework to filter the geological models and to only select models that can potentially cover the uncertain space. This framework is based on the mutual information for the calculation of the distance between the models and clustering for the grouping of similar models. Interactive visualization tasks have also been designed to make the whole process more understandable. Finally, we evaluated our results by comparing with the existent brute force approach.

2.2 Introduction

Uncertainty is related to poor knowledge of a phenomenon. In petroleum engineering applications, for instance, we have lots of uncertainties in all aspects of petroleum production phases.

¹Sahaf, Z., Hamdi, H., Maurer, F., and Costa Sousa, M. (2018). A visual analytics framework for exploring uncertainties in reservoir models. In In 9th International Conference on Information Visualization Theory and Applications (IVAPP 2018).

This is essentially due to a large number of unknowns that exist at any particular stage of exploration, development, and production workflow. In the exploration phase, which is the focus of this paper, the lack of knowledge in representing the measured data (e.g., due to noise), expressing the depositional settings, spatial configurations of the rock types, or mathematical uncertainty in representing the geology, are the key elements that largely impact the decision making process based on modeling [Caers, 2011].

Reservoir models are essential to portray the impact of uncertainty. A reservoir model is a 3D grid-based digital representation model of the subsurface composed of a large number of cells/voxels. Each cell has a location in space and a set of attributes describing geological properties (such as porosity and permeability).

Geostatistical methods are used to estimate attribute values of the cells where no information is available. The inherent uncertainties of these geostatistical models imply that the attribute values of a cell can be assigned to different values and still be consistent with known facts. Geologists capture the inherent uncertainty by creating a large number of models. Flow simulations then take the models as an input and determine the expected outcome on variables of interest (like overall oil production volume) over time. The large number of cells in the digital reservoir model and the computational cost of processing flow simulations (i.e., usually requiring several hours) prohibit a brute-force approach for conducting the numerical flow simulation for all possible models. Therefore, it is substantially favorable to carefully select a few models with great diversity that can reasonably represent the overall uncertainty [Idrobo et al., 2000].

Another critical requirement for domain experts is an efficient way to compare the 3D models in a large ensemble without running the costly flow simulations. This aspect can help identify how models are different or similar to each other spatially and visually. Using that, the users can find out which spatial areas have more contribution toward quantifying the uncertainty and thereby the oil production.

To address all these requirements, we have designed and developed a visual analytics frame-

work to identify the models that can potentially cover the uncertain space (that is referred to as “representative models”). The proposed process can resolve existing issues of previous studies. Current techniques like ranking [Ballin et al., 1992], random selection, or probability-based techniques [Rahim and Li, 2015], are all costly regarding computation. They are automatic processes preventing the domain experts from guiding the selection process. Moreover, they are not modular and target only some specific types of reservoirs [Yazdi and Jensen, 2014].

The first step in the proposed process is to establish a representative metric for calculating the (dis)similarity between a pair of 3D geological models. As such, a new distance measure has been designed based on the mutual information (MI) concept [Lin, 1998] [Goshtasby, 2012]. Distances are then employed within a clustering algorithm to create sets of similar models (i.e., models where simulation results are likely to be similar). In this state, cluster centers are identified as the default representative models, and the users can only run the simulation for this limited set of selected models. We show the accuracy of our selection method by comparing the actual flow simulation results of the selected models with the brute-force approach. A particular selection is accurate when the simulation results of all models in one cluster are very similar to each other. In addition, the representative models should cover a similar cumulative hydrocarbon production uncertainty range as the brute force approach. In summary, the main contributions of this paper are:

- Novel dis(similarity) metric for calculating pairwise distances between the 3D geological models (section 5).
- Analytical framework for uncertainty assessment of dynamic properties (e.g. oil production) that utilizes our proposed similarity metric (section 4 and 6).
- A visual analytics tool that supports the proposed framework and provides visual and interactive tasks for steering the uncertainty assessment process (section 8).

2.3 Related Work

2.3.1 Current Approaches for Selecting of Geological Models

Various methods are available for selecting geological realizations which can be broadly classified as random selection, ranking, probability distance-based realization reduction method, and clustering technique.

While randomly selecting a subset of realizations is a straightforward method for implementation, it may result in a wrong measure of geological uncertainty especially when the number of selected realizations is small. Ranking [Ballin et al., 1992] is the most common method for selecting geological realizations. This method arranges the geostatistical models based on an easily computable measure in an ascending/descending order and then selects the ones that have low, medium, and high values of that measurement. One of the major limitation of the existing ranking methods is that they rely significantly on the measure used. If the measure has a weak correlation with the production of the reservoir, then the selected models will not adequately represent the full set of realizations [Li et al., 2012].

Probability distance-based realization reduction/selection methods have also been recently investigated by some researchers [Rahim and Li, 2015]. In this approach, an optimization problem is solved to find an optimal subset that has similar statistical distribution characteristics to the superset of models. The main issue with these optimization problems is that they could be very complicated and time-consuming for a broad set of models, and in the presence of the outlier models, the optimization process might not converge. Clustering methods have also been proposed recently in the domain. For instance, [Scheidt and Caers, 2010] used simplified simulation results to compute the distance between the models to form a distance matrix. Then these distances were used to perform the clustering. The need of petroleum industry to address the geological uncertainty using a limited number of geological realizations necessitates designing an analytical framework that is computationally less expensive, dependent on the static properties of geological models rather than flow simulation results, visual and interactive, and capable of showing differences and similarities

between the models.

2.3.2 Visual Analytics Techniques for Multirun Data

The most similar dataset in computer science domain to the geological models in petroleum engineering is multirun models. In areas such as climate research and engineering, multirun data is often generated to study the variability of models and to understand the model sensitivity to specific control parameters [Kehrer and Hauser, 2013]. In general, multirun data stem from a type of process (like geostatistical algorithms in our case) that is repeated multiple times with varied parameter settings, leading to a large number of collocated data volumes [Wilson and Potter, 2009]. Since multirun data consists of a superset of volumetric models, their representation and analysis are challenging.

The representation of multirun data is somewhat new to the visualization community. It is a challenging task since the data is often high dimensional, multivariate, and large at the same time. Accordingly, one of the common ways is to aggregate the distributions of multirun data, by computing statistical summaries [Love et al., 2005]. Subsequently, the resulting data is visualized using mainly box plots [Kao et al., 2002], line charts [Demir et al., 2014], glyphs [Kehrer et al., 2011], or InfoVis techniques such as parallel coordinates or scatterplot matrices combined with statistics [Nocke et al., 2007].

On the analytics side of multirun data, statistical methods are among the first candidates to be used for reducing the data dimensionality. For example, [Kehrer et al., 2010] proposes a method to integrate statistical moments (mean, variance, skewness, and kurtosis) into the visual analysis of multirun data. Alternatively, mathematical and procedural operators are also used to transform the multirun data into some compact forms (e.g., streamlines, isosurfaces, or pseudocoloring) where existing visualization techniques are applicable [Love et al., 2005] [Fofonov et al., 2016].

Data mining techniques are also among the recent methods being used to explore the multirun data [Correa et al., 2009]. [Bordoloi et al., 2004] applied hierarchical clustering techniques to multirun data. In a recent work, Bruckner and Moller [Bruckner and Moller, 2010] presented a re-

sult driven exploration approach for physically-based multirun simulations. Each volumetric time sequence is first split into similar segments over time and then is grouped across different runs using a density-based clustering algorithm. This approach supports the user in identifying similar behavior across different simulation runs.

Our analytics approach fits into the data mining category since the similarity between ensembles needs to be discovered both effectively and visually. However, most of the proposed clustering approaches are based on the 2D ensembles such as images. Even though the 3D ensembles are available, their aggregations are used for the clustering task. In this work, we want to perform the clustering on the 3D ensembles directly without using their aggregation, and that requires an accurate definition of distance between geological 3D ensembles.

2.4 Requirement Analysis

Through extensive discussions with the reservoir engineers from our industry partners, we gathered the following required elements as engineering requirements needed for designing a useful analytics framework:

R1. Limited selection of reservoir models. How can we select few representative models from a super set of models?

R2. Low computational cost. How can we have a fast selection process? The reason is that reservoir engineers usually want to save time in the engineering tasks (running costly simulations) as much as possible.

R3. Flexibility on the reservoir properties used in the selection process. Depending on the type of reservoir, different reservoir properties are provided.

R4. Flexibility on the area of interest. How can we perform the selection process based on an area of interest in the reservoir model? Usually, only specific areas of reservoir model (e.g., areas around wells) are considered significant. Therefore, there is a need to perform the selection process based on a region of interest that the user selects.

R5. Illustration of reservoirs (dis)similarity distance. For instance, which area of models contribute more in the (dis)similarity value or how to spatially and visually observe the comparison between a set of 3D reservoir models.

2.5 Analytical Framework for Descriptive Uncertainty Assessment: Overview

An overview of our proposed process is represented in Figure 2.1. Initially, we have a set of 3D geological models (a). Our proposed block-based similarity metric is calculated for all pairs of models (b). The similarity values are then utilized to project models into a 2D space using multidimensional scaling techniques (c). Each point in 2D space corresponds to a 3D model. The distance between points in the projected space represents the similarity between models, the closer the points, the more similar the 3D models. The final step is to cluster the points and pick a representative model from each cluster (d). This is an iterative and interactive process, which depends on the size of the similarity block, clustering property, the number of clusters, etc. Each of these stages is explained in more detail in the subsequent sections.

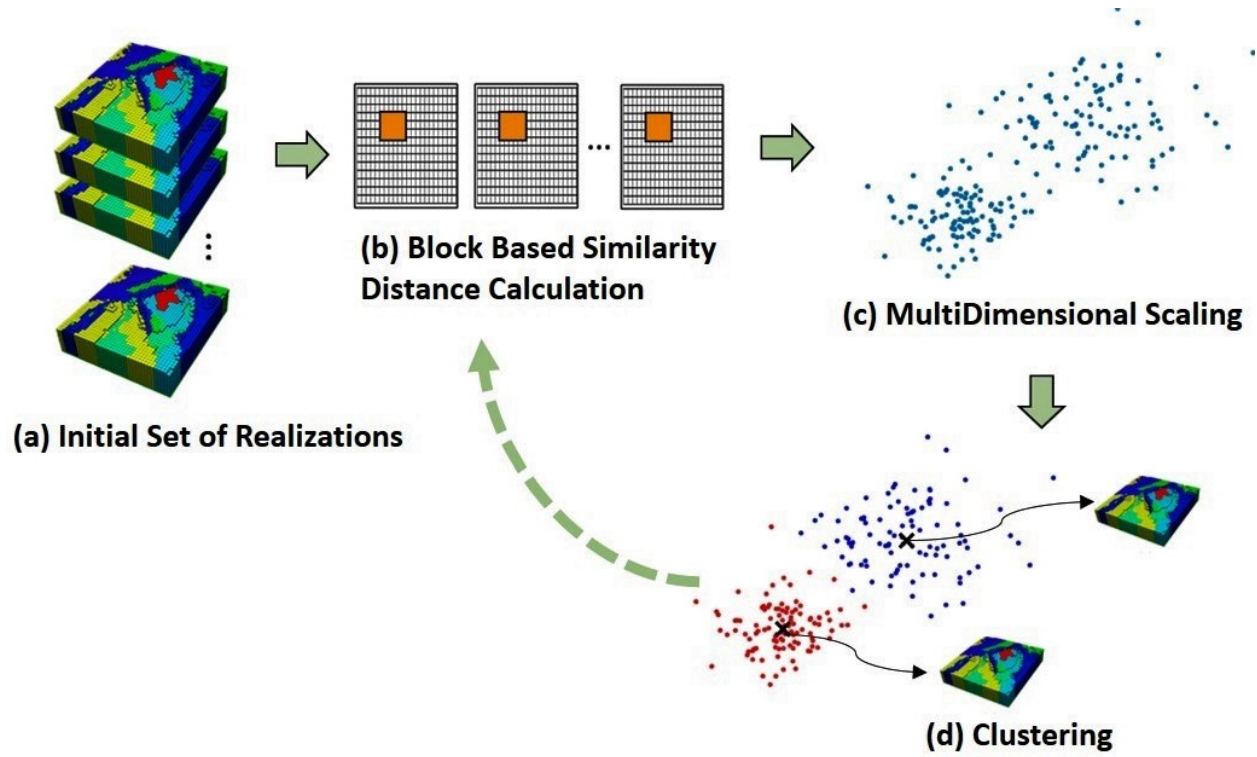


Figure 2.1: Overview of proposed filtering Process.

2.6 Calculation of (Dis)similarity Metric

A majority of reservoir simulation studies are performed on a single geological model. Therefore, 'distance' between reservoir models is somewhat new concept in that domain [Fenwick and Batycky, 2011]. Therefore it is essential to define a distance that reflects the requirement of the engineering tasks. Two models are called similar when they have a similar dynamic result (reservoir performance). The (dis)similarity distance can be calculated in a manner that should leverage two primary requirements. First, it has to be well correlated to the dynamic behavior of reservoir or flow response(s) interest [Scheidt et al., 2009]. Second, its calculation should not be very costly. According to our discussions with the domain experts and domain literature evaluation studies [Rahim and Li, 2015], static measures meet the two mentioned requirements and much preferred than dynamic mea-

asures. The reason is that static measures are simplified metrics designed to achieve a good correlation with the reservoir production performance variable of interest. For instance, Original oil-in-place(OOIP) is one of the critical terms calculated in reservoir simulation. It is calculated by the summation of the product of the following static properties volume (V), porosity (ϕ) and oil saturation (S) of cell c ($OOIP = \sum_c (V_c \phi_c (1 - S_c))$). Therefore, it shows how static properties are highly correlated with the dynamic (flow) terms. In addition to that, static measures are computationally much easier for evaluation when compared to reservoir flow simulation. It can be easily calculated for a broad set of realization (R2). In the next sections, we explain how these static measures use with our proposed similarity metric.

Reservoir models have 3D geometries with correlated spatial properties. Additionally, they can have some favorable 3D sub-structures (e.g., geological channels). Hence, any appropriate similarity measurement should be able to acknowledge these lateral geological heterogeneities (Figure 2.2). Therefore, we use a moving 3D template (block) approach to calculate (dis)similarity distance between a pair of models. The idea is to divide each 3D model into a set of smaller 3D blocks (Figure 2.2) where each block consists of a specific number of grid cells. The (dis)similarity measure is computed between corresponding blocks (templates) initially. Next, we take an average of the (dis)similarity values between all corresponding blocks. During this process, to reduce the bias of the fixed spatial position of blocks, we move the blocks in specific directions (x , y , z and diagonal) and distances (>1 and $<\text{block size}$). The final distance will be the average of (dis)similarity values in all the possible movements. For simplicity, a 2D representation of movement is shown in Figure 2.3. The yellow highlighted cells represent a prominent geological feature. Equation 2.1 shows how the final (dis)similarity is calculated between the two sample models with one movement and two states.

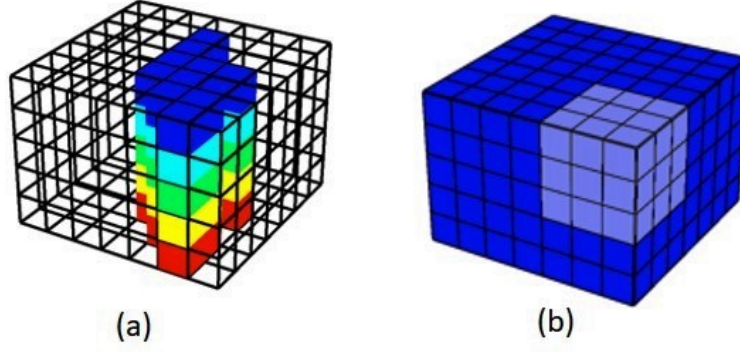


Figure 2.2: (a) A sample important 3D structure in the geological models. (b) A sample 3D block

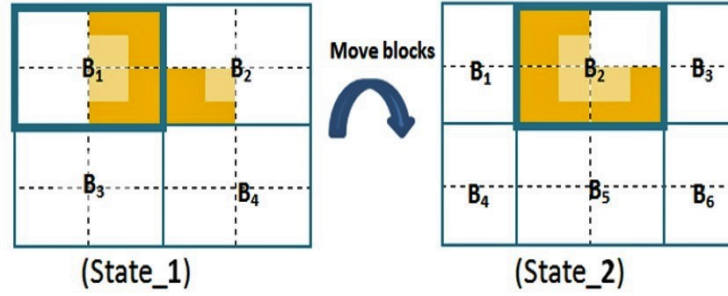


Figure 2.3: Representation of a sample favorable structure (highlighted in yellow), and movement of templates (dark blue frames).

$$MI = \text{MutualInformation}, Sim = \text{Similarity}, B = \text{Block}, M = \text{Model} \quad (2.1)$$

$$\begin{aligned}
 (\text{state_1})Sim(\textcolor{red}{M1}, \textcolor{blue}{M2}) &= \text{mean}(MI(\textcolor{red}{M1_B1}, \textcolor{blue}{M2_B1}), MI(\textcolor{red}{M1_B2}, \textcolor{blue}{M2_B2}), \\
 &\quad MI(\textcolor{red}{M1_B3}, \textcolor{blue}{M2_B3}), MI(\textcolor{red}{M1_B4}, \textcolor{blue}{M2_B4})) \\
 (\text{state_2})Sim(\textcolor{red}{M1}, \textcolor{blue}{M2}) &= \text{mean}(MI(\textcolor{red}{M1_B1}, \textcolor{blue}{M2_B1}), MI(\textcolor{red}{M1_B2}, \textcolor{blue}{M2_B2}), \\
 &\quad MI(\textcolor{red}{M1_B3}, \textcolor{blue}{M2_B3}), MI(\textcolor{red}{M1_B4}, \textcolor{blue}{M2_B4}), \\
 &\quad MI(\textcolor{red}{M1_B5}, \textcolor{blue}{M2_B5}), MI(\textcolor{red}{M1_B6}, \textcolor{blue}{M2_B6}))
 \end{aligned}$$

$$\text{TotalSim} = \text{mean}(Sim(\textcolor{red}{M1}, \textcolor{blue}{M2})_{\text{state_1}}, Sim(\textcolor{red}{M1}, \textcolor{blue}{M2})_{\text{state_2}})$$

Clearly, the similarity metric defined above can be used for any geological property (R3).

This can be specified interactively from the application interface based on domain expert knowledge. There are also scenarios that users need to consider multiple properties. In this scenario, (dis)similarity values are calculated separately for each property (using the proposed approach) initially. After then, all those values are averaged to determine the final (dis)similarity between two models.

2.6.1 Distance Specification

The next step in the similarity calculation process is to determine the distance between a pair of corresponding 3D blocks. Similarity-based approaches are widely used in different problems of science and engineering. A number of distance-based formulations have been proposed [Goshtasby, 2012], where Hausdorff and Euclidean distances are the most commonly used in the reservoir engineering domain. The latter one is being found effective more for images and 2D surfaces such as time lapse seismic maps [Huttenlocher et al., 1993], and Euclidean distance mostly takes care of linear correlations. Therefore, we propose to use Mutual Information (MI) as a relatively multi-purpose measure. MI is a popular information-theoretic measure of similarity which has been applied in many areas of visualization and graphics domain like image registration, multi-modality fusion and viewpoint selection [Bruckner and Möller, 2010] [Haidacher et al., 2008]. The major benefits of MI for our case are:

- 1) Applicability. It is applicable as long as the domain has a probabilistic model. This aspect allows the measure to be used in the domains where no similarity measure has previously been proposed (e.g., reservoir engineering domain).

- 2) Non-linear dependency detection. MI considers all types of dependencies (i.e., linear and non-linear) between two objects [Cover and Thomas, 2012]. The relationship between the property values in a pair of geological models could be non-linear, and MI considers all these types of dependencies.

- 3) Noise detection. Many studies show that MI is robust to alleviate the impact of noise than the other distances [Cole-Rhodes et al., 2003]. Reservoir modeling procedures can create outliers

in the simulated spatial structures. Therefore, it is critical not to be sensitive to the noise data.

To further observe the effectiveness of MI over other common distances in the domain, we create a simplified dataset as represented in Figure 2.4. This dataset can mimic some simplified channelized reservoir models. The first column (Figures 1, 4, 7) shows the original models, some noise is added to the models in the second column (Figures 2, 5, 8), and the models in the third column (Figures 3, 6, 9) are 90 degrees rotated. A reasonable distance should be less sensitive to noise, in a way that the distance between an original model and its noisy version should be minimal. On the other hand, the distance between an original model and its rotated version should be considerable, because they are indeed two different models. Figure 2.5 shows the distance calculation for three metrics: MI, Euclidean, and Hausdorff distances. The results show that Euclidean distance is not sensitive to rotation, and the Euclidean distance between a model (1) and its rotated version (3) is zero, and they are collocated in Figure 2.5. The similar pattern can be seen for the other pair of models ((4,6) and (7,9)). Although rotation is detected by Hausdorff distance and the rotated models are located far from the original models (see (1,4,7) vs (3,6,9) in Figure 2.5), noise is not detected by this distance and noisy models are considered as very different models and located far from the original models ((see (1,4,7) vs (2,5,8) in Figure 2.5). Finally, it can be seen how MI distance detects movements like rotation and also ignores the noise. For instance, models with their noisy version are located close to each other such as (4,5), (1,2), (7,8) in Figure 2.5. Moreover, on the other hand, movements like rotation are also captured perfectly (see how (6,3,9) in Figure 2.5 are located very far from the original models). These benefits of MI can highlight its effectiveness for many reservoir simulation studies. This is because, on one hand, even tiny movements can translate to a significant effect on the simulation result, and on the other hand, modeling errors can lead to creating some noisy structures in the geological models.

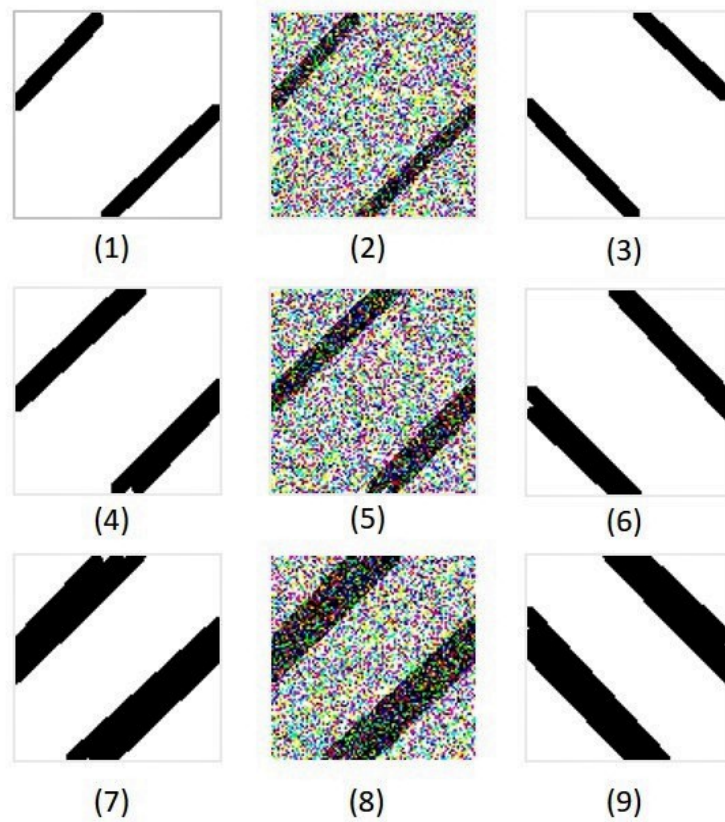


Figure 2.4: Dataset generated for evaluation of similarity distances.

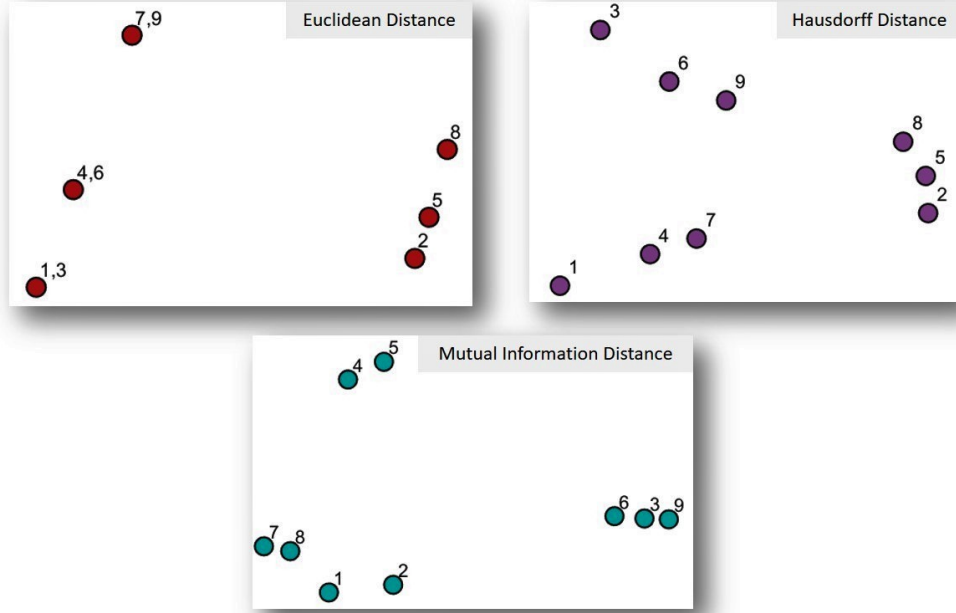


Figure 2.5: Comparison of different distance calculation methods including Euclidean, Hausdorff and Mutual Information.

2.6.2 Block Size Specification

In this research, we advocate a block wise approach for calculating the distance between the models. Such a block-wise strategy is widely used in image and video processing to exploit spatial and/or temporal locality and coherence. A critical aspect of this approach is to select a proper block size. As suggested by many researchers [Wang et al., 2008], block size should not be too small to sacrifice performance. Moreover, it should not be too large to ignore the locality and coherence feature of blocks. We use the concept of entropy, as suggested in [Honarkhah and Caers, 2010], to base the optimal block size. This optimal value is provided as a suggestion to the users in our designed application; however, users can change it based on their knowledge such as the use of correlation length as it is used in generating some geological patterns. Entropy measures the information contained in a message as opposed to the portion of the message that is determined. This concept, when applied to blocks in reservoir models, can determine the minimum information required to represent the whole model reliably. Hence, our method for optimal block selection is to

scan a reservoir model with different block sizes using our proposed moving block method. For each block size, we calculate the average (mean) entropy values of all blocks. Then, the entropy values are plotted for each block size. We did some numerical experiments for this algorithm including using different reservoir models, blocks with different sizes in each dimension, and blocks with equal sizes in all dimensions. Our empirical studies show the following essential trends in the specification of block size:

- In the first stages of increasing block size, the entropy sharply increases since the average number of information bits needed to encode the underlying patterns in the model is increasing.
- At a later stage where the block size has increased above the optimal block size, entropy increases at a much slower pace.
- In the stages that block size is close to the size of the original model, entropy stops increasing. The reason is that block contains repeated patterns and hence, the amount of carried information ceases to increase.

Therefore, according to these trends, the optimal size of the block is in the stage that entropy slowly increases as highlighted in Figure 2.6.

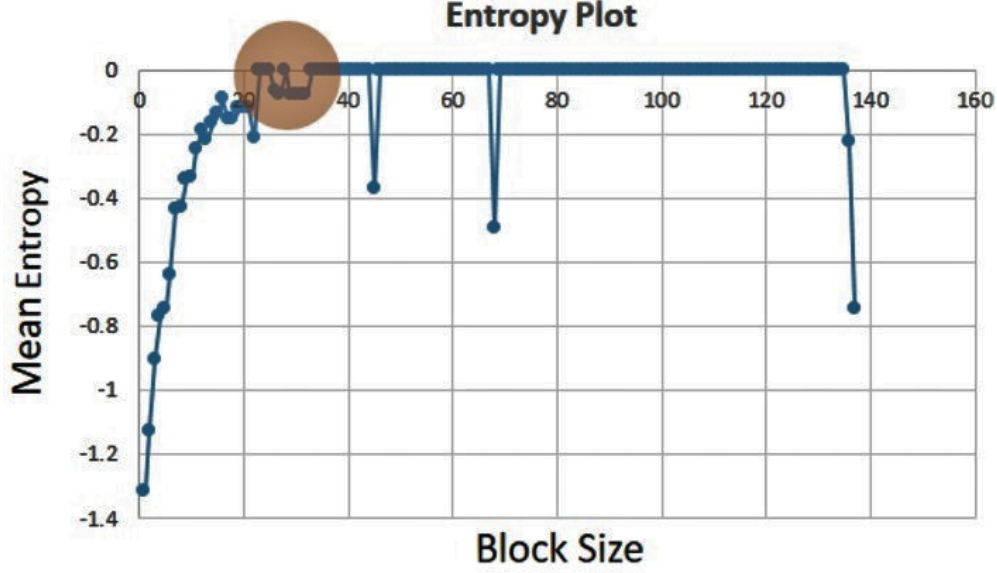


Figure 2.6: Mean entropy plot for different block sizes, with highlighted maximum entropy.

Figure 2.6 shows the entropy plot for increasing the block size (the scanning template) in the x direction. The size of the original model in the x dimension is 140, and from Figure 2.6, it can be seen that a block size with an x dimension around 25 to 30 (the highlighted area) is an optimal value. This is where the average entropy curve reaches its maximum for the first time. We can perform a similar procedure for the y and z dimensions, and get the optimal values for the other dimensions as well. In our case, for a model with size $140 \times 69 \times 9$, a block with size $23 \times 12 \times 2$ was found to be the optimal size.

2.7 Projection with Clustering

The calculated distances are utilized within a clustering algorithm to group similar models. Each cluster center is a default representative member of the containing cluster, which leads to our main requirement: to reduce the number of models needed for simulation (R1). The K-Means clustering (KMC) algorithm [Correa et al., 2009] is employed in this step because of its computational efficiency on large datasets. However, KMC suffers from a noticeable drawback. In the case where the data embeds a complex structure (e.g., data are non-linearly separable), a direct

application of KMC is not suitable because of its tendency to split data into globe-shaped clusters [MacKay, 2003]. To solve this problem, as suggested in [Shawe-Taylor and Cristianini, 2004], data will be mapped by a kernel transformation [Schölkopf et al., 1998] to a new space where samples become linearly separable. Although there are many available kernel functions in this study, we use radial basis function (RBF). To make the RBF kernel more general - that is not to be only the function of the Euclidean distance but also any other distances - the kernel is combined with multidimensional scaling (MDS) [France and Carroll, 2011] [Scheidt and Caers, 2010]. MDS is a classical approach that projects the original high dimensional space to a lower dimensional space, which can preserve the original distances. In the projected space, the spatial position is not critical; the crucial aspect is the distance between projected points. The closer points are to each other, the more similar they are based on the initially defined distance. The projection algorithm is summarized as follows:

1. Use MI to calculate the block-based distance $d(x_i, x_j)$ between each pair of models.
2. Use MDS to plot these locations in a low dimension, call these locations $x_{d,i}$ and $x_{d,j}$ with d the dimension in the MDS plot.
3. Calculate the Euclidean distance between $x_{d,i}$ and $x_{d,j}$.
4. Calculate the kernel function with given σ (Equation 2.2).

$$K_{ij} = K(x_i, x_j) = \exp\left(\frac{(x_{d,i} - x_{d,j})^T (x_{d,i} - x_{d,j})}{2\sigma^2}\right) \quad (2.2)$$

Other than having better and simpler visualization of projected models using kernel transformation, it has also a great benefit for clustering algorithm [Scholkopf and Smola, 2001]. K-means clustering works well for cases such as Figure 2.7.b, but goes wrong in complex cases such as Figure 2.7.a, where the variation of objects/points in the 2D plot is nonlinear. Therefore, it is frequently helpful to first transform the points using a kernel transformation, as shown in Fig-

ure 2.7.b, and then perform k-means clustering. This technique is called kernel k-means in the literature [Williams, 2002] [Dhillon et al., 2004].

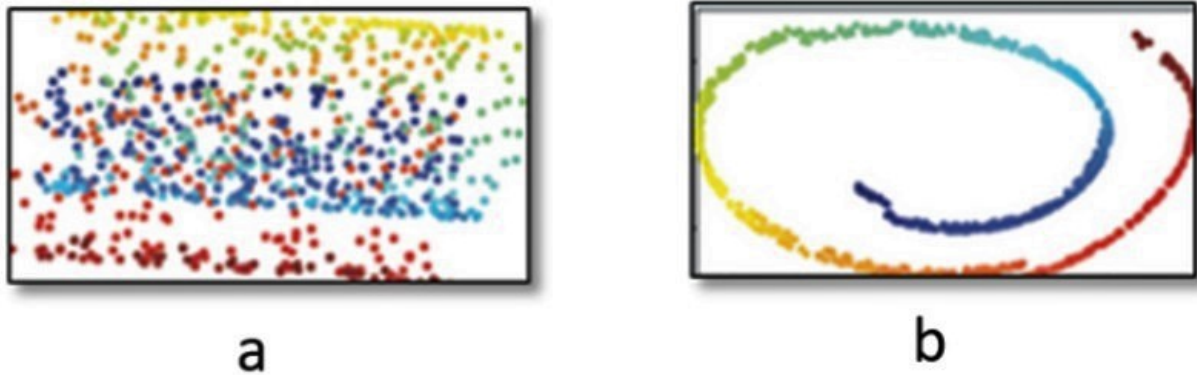


Figure 2.7: (a) represents projection with MDS, (b) represents projection with MDS using kernel methods. [Zhang et al., 2010]

The efficiency of kernel KMC in clustering the geological models is shown in Figure 2.8. It shows how the representation of data and clustering is different in two scenarios: projection with and without kernel transformation. It can be seen that the representation of data looks better and more importantly clustering results are more representative when a kernel transformation has been applied. Without kernel transformation, projected points are very close to each other, and that makes separation of clusters complicated. However, with kernel transformation, a well organized and linear structure can be seen in the results, and clusters are better represented and separated.

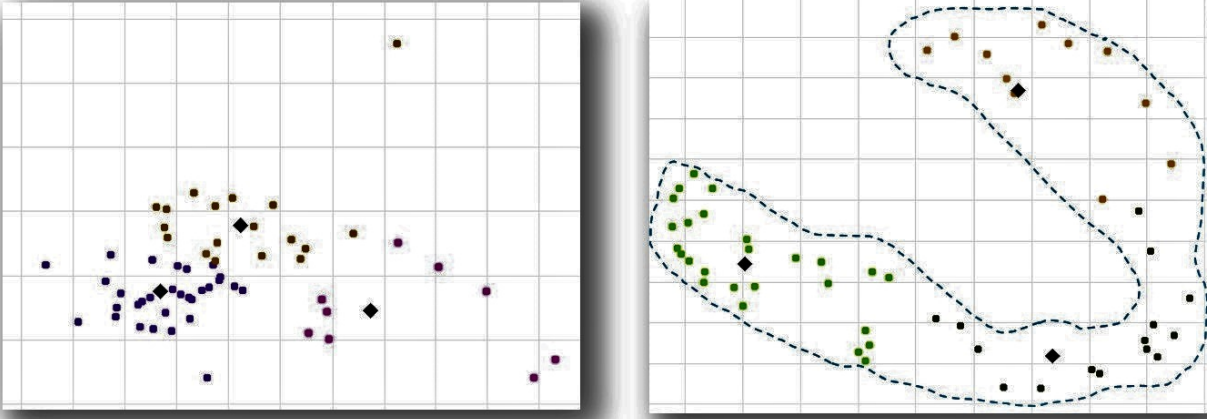


Figure 2.8: Difference between multidimensional scaling with (right) and without (left) kernel transformation on the case study dataset.

2.8 Evaluation

For evaluation of our proposed analytical framework, we compared our results with the current alternative process in the industry (i.e., to run flow simulation for all models individually). We run the complete flow simulation for all the models using the CMG reservoir simulator package, and plot the simulation results for the 'oil recovery factor' dynamic property (Figure 2.10). The plotted results show a range of uncertainty on the oil production and we expect that our cluster centers cover this range adequately. Regarding the datasets, our industry partner generated different datasets for us using different geostatistical algorithms and scenarios. The idea was to cover almost all different types of datasets in the domain. To evaluate the performance of our proposed analytical framework, they provided a various dataset with a different number of Cartesian models (15 to 100 models) and sizes (1000 to 100,000 cells). Therefore, we evaluated our process in all different scenarios. In the first simple scenario, 15 models were created in 5 groups, and they only changed 'facies' property in each group (Figure 2.9). Facies is an important geological property that reflects the rock type depositions. When we run the flow simulation for all the 15 models, and plot a

dynamic property (like oil recovery factor vs time), a range of uncertainty can be seen in the plotted curves (high, medium and low recovery factor). To capture this range of uncertainty with fewer of models, our proposed filtering framework is used to cluster models into three clusters. Cluster centers are shown with a star in Figure 2.11. The results show how cluster centers can represent the range of uncertainty. Highlighted curves in Figure 2.10 shows the cluster centers. In addition to that, our approach is much faster than the traditional brute-force approach. Depending on the complexity and size of the reservoir model, the execution time of a flow simulation could be different. In this case study, running a complete flow simulation takes around 5 minutes per model, that resulting in 75 minutes (an hour and a half) for all the models in total. However, our approach takes around only 3 minutes to calculate the distance between the models and generates the clustering result. Therefore, it can be seen that our approach is very time efficient in comparison to the existent techniques.



Figure 2.9: Different types of facies property that used for the creation of geological models.

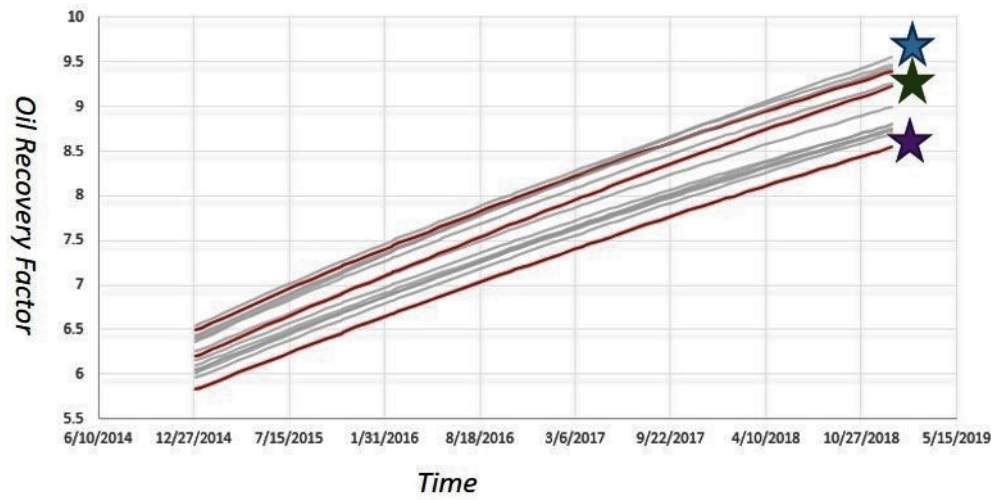


Figure 2.10: Simulation results of 15 geological models for Oil Recovery Factor property.

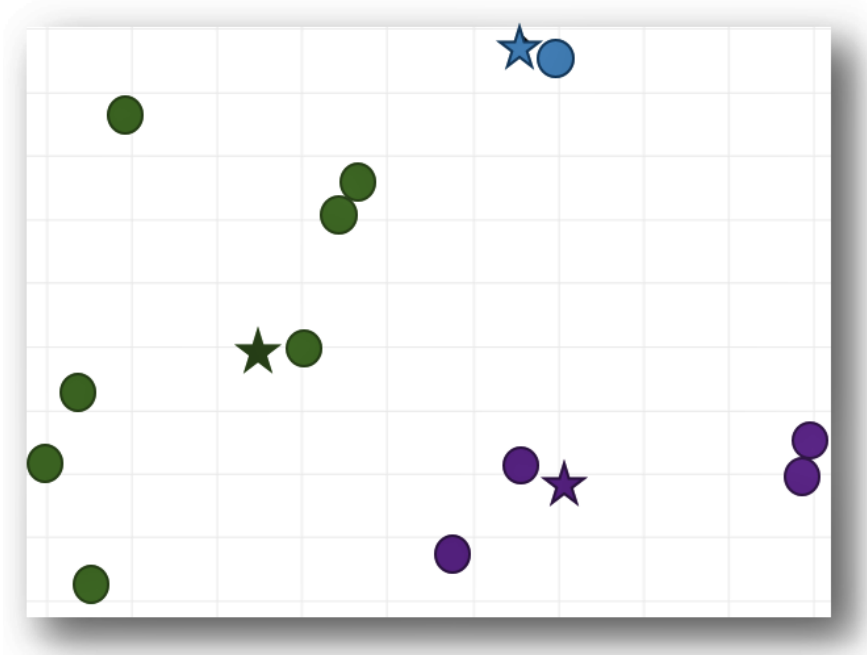


Figure 2.11: Clustering result for 15 geological models.

In another scenario, domain experts changed the value of all the properties (porosity, permeability, saturation, etc) and they created 100 models. The idea is to see how our proposed analytical framework performs for such scenarios. Figure 2.12 shows the flow simulation curves for all the

100 models. The range of uncertainty is much broader than the previous scenario. Our clustering result shows how this range of uncertainty can be represented by only six models (see the cluster center stars in Figure 2.13 and their corresponding curves in Figure 2.12). Similar to the previous case study, our approach had a very significant performance in comparison to the current brute force approach. The reason is that the flow simulations for all the 100 models took around 7 hours, while our approach generates the clustering results in only 45 minutes.

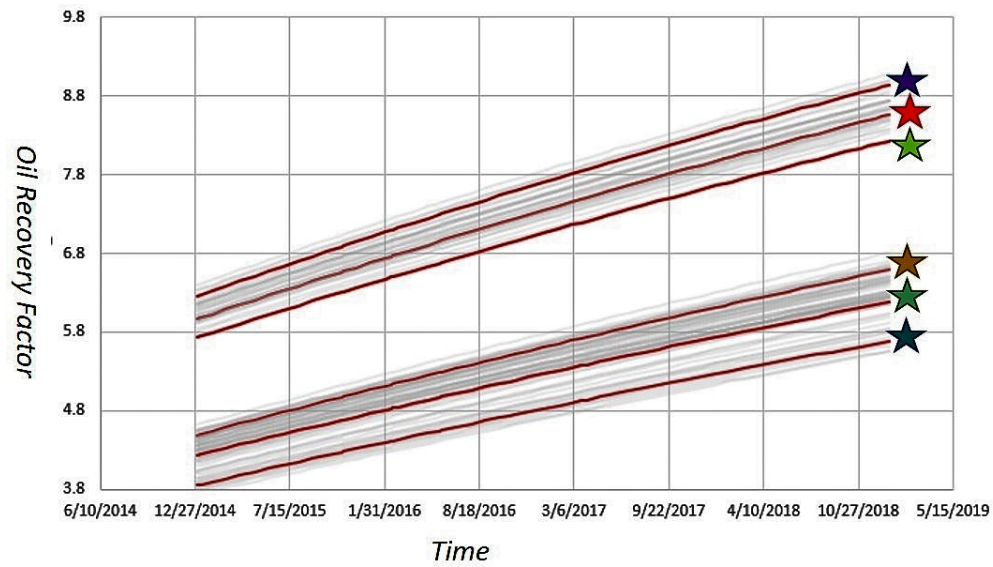


Figure 2.12: Simulation results of 100 geological models for Oil Recovery Factor.

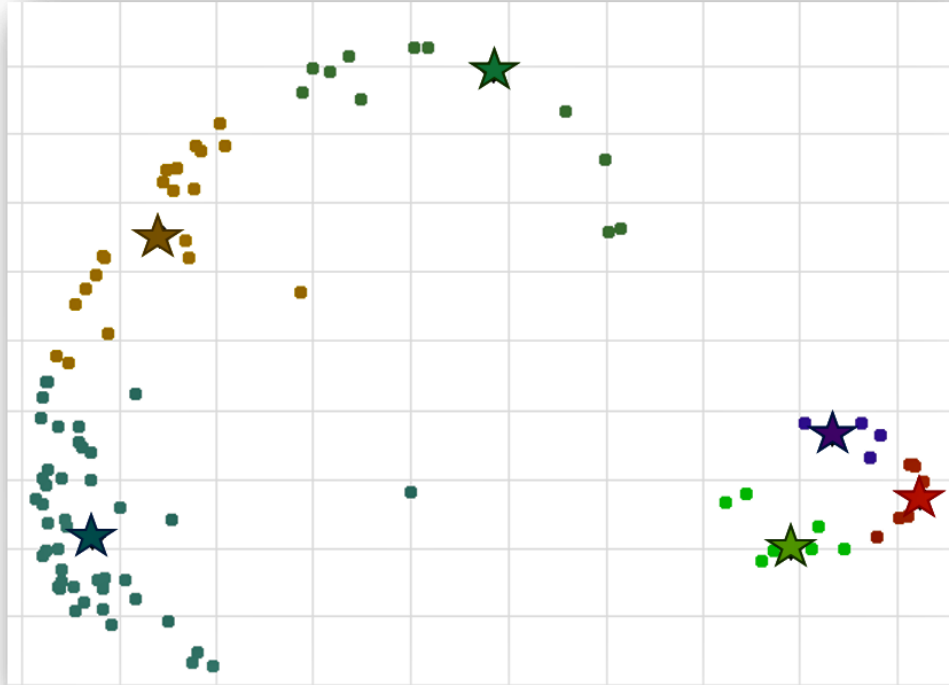


Figure 2.13: Clustering result for 100 geological models.

2.9 Visual Analytics Application

This selection process has been designed and developed in a visual analytics framework (Figure 2.14). It helps the users perform the selection process with a set of user-defined parameters and compare the models at different levels of details and views. The users can import any number of the models into the application. Each model can be visualized in 3D. The color scale shows the value of a selected property. Warm colors show the higher value of the selected property and in reverse for the cool colors. There are two main visual analytical processes in this prototype: selection process and comparison analysis.

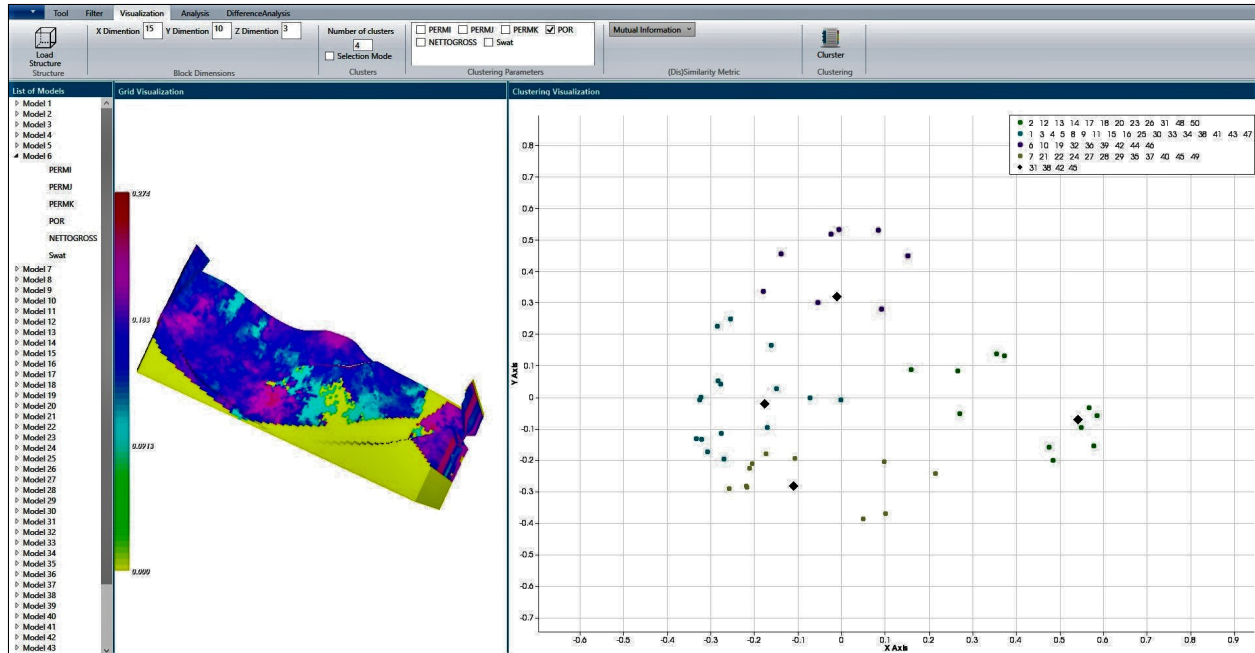


Figure 2.14: Projection and clustering of loaded modelsscal.

2.9.1 Selection Process

The selection process consists of two main steps: calculation of (dis)similarity and clustering. Two main parameters are specified by the users: block size and reservoir property(ies). The default optimal block size is calculated in the background (using the entropy-based approach mention in section 5.2) and is provided in the interface as a suggestion. However, users can also change that according to their knowledge of the reservoir. Block size is specified by three values for each 3D dimension: x, y, and z. Each of them can be changed by the users interactively. The other parameter that should be specified is the static reservoir property(ies) that are used for the distance calculation. Users can specify one or more number of properties (R3).

After that, the number of clusters should be specified by the user, that is determined based on user's budget and time for running flow simulation. The clustering outcome is presented on a diagram in the 2D view. Each point in the diagram corresponds to a projected geological model. The color legend on the 2D diagram shows the clustering results (Figure 2.14). In result, the

median member of each cluster is selected as the representative member of that cluster (R1). Since the calculated distances are mapped to the 2D view, it also helps the users to have an overall representation of models.

According to our discussions with domain experts, they need to perform selection process based on a specific region of a model. This is because they are sometimes dealing with very large reservoirs, and not the whole reservoir geometry is important to them. For instance, in a substantially large reservoir model, the engineers are usually interested in the areas around wells. To leverage this requirement, the users can freely sketch an arbitrary 3D area on the model. And then the selection process constrains the calculation of (dis)similarity only to that specific region (R4). (Figure 2.15)

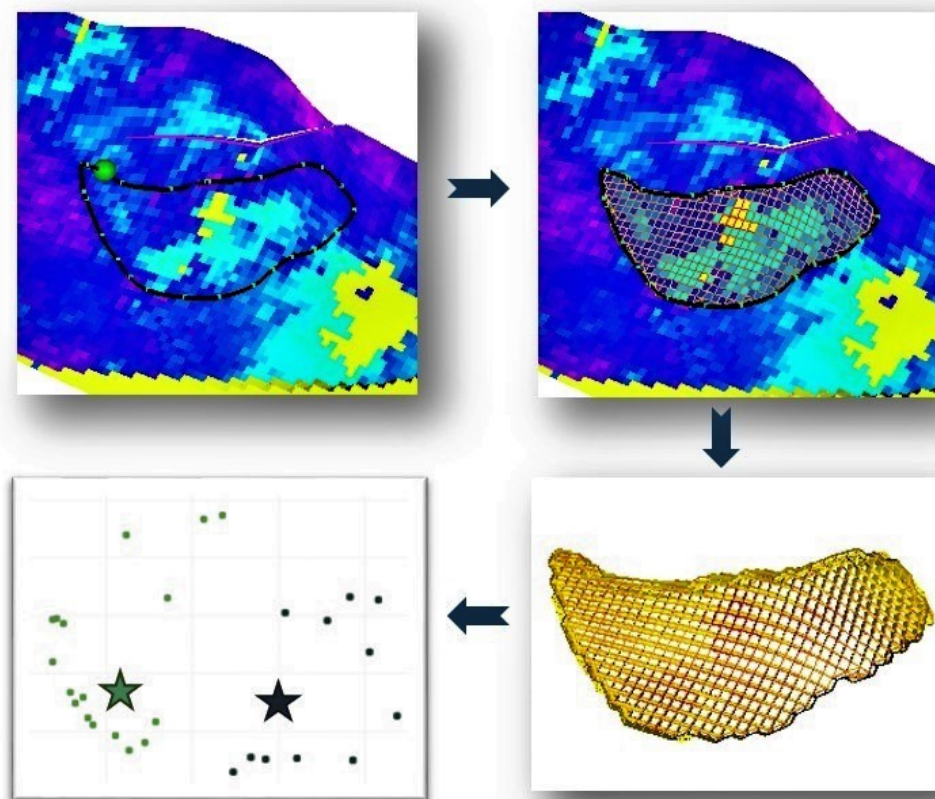


Figure 2.15: Filtering process for an arbitrary area of interest.

2.9.2 Comparison Analysis

In addition to the calculated distance values, the users also frequently need to get more detailed spatial information about the differences between models. For instance, the engineers might need to get some insights into the regions that the models show a significant difference. To provide this feature, users can select any number of models from the 2D view. A 3D similarity map is calculated and visualized for the specific selected models (R5). In the similarity map, the users can observe the local similarity between the models - i.e., which parts of the model contribute additional weights in the similarity and dissimilarity calculations. For instance, Figure 2.16 is a similarity map for four selected models. The color scale shows the amount of mutual information between all these four models. The results show that these models are very similar in the red and dark blue areas, and they are very different light green areas. This feature not only helps identify the important regions of models but also utilizes a useful feature of mutual information that can be calculated between multiple objects at the same time with multivariate mutual information techniques [Batina et al., 2011].

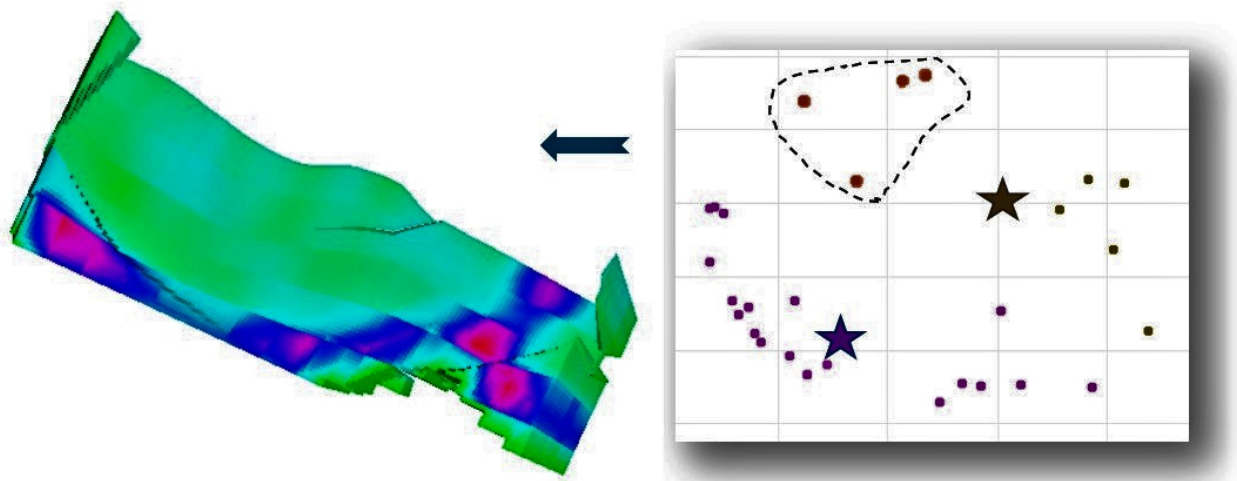


Figure 2.16: Similarity map for the selected models.

2.10 Conclusion and Future Works

In this paper, we introduced a new visual analytical framework for selecting a few representative models from an ensemble of geostatistical models that represents the overall production uncertainty. To achieve this purpose, a new block wise (dis)similarity metric was defined based on mutual information. This metric is projected to a lower dimension using the MDS technique, and then the new projected distance is used in a kernel KMC algorithm to group models based on their distances. The proposed workflow was evaluated using some datasets generated from various geostatistical algorithms. The results of the case studies show that our technique is accurate and efficient in comparison to the existent techniques. In the future, with the help of domain experts, we need to find more adequate parameters for uncertainty assessment of geological models. Moreover, regarding the application, we need to support comparing of clustering results, and in continuing that, provide more information to the users such as what is the best number of clusters, or what are the effective parameters. We will also further evaluate our application using a formal user study, that helps identify additional weakness and strengths of the current application and process.

2.11 Acknowledgment

We wish to thank the anonymous reviewers for their constructive comments, CMG (Computer Modelling Group Ltd.) for providing the reservoir data sets and Masoud Zehtabioskuie for his valuable help and feedback on the implementation of the application. This research was supported in part by NSERC.

Chapter 3

Clustering of Geological Models for Reservoir Simulation

Studies in a Visual Analytics Framework ¹

3.1 Introduction

Geological uncertainty is a common problem affecting the prediction of hydrocarbon production [Ballin et al., 1992]. Some of these uncertainties are reflected in the various rock types and their geometrics present in the subsurface of the earth which largely affect the future prediction [Caers, 2011]. To quantify such uncertainties, geostatistical methods are usually employed to generate multiple reservoir models honouring both soft and hard data (e.g. well-log, core and seismic). This ensemble of reservoir models represents an assessment of spatial uncertainty. The ensemble signifies a large number of equiprobable geological models. Naturally, the associated uncertainty in dynamic properties such as oil production can be assessed when a large number of models are generated and when the flow simulations were performed for all the models. This process is costly as a complex reservoir simulation can take days to complete. On the other hand, such a brute-force approach is not efficient as the simulation results of some sampled models can be very similar to each other. Therefore, it is critically important to carefully sample a few geostatistical models which can reasonably represent the overall uncertainty [Idrobo et al., 2000] [Fenik et al., 2009].

The goal of our research is to design and develop visual analytics techniques (Sun et al, 2013) to filter the geostatistical models and to only select the models that can potentially cover the uncertain space. The proposed technique can resolve the existing issues of previous studies. Current techniques like ranking (Deutsch et al; 1996), random selection, or probability-based techniques

¹Sahaf, Z., Hamdi, H., Maurer, F., Nghiem, L., and Sousa, M. C. (2016). Clustering of Geological Models for Reservoir Simulation Studies in a Visual Analytics Framework. In 78th EAGE Conference and Exhibition 2016.

[Li and Floudas, 2014], are costly in terms of computation. They are automatic processes preventing domain experts from guiding the selection process. Moreover, they rely mostly on either simplified fluid flow or unscaled geological properties that may not give correct results. That is because fluid phases might not follow the same flow paths in simplified models as they would in realistic models [Yazdi and Jensen, 2014].

The main contribution of this study is to introduce a novel visual analytics framework for reducing the computational cost in uncertainty assessment of dynamic properties by only utilizing the underlying static model properties while allowing the domain experts to improve the selection process interactively. Visual analytics methods allow decision makers to combine their human flexibility, creativity, and background knowledge with the enormous storage and processing capacities of today's computers to gain insight into complex problems. In our approach, the users can specify the reduction level interactively to filter out some models from an initial population of geo-statistical realizations. For instance, the users may want to select 10 or 20 models, depending on the scenario or requirement, among 100 available models. Moreover, the visualization techniques have been designed to represent a set of realizations, and to even interactively select a desired spatial 3D region (e.g. a region near a well) as an area of interest to perform the filtering process accordingly.

We designed a novel distance measure based on the mutual information (MI) [Goshtasby, 2012] [Lin, 1998] concept. Distances are computed between all the models and are then used by a multi-attribute clustering algorithm to create sets of similar models i.e. models whose simulation results will be potentially similar. We then pick one model from each cluster randomly for flow simulation. We show the accuracy of our selection method by comparing the simulation results of the selected models with the simulation results of all models in the initial population: a selection is accurate when the selected model simulation results are similar to the simulation results of all other models in the cluster.

3.2 Similarity Metric Calculation

One way to measure the similarity between a pair of 3D models is to represent the models in terms of two lengthy 1-Dimensional (1D) vectors and use a similarity metric (e.g. Euclidean distance) to compute the similarity between any two models. However, this approach is not entirely accurate for addressing this problem as geological models have 3D geometries with correlated properties. Additionally, the 3D models can have some favourable 3D sub-structures (e.g. channels) that any appropriate similarity measurement should be able to address their spatial resemblance (Figure 3.1). Therefore, we propose to use a moving 3D template to calculate the expected similarity between a pair of geostatistical realizations. The idea in this proposed method is to divide each 3D model into smaller 3D blocks (Figure 3.2) where each block consists of a specific number of cells. To calculate similarity between a pair of models, similarity values are computed between corresponding contained blocks (templates) initially. Next, we take an average of similarity values between all corresponding blocks of two models. During this process, in order to reduce the bias of the fixed spatial location of blocks, we move the blocks in certain directions (x, y, z and diagonal) and distances (> 1 and $< \text{block size}$). The final similarity value would be the average of similarity values in all the possible movements. For simplicity, a 2D representation of a movement is shown in Figure 3.3. The yellow highlighted cells represent an important feature. The figure shows that how the movement of templates can help better capture the similarity, and how the final similarity is calculated between two sample models with one movement and two states.

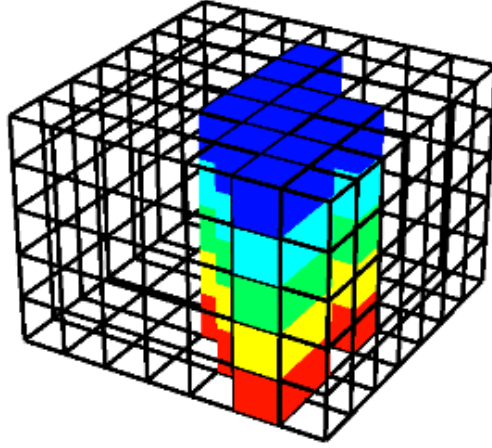
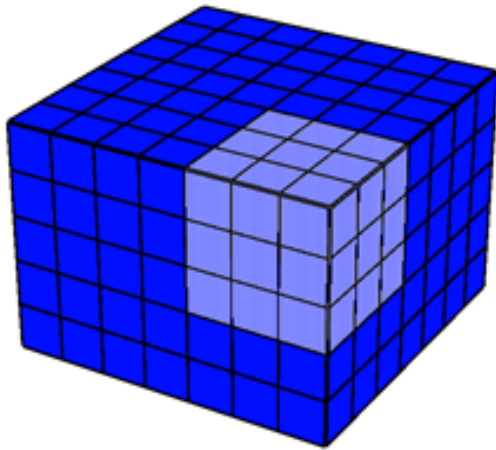
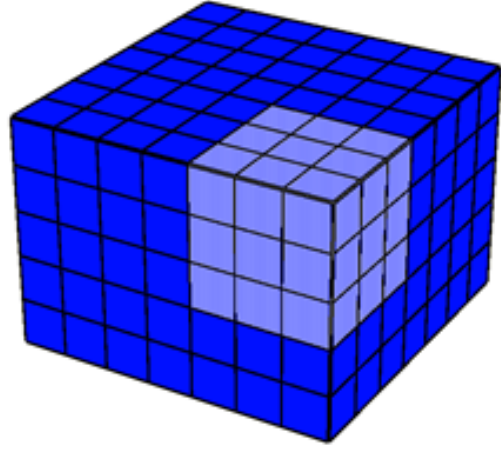


Figure 3.1: A sample important 3D structure in the geological models.



Model 1



Model 2

Figure 3.2: Representation of block based similarity calculation method for a sample block size $3 \times 3 \times 3$. This Figure represents corresponding blocks in two models.

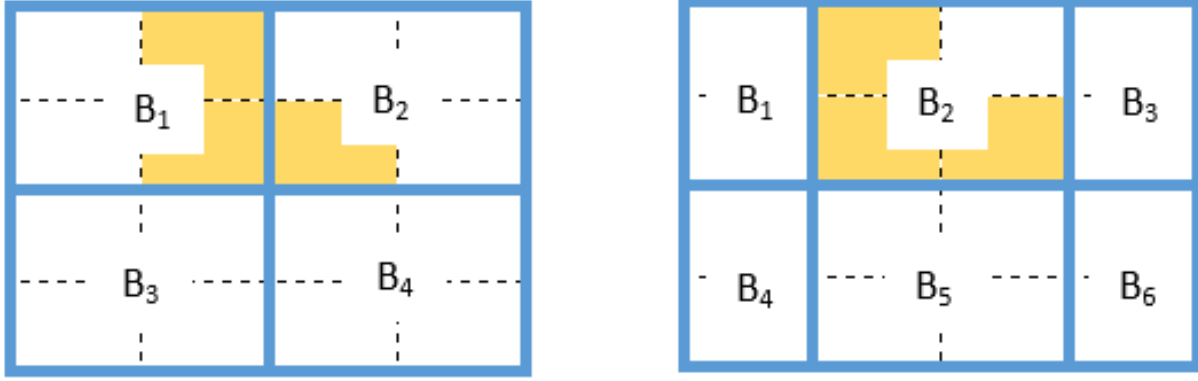


Figure 3.3: Representation of a sample important structure (highlighted in yellow), and see how similarity between important structures can be captured in a better way by movement of templates in possible directions. (B stand for Block).

The next step in the similarity calculation process is to determine the distance between a pair of corresponding 3D blocks. Noticing the main feature of geological models, each cell consists of several geological properties such as porosity or permeability. The relationship between the property values in a pair of models could be non-linear. Therefore, we use the information theory concepts to calculate the similarity based on the amount of shared information (i.e. mutual information) between a pair of blocks. Mutual information (MI) considers all types of dependencies (i.e. linear and nonlinear) between two objects. The formal definition of MI between two random variables X and Y , whose joint probability distribution is defined by $P(X,Y)$ is given by Equation 3.1. In this definition, $P(X)$ and $P(Y)$ are the marginal probability distributions of X and Y .

$$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) \frac{P(x,y)}{P(x)P(y)} \quad (3.1)$$

X and Y are the values of that property in a pair of models. MI can also be calculated between more than two variables. Our proposed algorithm is able to include multiple properties simultaneously.

3.3 Clustering

The similarity values are utilized within a clustering algorithm in order to group similar realizations. A representative realization is then randomly selected from each group, which help reduce the number of realizations that need to be simulated. The K-Means clustering (KMC) algorithm [Correa et al., 2009] is employed in this step because of its computationally efficiency on large data sets. However, KMC works fine with linear based distances while our proposed distance is based on MI which is non-linear. Therefore, the proposed MI-based similarity values are used to map all realizations on an Euclidean space using multidimensional scaling (MDS) method [France and Carroll, 2011] [Scheidt et al., 2009]. MDS is a classical approach that projects the original high dimensional space to a lower dimensional space, which can preserve the original distances. Since in most cases the structure of points in mapping space is not linear, we use some kernel methods [Scholkopf and Smola, 2001] [Smola and Schölkopf, 1998] to transform the Euclidean space into a new space, called the feature space. The goal of the kernel transform is that all points in this new space can behave more linearly, and therefore the KMC can provide more reasonable outcome. The visual and interactive representation of the original and the feature spaces, and their relationships are implemented in our proposed visual analytics framework.

3.4 Case Study

The proposed method has been applied to three case studies. In the first case study, three different correlation ranges are used to generate three facies distributions. For each case, five realizations have been generated by only changing the random seed in the Sequential Indicator Simulation algorithm (Deutsch, 2006); leading to having 15 realizations. In the other dataset, instead of changing the facies, the permeability and porosity distributions have been varied with a similar approach in the previous dataset; that yields another 15 realizations. Finally, in the third data set, all the permeability, porosity and facies distributions are changed, which results in a larger number of realizations (100). The latter dataset helps judge how our method works when a large number

of models is available. In all the three cases, the models dimension is $139 \times 48 \times 9$, and the available properties for all the models include porosity, permeability in all directions, net to gross, and water saturation.

3.5 Results

Due to the lack of space, we only show the results for the first case study. To validate our approach, we performed the flow simulations for all the models using CMG black oil simulator [CMG, 1980]. Figure 4 shows the plotted simulation results. The figure reveals that the simulation results for some models are very close to each other, while some others are completely different. This helps identify the expected clusters and use them during the evaluation process. From the simulation results, it is desired to select three representative realizations for the upper, middle, and the lower bounds (Figure 3.5). MDS is used to convert the MI similarity values to the Euclidean space. It can be seen from Figure 3.4 that there is a positive correlation between the original distance and the projected distance, which means that the distance between the original models (higher dimension) has been preserved in the Euclidean 2D space (lower dimension). After this step, KMC is applied to the projected domain. The accuracy of the results shown in Figure 4.11 is around 70%, which means that around 70% of models are clustered in the right group, and 30% are not. In terms of computational efficiency, running the simulation for all 15 models took one hour (4 to 5 minutes each); however, in the proposed application it took just 3 minutes to represent the clustering results, and representative models. Similar results have been derived for the two other datasets.

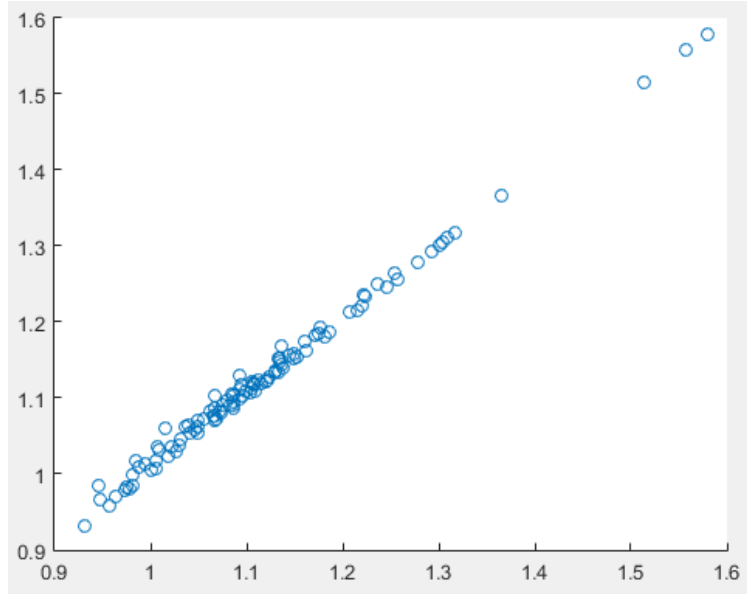


Figure 3.4: Correlation between original mutual information based distance and the projected distance using MDS technique.

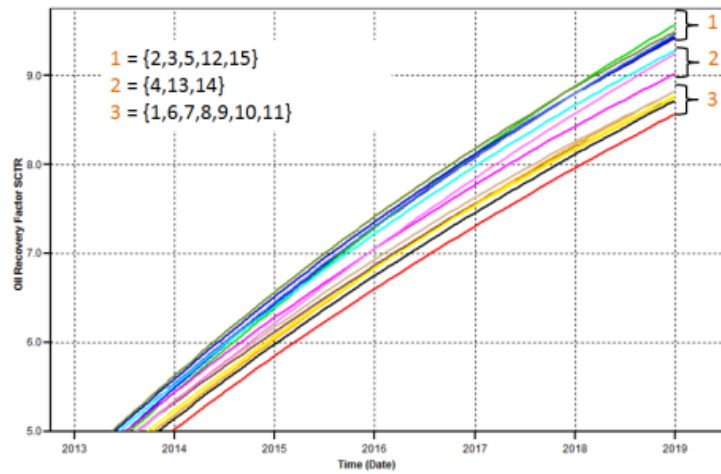


Figure 3.5: Simulation results for 15 realizations, along with expected clustering results.

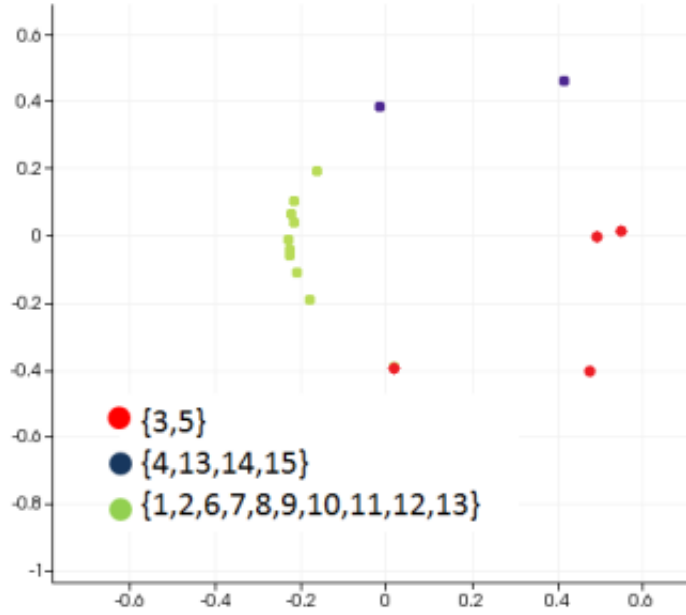


Figure 3.6: Actual results of clustering on 15 realizations, projected in 2D space using MDS.

3.6 Visual Analytics Framework

The detail of whole proposed algorithm has been implemented in a novel and unique visual analytics software framework that has three main elements including visualization, analytics, and interaction. The users can interactively specify the “number of clusters”, “block dimension”, and the “clustering properties” in this application. After clustering, a 2D scatterplot is represented where any point on the plot corresponds to a 3D geostatistical model. In this diagram, the closer points imply more similar models. To have a better understanding of spatial visual similarity between the models, the users can select the desired models by dragging and dropping the models on the 2D diagram, and observe the similarity map in the other 3D view. In the similarity map, the user can realize which parts of the models’ structures have higher contributions into the similarity calculations (Figure 3.7). Additionally, the framework has the ability to select a specific sector of the models for additional study. In Figure 3.8, it can be seen that clustering and projection results are calculated merely for the white highlighted area.

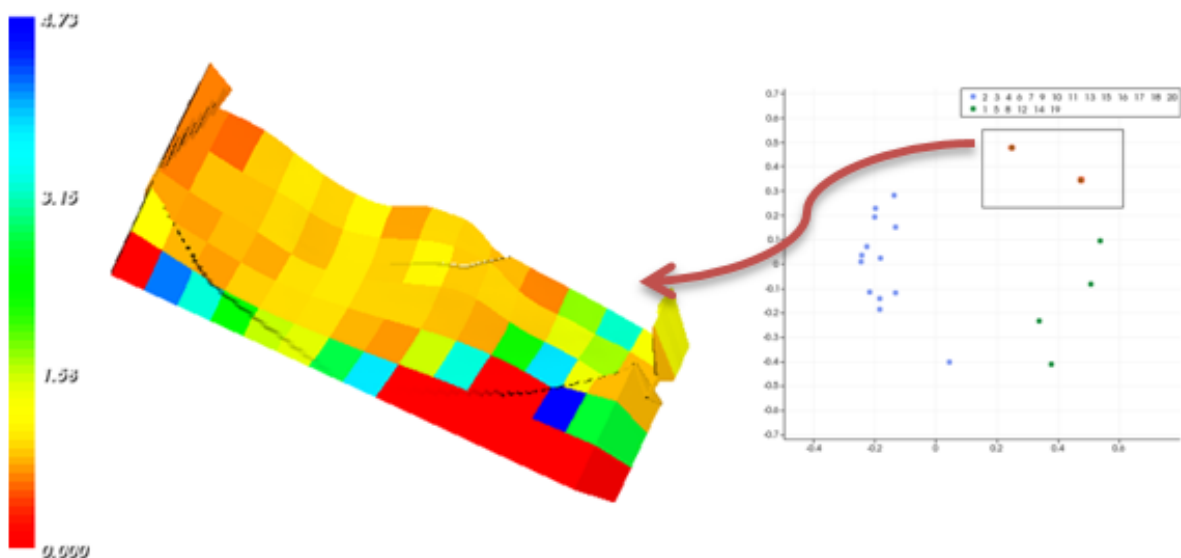


Figure 3.7: Visualization of spatial contribution of realizations into similarity.

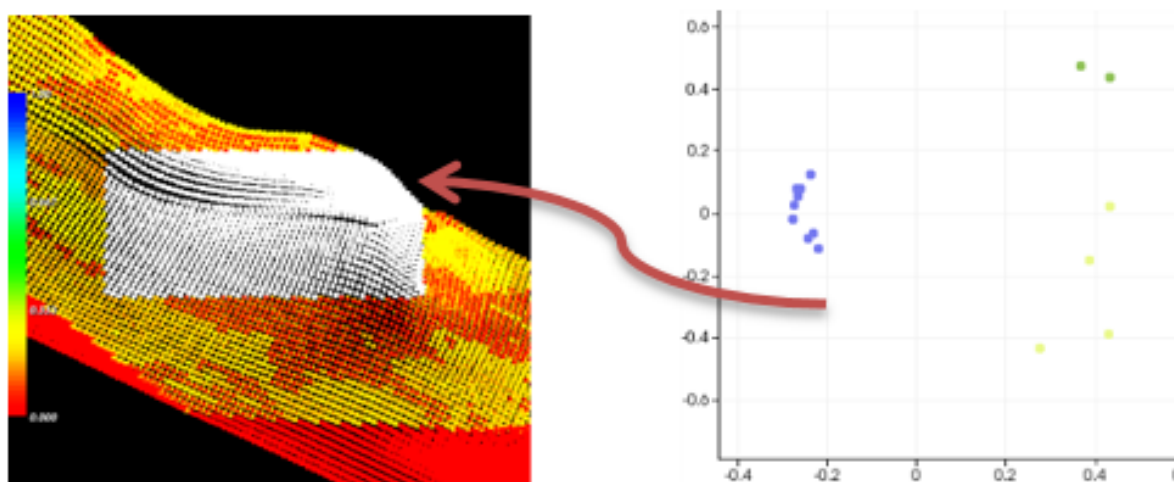


Figure 3.8: Calculating clustering based on a specific area of models which is selected interactively by users.

3.7 Conclusion

In this paper, we introduced a new visual analytics framework to select a few representative models from an ensemble of geostatistical models that can represent the overall production uncertainty. To achieve this purpose, a new similarity metric is defined based on mutual information. This metric is projected to a lower dimension using MDS technique and then the new projected distance is used in a KMC algorithm to group models based on their similarity. The whole process is implemented in a visual and interactive framework. The proposed workflow was exemplified using some datasets generated from various geostatistical facies realizations using different variogram correlation lengths. The results on the case studies show that our technique is 70% accurate and much more time efficient in comparison to the existent techniques. The method is being enhanced for more accurate clustering.

Chapter 4

Filtering Geological Realizations for SAGD (Extended Version)

1

4.1 Abstract

Steam Assisted Gravity Drainage (SAGD) process is one of the most viable thermal recovery methods for heavy oil reservoir. In order to assess the impact of geological uncertainty in such reservoirs, a large number of equally-probable models, called realizations, are generated. Frequently, each realization is used with a flow simulation process to obtain the degree of uncertainty in reservoir production. This brute force approach is a cumbersome process, since many of the realizations may exhibit nearly equal flow performance. Additionally, the simulation process can be very time consuming for various scenarios such as thermal models or the models with large number of grid cells. In this paper, we introduce a visual analytics framework for filtering the realizations for SAGD and select the ones that potentially represent distinct flow performances. This framework is based on calculating the (dis)similarity distances between all pairs of the realizations and then clustering the models using the computed distances. Each cluster contains an ensemble of models that have potential similar flow performance. The results for a case studied in this paper, show that our technique is very efficient in identifying the clustered models and can closely match the brute force approach with minimal computational time.

¹Sahaf, Z., Hamdi, H., Maurer, F., Nghiem, L., Chen, Z., and Sousa, M. C. (2017). Filtering Geological Realizations for SAGD. Submitted to First Break Journal <http://fb.eage.org/>

4.2 Introduction

A substantial proportion of the worldwide hydrocarbon resources includes heavy oil and bitumen resources [Birol et al., 2010]. Heavy oil reservoir fluids have high specific gravity and viscosity compared to the conventional oil reservoirs [Meyer et al., 2007]. The successful recovery from these resources is based upon developing a mechanism that displaces the heavy oil in the reservoir [Ancheyta and Speight, 2007]. Usually, special enhanced oil recovery techniques such as thermal methods are introduced to augment the hydrocarbon production by reducing the viscosity. The majority of thermal recovery technologies are based on steam injection processes through vertical or horizontal well [Le Ravalec et al., 2009] such as Steam Assisted Gravity Drainage (SAGD) which was first introduced by Butler in 1991 [Butler, 1991]. The SAGD implementation involves drilling two parallel horizontal wells at the bottom of a usually unconsolidated sandstone reservoir that includes a production well and a steam injection well located 5 to 10 m above the producer (Figure 4.1). At the initial stages of the SAGD process, both injector and producer are circulated with and heated by high temperature steam until the formation in the region around the well-pair warms up. As a result, the viscosity of oil reduces and the hydraulic communication between the injector and producer is established. After this initial heating period, the lower well at the base of the reservoir is placed on production while the steam injection continues in the upper well to replace the drained oil in the reservoir. In this manner, the steam plume expands around the injector and mobilizes the impacted oil. After condensation, the cooled steam and the mobilized oil are both moves toward the production well.

Butler [Butler, 1991] derived a formula for estimating the SAGD recovery in a homogeneous reservoir based on the moving boundary concept. In his approach, the region that is enclosed by the growing interface between the steam and the intact heavy oil (or bitumen), is approximated by a reverse cone. This region, which is identified at any time by a certain temperature cutoff, is called the steam chamber. Despite the proximity of wells, the SAGD mechanism allows the steam chamber to form and expand gradually where the mobilized oil will be drained from a larger region. In

order to have maximum steam contact with the reservoir, it is important to have a good vertical communication within the reservoir components. For instance, the presence of shale and other heterogeneities can drastically affect the SAGD performance and feasibility [Yazdi and Jensen, 2014]. Shales usually baffle the steam rise and prevent the formation of a uniform steam chamber, hindering the downward oil drainage into the production well [Crain, 2013].

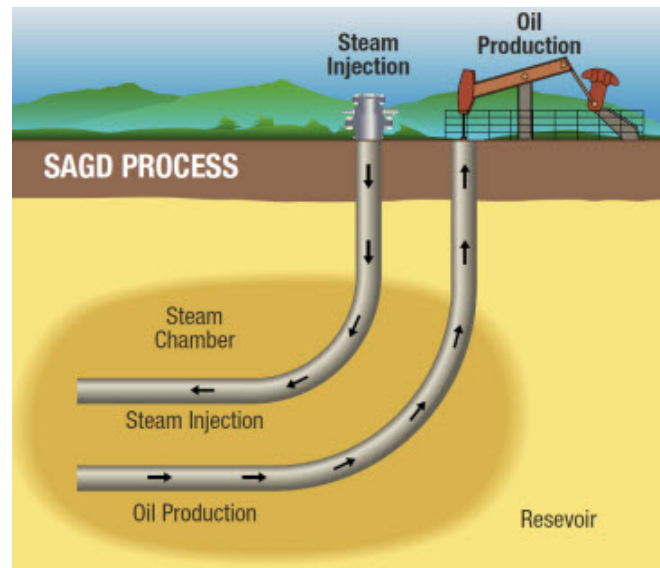


Figure 4.1: Workflow of SAGD process [AlbertaEnergy, 2009].

Geostatistical reservoir simulation techniques are used to generate an estimate of spatial configuration of the geological facies as well as petrophysical properties to honor the available local data (e.g. log and core data). Since the nature of stochastic simulation is statistical, there are always some uncertainties in distributions of the reservoir properties within the gridded reservoir volume. Therefore, many equally-probable reservoir models (i.e. realizations) can be generated by the geostatistical algorithms to reflect our lack of knowledge in representing the geology. These realizations have distinct spatial configurations but share the same statistics extracted from available data. Nevertheless, none of these realization is exactly the same as that of the real reservoir itself. The difference between several realizations can be assessed to obtain a measure of uncertainty in the geological and physical properties of the reservoir that is reflected in the realizations

[Caers, 2011]. Each realization is used with a flow modeling process to obtain the degree of uncertainty in reservoir performance by plotting the production rates, cumulative productions, or the fluid recoveries [Ballin et al., 1992]. Clearly, running the flow simulations for all of the possible realizations can be a cumbersome and wasteful, since many of the realizations may exhibit nearly equal flow performance. Therefore, it is crucial to identify which realizations will exhibit different flow performance prior to the flow simulation process.

The goal of this research is to customize our previously proposed visual analytics framework [Sahaf et al., 2016] [Sahaf et al., 2018] for the SAGD process. Within this framework, users can visually and interactively filter the geological models and only select the ones with potentially different flow performance. This framework is based on calculating a (dis)similarity distance between all pairs of models and clustering the models using the computed distance. Each cluster contains models that have potential similar flow performance that is a good candidate for flow simulation.

The proposed technique in this paper can resolve the existent issues of other techniques in literature that are computationally costly and frequently lack any interaction component and visual selecting features. The main contribution of this study is to introduce a novel visual analytics framework and an efficient technique for filtering the equi-probable SAGD models using static properties such as porosity and permeability. We take the advantage of visual analytics techniques to allow the decision makers to engage their knowledge into the filtering process [Sun et al., 2013].

4.3 Related Work

Given the importance of selecting representative models for uncertainty analysis using geostatistical realizations, several strategies have been proposed in the literature, that can be broadly classified into random selection, ranking, probability based techniques and clustering.

In the random approach, engineers randomly choose a few realizations for performing the flow simulations. In this manner, models that exhibit similar performances might be selected and therefore, the correct range of production uncertainty is not adequately represented. Ranking is

a superior method that arranges the geostatistical models based on an easily computable measure in an ascending/descending order and then selects the ones with high (P10), medium (P50) and low (P90) production responses [Deutsch et al., 1996]. The selection of geological realizations is accurate, if the selected ranking measure is highly correlated with the production performance parameters. In addition to that, many of the proposed measures are based on choosing a single static property that is a subjective criterion set by the decision makers. The whole process demands a time consuming manual analysis [Meira et al., 2017].

Some researchers (e.g. [Rahim, 2015]) have also recently investigated other approaches such as probability distance-based realization reduction/selection methods. This method minimizes the probability distance between the original distribution of the superset of realizations and that of the reduced subset of realization. As a result, a smaller subset of realizations is selected. The main issue with these optimization problems is that they could be very complicated and time-consuming for a large set of models, and in the presence of the outliers, the optimization process might not converge [Meira et al., 2017].

Besides the mentioned approaches, several authors proposed clustering based techniques. For instance, Scheidt et al. ([Scheidt and Caers, 2010]) and Singh et al. ([Singh et al., 2014]) proposed a number of specific distance metrics based on the simplified simulation results (e.g. streamline simulations). Then these distances are used to perform the clustering. Consequently, the models in each cluster are considered to be similar in terms of the flow behavior. The type of distance which is calculated in these techniques are mostly cell based and not consider the 3D structure of the facies in the reservoir models.

Considering the limitations of the existent techniques, we propose an analytical framework that is computationally less expensive, which is only dependent on the static properties of geological models rather than flow simulation results. In addition, our approach is accompanied by the visual and interactive components, which is capable of showcasing the differences and similarities between the models. This framework would then help the users to explore the geological uncertainty

using a limited number of geological realizations

4.4 Proposed Visual Analytics Framework

We have previously proposed a visual analytics framework for filtering the conventional reservoir realizations [Sahaf et al., 2016]. Figure 4.2 illustrates an overview of our proposed framework. Initially, we have a set of 3D geological models. Pair-wise distances are calculated between all pairs of models using a block based similarity metric. The calculated distances are then utilized to project models into a lower space (2D) using multidimensional scaling techniques [France and Carroll, 2011]. Each point in the 2D space corresponds to a 3D model. This 2D projected space helps to have an overall representation of the superset of models, in a way that the closer the points, the more similar the 3D model are. In the next step, the initial ensemble of the models is clustered into different groups based on their similarities. Prior to applying clustering, kernel transformation methods [Dhillon et al., 2004] are used to align the projected points in a more linear and simpler way. This is mainly because clustering works more efficiently in linear spaces [Caers, 2011] [Sahaf et al., 2018].

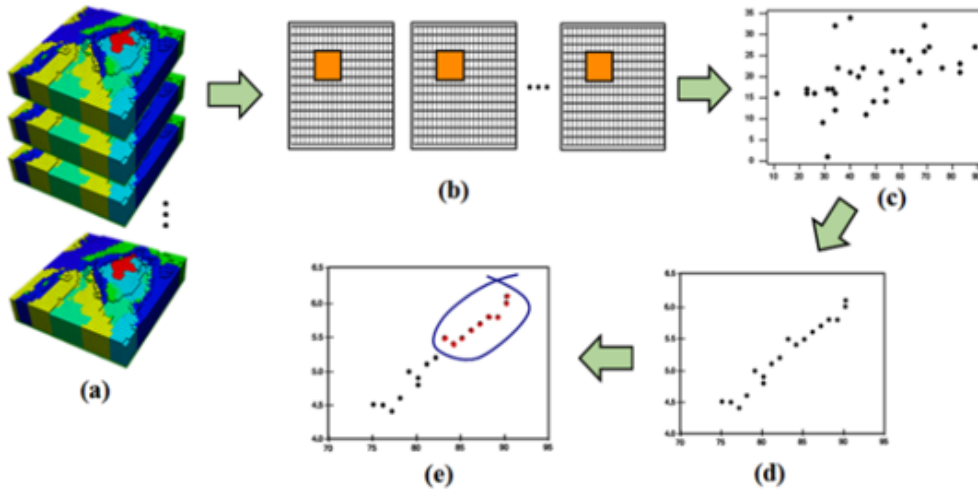


Figure 4.2: Proposed visual analytics framework for selecting of reservoir models [Sahaf et al., 2018].

4.5 Distance Calculation

Definition of distance is very critical in measuring distance between the geological models. Many of the related reservoir simulation studies (like ranking) are performed based on a single geological model. Therefore, the 'distance' between reservoir models is not very popular in petroleum engineering [Fenwick and Batycky, 2011]. Two models are called similar, when they have similar dynamic result (reservoir performance). There are two essential requirements for defining a distance between two geological models. First, it has to be well correlated with the dynamic behavior of reservoir [Scheidt and Caers, 2008]. Second, the associated calculations should not be very costly. Although the first requirement shares the same ground with the ranking methods, some studies have already shown that the distance-based methods are more superior and can provide more effective and robust solutions [Scheidt et al., 2009].

In a recent study, we proposed an efficient and robust block based approach for the calculation of distance between geological realizations. In this technique, we use a 3D gridded block as a moving template to sweep the entire reservoir model (as shown in Figure 4.3). At each moving step, (dis)similarity distance between the corresponding blocks in the two models is calculated. The final distance is an average of the (dis)similarity values between all the corresponding blocks. The distance metric which is used in this study is Mutual Information (MI). MI is a popular information-theoretic measure of similarity (or dissimilarity) which has been applied in many areas of visualization and graphics domain like image registration, multi-modality fusion and view point selection [Bruckner and Möller, 2010] [Haidacher et al., 2008]. In our recent study [Sahaf et al., 2018], we showed the effectiveness of MI in comparison to the other well-known distances (e.g. Euclidean and Hausdorff distances). Figure 4.4 shows a summary of our distance calculation between two different models.

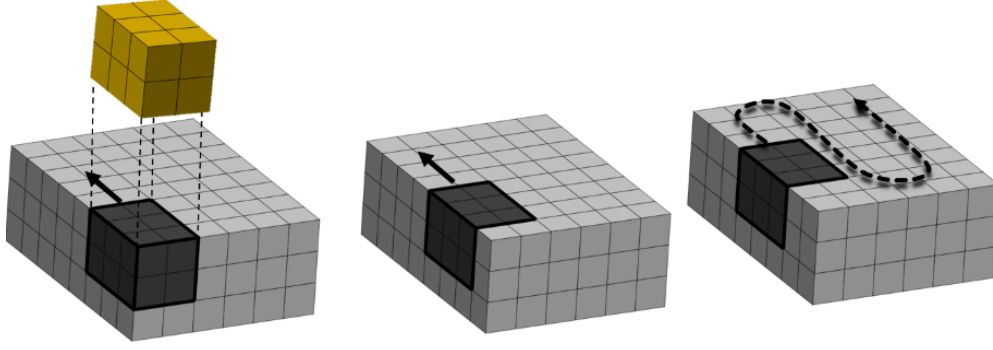


Figure 4.3: Block based approach for calculating distance between realizations.

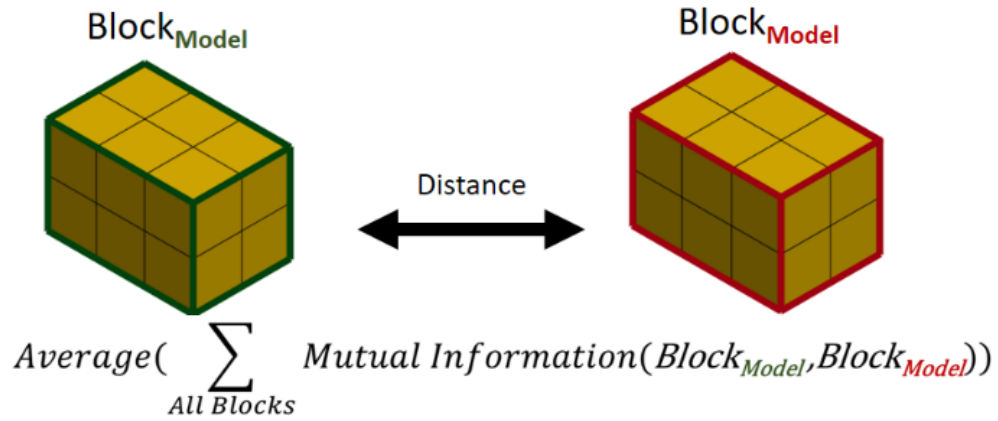


Figure 4.4: Calculation of distance between two blocks.

The size of the scanning template is an essential component of our proposed distance measurement. An entropy based approach is used to calculate the best possible size for the subject realizations [Sahaf et al., 2017].

In this process, the user can specify any number of static properties (such as porosity, permeability, saturation, etc.) to be incorporated in the distance calculation process. With the selection of multiple static properties, distances are first calculated individually for each property, and the final distance is the average of all the distances measured for all of the selected properties [Sahaf et al., 2018].

4.6 Projection with Clustering

At this point, our aim is to use the calculated distances within a clustering algorithm in order to identify the similar models. Consequently, we have cluster centers that are the default representative members of the superset of realizations. That leads to our main requirement - that is to reduce the number of models needed to be simulated. K-means [Correa et al., 2009] is one of the most popular clustering algorithms that is proved to be very efficient on large datasets. However, a major drawback to K-means is that it cannot separate clusters that are non-linearly separable in input space [MacKay, 2003]. Kernel K-means [Schölkopf et al., 1998] has emerged to tackle such problems [Shawe-Taylor and Cristianini, 2004]. In this approach, kernel transformations like Radial Basis Function (RBF) are used to map the projected points to a new space where points become linearly separable [France and Carroll, 2011] [Scheidt and Caers, 2010]. Finally, clustering is performed on this linear space. Other than having better and simpler visualization of projected models using kernel transformation, it provides more accurate clustering results [Scholkopf and Smola, 2001], as shown in Figure 4.5. It can be seen that the representation of data looks better when a kernel transformation has been applied. Without kernel transformation, projected points are very close to each other, and that makes the clustering very complicated. However, with kernel transformation, a well-organized and linear structure can be seen in the results, and clusters are better represented and separated.

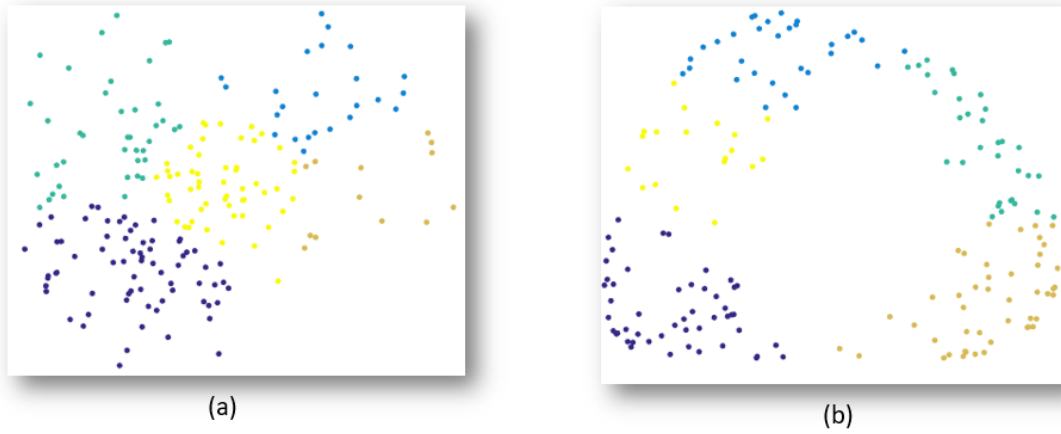


Figure 4.5: Clustering results of projected models with (b) and without (a) kernel transformation.

4.7 Extended proposed framework for SAGD models

This process works well for a number of conventional reservoir model scenarios as those outlined by Sahaf et al. ([Sahaf et al., 2016]). However, for the SAGD simulation models, this process cannot be applied directly without additional considerations to the heterogeneities specific to the near wellbore areas where we have a steam chamber of limited extent. This point is shown in Figure 4.6, where the black straps correspond to the distribution of shale lenses in three distinct realizations. If we apply the proposed filtering process directly on these three models, it groups model (a) and (b) in one cluster (they are identified as similar models), and model (c) in another different cluster. The reason is that the shale orientations are similar in models (a) and (b), except that they are connected in model (b). On the other hand, model (c) has a totally different shale orientation. However, when we run the flow simulations for all the models, the flow simulation results reveal that the performance of the model with continuous shale patches (b) is completely different from the case with discontinuous shale patches (a). Additionally, it can be also seen that the flow behavior of models (a) and (c) are very similar to each other (Figure 4.6).

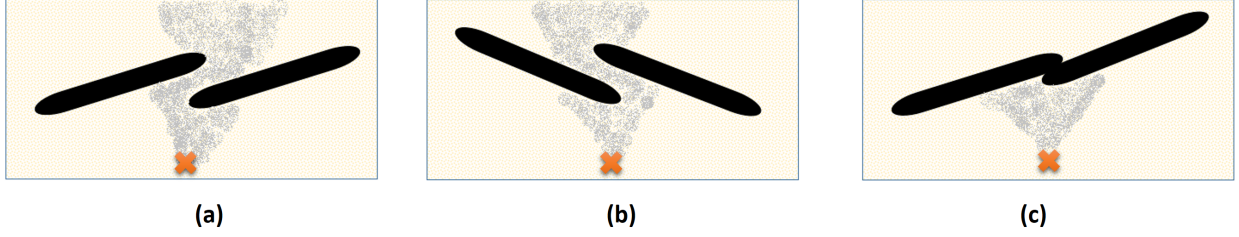


Figure 4.6: Simulation result and steam chamber for three different SAGD based reservoirs with varying shale continuity. (Black areas show sample of shale bodies. Smoky areas show the steam chamber and the orange cross shows the steam injection well.)

This simple example indicates that our proposed visual analytics framework may not be directly applicable to the reservoir realizations of the processes such as SAGD which only concern particular areas of the reservoirs that are affected by the steam chamber. The principal reason is that shale continuity is an important factor in shaping the steam chamber and when we include all the non-informative reservoir areas the contribution of the local differences on the similarity measure will be reduced. Therefore, it is concluded that the near-well region and more specially the steam chamber area is the most crucial area in our analysis process. Therefore, it is required to perform our proposed algorithm locally with only focusing to some areas close to the well locations and exclude the far areas where they have minimal impact on the SAGD performance. In order to extract this areas, we propose to run a single flow simulation for a homogenous model without any shale or barrier (Figure 4.7 - a). This task assist in depicting the general extent and location of the steam chamber (Figure 4.7 - b) and helps to exclude the areas outside this region in the subsequent steps of our framework (the red highlighted area in Figure 4.7- c). In the next stage, pair-wise distances are calculated considering only the highlighted extracted steam chamber areas (Figure 4.7s - d). The calculated distances are then used to map the models into the lower space using the MDS projection technique (Figure 4.7 - e). Kernel transformations are applied, in the next step, to linearize the space (Figure 4.7 - f). Finally clustering is performed on the linearized

space (Figure 4.7 - g). The entire customized process is illustrated in Figure 4.7.

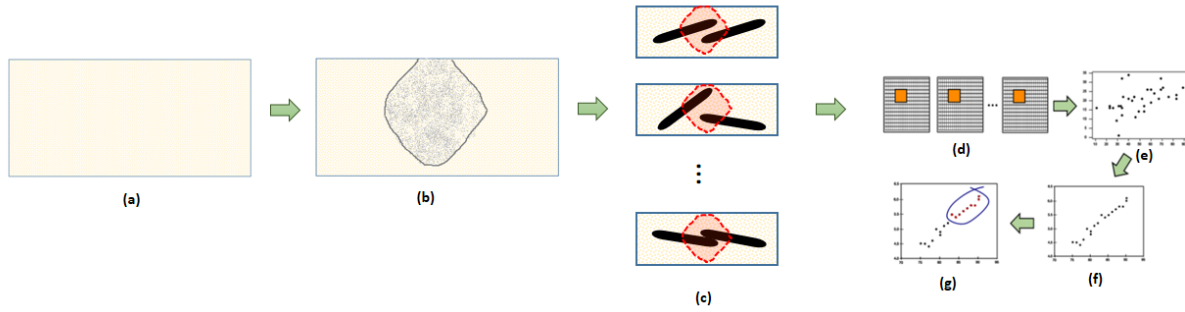


Figure 4.7: Proposed custom visual analytics process based on steam chamber area. (a) Original reservoir model, (b) steam chamber area, (c) highlighted steam chamber area for all the realizations (d) distance calculation only for the steam chamber area, (e) projection of models into lower space, (f) transformation of models into new space using kernel functions, (g) clustering of models in the new space.

4.8 Result

In this section we apply the clustering approach to a case study of five realizations for a SAGD process. The models represent different shale distributions with fixed porosity and permeability values for each facies. The reservoir model has $249 \times 1 \times 69$ grid cells in x, y and z directions respectively.

The subject models in this study are illustrated in Figure 4.8. To validate our approach, we performed the complete flow simulation for all the models. Figure 4.9 represents the plotted simulation results. It can be seen that the simulation results for some models are very close to each other, while some others are completely different. This assists in identifying the expected clusters for the evaluation process. From the simulation results, it is desired to select two representative realizations for the upper and lower bounds (Figure 4.9).

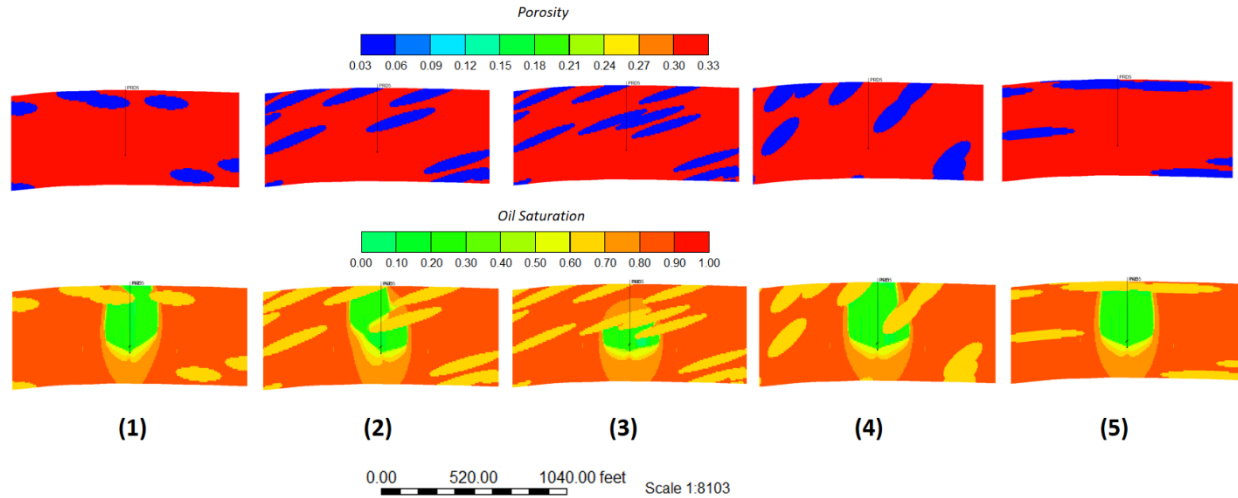


Figure 4.8: Illustration of our case study with five reservoir models. First row shows the actual model and the second row shows the simulation results of models. Green areas in the second area show the steam chamber area.

According to the proposed framework, we run a flow simulation for a model without any shales. Then the maximum steam chamber area is extracted as shown in Figure 4.10. The same area is used for all the models to calculate the distances. Porosity and permeability are two critical static properties in this case study. In our case, porosity and permeability are correlated. Therefore, either of permeability or porosity can be used for the calculation of distances. The main reason is that the distribution of property values is important for our MI based distance, rather than the actual values. For instance, the distance would be similar either using porosity values (0.03 to 0.33 range) or permeability values (0.001md to 4519md range). The calculated similarity values are used to project models into a 2D space. Finally, clustering algorithm is applied on the points (models) in the 2D space. In this space, the relative positioning of the models is important (i.e. how close or how far they are from each other), rather than their actual positions in the space. The clustering results in Figure 11 shows how clustered models match with the expected clustering results that we got from the flow simulations. For instance, model 3, which has the lowest production rate

(Figure 4.9) is clustered in a separate cluster. In the projected space, this model is located far from other models. This favourable clustering is resulting from the use of clustering only in informative regions, which only include the areas inside the maximum steam chamber. If we apply the filtering process on the entire model (Figure 4.11; Right) and compare the results with local approach (i.e. within the steam chamber) and the expected closures from the simulations, (Figure 4.11; Left) we can immediately identify the importance of the local scheme. Results in Figure 4.11 (Right) reveal how grouping of models is unsatisfactory when the entire model is considered for clustering. For instance, the results from flow simulations show that although model 4 has the highest production rate and it is very close to models 1, 2 and 5, it is erroneously clustered in a separate cluster in Figure 4.11 (Right), which is not acceptable. Moreover, the clustering results should have indicated that model 3 to be very different than the other models, and again this expectation has not been produced using the clustering applied on the entire model areas. All these findings confirm how steam chamber area is important in SAGD models, and our proposed framework should be modified accordingly to consider that area. In addition to providing more accurate results, the local filtering can also increase the computational performance of the clustering algorithm itself.

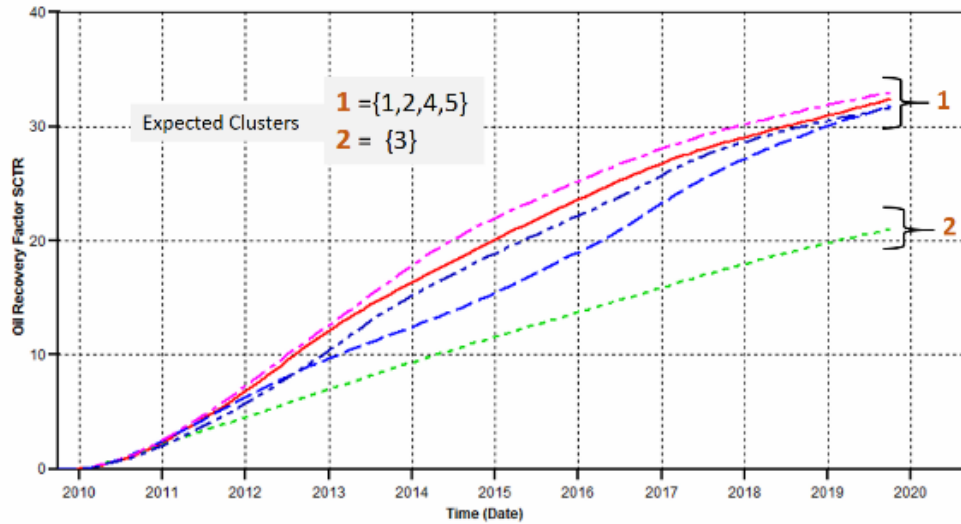


Figure 4.9: Simulation results for five different realizations for SAGD process.

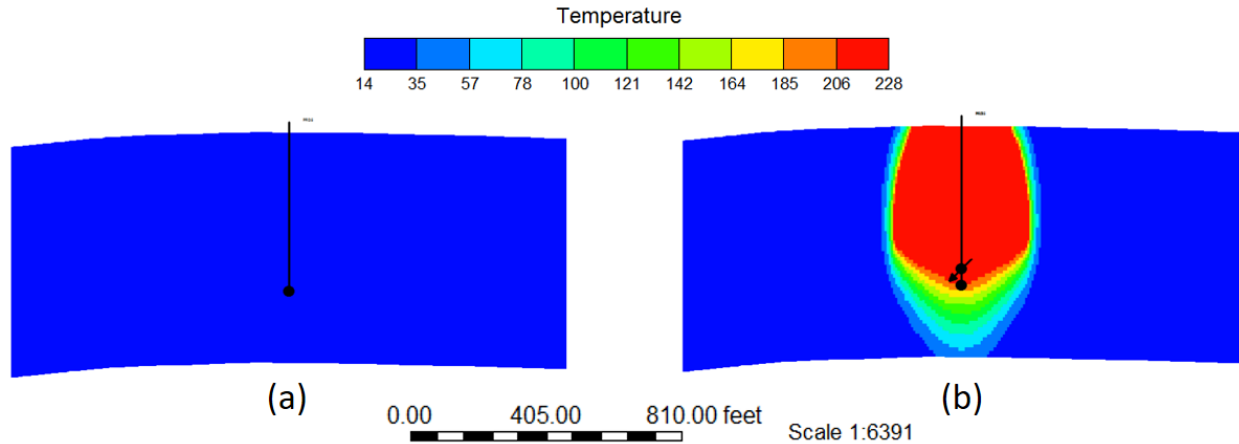


Figure 4.10: (a) Representation of a reservoir model without any shale structure. (b) Maximum steam chamber area that is extracted by running simulation for that reservoir model.

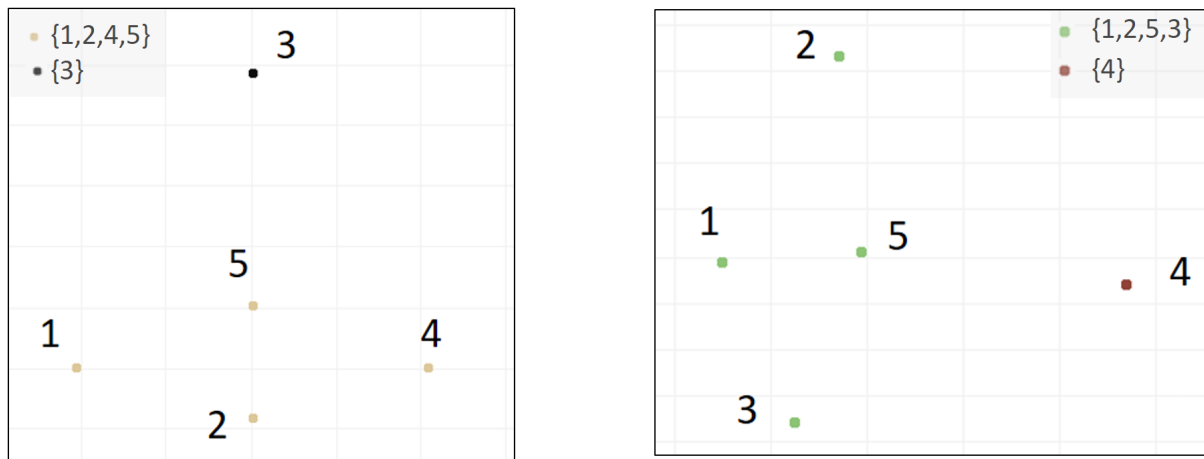


Figure 4.11: (Left) Clustering result with considering only the steam chamber area of models. (Right) Clustering result on the entire area of models.

4.8.1 Computational Time

Flow simulations are usually very costly for the SAGD models. Depending to the size and complexity of the reservoir models, it can take weeks to months for a single flow simulation to com-

plete. In this case study, we used 2D models to save the computational time. Each flow simulation for any of the 2D models took around one hour to complete. However, our framework generates clustering results with the representative models in less than 20 seconds, which is significantly different from the brute-force approach (running simulation for all the models).

4.9 Conclusion

In this paper, we introduced a visual analytics framework for filtering the realizations in SAGD process. This framework helps to select a few representative models from an ensemble geostatistical models that can represent the overall production uncertainty. The engineer can use the cluster centres to perform the simulations and finds the bounds of the production uncertainty without the requirement to simulate all realizations. To achieve this, we employed our previously proposed visual analytics and customized it in a way that only the steam chamber area would be considered for all the models. This steam chamber area is extracted by running an additional flow simulation for a case without any shale or barrier area. The results on the case study show that our technique is well matched with the brute force approach, which is based on running the flow simulations for all realizations. In addition to the accuracy of results, the whole filtering process would be faster, since only a small important area of realizations is considered.

Chapter 5

An ROI visual-analytical approach for exploring uncertainty in reservoir models ¹

5.1 Abstract

Uncertainty in reservoir geological properties has a major impact on reservoir modeling and operations decision-making, and it leads to the generation of a large set of stochastic models, called geological realizations. Flow simulations are then used to quantify the uncertainty in predicting the hydrocarbon production and to make optimal decisions in reservoir development. However, reservoir flow simulation is a computationally intensive task due to the existence of a large number of grid blocks available in the models. Normally only a small number of realizations are chosen from a large superset for the flow simulation model to describe the future production uncertainty. In our recent paper, we proposed a visual based analytical framework to select a few models from a large ensemble of geological realizations. In the proposed framework, all pair-wise distances are calculated between the realizations using a novel designed distance measurement. The calculated distances are then used within a clustering technique to group realizations into similar clusters. Cluster centers are the representative geological realizations that can provide different simulation results. In this paper, we extend our prior framework by introducing a region of interest selection method using the variance concept, that helps to perform the entire selection analysis only based on a specific portion of the reservoir, while gaining even more performance. The effectiveness of region of interest selection techniques is shown on two case studies. For evaluation of our visual analytical framework, we perform a complete user study with engineers and geologists. User stud-

¹Sahaf, Z., Cabral Mota, R., Hamdi, H., Costa Sousa, M. and Maurer, F. (2018). An ROI Visual-analytical Approach for Exploring Uncertainty in Reservoir Models. Submitted to the Journal of Communications in Computer and Information Science (Springer), <https://www.springer.com/series/7899>

ies involve certain tasks that cover all the key aspects of the framework. User feedback suggests that the usefulness, usability and visual interactivity of our application are the key strengths of our approach.

5.2 Introduction

Geological properties are important parameters used in the oil extraction processes from reservoirs. These parameters influence the production performance of the reservoirs. Geological uncertainty exists because of the lack of knowledge to exactly describe the geological properties of every section of the reservoir. Techniques such as well logging and core analysis can give some ideas about the local geological properties of interest areas of the reservoir. However, the inter-well geological properties of many uncored areas will be yet unknown. As a result, the geological uncertainty will always exist in reservoir description workflows [Caers, 2011].

Reservoir performance can be quantified by the flow simulation, which provides the production time series such as the cumulative oil production (COP) rate and the net present value (NPV). The flow production parameters strongly depend on the underlying geological properties of the reservoir. It is very important to quantify the impact of geological uncertainty on reservoir modeling procedures. Otherwise, the model may give an unreliable assessment of production capacity of the reservoir. To represent the geological uncertainty, multiple geological realizations are usually generated using geostatistical tools [Goovaerts, 1998] (called geological realizations) so as to obtain a broad range of possible geological properties for a subject reservoir. However, reservoir flow simulations cannot be run for all of the possible realizations due to the significant computer processing time. Therefore, in practice, only a small number of geological realizations are chosen to perform reservoir simulations to describe the geological uncertainty [Idrobo et al., 2000]. However, all the current selection techniques [Ballin et al., 1992] [Scheidt and Caers, 2010] are the one-time automated processes, that lack any visualization prospective and more importantly they do not incorporate the user's knowledge into the selection process.

In a recent paper [Sahaf et al., 2018], we proposed an analytical process for selection of geological models. Our approach relies on a pair-wise distance based clustering technique that groups the geological realizations based on their similarities. There are limited comprehensive distances defined in the domain for the geological models. Therefore, we proposed a block based distance calculation method, which employs the mutual information concept [Goshtasby, 2012] [Lin, 1998]. Thereafter, dimensionality reduction techniques are employed to calculate the distances required for mapping the reservoir models into lower spaces (like a point in a 2D or 3D space). Clustering would be then performed on the lower space. Each cluster contains similar geological realizations, and cluster center is the representative of that cluster. In order to have more accurate results (in terms of distance calculation between the reservoir models and clustering of the models) and better performance (in terms of time), three interactive region of interest (ROI) selection techniques are added to our existent framework to let the users perform the analysis on particular areas of the reservoir model. The entire process has been designed in a visual interactive process and has been developed in a virtual reality environment, which can enhance the incorporation of the user's knowledge in the entire process.

The reminder of the paper is organized as follows: section 2 provides an overview of geological realization techniques in the engineering domain and also the current visual interactive processes for similar datasets, section 3 provides an overview of the proposed framework, section 4 provides the details of variance model calculation and the proposed selection techniques, section 5 provides the details of distance calculation, section 6 explains about the clustering technique, section 7 explains how the variance model is interacted with the clustering results, section 8 provides case studies and the results, section 9 provides the user study evaluations and finally conclusions are presented in section 10.

5.3 Related Work

5.3.1 Current Approaches for Selection of Geological Realizations

Various methods are available for selecting geological realizations. Random selection of realizations is one of the easiest methods; however, it can not give a correct measure of geological uncertainty [Yazdi and Jensen, 2014]. Ranking [Ballin et al., 1992] is the most common method for selecting geological realizations, that arranges the geostatistical models based on an easily computable measure in an ascending/descending order and then selects the ones with low, medium, and high values of that measurement. Ranking methods suffer from a major drawback, that they rely greatly on the type of the measure that is used. Distance based methods have been recently investigated by some researchers [Rahim and Li, 2015] [Scheidt and Caers, 2010], that they try to select few representative models using the similarity distances between the models. The need of petroleum industry to address the geological uncertainty using a limited number of geological realizations necessitates designing an analytical framework that is computationally less expensive that is dependent on the static properties of geological models rather than flow simulation results. In addition to enhance the exploration of the geological model and perform some specific engineering tasks on the model, It is also highly favorable to create a visual and interactive environment capable of showing differences and similarities between the models.

5.3.2 Visual Analytics Techniques for Multi-run Data

The most similar dataset in computer science domain to the geological models in petroleum engineering is multirun data. In general, multirun data stems from a type of process (like geostatistical algorithms in our case) that is repeated multiple times with different parameter settings, leading to a large number of collocated data volumes [Wilson and Potter, 2009]. Since multirun data consists of a superset of volumetric models, their representation and analysis are challenging [Kehrer and Hauser, 2013].

Accordingly, one of the common ways for visualization of such datasets is to aggregate the dis-

tributions of multirun data, by computing statistical summaries [Love et al., 2005] and representing them mainly by box plot [Kao et al., 2002], line chart [Demir et al., 2014], glyphs [Kehrer et al., 2011], or InfoVis techniques such as parallel coordinates or scatterplot matrices [Nocke et al., 2007]. For the analysis of multirun data, statistical methods (such as mean, variance, skewness, and kurtosis) are vastly used to reduce the data dimensionality [Kehrer et al., 2010]. Alternatively, mathematical and procedural operators are also used to transform the multirun data into some compact forms (e.g. streamlines, isosurfaces, or pseudocoloring) where existing visualization techniques are applicable [Love et al., 2005] [Fofonov et al., 2016]. Data mining techniques such as clustering are also among the recent methods being used to explore the multirun data [Correa et al., 2009] [Bordoloi et al., 2004] [Bruckner and Moller, 2010], that try to identify similar behavior across different simulation runs. Most of these analytics researches are on the 2D ensembles such as 2D images. For the 3D ensembles, their aggregations are usually used for the analysis tasks, which we want to avoid in this research, due to the importance of 3D structures in the reservoir models.

5.4 Analytical Framework

An overview of our extended process is represented in Figure 5.1. Initially, we have a set of 3D geological models (a). Variance is calculated per cell for a specific geological property that is called “variance model” in this study. (b). Different selection techniques are used with the variance model to select the ROIs (c). The proposed block based similarity metric is calculated for all pairs of models considering only the user selected ROIs (d). The similarity values are then utilized to project models into a 2D space using multidimensional scaling techniques (e). Each point in the 2D space corresponds to a 3D model. The distance between points in the projected space represents the similarity between the models; the closer the points, the more similar the 3D models are. Cluster centers are selected by default and considered as the representative models for each cluster. These models are the candidates for running the flow simulations and uncertainty assessment (f). Cluster centers are expected to be very different from each other. However, it is

very important to show how models are different spatially (i.e. in terms of which areas model are different). This is because the spatial variation of the properties can have a large impact on the flow simulation results. On the other hand, users might want to select other models than the cluster centers. To leverage these requirements, an ROI analysis method is proposed. As such, the variance is calculated only for the selected models and is projected on the cells that are within the ROI. If the selection of a new model increases the variance, it is indicative of a very distinct model. However, if only minimal changes are observed in the variance model, it is inferred that the selected model is potentially similar to the other models in the cluster. More importantly, the visual inspection of variance changes on the cells of ROI helps the users to easily observe if a different model has been selected. The entire process has been designed and developed in a virtual reality environment. Each of these stages are explained in more detail in the subsequent sections which can also be inspected visually in the prepared video².

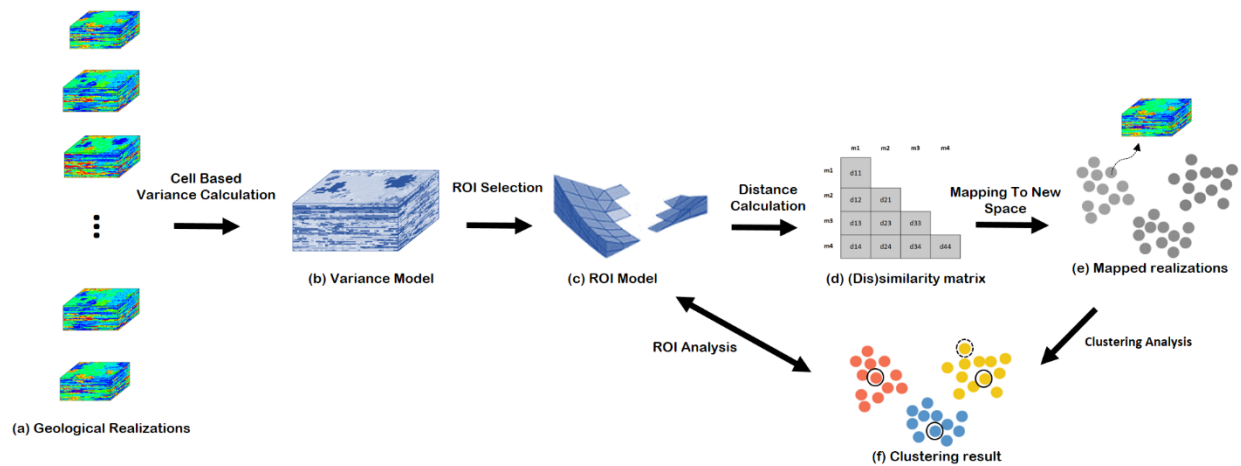


Figure 5.1: An overview of our extended visual analytical framework.

²<https://vimeo.com/251921649>

5.5 Variance Model and ROI Selection

Reservoir models are usually very large and consist of millions of cells. The models often embed complicated structures with non-uniform layers. In many circumstances, the engineers need to only focus on specific areas of the models for particular studies such as local history matching (i.e., calibration). The calibration processes are aimed to find a geological model that can best match the dynamic data. However, very often, the outcome of this process is not entirely satisfactory as the dynamic production of data of certain wells might not match the observed data. Another example is the 4D seismic history matching where the local discontinuities (e.g. faults or permeability variations) can cause the spatial measured seismic response not to match the simulated seismic data. In all these examples, the lack of large conditioning data is meant to having a large number of models that can equally produce the same response (i.e. non-uniqueness). As such, the focus of the engineer would be on specific spatial areas (e.g. around wells, aquifer, anticline, etc.) across many equiprobables models to fine tune the local unknown to achieve a reasonable match. Depending on the type of the reservoir, these regions of interest can be different. To facilitate the improvement of this local investigation processes, we provide three different selection methods: filtering, group selection, and single cell selection. They all help particular requirements of the users by interactive and visual selection of ROIs for a subject reservoir model. All these selection techniques are performed on the calculated variance model. In this variance model, variance is calculated per cell over all ensemble of geological models as shown in Figure 5.2. Figure 5.3 shows a sample of variance model for one of our case studies. Red and blue areas illustrates high and low variance areas subsequently, that shows the realizations are more different in red areas (lack of certainty).

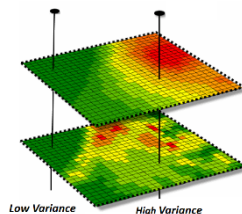


Figure 5.2: Variance calculation between geological realizations.

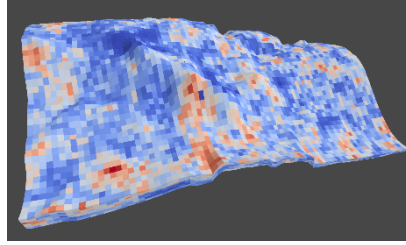


Figure 5.3: A sample of calculated variance model.

In the filtering selection technique, the user specifies a variance threshold interactively. As a result, the cells below the specified threshold are filtered out (Figure 5.4). As mentioned in the previous section, the areas with large variability indicate the lack of knowledge, that is resulted from the lack of enough spatial constraining data. In contrast to the low variance areas, these high variance areas contribute more in detecting differences between models. Therefore, it is very important to inspect these areas by filtering out low variance areas, as shown in Figure 5.5.

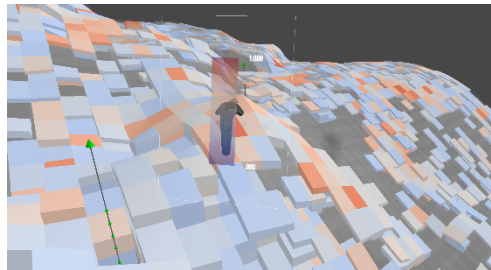


Figure 5.4: Interactive filtering of the variance model.

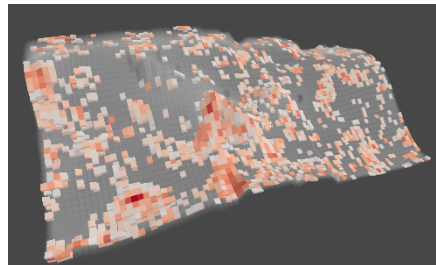


Figure 5.5: Filtered out variance model.

In the group selection technique, a group of cells can be selected using a visual and interactive

box widget. This box can be created dynamically in any size, in any location and as many times as needed. Using this technique, users can select multiple collocated cells (as the ROI) with one operation. As an example, in Figure 5.6 the user tries to create a box in a high variance area. Consequently, a large box is selected to cover the area as shown in Figure 5.7. Multiple boxes can be also created as shown in Figure 5.8.

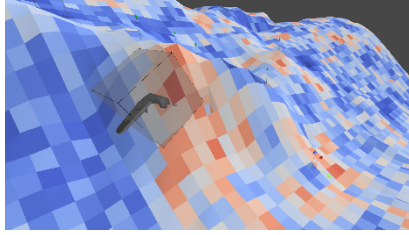


Figure 5.6: Interactive creation of box for the group selection.

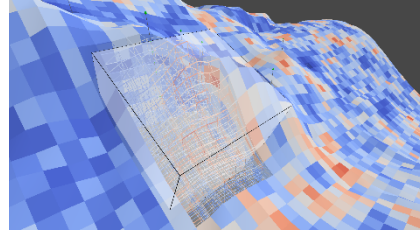


Figure 5.7: A sample representation of group selection using the interactive box.

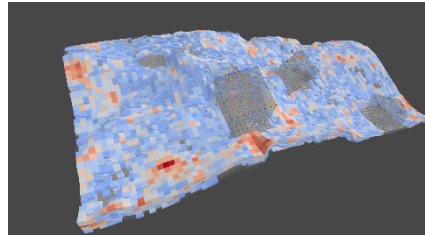


Figure 5.8: Multiple creation of boxes for selection of the ROIs.

Last but not least, is the single cell selection technique, where a cross sectional view is provided to the users, where they can select individual cells. Using the cross sectional view, the users can observe internal areas of the reservoirs, and select important cells with high variance. Figure 6.9 shows an example where that user tries to select cells around a subject well.

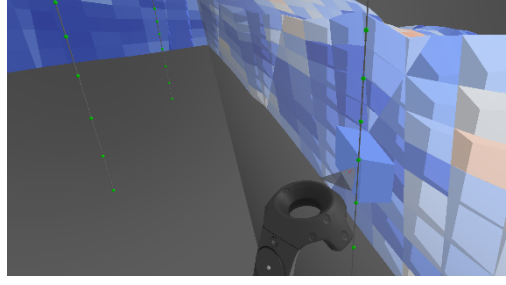


Figure 5.9: Single cell selection around a well.

The final ROI is the union of all the user selected areas (group selection and filtering) and cells (single selection), that can be subsequently used for the calculation of distance between the realizations. A sample ROI is shown in Figure 5.10.

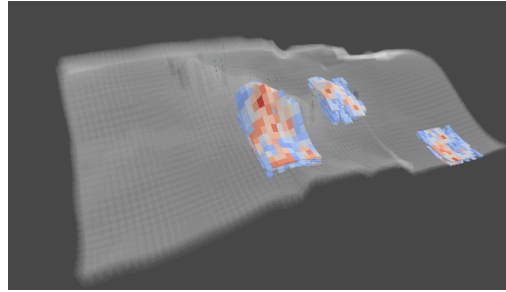


Figure 5.10: A sample ROI selection.

5.6 Distance Calculation

Many of reservoir engineering tasks are usually performed on a single geological model. Therefore, “distance” between reservoir models is rather a new concept in that domain [Fenwick and Batycky, 2011] and it is important to define a distance that reflects the requirement of the engineering tasks. Two models are called similar, when they have similar dynamic result (reservoir performance).

In our recent study, we proposed a block based approach for calculation of distance between geological realizations [Sahaf et al., 2018]. In this technique, we use a 3D block as the moving template, where each block consists of a few number of grid cells. This 3D block sweeps the entire reservoir model (as shown in Figure 5.11), and at each step, the (dis)similarity distance is calculated

between the corresponding blocks in two models. Finally, we take the average of the (dis)similarity values between all the corresponding blocks. The distance metric which is used in this study is Mutual Information (MI). In [Sahaf et al., 2018], we performed a thorough evaluation study on the MI and showed the effectiveness of MI distance on the geological realization in comparison to the other well-known distance (e.g. Euclidean and Hausdorff distances). More importantly, our proposed distance metric was successfully applied to a number of specific reservoir engineering problems [Sahaf et al., 2016] [Sahaf et al., 2017]. Figure 5.12 shows a summary of our distance calculation between two different models.

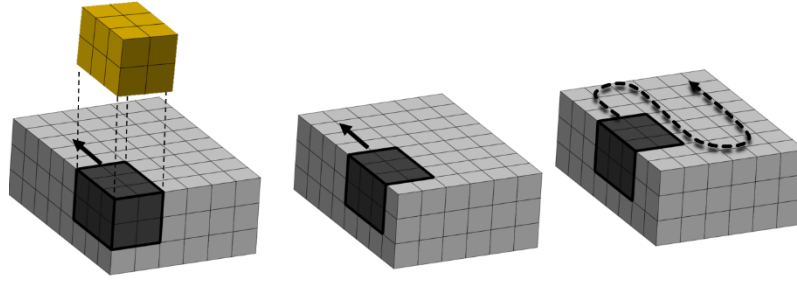


Figure 5.11: Block based approach for calculating distance between realizations.

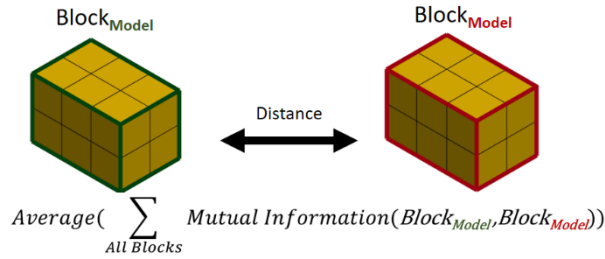


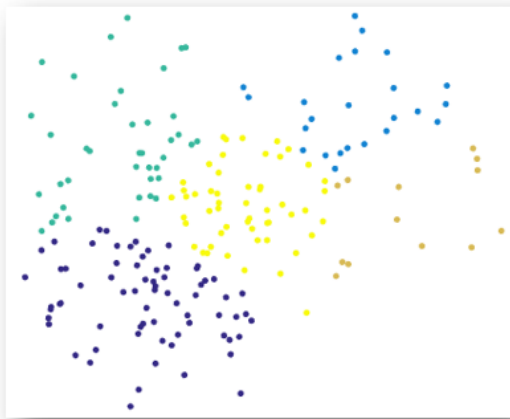
Figure 5.12: Calculation of distance between two blocks.

Block size is an essential component of our proposed distance measurement. We use the concept of entropy, as suggested in [Honarkhah and Caers, 2010], to base the optimal block size. This optimal value is provided as a suggestion to the users in our designed application; however, the users can change the size of block based on their knowledge such as the use of correlation length [Mela and Louie, 2001] as it is used in generating some geological patterns.

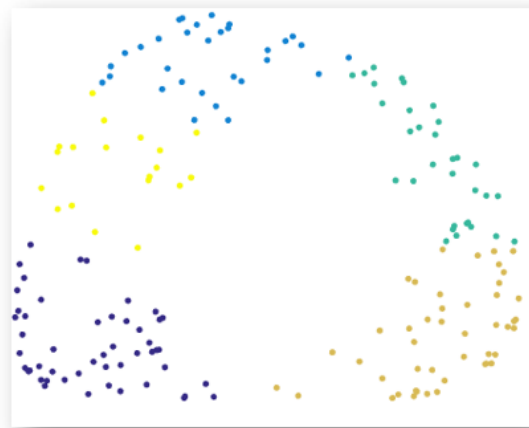
5.7 Projection with Clustering

The calculated distances are utilized within a clustering algorithm in order to group similar models. Each cluster center is a default representative member of the containing cluster, which leads to our main requirement - that is to reduce the number of models needed to be simulated. The K-Means Clustering (KMC) algorithm [Correa et al., 2009] is employed in this step because of its computational efficiency on large data sets. However, KMC suffers from a noticeable drawback. In the case where the data embeds a complex structure (e.g. data are non-linearly separable), a direct application of KMC is not suitable because of its tendency to split data into globe-shaped clusters [MacKay, 2003]. In order to solve this problem, as suggested in [Shawe-Taylor and Cristianini, 2004], data will be mapped by a kernel transformation [Schölkopf et al., 1998] to a new space where samples become linearly separable. The crucial aspect in the new space is the relative distance between projected points. The closer points are to each other, the more similar they are based on the original defined distance. Other than having a better and simpler visualization of the projected models using kernel transformation, it has also a great benefit for clustering algorithm that is providing more accurate results [Scholkopf and Smola, 2001]. This technique is called kernel k-means in the literature [Williams, 2002] [Dhillon et al., 2004].

The efficiency of kernel KMC in clustering the geological models is shown in Figure 5.13 and Figure 5.14. They both show how the representation of data and clustering is different in two scenarios: projection with and without kernel transformation, one in 2D view and the other one in 3D view. In both views, it can be seen that the representation of data looks better and more importantly clustering results are more representative when a kernel transformation has been applied. Without kernel transformation, projected points are very close to each other, and that makes separation of clusters complicated. However, with kernel transformation, a well-organized and linear structure can be seen in the results, and clusters are better represented and separated.

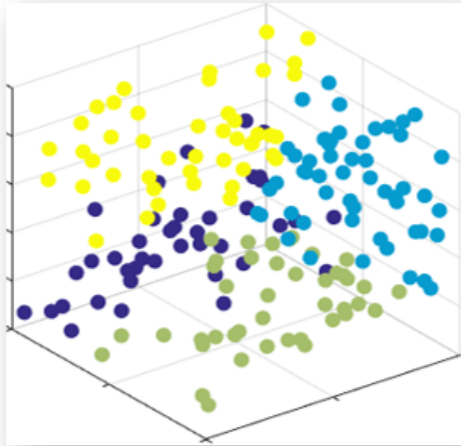


(a)

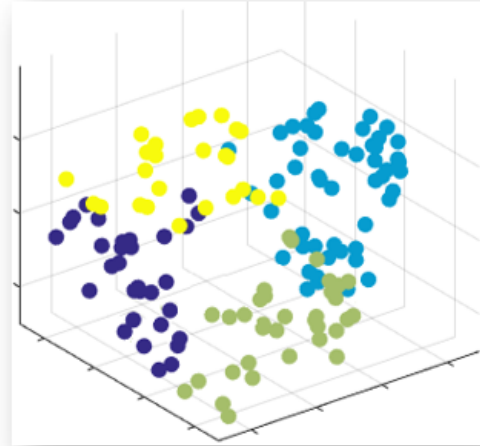


(b)

Figure 5.13: 2D view of clustering results of projected models with (b) and without (a) kernel transformation.



(a)



(b)

Figure 5.14: 3D view of clustering results of projected models with (b) and without (a) kernel transformation.

5.8 ROI Analysis

As mentioned in the previous section, cluster centers are the default selected models that represent the variability in the entire set of subject ensemble of reservoir models. These representative models would be different in terms of simulation result, and tend to cover the entire range of the production variability. However, users might want to run additional simulations for other different models to have a better understanding of the estimated uncertainty range by the cluster members. The first potential candidates are the outlier models. For instance, in Figure 5.15, although the clusters centers are well calculated (annotated with stars), there are some models that are located very far from the cluster centers and look different (highlighted with black circles). Generally, ROI analysis is a proposed component that helps users explore other different models. In this analysis, initially users see the variance model calculated only between the centers for the ROI (Figure 5.15). In the next step, users can interactively select other different models (can be outliers), and the variance model is re-calculated to include the additional selected models. If the calculated variance increases (in the areas that user prefers) (Figure 5.16 - (b)), it shows that the selected model is a different model (comparing to the previous selected ones). However, if the variance decreases (Figure 5.16 - (a)), it shows that the selected model is similar to the previous ones, and there is no benefit in running an additional simulation for this model, because it is likely that the production uncertainty range would not be changed.

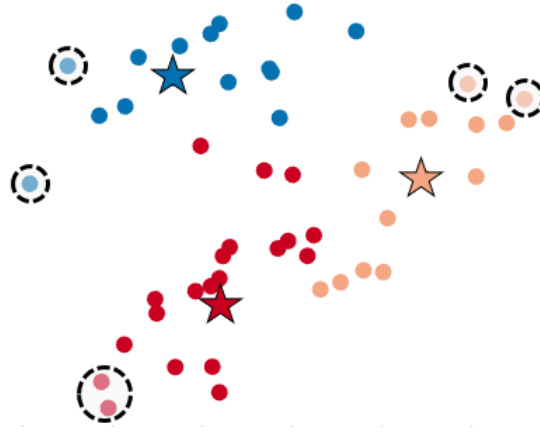


Figure 5.15: Representation of some outsider models in each cluster that can be considered as potential different models.

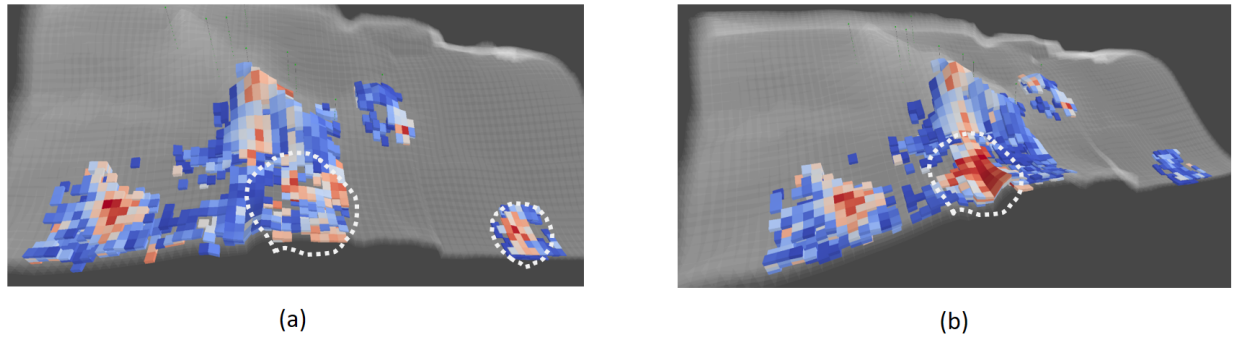


Figure 5.16: ROI analysis with selection of the most outsider model (a) and the second most outsider model (b), and observation of variance changes in each scenario.

5.9 Results (with Case Studies)

For evaluation of our proposed extended analytical framework, we first compare our results with the current alternative process in the industry, i.e. to run flow simulation for all the models individually. We run the complete flow simulation for all the models using the CMG reservoir simulator package [CMG, 1980], and plot the simulation results for the 'oil recovery factor'. The plotted re-

sults show a range of uncertainty on the oil production and we expect that our cluster centers cover this range adequately. In the second phase of evaluation, our engineer colleagues select an ROI on the reservoir model. The pair-wise distances are then calculated between the reservoir models only considering the ROI regions. According to our framework, models are then clustered into similar groups. Finally, effectiveness of ROI selection is shown by comparing clustering results with and without the ROI selection. In terms of the datasets, our industry partner generated different datasets for us using different geostatistical algorithms and scenarios. The idea is to cover almost all different types of datasets in the domain. In the first case study, our engineer colleagues generate 20 realizations by changing the seed in the variogram of permeability property. Other geological properties are remained unchanged. The size of realizations are 26,000 cells. All the realizations were conditioned with the same well log data. Our entropy based algorithm [6] suggests $15 * 10 * 3$ as the best block size for calculating similarity distance between realizations. Figure 5.18 shows the flow simulation curves for all the 20 models. A good clustering result should cover this range of uncertainty. Figure 5.17 shows the clustering results in the scenario that no ROIs are selected, and the clustering is performed on the entire model. It can be seen that the range of uncertainty is not covered completely. However, when some appropriate ROIs are selected (like high variance areas and well areas) as shown in Figure 5.21, clustering results cover the entire range of uncertainty better (Figure 5.20). In addition to that, the clusters are more coherent (see how clusters are spread without ROI selection) and better reflect the distances of models (Figure 5.19). On the other hand, the whole process is much quicker with the ROI selections. The reason is that less number of cells are involved and hence the calculations is naturally faster. In this specific example, without ROI selection, the entire process takes around 3 minutes, while with the ROI selections, only 30 seconds takes to generate the clustering results.

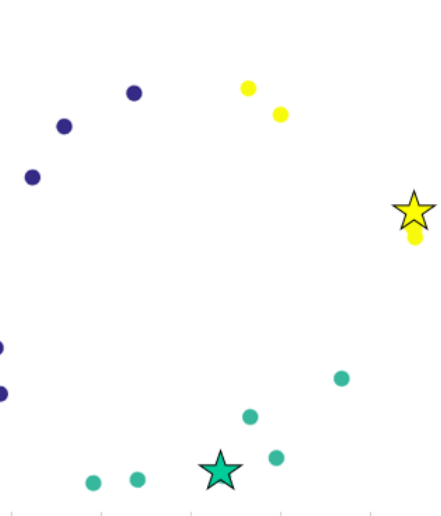


Figure 5.17: Clustering result without the ROI selection (20 models).

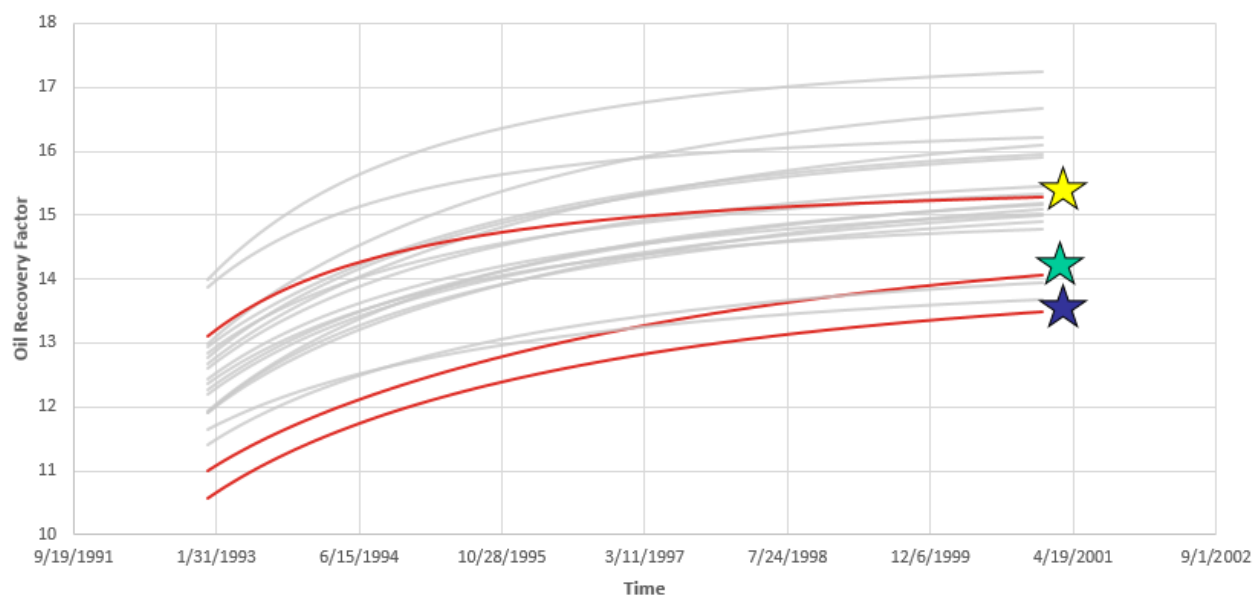


Figure 5.18: Representation of cluster centers on the actual simulation plots (20 models and without the ROI selection).

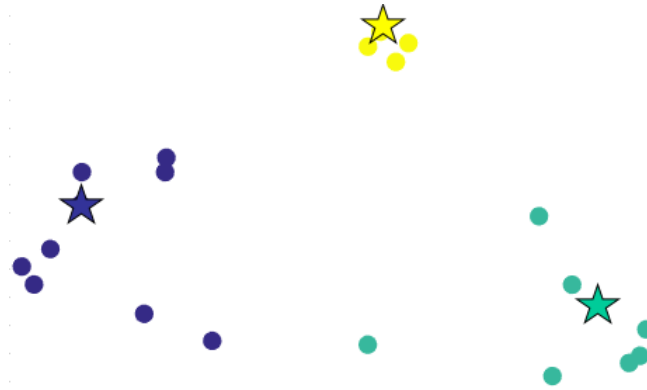


Figure 5.19: Clustering result with the ROI Selection (20 models).

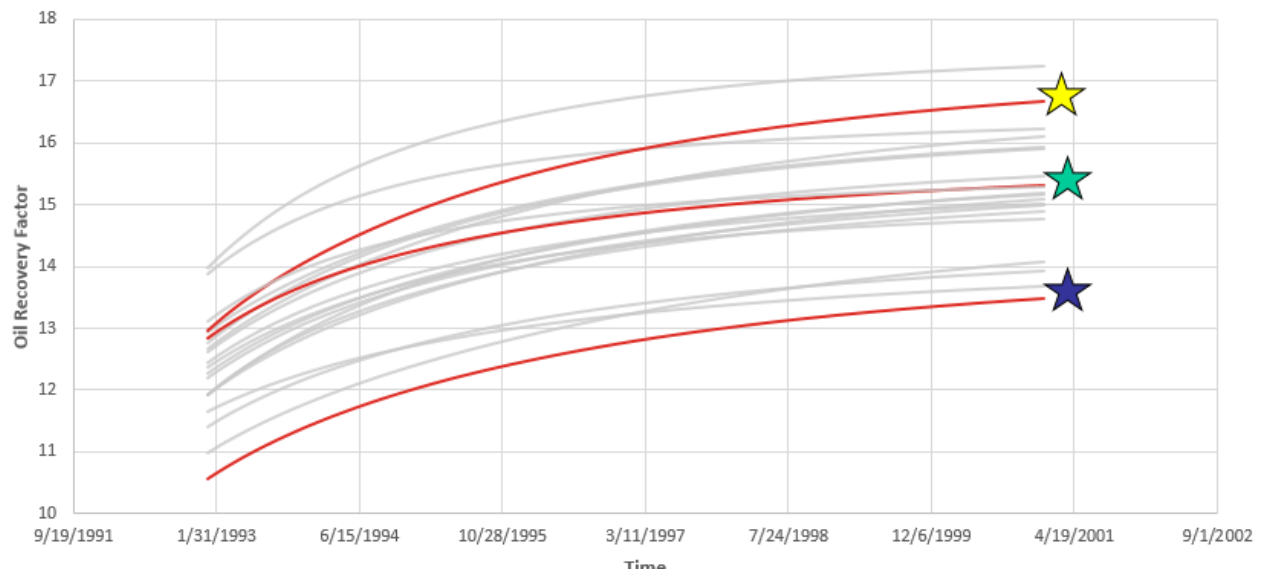


Figure 5.20: Representation of cluster centers on the actual simulation plots (with the ROI selection and 20 models).

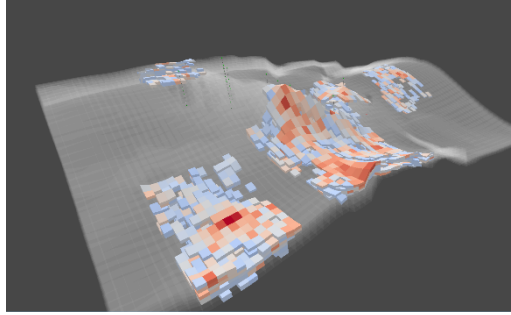


Figure 5.21: ROI selection for the reservoir (20 models).

In the second case study, 100 models are generated as a result of changing the variogram of all the available geological properties (porosity, permeability, net-to-gross and water saturation). Each realization consists of 60,000 cells. Figure 5.23 shows the flow simulation curves for all 100 models which represent a more diverse range of uncertainty in this case study. Similar to the previous example, clustering results are calculated with (Figure 5.22) and without (Figure 5.24) ROI selection scenarios (ROI is shown on Figure 5.26), and similar results are received. The highlighted cluster centers on the actual simulation plots show that the range of uncertainty is better covered with the ROI selection (Figure 5.25). In addition to that, without ROI selection, many models are detected to be very similar to each other and collocated in the projected view. However, with the ROI selection, the differences of models are better represented. In terms of the time-wise performance, with the ROI selection, the entire process took around 3 minutes, while without the ROI selection, it took around 2 hours to produce the clustering result for 100 models.

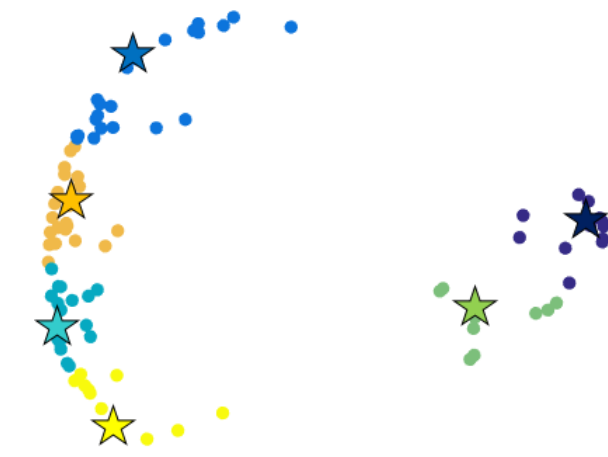


Figure 5.22: Clustering result without the ROI selection (100 models).

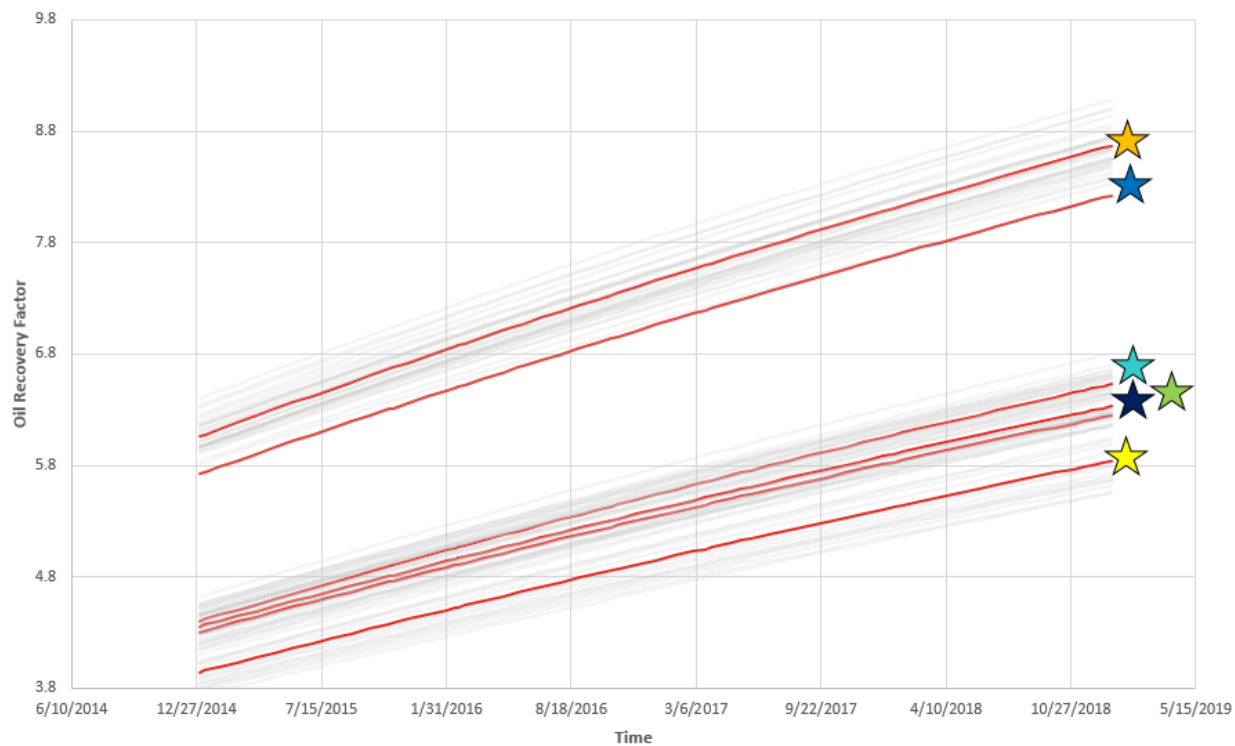


Figure 5.23: Representation of cluster centers on the actual simulation plots (100 models and without the ROI selection).

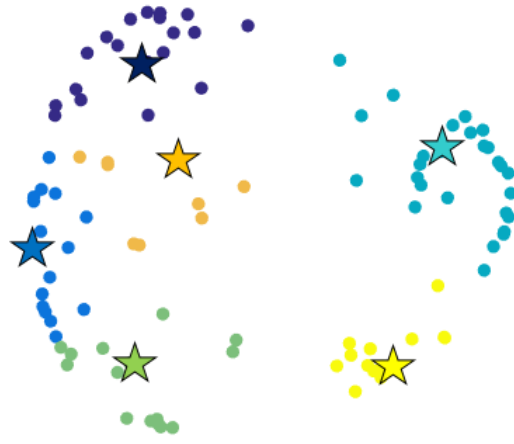


Figure 5.24: Clustering result with the ROI Selection (100 models).

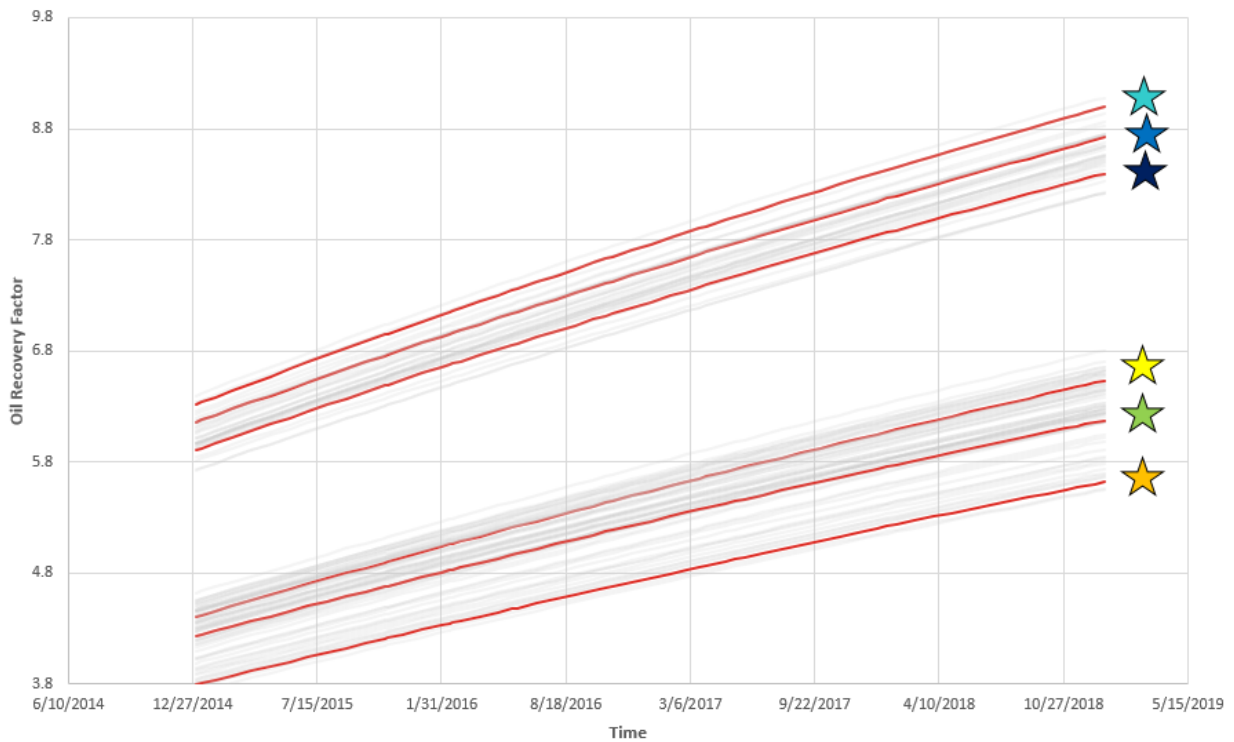


Figure 5.25: Representation of cluster centers on the actual simulation plots (with the ROI selection and 100 models).

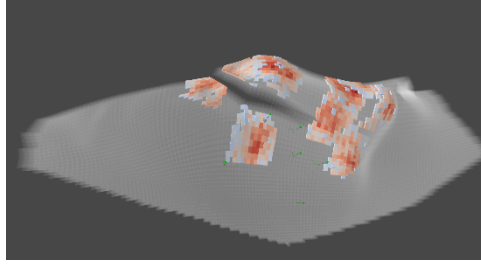


Figure 5.26: ROI selection for the reservoir (100 models).

5.10 User Evaluations (Result of User Studies)

We evaluated our proposed visual interactive analytical framework in a series of formal user evaluations. We conducted 12 sessions with external (industry) and internal (graduate students with experience at the industry) reservoir engineering experts. Each of these evaluation sessions lasted around 90 minutes.

Sessions started with a brief introduction to the goals of the study the subjective and qualitative evaluation of the visual interactive framework for the selection of geological realization and the workflow of our proposed framework and also a brief interview to better delineate the participant's fields of expertise, as well as previous experience within the domain. Two specific tasks were designed for each session. Each task was started with a demo session, in which we introduced the goal of the task followed by the required interactions and steps needed for accomplishing the task. We then invited the participants to try the task out to be familiar with the VR environment, in order to be able to perform the task independently. After this, users were informed to accomplish the specified task. After each task, participants were asked to express their ideas about the task, reflecting the usefulness of task, potential problems and suggestions for improvement. Sessions were recorded, for posterior qualitative analysis. Figure 5.27 shows screenshots from some of our user study evaluation sessions, while users are performing the mentioned tasks. A short video is also prepared to show a summary of user studies .

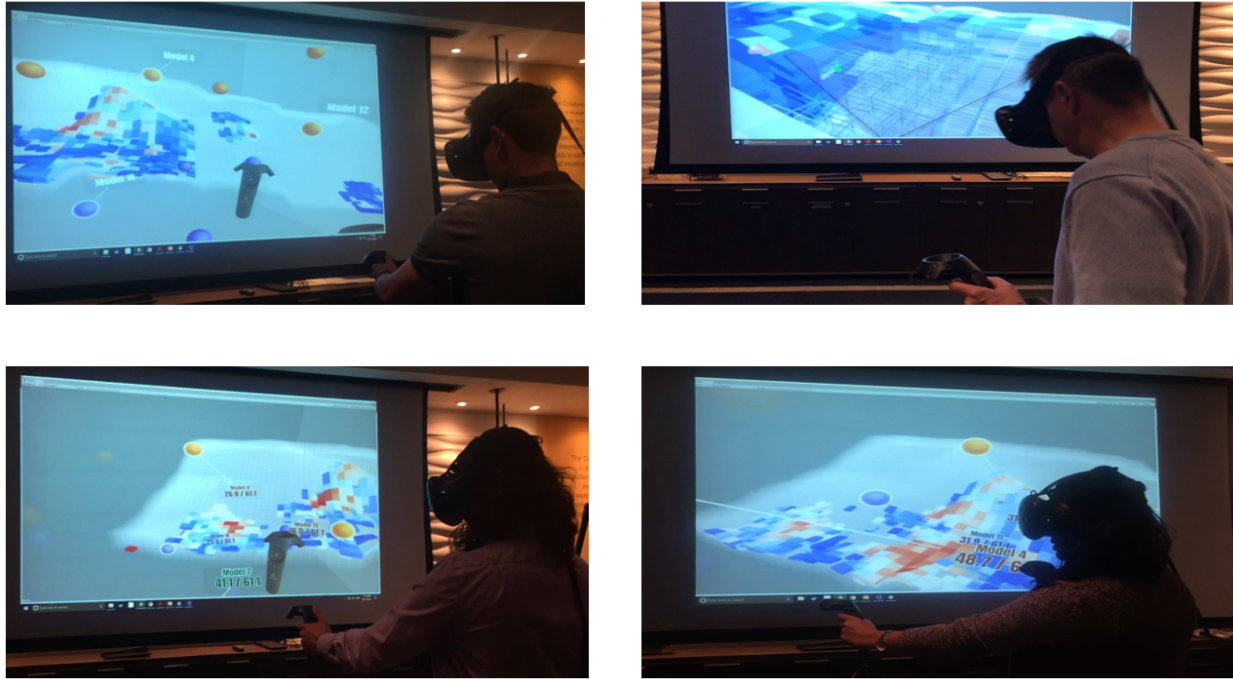


Figure 5.27: Representation of cluster centers on the actual simulation plots (with the ROI selection and 100 models).

5.10.1 Participants

Four females and eight males participated in this study. Participants' ages were ranged from 25 to 50 years old. All participants had at least a bachelor's degree in petroleum related areas. In addition, they were either working in the related industry or having some past industry experience in their field as interns or full time employees.

Although all participants were familiar with the specific reservoir models, they had slightly different backgrounds (such as specializations in reservoir simulation, oil production history matching, drilling engineering, geophysics, and so on [Cosentino, 2001]). We believe that factor did not compound our evaluation but rather contributed in allowing reflections on different point of views, and helped in diversifying and enriching the possible usage perspectives of our proposed framework. Despite this variety, we were still able to perceive many common opinions, which

we collected and analyzed for each of the components in our proposed framework. In the next subsections, we explain each task in detail and discuss some of the results.

5.10.2 ROI Selection Task

In this task, the three proposed selection techniques were introduced to the users: filtering, group selection and single cell selection. Users were requested to select ROI for a reservoir model using the three selection techniques. The main goals of this task were to recognize which regions of a reservoir model are important and to find out the usefulness of the selection techniques. As stated in the previous sections, ROI selection is performed on the variance model, which is calculated based on the realizations in a reservoir ensemble. When inquired about the common ROIs of a reservoir model, we received a list of common opinions, as the following:

- High variance areas. Almost all the participants selected the high variance areas (that are shown in colors that are more reddish). The main reason was that they believed that these areas can contribute more in different simulation results. Some of the participants tended to select all the high variance areas, while some others focused more on the centered areas. Centered areas or areas around wells are believed to be more important, since they have more impact on the oil production results.
- Well areas. The second common selected areas were areas around wells. The reason is that they believe that well areas have the most impact on the oil production results.
- Heterogeneous areas. Other than the two mentioned common areas, some of the participants (6 out of 12) focused on other heterogeneous areas like fractures, faults, hills, aquifers, etc. They believed that the heterogeneous areas can effect on the flow paths and therefore could generate different simulation results.

Among the three selection methods that participants had for the selection of ROI, they preferred the filtering and group selection options. Single cell selection was pointed out to be less usable,

since “*Reservoir models are usually very large, and they contain very thin cells. And that makes single cell selection time consuming and tedious*”. However, few of them believed that single cell selection is helpful for selection of cells in the well paths and other cells in upscaled reservoir models. The improvement suggestions that we received for each section are as follows:

- Filtering. Filtering was the most familiar component in our framework, therefore, not very important improvement comments received for this section, except one: “*sometimes I need to specify a range of variance to show the cells only within that range. It would be good if I can edit the range of variance on the filtering widget.*”
- Group selection. Although it was a new way of selection for many participants, they liked this type of interaction with the box. They found it very smooth, user friendly, and useful for selection of a group of cells. We also received some additional improvement comments, like “*instead of square shape, maybe other form of widgets can be used like circles or hexahedron*” or “*free form sketches on the 2D cross sections can be very useful*”.
- Single cell selection. We designed a dynamic cross sectional view for single cell selection. This dynamic view follows the participants head movements, wherever user moves or looks, that part of the reservoir cuts and user can perform the single selection with the hand controllers. If user has many head movements, that view changes a lot and makes the view confusing. Therefore, one improvement comment that we received was: “*it’s better to fix the view and then do the single cell selection*”. Due to the large number of cells, one of the good comments that we received for faster single cell selection was: “*an idea of brushing can be used for single cell selection, that user can quickly sweep several cells and perform the selection*”.

5.10.3 Analysis Task

In this task, mapping and clustering are performed for a specific ROI. Both the ROI and the clustering results are then represented to the users in a single view. Cluster centers are selected by default in the initial view and the ROI shows the variance calculation only for the centers. Users were then inquired to make decisions about the selection of other different models for each cluster. For each cluster, they were asked to first find and select the outlier of the cluster (the farthest node from the center of cluster). Then, they have to visually inspect if the selected model is different enough or not. If not, they would inspect the second most outsider, and they continue this process until they find different models in each cluster.

Generally, most of the participants were found the application more enjoyable and useful in this second task, since they believed that they get used to the application environment. Using the distance and visual inspection of models, most of them selected the right different models. Therefore, they found the whole idea of interactive model selection with ROI and distance inspection, important and usable: *“It is very beneficial that I can visually inspect if a different model is selected or not”* or *“the visual inspection of changes is very important, since I can better focus on changes in particular areas. For instance, changes in the surrounding of the reservoir is not important to me, but changes around wells are important”* or *“it’s very good that I can see where the changes happen”*. Some of the common issues and improvement tasks that we received for this task were: *“Interior changes cannot be seen immediately like the changes on the surface”*, *“reduce the amount of information that are shown in inspection of model distance in the node diagram”* and *“size of spheres in the node diagram can be changed according to the distance, that I can inspect the distances easier”*.

5.11 Conclusion

In this paper, we extended our previously proposed visual analytics framework, that is used for selecting a few representative models from an ensemble of geostatistical models that represents

the overall production uncertainty. The framework is extended to include ROI selection concept, and the entire selection process can be applied only on particular areas of the reservoir models. To achieve this purpose, first a variance model is calculated based on the entire ensemble, then three selection methods are proposed for the selection of ROIs: filtering, group selection and single cell selection. Then, models are grouped based on their similarity distances. Center of each group is selected by default, as the suggested models for reservoir simulations. ROI analysis, in the next step, would help users to select more different models. The proposed workflow was first evaluated with one case study to show the importance of ROI selection. In the second round of evaluation, we performed formal user studies with the domain experts. Many improvement ideas were received in the evaluation sessions, and in the future studies, we are going to work on those ideas to enhance our framework and address the weaknesses, such as providing more selection techniques to give users more options for selecting ROIs, designing analytical visual tasks to be able to better see the interior areas of the reservoir models, and providing special views for comparison of clustering results for different ROIs.

5.12 Appendix - User Interactivity

In this paper, we proposed a visual interactive process for analyzing the range of uncertainty in a reservoir ensemble. In this process, users should follow a series of steps (or tasks) to explore the range of uncertainty. User's knowledge is very important to be incorporated in certain steps of the process which helps to improve the process results. On the other hand, some other steps are performed automatically according to the user's input. The specification of tasks are explained in table Table 5.1.

| | Task | User Interactiviy vs Automatic | Task Outcome |
|--|------|--------------------------------|--------------|
|--|------|--------------------------------|--------------|

| | | | |
|---|--|---|--|
| 1 | Import reservoir models into the application | Users can select how many models should be imported to the application, and also they can choose how many geological property should be imported per model. | The requested number of model and geological properties are imported into the application. |
| 2 | Region of interest (ROI) selection | This is an optional step. Users can choose to select important regions to be incorporated in the analysis process. This is a user interactive step, that users can utilize any of the provided selection approaches (filtering, single cell selection, and group selection) to extract few regions of interest. | The region of interest areas are highlighted and represented to the users. |
| 3 | Distance calculation | Users can select block size and geological properties to be used for the distance calculation. Considering the user's input, the rest of required of calculations (block scanning and mutual information) is been done automatically. | All the pairwise distance between the reservoir models are calculated. |

| | | | |
|---|------------|---|--|
| 4 | Clustering | <p>Users can select number of clusters. Then, using the calculated distances in the previous step, clustering algorithm is performed automatically to group the models.</p> | <p>Reservoir models are grouped according to the requested number of clusters. All the models are projected into a lower space (2D or 3D), positioned according to their pairwise distances and finally colored according to the belonged clustering group. Representative models (cluster centers) are shown with a different icon.</p> |
|---|------------|---|--|

| | | | |
|---|---------------------|--|--|
| 5 | Clustering analysis | In this interactive step, users can evaluate the representative models to check if different models have been retrieved or not. To achieve that, they should first interactively select and add/remove models into the set of representative models. Variance (or multi-mutual information) is calculated automatically between the selected representative models and then visualized on the model structure. More high variance areas means more different models, and vice versa. Therefore, users can visually inspect if a heterogeneous set of models were obtained or not | The outcome is a set of different representative models. |
|---|---------------------|--|--|

Table 5.1: Involved tasks in the proposed uncertainty exploration process and the level of user interactivity in each of them.

Chapter 6

VR application for the uncertainty analysis of reservoir ensembles

6.1 Introduction

With an increase in modeling and computational methodologies, software and hardware resources, the aspect of uncertainty has received increased attention in the Earth Sciences (e.g.[Wu et al., 2006] [Brown and Heuvelink, 2007] [Caers, 2011]). In the previous chapter, I explained about the uncertainty that involves in the geological modeling, due to the limited amount of available information about the geological properties of every section of a realistic reservoir. A typical approach for addressing the uncertainty in reservoir models is to first use well logging techniques to obtain some idea about the geological properties of few areas of the reservoir. However, the information about the geological parameters of many areas would still be unknown. Therefore, statistical techniques such as geostatistical algorithms are used to quantify the possible values of geological properties for the unknown areas, which results into the generation of many equiprobable reservoir models, called geo-realizations (considering all the possible combination of geological properties). It should be noted that none of these realizations is exactly the same as the real reservoir itself. Therefore, to investigate the uncertainty in the geological properties and also in the reservoir flow performance, the flow simulation needs to be run for all the realizations. This can be a very time-consuming and wasteful process since many of the realizations could be very similar to each other and provide nearby equal flow performance. Additionally, depending on the size of the realizations, flow simulation can take hours to even months. Therefore, it is very essential to be able to obtain the range of uncertainty with few numbers of reservoir models.

In the previous chapters, I proposed a visual analytical framework for exploring the range of

uncertainty with few numbers of reservoir models. This framework helps to find out the distances between the models, observe which models are (dis)similar and extract the few different ones that are potentially cover the range of uncertainty.

In this chapter, I would explain about the developed visual interactive application that supports the proposed analytical framework. This is an essential requirement since many algorithms and methodologies are constantly evolving every day, however, not many software applications are provided to support those techniques, and allows the general practitioners and analysts access to them. On the other hand, visual analysis of reservoir ensembles is a very challenging task and cannot be performed by the existing domain applications (such as CMG [CMG, 1980], Petrel [Schlumberger, 2017], etc.). The initial reason is that the reservoir ensembles contain a large set of 3D reservoir models, that they all have an identical physical 3D structure, but they are different in terms of geological properties assigned to the cell blocks (smallest building piece of reservoir models), as shown in Figure 6.1. The other reason is that the reservoir models are sometimes different in terms of only small areas, and the differences between models cannot be easily seen visually. Moreover, the existing application can illustrate only one or two reservoir models, and they usually don't have any technique for visualizing ensembles. The last known issue is that many statistical and analytical algorithms have been proposed to summarize the ensembles into numbers, but few of them tried to keep the relationship between the numbers and the actual ensembles. This becomes more serious when 3D ensembles are involved, and it is always needed to illustrate the analytical results on the actual 3D structure of the ensemble members.

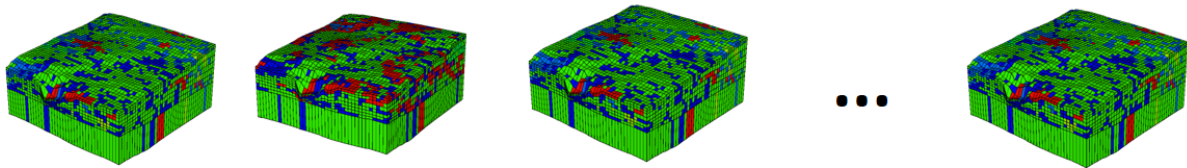


Figure 6.1: An ensemble of 3D reservoir geological models.

Considering all the mentioned requirements, the main purpose of this application is to involve the users (domain experts) in all the steps of uncertainty exploration process and try to provide more enhanced results by utilizing the user's knowledge and experience about the datasets. Different visual interactive components have been provided to the users to explore the spatial similarity and differences between the models, discard similar realizations, extract few different numbers of realizations, and finally validate their findings. More importantly, they can perform all these operations visually and interactively in a Virtual Reality (VR) environment.

6.2 Review of the uncertainty exploration process

An overview of the proposed uncertainty exploration framework is shown in Figure 6.2. The framework process starts with a set of geological realizations. In the next step, the variance is calculated per cell between all the realizations for a user selected geological property, and the outcome is called variance model. This variance model is been used in the next component, which is the Region of Interest (ROI) selection, that helps to discard low variance areas using the provide selection techniques. Next, all the pair-wise distances are calculated between the realization considering only the selected ROIs. Projection techniques are then used to project the models into points in a lower space (2D or 3D) using the calculated distance. The relative distance between the points reflects how much models are similar or different. Clustering techniques are then used in the projected space to group the models based on their distances. Finally, cluster centers are the representative models that are potentially able to cover the range of uncertainty. In addition to that, the user can also add more models into the set of representative models and visually verify if a different and heterogeneous model is been selected or not. This part is handled by calculating the variance between the selected the models and then visualizing the calculations spatially on the ROI. Therefore the user can add more models into the selection set and then visually inspect if the variance is increased (different model is been added) or not (the same model as the other ones in the set is been selected).

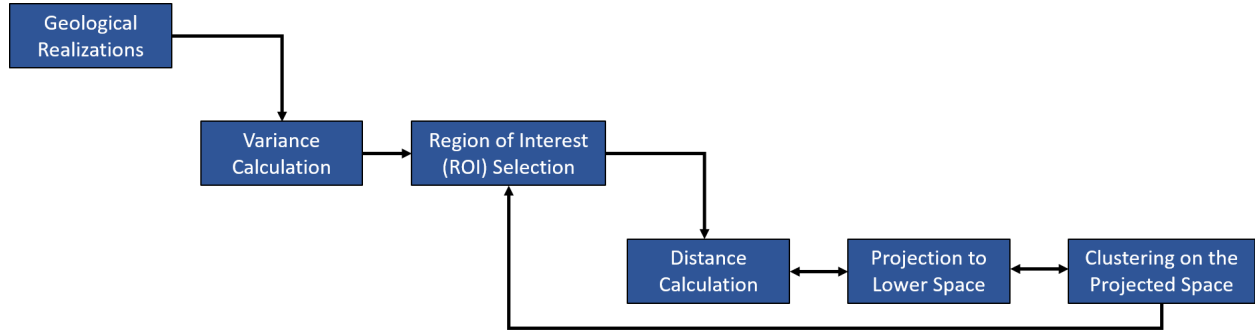


Figure 6.2: An overview of the proposed analytical framework for the uncertainty exploration of reservoir ensembles.

6.3 Application design

The application has been designed according to the proposed analytical framework. Figure 6.3 represents the three main layers of the application with the associated used technologies. The implemented application runs in a VR environment using HTC Vive [HTC, 2017] headset and two tracked hand controllers. The majority of the application has developed with the C# .NET and the VR user interface has built in the unity framework. This application has some sophisticated analytical components (such as mutual information calculation, multi dimensional scaling, clustering, and kernel transformations) which were implemented in MATLAB using the MATLAB compiler where it create .NET dlls from the MATLAB scripts. The .NET dlls were then used with the rest of C# components in the underlying framework. The C# components were then served to the user interface components in the unity framework.

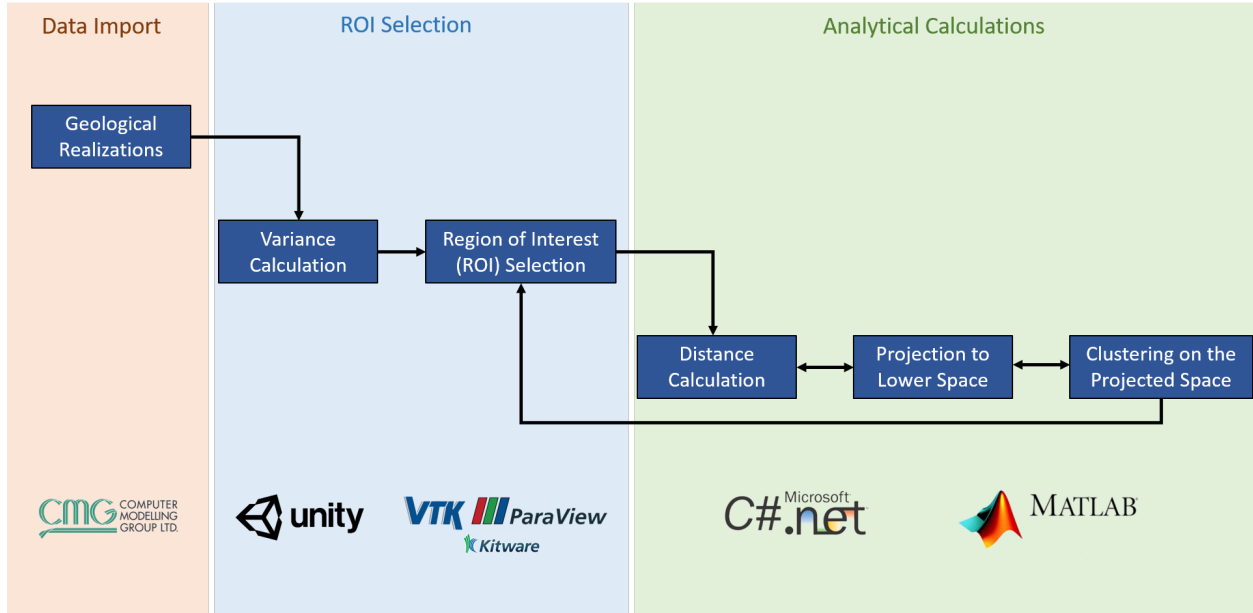


Figure 6.3: An overview of the visual analytics application design with the utilized technologies.

6.3.1 Data Import

In this layer, data, which is the ensemble of 3D reservoir models, is imported into the application. Since all the reservoir models have an identical structure, one structure file is been loaded into the application. However, models are different in terms of the geological property values. Therefore, depending on the number of properties, each model has one or more numbers of property data and they are all entered into the application. The original data format of reservoir ensembles is the CMG format. The structure of CMG file format is shown in Figure 6.4. It contains one single file (.dat) for specification of reservoir 3D structure and one file (.inc) for each property in each ensemble. CMG has a dedicated licensed reader for accessing and reading those files. However, to provide a free and quicker access to the ensemble data, we converted the CMG data format to the vtk format. VTK is a simple and XML based format, which supports random access, parallel I/O, portable data compression. VTK datasets are classified into one of two categories: structured and unstructured. We are using the structured dataset which is a topologically regular array of

cells such as pixel and voxels (e.g., image data) or quadrilateral and hexahedra (e.g., structured grid). All of the VTK XML file types are valid XML documents. The document-level element is `VTKFile`, which contains one element whose name corresponds to the type of dataset the file describes. This is referred as dataset element, which is one of `ImageData`, `RectilinearGrid`, `StructuredGrid`, `PolyData`, or `UnstructuredGrid`. The dataset element contains one or more `Piece` element, each describing a portion of the dataset. Together, the dataset element and `Piece` elements specify the entire dataset.

Each piece of a dataset must specify the geometry (points and cells) of that piece along with the data associated with each point or cell (Figure 6.6). For instance, we are using unstructured grid type for our dataset and it has the following structure:

```
<VTKFile type="UnstructuredGrid">
  <UnstructuredGrid>
    <Piece NumberOfPoints="#" NumberOfCells="#">
      <PointData><!-- Data assigned to each point --></PointData>
      <CellData><!-- Data assigned to each cell --></CellData>
      <Points><!-- Point coordinates --></Points>
      <Cells><!-- Sequence of cells --></Cells>
    </Piece>
  </UnstructuredGrid>
</VTKFile>
```

Every dataset describes the data associated with its points and cells with `PointData` and `CellData` XML elements as follows:

```
<!-- Temperature is the active scalar array -->
<PointData Scalars=Temperature>
  <DataArray Name=Velocity />
  <DataArray Name=Temperature />
  <DataArray Name=Pressure />
```

</PointData>

Figure 6.5 represents a schematic structure of vtk file format. It can be seen that the geometry and property data are all integrated into one file. Although point data is been duplicated in all the files, it is been read only once in the application program.

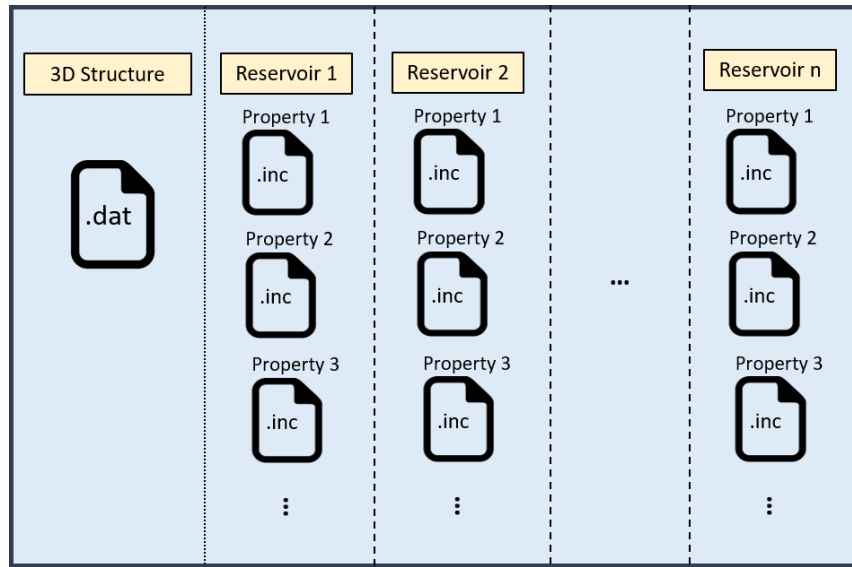


Figure 6.4: Structure of a CMG data file format.

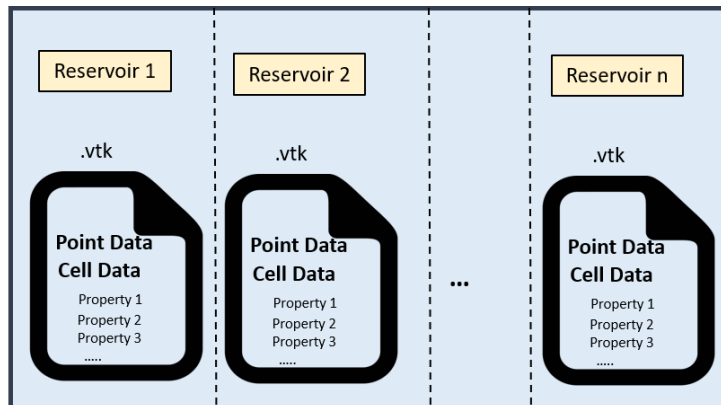


Figure 6.5: Structure of a VTK data file format.

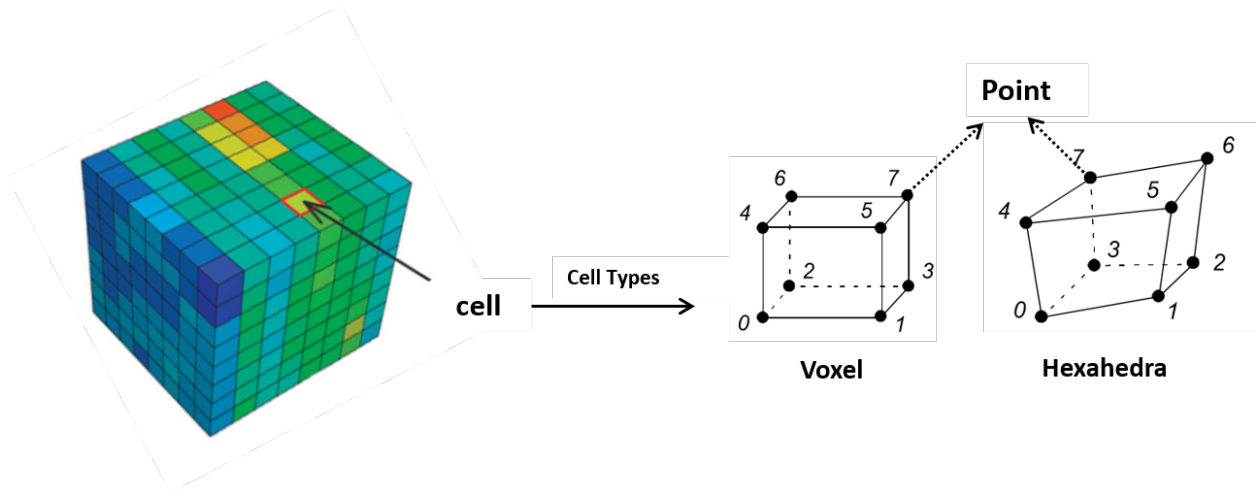


Figure 6.6: Representation of cell (structure and types) and point in the reservoir 3D geometry.

6.3.2 ROI Selection

Once the ensemble data (a set of reservoir models) is imported into the application, the per cell variance is calculated for a user selected geological property and then represented on the reservoir geometry, called variance model. High variance areas are shown with more warm colors and represent the areas with larger uncertainty in the value of a geological property. However low variance areas are shown with more cold colors and represent the areas with more certainty, where the value of properties are either same or similar to each other within the entire ensemble. Therefore, the idea of variance model and ROI selection is to allow users to filter our certain areas and focus more on the uncertain areas. The reason is that the uncertain areas contribute more to the specification of the final uncertainty range and also the differentiation of reservoir models.

Three different selection techniques are proposed and provided to the users to be able to select important regions in the reservoir. Selection techniques are designed at different granularities: filtering, group selection, and single cell selection; and they are all performed on the calculated variance model. Users can utilize any combination of the provided techniques to select some region of interests, to be used in the next component, which is the calculation of pair-wise distances

between the reservoir models.

In a typical process, users may first filter out low variance areas using the filtering method, where they need to specify a variance threshold (cells with less variance than the specified threshold would be removed in the reservoir presentation). Practically, the user can enable the filtering panel by the left controller. The filtering panel is an interactive color bar that shows the range of uncertainty and the assigned colors. To specify the minimum threshold, the user can use the right controller, then slide interactively through the color bar range and immediately observe the result (filtered out cells) on the 3D reservoir (Figure 6.7).

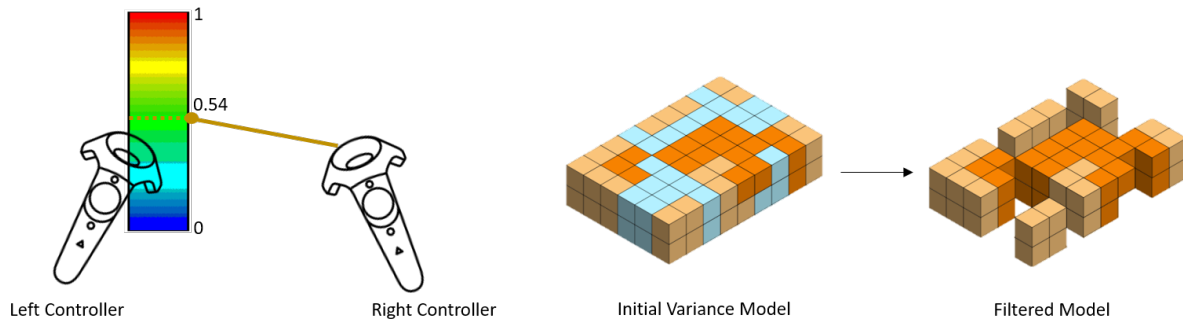


Figure 6.7: Filtering process for removing the low variance cells. (Left) Represents how the controllers are used to perform filtering. (Right) Represents filtering on a sample data. Blue cells are the representative of the low variance cells. For instance, when variance threshold is set to 0.5, such the low variance cells are discarded.

In result of the filtering operation, low variance cells are filtered out, and the user subsequently might see some separate sparse areas. Therefore, following the filtering, the user might carry out group selection to select one more of those areas. Group selection, called also box selection, helps the users to select multiple cells simultaneously using an interactive box widget. Practically, users first locate an important area and then create a 3D box around the spotted area using the controller trigger button. The created box continues to grow as long as the user keeps pressing the trigger button (Figure 6.8 - right). The box location and orientation can be changed by pressing the

trackpad. Once the position, orientation and size of the box are finalized (by releasing the trigger and trackpad buttons (Figure 6.8 - left)), all the cells that are happened to be inside the box are wire framed to represent that they are selected. Multiple boxes can be created with different size and orientations in different areas of the reservoir. And finally, boxes can also be deleted with the following steps: put the controller inside the box, grab the box with keep pressing the trackpad, mimic a “throw away” gesture and then finally release the trackpad.

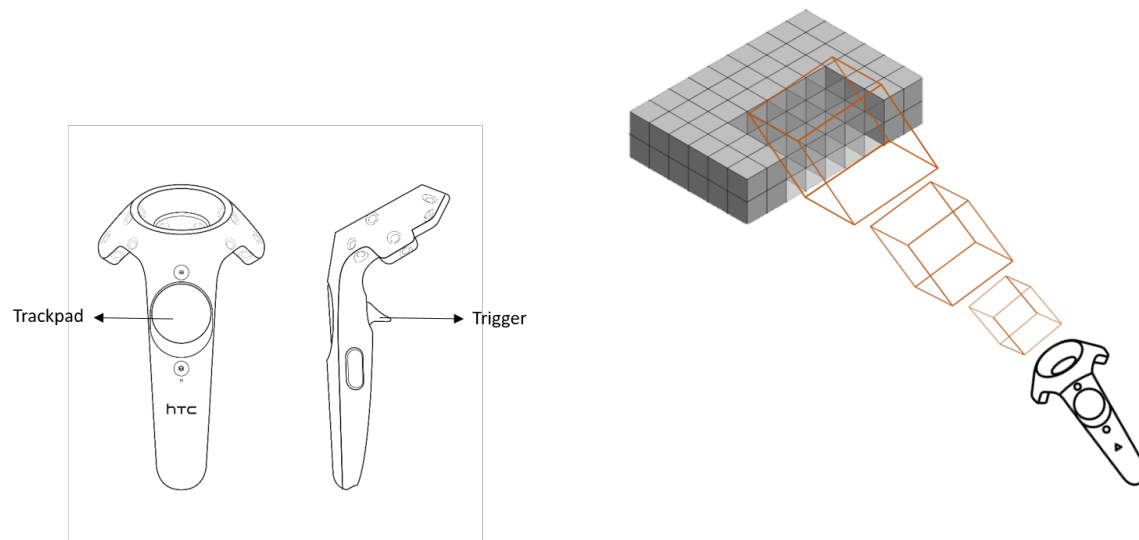


Figure 6.8: Group selection process. (Left) Represents the important controller keys that are used during the group selection process. (Right) Represents the creation of a sample box for selecting a group of cells.

Lastly, the users can select single cells for fine-tuning group selections. To achieve that, a 3D cross-sectional view is provided to users to be able to walk through the reservoir and observe the internal areas. This cross-sectional view helps to have a more detailed view of the reservoir. While having a close view of the reservoir, users can simply target a cell, put the controller inside the cell (cell is hovered with this action), press and release the trigger button and the cell is selected (wire framed) subsequently (Figure 6.9). The same action can be performed for unselecting single cells.

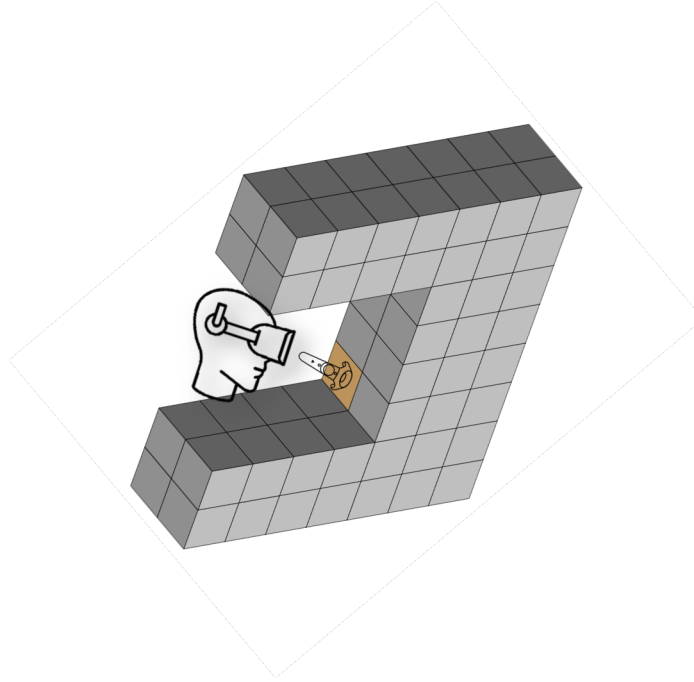


Figure 6.9: Single cell selection process, which shows that users can look inside the model and select any cells by putting the controller inside the cell and pressing the trigger button.

6.3.3 Analytical Calculations

In result of using the proposed selection methods, users select some regions of interest. Those selected areas are then transferred to the analytical component, where used initially to calculate pair-wise distances between all the reservoir models. The calculated distances are then projected to the points in the 3D space, and clustering is performed on the projected 3D points. Each cluster has a group of points and a cluster center which illustrates with a different color than the other clusters (Figure 6.10 - left). Using the controllers, users can hover the 3D points and receive more information about that hovered point such as which model is mapped to this point and also the distance between the hovered point and the cluster centers (Figure 6.10 - right).

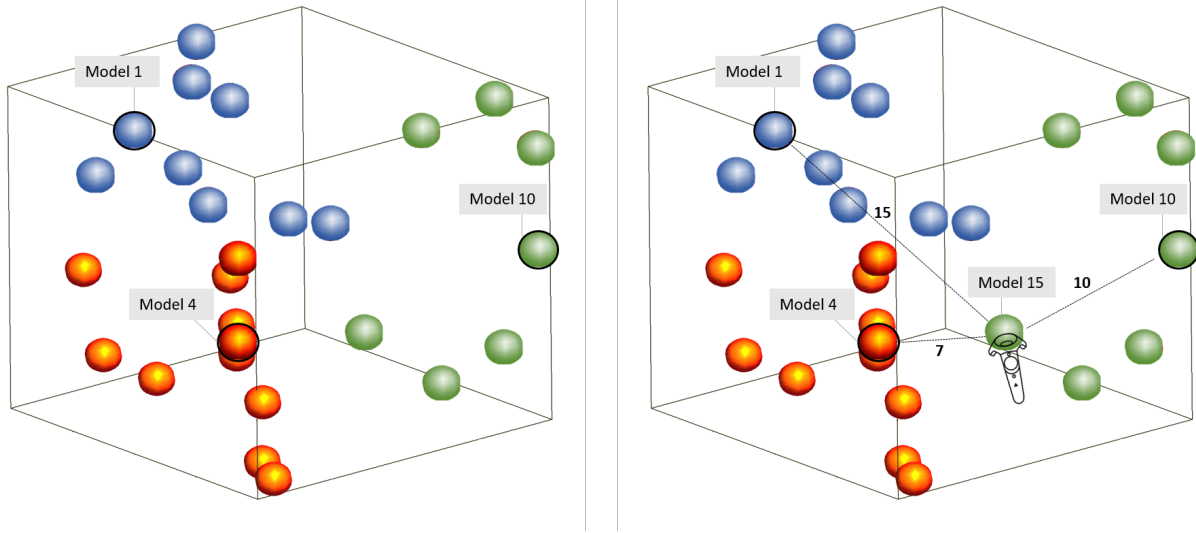


Figure 6.10: Representation of cluster nodes for three clusters. (Left) Highlighted models show the center of clusters, which are the representative models. (Right) Process of hovering and selecting outlier models and inspecting the distances.

As shown in Figure 6.10, each cluster is shown in a different color and cluster centers are initially highlighted with small billboards that depict the information about the projected model. Additionally, selected ROIs are also represented in another view and are color-coded with the calculated variance values between only the cluster centers. According to our proposed analytical framework, cluster centers are the representative model of the entire ensemble and they potentially cover the range of uncertainty. Therefore, users run the flow simulation only for the representative members to obtain the range of uncertainty. However, other than the cluster centers, there might be other different models that contribute to the specification of the range of uncertainty. One of the best candidate ones is the cluster outliers, models that are located very far from the center. These outlier models are potentially very different from the other models and contribute to the better exploration of the uncertainty. To validate this hypothesis, users can perform the following steps: visually inspect the outlier models, hover and select them with the controller, this model is added

to the set of representative models, variance is calculated between all the representative models (including the newly added one), and finally visualized on the previously region of interests. By visually inspecting the variance of the selected models on the region of interests, users can investigate easily if the newly added model increased the variance (a different model is added) or not (the same model is added). Figure 6.11 shows an example of the mentioned analysis process, for a sample ROI that shows initially the variance between the center members, then two models are tried to be added to the representative members. It can be visually and easily seen that when Model A is added, the variance is not changed dramatically, which shows that the Model A is similar the other representative models and it has no benefit to add it to the representative members. However, on the other hand, when Model B is added, the variance is increased a lot (more warm color areas is added) and it shows that Model B is very different than the existent representative models and it could contribute a lot in the specification of uncertainty.

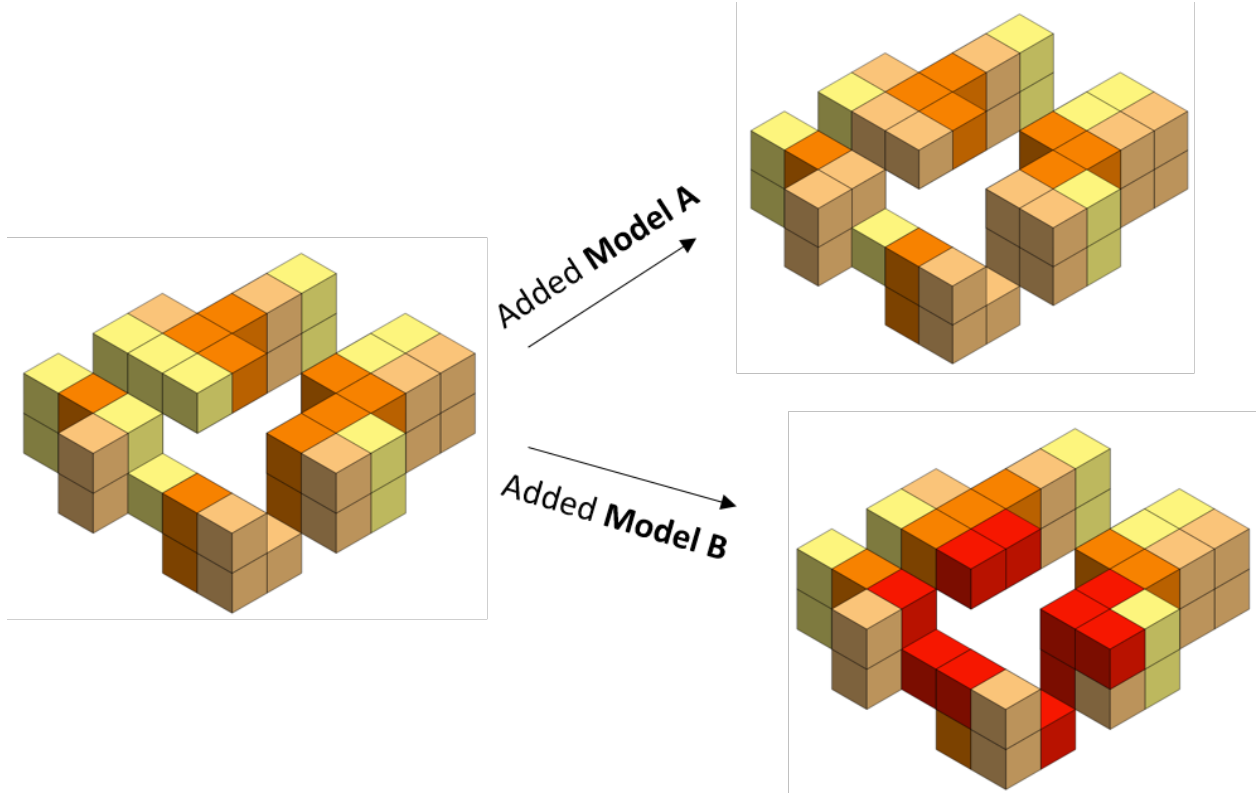


Figure 6.11: ROI analysis process for a sample scenario, which shows how selection of outlier models can change the overall variance of the representative models.

6.4 Evaluation

In order to evaluate the application, we performed formal user studies with twelve participants, who have a rich background, experience or knowledge in reservoir engineering domain. The main purpose of this user study is to assess how well the application can support the proposed analytical framework from the user's perspective in terms of the following items:

- Understandability of the proposed uncertainty exploration process within the implemented application.
- Practicability of the VR environment (with the headset and hand controllers), how

easy is the user interactions with these mediums.

- Usefulness of the proposed visual interactive tasks.
- Applicability of the application in comparison with the traditional desktop applications.

The study was performed with 12 individuals including eight males and four females, ages ranging from 24 to 58 years old. All the participants had the experience of working professionally at the reservoir engineering industry, and all of them had at least B. Sc. Degree and some had graduate degrees in one of the related fields of Petroleum Engineering. Each evaluation session lasted about 90 minutes. Participants were initially asked to fill up a pre-questionnaire to disclose some background information: professional and academic titles, age, previous experience with similar reservoir engineering application, level of comfort with new technologies, etc. This was followed by an introduction of the overall uncertainty analysis workflow supported by the developed application.

Following that, we performed two main tasks: selection of ROIs using the three provided selection methods and exploration of the set of representative models. Each task consists of an introduction about the task goal, a demo to perform a simplified version of the task with a different dataset, the task is performed by the participant, and finally, a semi-structured interview to ask about the performed task. The entire session is video and audio recorded for further qualitative analysis.

The first task was about the selection of ROIs for one of our case studies, a reservoir ensemble with 20 reservoirs, and each reservoir has 26000 cells and more than 20 wells. Users were allowed to select some important regions using the three provided methods: filtering, group, and single cell selection.

Almost all the participants preferred to begin selection with filtering because they found it more familiar (according to their experiences) and easy to perform. Next, five out of twelve participants mentioned group selection to be very easy to learn and remember; however, they also stated that

it takes some time to get familiar with the controller buttons. Regarding that, we saw participants performing correct initial actions for creating the box, but after a while, they had some confusions between using the trigger (for creating) and the trackpad (for grabbing) buttons. It seemed that some of the participants preferred to use the trigger button for grabbing (instead of the trackpad buttons). This can be because of the natural feeling that they had with the trigger button, which gives a sense of taking some object, and this can be adjusted in the later versions. This confusion happened for the first or second try of creating the box, and all the participants get familiar with the buttons soon. While very few participants had some little issue of grabbing the box for the deleting, almost all the participants found the “throw away” gesture for deleting the box very intuitive and natural. The reason is that the action completely matches with our real actions for discarding something that is not needed anymore.

Seven participants used single cell selection as the last step in the selection process (the other participants did not see the necessity of using it). Although they found it very useful for the reservoir with larger cell blocks, they had some difficulty in the inspection of the internal areas of the reservoir using the provided cross-sectional views. The reason is that the cross-sectional views are designed to look where the user’s headset is looking at. So, the views are changing fast with the head movements and that makes confusion.

When we asked users for the ideas for improvement, we got the following feedback:

- Flexible changes of the box widget size. During the box selection, eight participants asked if the size of the box can be changed after the creation or not. In that regard, one of the participants suggested the idea of “face extrusion”, which the user can change the size of the box by extruding any of the box faces in any direction that user wants.
- Fixed cross-sectional views. Five participants suggested the idea of fixed cross-sectional views to investigate the internal areas of the reservoir, rather than views that are changing by the head movements (where the user is looking at).

- Faster single cell selection. Some of the users found the single cell selections a bit slow for coarse reservoirs. Therefore, two of the participants suggested faster approaches like brushing ideas for the selection of multiple cells with a quicker action.

For the second task, a pre-processed clustering result was provided to the users for a specific ROI. Then we wanted users to inspect each cluster and try to find outlier models that can be potentially added to the set of representative models. Practically, what they needed to do is to look into the cluster nodes, find the furthest node visually, hover and select it, and then visually inspect the variance of ROI (to observe the impact of the new model on the overall variance). Generally, all the users performed better in this second task, because they get more familiar with the environment and the hand controllers. From the observations, we noticed that most of the participants (11 out of 12) vastly used the two controllers simultaneously during their activity time (i.e., grabbing the cluster nodes box while looking into the ROI to validate the outlier candidate). These series of actions are well aligned with our design goals.

When we asked users for the ideas for improvement for this task, we got the following feedback:

- Provide more quantitative data in the visual variance analysis whenever new cluster nodes are (un)selected. The participants (nine out of twelve) who have the experience of traditional software applications propose this. For instance, some of the participants suggested a histogram calculation for representing the distribution of cells within the ROI.
- Better inspection of variance changes for the internal areas. Some of the participants suggested the idea of using cross-sectional view at this step for inspection of variance changes for the internal area of the ROIs.
- Quicker inspection of changes in the variance. This requirement is proposed by some of the users for better inspection of changes (when nodes are (un)selected) in

large reservoir ROIs, for instance, a method to automatically draw attention to areas with substantial alternations.

6.5 Conclusion

This chapter presented the design and assessment of a software application in the VR environment for the uncertainty analysis of geological ensembles. I explained about the design of the application, which includes contained components with their rationality, the relationship of the components with the proposed analytical framework, flow of the application, and the utilized technologies in the application. The major goal of this application is to enhance the user's performance and involvement while exploring uncertainty in reservoir ensembles using the previously proposed framework. The developed tool was evaluated within the formal user study sessions with twelve domain experts; soliciting their judgments with regards to usability, usefulness, as well as ideas for improvement of the tool to further facilitate the intended workflow. Feedback was collected during the evaluation through a series of semi-structured interviews. Results of our qualitative evaluation are promising, support our design goals, and have advanced our understanding of the needs and expectations for this application. We plan to extend the capabilities of our tool using some suggestions provided by participants. We also consider supporting multiple users in the same virtual environment as this would be useful for facilitating collaborative work commonly carried out in the oil industry. We hope our work will provide insights into the design and uncovering of gaps or opportunities for future researches, either in petroleum engineering or other domains.

Chapter 7

Conclusion and Future Work

In this thesis, I have presented some important contributions to the visual analytics aspect of uncertainty assessment of 3D ensembles, with a specific case in geoscience and reservoir engineering domain.

In Chapter 2, I proposed a visual analytical framework for exploring uncertainty within a set of 3D reservoir models. The main purpose of this framework was to find few number of models that are different in terms of the reservoir flow behavior and cover the uncertainty range of the whole ensemble. To achieve this goal, first, I proposed a new approach for calculation of pair-wise distances between 3D reservoir models. The calculated distances were then used within the projection techniques (i.e. multidimensional scaling) to map the 3D models into points in lower spaces (such as 2D or 3D), which resulted in a better overall representation of the ensemble uncertainty. Thereafter, clustering algorithms were performed on the projected space to group models based on their similarities. Cluster centers are considered as the representative models of the entire set of models, in a way that they are different in terms of the dynamic behavior and they can cover the range of uncertainty. Therefore, instead of running costly flow simulations for all the models (for finding the uncertainty range), it is sufficient to just run the flow simulations for the representative members and save a considerable amount of time. In chapter 3, I provide a complete case study to showcase the applicability of my proposed visual analytics approach for a conventional oil reservoir, where an ensemble of 3D pixel-based geostatistical realizations were generated that were different in spatial petrophysical properties (i.e porosity and permeability). In this case study, the representative models (the outcome of the proposed framework) was compared with the brute force approach (running flow simulation for all the members), and the results highlighted how the representative models are enough to cover the uncertainty range without running additional flow

simulations.

As a second case study, in chapter 4, I discussed an adapted version of the proposed framework for the heavy oil reservoirs implementing SAGD as a production recovery mechanism. The framework was adapted to only consider a specific area of the reservoir model for calculation of pair-wise distances. The reason for this modification was that steam chamber area is one of the essential areas for the SAGD reservoirs, and considering the entire reservoir into the calculations could lead to misleading results. The adapted framework successfully applied on the SAGD reservoirs and similar to the previous case study, the representative models were evaluated when compared with the actual simulation results, to observe if the range of uncertainty is fully covered or not. Chapter 6 was included to generalize the idea of “important areas” in a way that three visual interactive selection techniques were provided to the users, that they can select any areas of the reservoir model as the regions of interest (ROI). This extended framework was designed and developed in a virtual reality environment, that is explained in detail in Chapter 7. A complete formal user study was performed with the geologists and reservoir engineers to evaluate this extended framework in terms of analytics, interaction, and visualization.

From the extensive research that I performed in this domain and the application of the proposed framework on different case studies and also the feedback that I received during the evaluation user study sessions, I found out several foresaw future work that can be categorized into two directions: complementary and new directions.

The complementary future work are among the short-term plans, and they could largely enhance the proposed framework. I have listed some of the important ones below:

- Many parameters were designed in the framework that user should specify depending on the scenario. However, I received this feedback that the framework should always have a suggestion for the user-defined parameters. I worked on some of these items such as best block size or cluster centers. However, there are some left that can be performed in the future to better improve the framework. These might

include some techniques for selecting the adequate number of clusters, best clustering property, best dimension for reducing the dimensionality or default important ROIs.

- Over extensive evaluations, I realized that the use of RBF kernels can provide better mapping characteristics which resulted in noticeably better clustering outcome [Dhillon et al., 2004] [Scheidt et al., 2009]. I used RBF kernels to map the calculated pair-wise distances into a lower space for better overall representation of models. Additionally, clustering algorithms perform best in linear spaces. However, in the course of evaluating the framework it was noticed that the RBF kernel sometimes could not perform well when the projected points were very scattered and therefore, the kernel could not properly align the points linearly. Therefore, one future direction is to try other non-linear kernels like Laplace kernel, Polynomial kernel, etc [Souza, 2010].
- From the application perspective both desktop and VR environment, I received many constructive feedback that most of them were presented in Chapter 6 and 7 as part of the user study evaluations. The important ones, in terms of visualization, can be summarized as a better representation of the interior area of the reservoir models, illustration of node cluster graph with different node sizes (in a way that size of nodes are varying according to the pair-wise distances), and visualization of distance between multiple models. In addition, there are several substantial future directions in terms of better user interactivity such as using free-form boxes for multiple cells selections, utilizing brushing techniques for single cell selections, and having a relatively fixed cross sectional view for inspecting the interior area of the reservoir models.
- Regarding the usage of the proposed framework, one of the important future work could be to integrate the framework into the simulation applications and create a

closed-loop framework between the uncertainty assessment and simulation. As such the users can utilize the selection framework, select the few representative models, send them to the simulation framework and finally verify if the selected models could well cover the uncertainty range. Otherwise, they can come back to the selection framework and select additional models, and repeat this process as much as they want.

The new direction future work are more complex than the other category and are among the long-term plans. They are somehow considered as the alternative approaches for tackling the same underlying problem (uncertainty assessment in the ensembles). Some of the important ones are listed as follows:

- One of the essential steps of the proposed framework is to project the entire set of models in a lower 2D or 3D space, according to the calculated pair-wise distances. In the lower space, a better overall representation of the entire ensemble is achieved. However, on the other hand, this dimensionality reduction approach causes to lose some information, that is somehow unknown, and it's not clear if some important dimensions (or information) is missed. Therefore, one important future direction is to investigate more about different dimensions of the problem, and try to visualize the additional important dimensions as well. Considering the fact that more than three dimensions achieved, then there would be the challenge of visualizing data in 4D, 5D or higher dimensions, that needs to be carefully researched.
- One of the essential analytical components of the proposed selection framework is the clustering algorithm. And specifically, I focused more on the K-means and kernel K-means algorithms. All these approaches are considered as the unsupervised learning techniques. However, as a future direction, supervised learning approaches [Caruana and Niculescu-Mizil, 2006] can be also investigated and compared with the unsupervised ones for the specific scenarios that are available in this research.

To apply them, it should be further investigated which and how the involved attributes can be learned and used for the classifications. Some of the suitable candidates for the investigation can be Support Vector Machines (SVM), Decision Tree, and Neural Networks.

- Generalization and utilization of the proposed framework for the similar type of ensembles. 3D ensembles can be seen vastly in other domains such as medical, engineering, etc [Bruckner and Moller, 2010] [Wilson and Potter, 2009]. Most of the literature researches focus on the 2D ensembles and less can be seen for the 3D ensembles. Therefore, there are still many challenges in the analysis and visualization of the 3D ensembles. In this thesis research, I could provide a rich framework for the uncertainty assessment of the reservoir ensembles in simulation studies. However, one significant future direction could be to adapt similar framework for other type of ensembles.

Appendix A

Evaluation

A.1 Evaluation of clustering results with quantitative measurements

As shown in all the papers, in order to evaluate clustering results, cluster centers are retrieved and tried to assess if they can cover the range of uncertainty. As a secondary approach, I did a quantitative evaluation to assess the accuracy of clusters. In this approach, expected cluster groups are first obtained from the actual simulation results. In the next step, actual cluster groups are compared to the expected cluster groups and tried to find out how many models are clustered in the right/wrong group. An example of such a quantitative measurement is shown in Figure A.1. In this example, two expected cluster groups are identified and colored in different colors. Red models in the top tier are the representative of the high uncertainty and blue models in the bottom tier are the representative of the low uncertainty. Figure A.1 - (c) illustrates the actual cluster groups and it can be visually seen that few models are clustered in the wrong cluster. In all the case studies that I did in this thesis, the range of accurate correct clusters is between 70% to 90%. However, almost in all the case studies, the range of uncertainty is been retrieved properly.

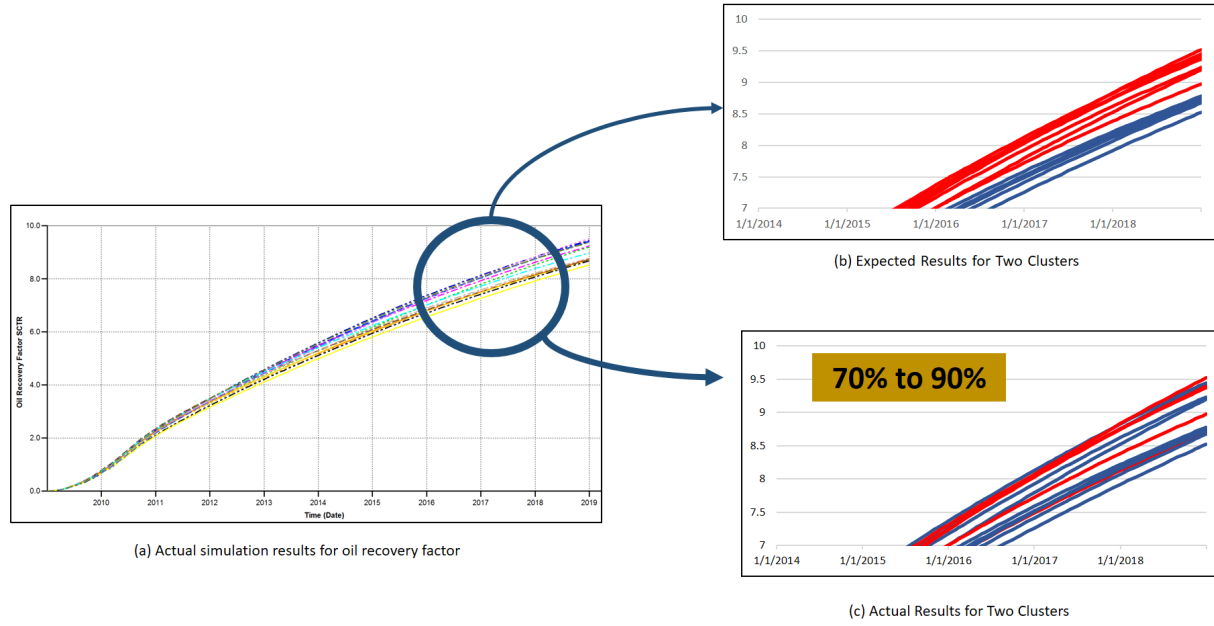


Figure A.1: Quantitative measurements for evaluating clustering results. (a) Actual simulation results for the entire ensemble shown for the oil recovery factor property. (b) Represents the expected clusters in two top and bottom tiers. (c) Illustrates the actual clustering groups, which can be seen that some of the clusters are grouped in the wrong clusters.

Case studies were generated with different specifications and conditions to make sure about the accuracy of results (70% to 90% correct cluster results). On the other hand, the aim of those case studies is to generate almost all the main types of reservoirs. One of the primary case studies that I studied in this thesis initiated with three reservoir realization models that have different facies distribution (A.2). In the next steps, more realizations were generated from each of those three ones, by changing seed in the geostatistical algorithms, leading to an ensemble of realizations. In this particular ensemble, my proposed analytical clustering based framework is expected to group realizations according to their facies differentiation.

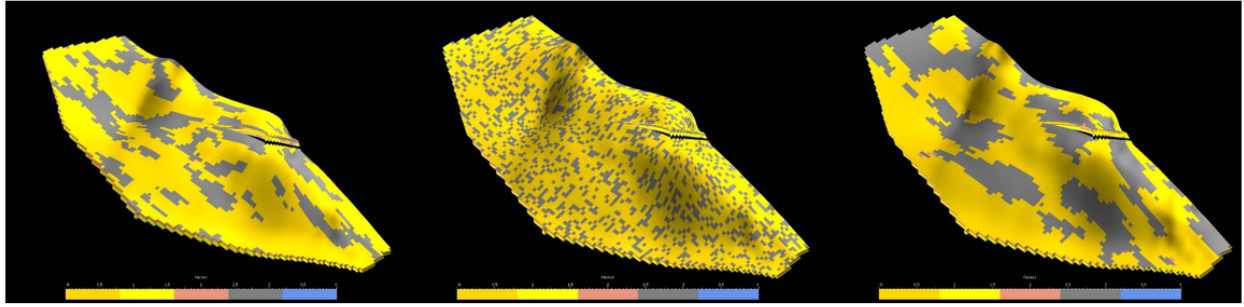


Figure A.2: Different facies distribution that are used to generate different realizations.

Same as the previous case study that initiated with few different facies distributions, more case studies were generated for other important static geological properties, such as porosity and permeability. In these case studies, instead of facies, the ensemble generation process started with one single facies distribution and different porosity or permeability distributions. Other realizations were generated from those major distributions by changing seed in the geostatistical algorithms, leading to an ensemble of realizations that they are different in terms of the porosity and permeability, and the analytical framework is expected to capture that. These case studies are called controlled case studies, that the focus is to create the ensemble of realizations in a very controlled process, that only a few parameters are changed and the goal is to observe the change of those controlled parameters in the reservoir ensembles. FigureA.3 represents a sample controlled case study, that the realizations are only different regarding the control parameter (Porosity), and they are all the same in terms of the other properties (Net to gross and water saturation). In these scenarios, usually, the control parameter (Porosity) is selected as the clustering property, because it contributes more to the differentiation of the models.

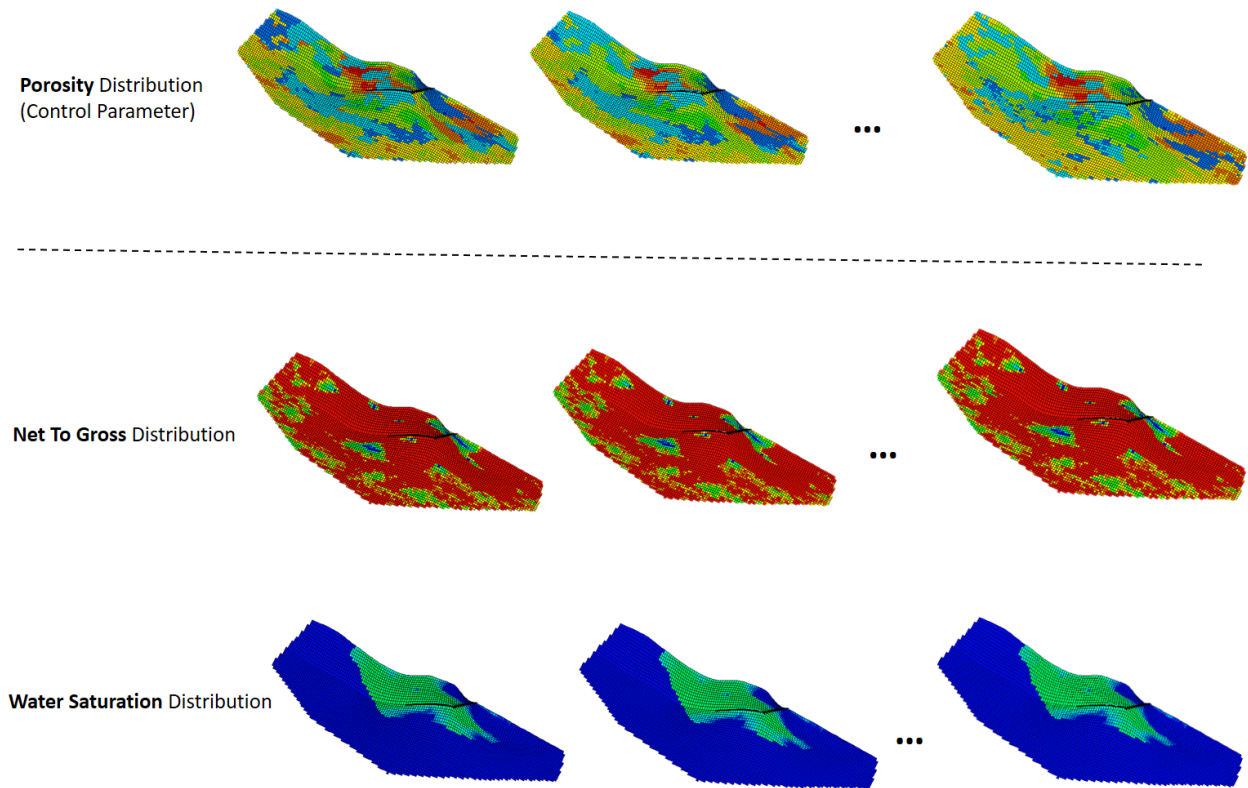


Figure A.3: An example of controlled case study that reservoir models are only different in terms of the controlled parameter.

Although controlled case studies are a good start for the evaluation of particular changes in the ensembles, they are different from the real case studies. Since, unlike the controlled case studies, usually more than one property is changed in the real case studies. An example of such case studies is represented in FigureA.4, wherein realizations are very different from each other without any clear specific pattern regarding all the static properties. I also applied my framework on two different real case studies, one is a lab created case study that more than one property (facies, porosity, and permeability) were changed at the same time in the reservoir ensembles, and the other one is indeed a real case study, Brugge field, which is a public data from a real reservoir in Belgium where almost all the static properties were changed in the realizations. My proposed

analytical framework was applied to all these case studies and generated the cluster groups with the expected accuracy range of 70% to 90%.

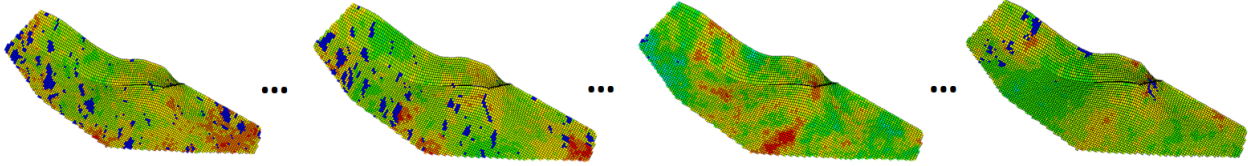


Figure A.4: Porosity distribution in a sample real reservoir ensemble that multiple geological properties are changed at the same time.

The goal of clustering algorithms is to increase the intra-cluster similarity (items within a cluster are similar) and decrease inter-cluster similarity (items from different clusters are dissimilar). *Purity* is one of the well-known and transparent criteria that is used for measuring the quality and accuracy of clusters [Schütze et al., 2008]. Purity is calculated simply by counting the number of correctly assigned items and dividing by the number of items A.1.

$$Purity(C) = \frac{\sum_{i=1}^k T_{c_i}}{N} \quad (A.1)$$

In the above formula, C is the set of clusters, k is the number of clusters, T is the set of correctly classified items which is calculated per cluster, and N is the number of elements. For instance, FigureA.5 represents a sample clustering for three different shapes. Majority class and number of members of the majority for the three clusters are: *cross*, 5 (cluster 1); *circle*, 4 (cluster 2); and *diamond*, 3 (cluster 3). Purity is $(1/17) * (5 + 4 + 3) = 0.71$. Therefore, it can be inferred that the clusters are 70 percent accurate.

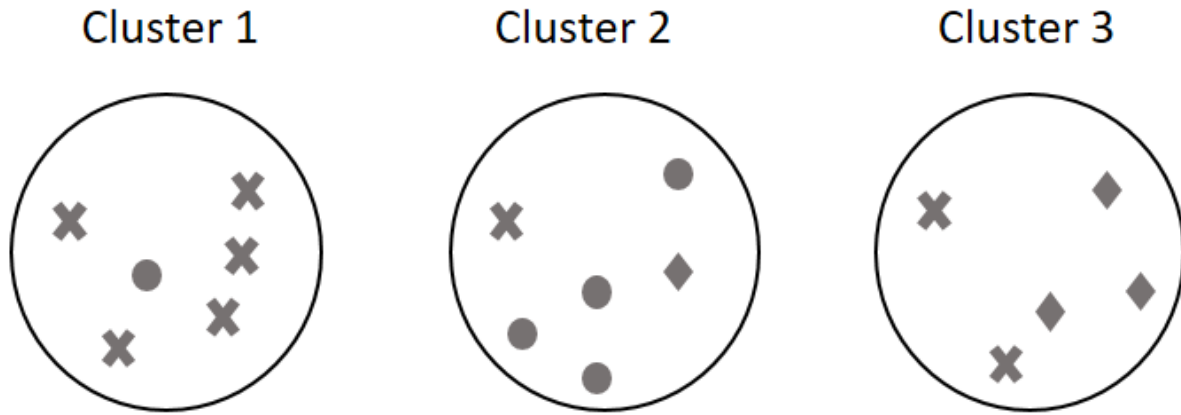


Figure A.5: Sample clustering for three different shapes.

Similar to the mentioned example, I used the purity criteria to calculate the accuracy of clusters. Expected or true clusters are extracted from the result of actual flow simulation results. FigureA.6 represents the actual flow simulation results for one of the studied case studies. Expected clusters are colored differently according to the contained tiers (top, middle, bottom). Those tiers represent the different trends of oil recovery factor within the ensemble. FigureA.7 shows the expected and actual cluster results. Colored numbers are the index of realizations (or models). Actual simulation results show the number of correctly clustered items, according to the majority of items in each cluster. Therefore, the purity of this clustering result is $(4 + 2 + 5) / 15 = 0.73$, which means that the accuracy of clusters are 73 percent.

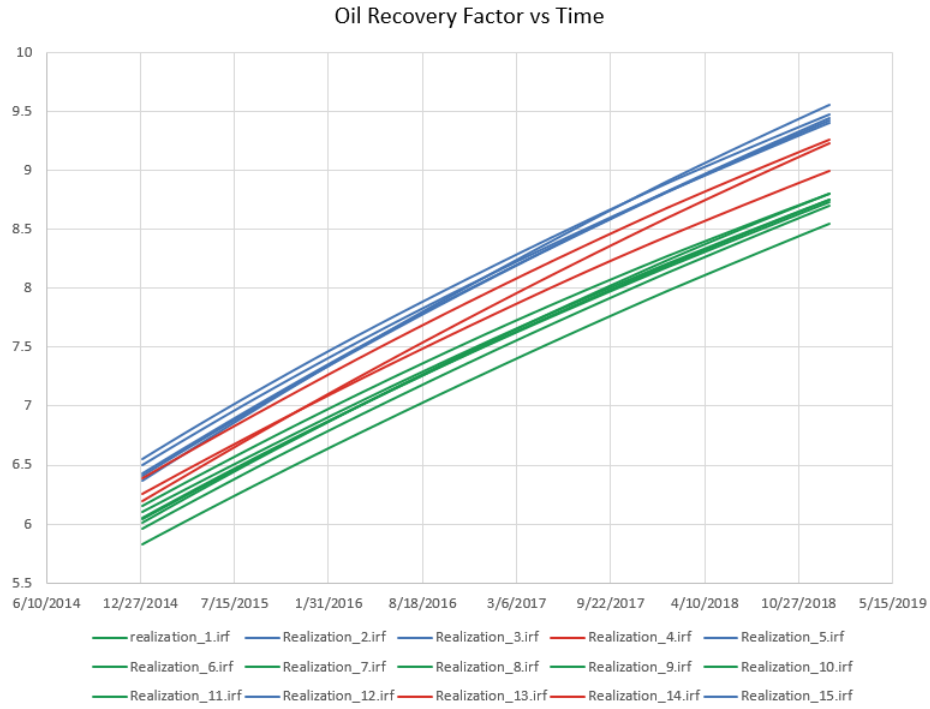


Figure A.6: Representation of expected clusters using the actual flow simulation results.

| Expected Cluster Results | | | Actual Cluster Results | | |
|--------------------------|-------------|-------------|------------------------|-------------|-------------|
| Cluster 1 | Cluster 2 | Cluster 3 | Cluster 1 | Cluster 2 | Cluster 3 |
| Top Tier | Middle Tier | Bottom Tier | Top Tier | Middle Tier | Bottom Tier |
| 2 | 4 | 1 | 2 | 5 | 6 |
| 3 | 13 | 6 | 3 | 13 | 7 |
| 5 | 14 | 7 | 5 | 14 | 8 |
| 12 | | 8 | 12 | 11 | 9 |
| 15 | | 9 | 15 | 1 | 10 |
| | | 10 | 4 | 2 | 5 |
| | | 11 | | | |

← Correctly clustered items

Figure A.7: Comparison of expected and actual cluster results and calculation of the correctly clustered items in each cluster.

Appendix B

Clustering algorithm

B.1 Choice of clustering algorithm

There are two main approaches of clustering available in data mining field including hierarchical and K-means clustering. Hierarchical clustering can be done in bottom-up way or top-bottom. Bottom-up hierarchical clustering that also called agglomerative clustering, starts with each example in its own singleton cluster. At each time-step, greedily two most similar clusters are merged. This process continues until there is a single cluster of all examples. Top-down clustering that also called divisive clustering, starts with all examples in the same cluster. At each time step, the outsiders from the least cohesive clusters are removed. This process stops when each example is in its own singleton cluster (Figure B.1). Agglomerative clustering is more popular and simpler than divisive.

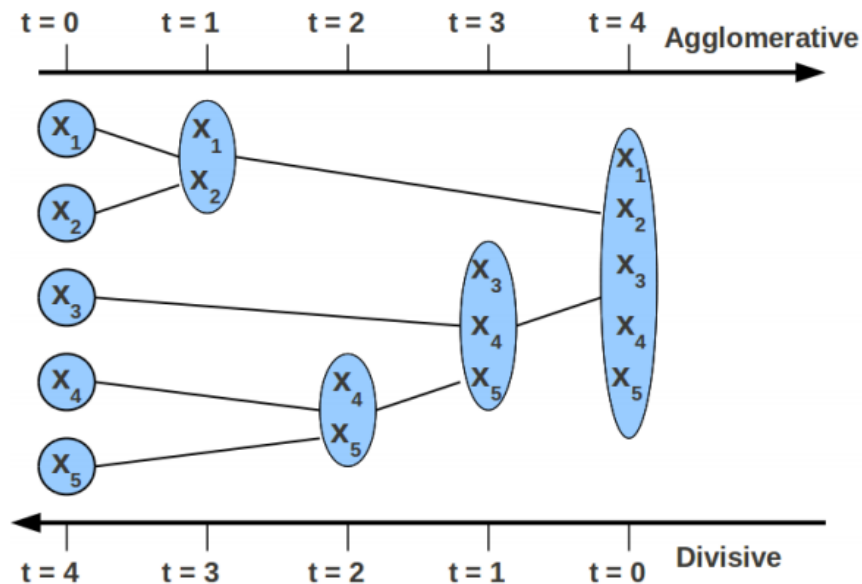


Figure B.1: Representation of hierarchical clustering.

In contrary of hierarchical clustering that it does not need to know the number of clusters during the clustering process, in K-means clustering which is a popular clustering, the number of clusters (K) has to be specified initially. K cluster centers have to be initialized either randomly anywhere in the domain or choose any K examples as the cluster centers. Then each of the examples has to be assigned to its closest centers, and after then new cluster centers are recomputed accordingly. This process repeats until there is no convergence. A possible convergence criteria could be “cluster centers do not change any more”.

I tried both approaches in this thesis research to cluster reservoir models. The accuracy of k-means cluster groups (according to the expected cluster results) is better than hierarchical ones. In addition to that, I found out two other problems with hierarchical clustering including a) it is computationally expensive and it works fine with relatively small data sets, and b) if a mistake happens in early stages of clustering it cannot be corrected. These two problems have been resolved by k-means clustering. K-means is computationally efficient and it can be applied on large data sets, and at each iteration of clustering, it tries to improve the previous iteration results. Therefore, I decided to apply k-means clustering instead of hierarchical clustering for the data sets that are used in this research.

Appendix C

User Study Evaluations

C.1 Consent Form



Name of Researcher, Faculty, Department, Telephone & Email:

Zahra Sahaf, Faculty of Science, Department of Computer Science, [REDACTED]

Supervisor:

Mario Costa Sousa, Department of Computer Science

Title of Project:

Filtering and clustering of geological model ensembles in a visual framework for reservoir simulation studies

Sponsor:

Natural Sciences and Engineering Research Council of Canada (NSERC)

This consent form, a copy of which has been given to you, is only part of the process of informed consent. If you want more details about something mentioned here, or information not included here, you should feel free to ask. Please take the time to read this carefully and to understand any accompanying information.

The University of Calgary Conjoint Faculties Research Ethics Board has approved this research study.

Participation is completely voluntary, anonymous and confidential. You are free to discontinue participation at any time during the study.

Purpose of the Study

We would like to research the benefits of a new visual analytical process for filtering of reservoir geological model ensembles. Within this process, users can cluster geological models based on their similarities. In addition to that, they can observe how models are similar and different to each other.

What Will I Be Asked To Do?

You will be asked to work with our proposed process and tasks, and provide feedback to us.

For the hosted sessions, there will be a 15-minutes introduction to the application and its particular forms of interaction. Participant will be expected to perform some simplified tasks related to the designed process. This might take up to 30 minutes. Afterwards, participants are interviewed using a structured interview format to get their feedbacks. This discussion will take 10 to 15 minutes.

The tasks (and the interviews) will be running in individual setting, and you as a participant can decline to answer any questions.

You might agree to have their photo taken, and/or be videotaped. The photos and/or videos collected from the experiments may be used in publications, presentations in the public domain for academic purposes only. Only your physical likeness will be used in this media. Your name or any other identifying information will be kept confidential.

You may withdraw from the study at any time. If you wish to withdraw, the session will be stopped immediately and you will be thanked. Any possible questions you may have will be answered, and the data collected from you will be destroyed and not be included in the analysis of the study.

What Type of Personal Information Will Be Collected?

To participate in this study, you might agree to have your photo taken and/or videotaped. In this case, you allow the researchers to use your physical likeness in publications or in the public domain, in photo and/or video format.

Please note that, where intended reporting of photographed or videotaped images includes public display, the researchers will have no control over any future use by others who may copy the images and repost them in different formats or contexts, including online.

| | |
|--|--|
| I agree to have my photo taken during this experiment; but not to be used for public display | Yes: <input type="checkbox"/> No: <input type="checkbox"/> |
| I agree to have my photo taken during this experiment; but to be used for public display | Yes: <input type="checkbox"/> No: <input type="checkbox"/> |
| DO NOT agree to have my photo taken during this experiment. | Yes: <input type="checkbox"/> No: <input type="checkbox"/> |
| ----- | |
| I agree to be videotaped taken during this experiment; but not to be used for public display | Yes: <input type="checkbox"/> No: <input type="checkbox"/> |
| I agree to be videotaped taken during this experiment; but to be used for public display | Yes: <input type="checkbox"/> No: <input type="checkbox"/> |
| DO NOT agree to be videotaped taken during this experiment. | Yes: <input type="checkbox"/> No: <input type="checkbox"/> |

Are there Risks or Benefits if I Participate?

There are no foreseeable risks, harms, or inconveniences to you.

What Happens to the Information I Provide?

Participation is completely voluntary. You are free to discontinue participation at any time during the study. Any dialogue between the researcher and you as a participant may be used for presentation and/or publication of results. Pictures taken by the computer or video recorded by the camera may be used for presentation and/or publication of results, only if you granted the acquisition of this material. The anonymous and identifiable data will be stored securely for at least five years on a computer disk, at which time, it will be permanently erased. Should the data prove exceptionally applicable to future research or to inform a PhD project, it may then be stored up to the point of data analysis. The computer will be password protected and only the researchers will have access to your information. Other physical data (such as paper notes and your business cards) will be kept in a safe place.

Please note that if you have provided permission for the public display of video-recorded or photographed images, no meaningful anonymity can be provided and you will be widely recognizable as a study participant.

Signatures

Your signature on this form indicates that 1) you understand to your satisfaction the information provided to you about your participation in this research project, and 2) you agree to participate in the research project.

In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. You are free to withdraw from this research project at any time. You should feel free to ask for clarification or new information throughout your participation.

Participant's Name: (please print) _____

Participant's Signature: _____ Date: _____

Researcher's Name: (please print) _____

Researcher's Signature: _____ Date: _____

Questions/Concerns

If you have any further questions or want clarification regarding this research and/or your participation, please contact:

Zahra Sahaf,
Department of Computer Science,

Roberta Cabral Mota,
Collaboration Center,

Mario Costa Sousa,
Department of Computer Science,

If you have any concerns about the way you've been treated as a participant, please contact the Research Ethics Analyst, Research Services Office, University of Calgary at _____ email _____
_____. A copy of this consent form has been given to you to keep for your records and reference. The investigator has kept a copy of the consent form.

C.2 Pre user study questionnaire

Pre-study Questionnaire

1) Name:

2) Age:

3) Current professional/academic title:

4) Professional background/education and experience (include no. years):

a. Academic training (degrees, specializations):

b. Professional training (with professional positions):

5) Res. engineering software used in the past, for how long, and for which role.

6) How comfortable are you when using new interaction mediums/ new software? –

(Scale 1-5; 1: Not at all comfortable, 5: Very comfortable)

C.3 After user study questionnaire

Please answer the following questions in relation to your experience in the tasks you have just completed. These questions relate to the thoughts and feelings you may have experienced during the event. There are no right or wrong answers. Think about how you felt during the event and answer the questions using the rating scale below. Circle the number that best matches your experience from the options to the right of each question.

| | Strongly disagree | | | | | Strongly agree | | | |
|--|----------------------|---|---|---|---|-------------------|---|--|--|
| 1. It helps me be more effective. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 2. It is easy to use. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 3. I learned to use it quickly. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 4. I am satisfied with it. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 5. It helps me be more productive. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 6. I quickly became skillful with it. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 7. It is user friendly. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 8. It is useful. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 9. I don't notice any inconsistencies as I use it. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 10. It is wonderful. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 11. Using it is effortless. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 12. It is simple to use. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 13. It gives me more control over the activities in my life. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 14. I would recommend it to a friend. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 15. It requires the fewest steps possible to accomplish what I want to do with it. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 16. It is fun to use. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 17. I easily remember how to use it. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 18. I can use it successfully every time. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 19. Both occasional and regular users would like it. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 20. It makes the things I want to accomplish easier to get done. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 21. It is easy to learn to use it. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 22. It works the way I want it to work. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 23. I can use it without written instructions. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 24. It is flexible. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 25. It saves me time when I use it. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 26. It is pleasant to use. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 27. It meets my needs. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 28. I can recover from mistakes quickly and easily. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 29. It does everything I would expect it to do. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 30. I feel I need to have it. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |

Appendix D

Copyright

CONSENT TO PUBLISH and COPYRIGHT TRANSFER

For the mutual benefit and protection of Authors and Publishers, it is necessary that Authors provide formal written Consent to Publish and Transfer of Copyright before publication of the Book. The signed Consent ensures that the publisher has the Author's authorization to publish the Contribution.

Conference: **IVAPP 2018 - International Conference on Information Visualization Theory and Applications.**

Place/Date: **Funchal, Madeira, Portugal; 27 - 29 January, 2018.**

Book Title: .

Edited by: .

Publisher: **SCITEPRESS.**

Paper number: **18.**

Title of the contribution: **A Visual Analytics Framework for Exploring Uncertainties in Reservoir Models.**

Author (**name** and address):

Zahra Sahaf.

Canada.

It is herein agreed that:

The copyright to the contribution identified above is transferred from the Author to the Science and Technology Publications, Lda (SCITEPRESS). The copyright transfer covers the exclusive, sole, permanent, world-wide, transferable, sub licensable and unlimited right to reproduce, publish, transmit and distribute the contribution, including reprints, translations, photographic reproductions, microform, electronic form (offline, online), or any other reproductions of similar nature, including publication in the aforementioned book or any other book. SCITEPRESS is also entitled to carry out editorial changes in the contribution with the sole purpose of enhancing the overall organization and form of the contribution. The Author retains the rights to publish the contribution in his/her own web site or in his/her employer's web site, as long as it is clearly stated that the contribution was presented at IVAPP 2018 and a link to the IVAPP web site is made available there.

The Author warrants that his/her contribution is original, except for such excerpts from copyrighted works as may be included with the permission of the copyright holder and author thereof, that it contains no libelous statements, and does not infringe on any copyright, trademark, patent, statutory right, or propriety right of others. The Author signs for and accepts responsibility for releasing this material on behalf of any and all co-authors.

In return for these rights:

The publisher agrees to have the identified contribution published, at its own cost and expense, in the conference proceedings.

The undersigned hereby gives permission to SCITEPRESS to have the above contribution published.

Date: **06 November, 2017**

Author's Signature:





FACULTY OF SCIENCE

Department of Computer Science
ICT 601
2500 University Drive NW
Calgary, AB, Canada T2N 1N1

From: Mario Costa Sousa and Frank Maurer
Subject: Letter of Permission, Zahra Sahaf's PhD thesis






February 23, 2018

Dear Co-Authors,

On behalf of our PhD candidate, Ms. Zahra Sahaf, we would like your permission to reprint in her manuscript-based dissertation format the contents from the following paper:

Sahaf Z*, Hamdi H, Maurer F, Nghiem L, Costa Sousa M. (2016).
Clustering of Geological Models for Reservoir Simulation Studies in a Visual Analytics Framework.
78th EAGE Conference and Exhibition 2016. Session: Quantifying and Managing Uncertainty in Reservoir Modelling
Vienna, Austria. 2016/05-06
DOI 10.3997/2214-4609.201601143

If this meets with your approval, please sign this letter below and return it to my e-mail [REDACTED].

| | | |
|---|---|---|
|  Zahra Sahaf PhD Candidate Computer Science University of Calgary |  Hamidreza Hamdi Postdoctoral Fellow Computer Science University of Calgary |  Frank Maurer Professor Computer Science University of Calgary |
|  Long Nghiem Vice President Research & Development Computer Modelling Group Ltd. Calgary |  Mario Costa Sousa Associate Professor Computer Science University of Calgary | |



FACULTY OF SCIENCE

Department of Computer Science
ICT 601
2500 University Drive NW
Calgary, AB, Canada T2N 1N4

From: Mario Costa Sousa and Frank Maurer
Subject: Letter of Permission, Zahra Sahaf's PhD thesis

February 23, 2018

Dear Co-Authors,

On behalf of our PhD candidate, Ms. Zahra Sahaf, we would like your permission to reprint in her manuscript-based dissertation format the contents from the following papers:

Sahaf Z*, Hamdi H, Maurer F, Nghiem L., Chen Z., Costa Sousa M. (2017).
Filtering Geological Realizations for SAGD.
79th EAGE Conference and Exhibition 2017. Session: Decision Risk Analysis (EAGE/SPE)
Paris, France. 2017/06
DOI 10.3997/2214-4609.201701313

Sahaf Z*, Hamdi H, Maurer F, Nghiem L., Chen Z., Costa Sousa M. (2017).
Filtering Geological Realizations for SAGD.
Extended version of the EAGE 2017 paper above submitted to *First Break Journal* (<http://fb.eage.org/>)

If this meets with your approval, please sign this letter below and return it to my e-mail [REDACTED].

Zahra Sahaf

PhD Candidate
Computer Science
University of Calgary

Hamidreza Hamdi

Postdoctoral Fellow
Geoscience
University of Calgary

Frank Maurer

Professor
Computer Science
University of Calgary

Long Nghiem

Vice President
Research & Development
Computer Modelling Group Ltd.
Calgary

Zhangxing Chen

Professor
Chemical & Petroleum Engineering
University of Calgary

Mario Costa Sousa

Associate Professor
Computer Science
University of Calgary

Bibliography

- [Agada et al., 2017] Agada, S., Geiger, S., Elsheikh, A., and Oladyskhin, S. (2017). Data-driven surrogates for rapid simulation and optimization of water injection in fractured carbonate reservoirs. *Petroleum Geoscience*, 23(2):270–283.
- [AlbertaEnergy, 2009] AlbertaEnergy (2009). Oil sands.
- [Ancheyta and Speight, 2007] Ancheyta, J. and Speight, J. G. (2007). *Hydroprocessing of heavy oils and residua*. CRC Press.
- [Ballin et al., 1992] Ballin, P., Journel, A., Aziz, K., et al. (1992). Prediction of uncertainty in reservoir performance forecast. *Journal of Canadian Petroleum Technology*, 31(04).
- [Batina et al., 2011] Batina, L., Gierlichs, B., Prouff, E., Rivain, M., Standaert, F.-X., and Veyrat-Charvillon, N. (2011). Mutual information analysis: a comprehensive study. *Journal of Cryptology*, 24(2):269–291.
- [Birol et al., 2010] Birol, F. et al. (2010). World energy outlook 2010. *International Energy Agency*, 1(3).
- [Bordoloi et al., 2004] Bordoloi, U. D., Kao, D. L., and Shen, H.-W. (2004). Visualization techniques for spatial probability density function data. *Data Science Journal*, 3:153–162.
- [Brown and Heuvelink, 2007] Brown, J. D. and Heuvelink, G. B. (2007). The data uncertainty engine (due): A software tool for assessing and simulating uncertain environmental variables. *Computers & Geosciences*, 33(2):172–190.
- [Bruckner and Möller, 2010] Bruckner, S. and Möller, T. (2010). Isosurface similarity maps. In *Computer Graphics Forum*, volume 29, pages 773–782. Wiley Online Library.
- [Bruckner and Moller, 2010] Bruckner, S. and Moller, T. (2010). Result-driven exploration of simulation parameter spaces for visual effects design. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1468–1476.
- [Butler, 1991] Butler, R. M. (1991). Thermal recovery of oil and bitumen.
- [Caers, 2011] Caers, J. (2011). *Modeling Uncertainty in the Earth Sciences*. Wiley Online Library.
- [Cannon, 2018] Cannon, S. (2018). *Reservoir Modelling: A Practical Guide*. John Wiley & Sons.
- [Caruana and Niculescu-Mizil, 2006] Caruana, R. and Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning*, pages 161–168. ACM.

- [CMG, 1980] CMG (1980). Imex, black oil reservoir simulator. <https://www.cmgl.ca/imex>.
- [Cole-Rhodes et al., 2003] Cole-Rhodes, A. A., Johnson, K. L., LeMoigne, J., and Zavorin, I. (2003). Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. *IEEE transactions on image processing*, 12(12):1495–1511.
- [Correa et al., 2009] Correa, C. D., Chan, Y.-H., and Ma, K.-L. (2009). A framework for uncertainty-aware visual analytics. In *IEEE VAST*, pages 51–58.
- [Cosentino, 2001] Cosentino, L. (2001). *Integrated reservoir studies*. Editions Technip.
- [Cover and Thomas, 2012] Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.
- [Crain, 2013] Crain, E. (2013). Welcome to crains petrophysical handbook. *Online Shareware Petrophysics Training and Reference Manual*, url <http://www.spec2000.net>, Accessed.
- [Dake, 1983] Dake, L. (1983). Fundamentals of reservoir engineering., 1978. *United States of America: Elsevier*.
- [Demir et al., 2014] Demir, I., Dick, C., and Westermann, R. (2014). Multi-charts for comparative 3d ensemble visualization. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2694–2703.
- [Deutsch et al., 1996] Deutsch, C. V., Srinivasan, S., et al. (1996). Improved reservoir management through ranking stochastic reservoir models. In *SPE/DOE Improved Oil Recovery Symposium*. Society of Petroleum Engineers.
- [Dhillon et al., 2004] Dhillon, I. S., Guan, Y., and Kulis, B. (2004). Kernel k-means: spectral clustering and normalized cuts. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 551–556. ACM.
- [Fenik et al., 2009] Fenik, D., Nouri, A., Deutsch, C., et al. (2009). Criteria for ranking realizations in the investigation of sagd reservoir performance. In *Canadian International Petroleum Conference*. Petroleum Society of Canada.
- [Fenwick and Batycky, 2011] Fenwick, D. and Batycky, R. (2011). Using metric space methods to analyse reservoir uncertainty. In *Proceedings of the 2011 Gussow Conference*.
- [Fofonov et al., 2016] Fofonov, A., Molchanov, V., and Linsen, L. (2016). Visual analysis of multi-run spatio-temporal simulations using isocontour similarity for projected views. *IEEE transactions on visualization and computer graphics*, 22(8):2037–2050.
- [France and Carroll, 2011] France, S. L. and Carroll, J. D. (2011). Two-way multidimensional scaling: A review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(5):644–661.

- [Frster et al., 2006] Frster, A., Norden, B., Zinck-Jrgensen, K., Frykman, P., Kulenkampff, J., Spangenberg, E., Erzinger, J., Zimmer, M., Kopp, J., Borm, G., et al. (2006). Baseline characterization of the co2sink geological storage site at ketzin, germany. *Environmental Geosciences*, 13(3):145–161.
- [Goovaerts, 1998] Goovaerts, P. (1998). Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. *Biology and Fertility of soils*, 27(4):315–334.
- [Goshtasby, 2012] Goshtasby, A. A. (2012). *Image registration: Principles, tools and methods*. Springer Science & Business Media.
- [Haidacher et al., 2008] Haidacher, M., Bruckner, S., Kanitsar, A., and Gröller, M. E. (2008). Information-based transfer functions for multimodal visualization. In *VCBM*, pages 101–108.
- [Hibbard et al., 2002] Hibbard, B., Böttinger, M., Schultz, M., and Biercamp, J. (2002). Visualization in earth system science. *ACM SIGGRAPH Computer Graphics*, 36(4):5–9.
- [Honarkhah and Caers, 2010] Honarkhah, M. and Caers, J. (2010). Stochastic simulation of patterns using distance-based pattern modeling. *Mathematical Geosciences*, 42(5):487–517.
- [HTC, 2017] HTC (2017). Htc vive. <http://www.vive.com>.
- [Huttenlocher et al., 1993] Huttenlocher, D. P., Klanderman, G. A., and Rucklidge, W. J. (1993). Comparing images using the hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*, 15(9):850–863.
- [Idrobo et al., 2000] Idrobo, E. A., Choudhary, M. K., Datta-Gupta, A., et al. (2000). Swept volume calculations and ranking of geostatistical reservoir models using streamline simulation. In *SPE/AAPG Western Regional Meeting*. Society of Petroleum Engineers.
- [Kao et al., 2002] Kao, D., Luo, A., Dungan, J. L., and Pang, A. (2002). Visualizing spatially varying distribution data. In *Information Visualisation, 2002. Proceedings. Sixth International Conference on*, pages 219–225. IEEE.
- [Kehrer et al., 2010] Kehrer, J., Filzmoser, P., and Hauser, H. (2010). Brushing moments in interactive visual analysis. In *Computer Graphics Forum*, pages 813–822. Wiley Online Library.
- [Kehrer and Hauser, 2013] Kehrer, J. and Hauser, H. (2013). Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE transactions on visualization and computer graphics*, 19(3):495–513.
- [Kehrer et al., 2011] Kehrer, J., Muigg, P., Doleisch, H., and Hauser, H. (2011). Interactive visual analysis of heterogeneous scientific data across an interface. *IEEE Transactions on Visualization and Computer Graphics*, 17(7):934–946.

- [Le Ravalec et al., 2009] Le Ravalec, M., Morlot, C., Marmier, R., and Foulon, D. (2009). Heterogeneity impact on sagd process performance in mobile heavy oil reservoirs. *Oil & Gas Science and Technology-Revue de l'IFP*, 64(4):469–476.
- [Li et al., 2012] Li, S., Deutsch, C. V., and Si, J. (2012). Ranking geostatistical reservoir models with modified connected hydrocarbon volume. In *Ninth International Geostatistics Congress*, pages 11–15.
- [Li and Floudas, 2014] Li, Z. and Floudas, C. A. (2014). Optimal scenario reduction framework based on distance of uncertainty distribution and output performance: I. single reduction via mixed integer linear optimization. *Computers & Chemical Engineering*, 70:50–66.
- [Lin, 1998] Lin, D. (1998). An information-theoretic definition of similarity. In *ICML*, volume 98, pages 296–304. Citeseer.
- [Love et al., 2005] Love, A. L., Pang, A., and Kao, D. L. (2005). Visualizing spatial multivalue data. *IEEE Computer Graphics and Applications*, 25(3):69–79.
- [MacKay, 2003] MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.
- [Matkovic et al., 2005] Matkovic, K., Jelovic, M., Juric, J., Konyha, Z., and Gracanin, D. (2005). *Interactive visual analysis and exploration of injection systems simulations*. IEEE.
- [Meira et al., 2017] Meira, L., Coelho, G., Silva, C., Schiozer, D., Santos, A., et al. (2017). Rmfinder 2.0: An improved interactive multi-criteria scenario reduction methodology. In *SPE Latin America and Caribbean Petroleum Engineering Conference*. Society of Petroleum Engineers.
- [Mela and Louie, 2001] Mela, K. and Louie, J. N. (2001). Correlation length and fractal dimension interpretation from seismic data using variograms and power spectra. *Geophysics*, 66(5):1372–1378.
- [Meyer et al., 2007] Meyer, R. F., Attanasi, E. D., and Freeman, P. A. (2007). Heavy oil and natural bitumen resources in geological basins of the world. Technical report.
- [Nocke et al., 2007] Nocke, T., Flechsig, M., and Bohm, U. (2007). Visual exploration and evaluation of climate-related simulation data. In *2007 Winter Simulation Conference*, pages 703–711. IEEE.
- [Rahim, 2015] Rahim, S. (2015). *Reservoir Geological Uncertainty Reduction and Its Applications in Reservoir Development Optimization*. PhD thesis, University of Alberta.
- [Rahim and Li, 2015] Rahim, S. and Li, Z. (2015). Reservoir geological uncertainty reduction: an optimization-based method using multiple static measures. *Mathematical Geosciences*, 47(4):373–396.

- [Sahaf et al., 2018] Sahaf, Z., Hamdi, H., Maurer, F., and Costa Sousa, M. (2018). A visual analytics framework for exploring uncertainties in reservoir models. In *9th International Conference on Information Visualization Theory and Applications (IVAPP 2018)*.
- [Sahaf et al., 2017] Sahaf, Z., Hamdi, H., Maurer, F., Nghiem, L., Chen, Z., and Sousa, M. C. (2017). Filtering geological realizations for sagd. In *79th EAGE Conference and Exhibition 2017*.
- [Sahaf et al., 2016] Sahaf, Z., Hamdi, H., Maurer, F., Nghiem, L., and Sousa, M. C. (2016). Clustering of geological models for reservoir simulation studies in a visual analytics framework. In *78th EAGE Conference and Exhibition 2016*.
- [Scheidt and Caers, 2008] Scheidt, C. and Caers, J. (2008). Uncertainty quantification using distances and kernel methods—application to a deepwater turbidite reservoir. Technical report, Technical report, Stanford University, Energy resources Engineering Dep.
- [Scheidt and Caers, 2010] Scheidt, C. and Caers, J. (2010). Bootstrap confidence intervals for reservoir model selection techniques. *Computational Geosciences*, 14(2):369–382.
- [Scheidt et al., 2009] Scheidt, C., Caers, J., et al. (2009). Uncertainty quantification in reservoir performance using distances and kernel methods. *SPE Journal*, 14(04):680–692.
- [Schlumberger, 2017] Schlumberger (2017). Petrel software platform. <http://www.software.slb.com/products/petrel>.
- [Schölkopf et al., 1998] Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5):1299–1319.
- [Scholkopf and Smola, 2001] Scholkopf, B. and Smola, A. J. (2001). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.
- [Schütze et al., 2008] Schütze, H., Manning, C. D., and Raghavan, P. (2008). *Introduction to information retrieval*, volume 39. Cambridge University Press.
- [Shawe-Taylor and Cristianini, 2004] Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel methods for pattern analysis*. Cambridge university press.
- [Singh et al., 2014] Singh, A., Maučec, M., Carvajal, G., Mirzadeh, S., Knabe, S., Al-Jasmi, A., and El Din, I. (2014). Uncertainty quantification of forecasted oil recovery using dynamic model ranking with application to a me carbonate reservoir. In *IPTC 2014: International Petroleum Technology Conference*.

- [Smola and Schölkopf, 1998] Smola, A. J. and Schölkopf, B. (1998). *Learning with kernels*. GMD-Forschungszentrum Informationstechnik.
- [Souza, 2010] Souza, C. R. (2010). Kernel functions for machine learning applications. *Creative Commons Attribution-Noncommercial-Share Alike*, 3:29.
- [Sun et al., 2013] Sun, G.-D., Wu, Y.-C., Liang, R.-H., and Liu, S.-X. (2013). A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. *Journal of Computer Science and Technology*, 28(5):852–867.
- [Terry et al., 2013] Terry, R. E., Rogers, J. B., and Craft, B. C. (2013). *Applied petroleum reservoir engineering*. Pearson Education.
- [Wang et al., 2008] Wang, C., Yu, H., and Ma, K.-L. (2008). Importance-driven time-varying data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1547–1554.
- [Wellmann and Regenauer-Lieb, 2012] Wellmann, J. F. and Regenauer-Lieb, K. (2012). Uncertainties have a meaning: Information entropy as a quality measure for 3d geological models. *Tectonophysics*, 526:207–216.
- [Williams, 2002] Williams, C. K. (2002). On a connection between kernel pca and metric multidimensional scaling. *Machine Learning*, 46(1-3):11–19.
- [Wilson and Potter, 2009] Wilson, A. T. and Potter, K. C. (2009). Toward visual analysis of ensemble data sets. In *Proceedings of the 2009 Workshop on Ultrascale Visualization*, pages 48–53. ACM.
- [Wong and Bergeron, 1994] Wong, P. C. and Bergeron, R. D. (1994). 30 years of multidimensional multivariate visualization. In *Scientific Visualization*, pages 3–33.
- [Wu et al., 2006] Wu, J., Jones, B., Li, H., and Loucks, O. L. (2006). *Scaling and uncertainty analysis in ecology*. Springer.
- [Yazdi and Jensen, 2014] Yazdi, M. M. and Jensen, J. L. (2014). Fast screening of geostatistical realizations for sagd reservoir simulation. *Journal of Petroleum Science and Engineering*, 124:264–274.
- [Zhang et al., 2010] Zhang, J., Huang, H., and Wang, J. (2010). Manifold learning for visualization and analyzing high dimensional data. *IEEE*.