UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

GUSTAVO JANDT FELLER

Visualization of Geochemical Simulation Ensembles

Thesis presented in partial fulfillment
of the requirements for the degree of
Master of Computer Science

Advisor: Carla Maria Dal Sasso Freitas

Porto Alegre
December 2018

CIP — CATALOGING-IN-PUBLICATION

Jandt Feller, Gustavo

Visualization of Geochemical Simulation Ensembles / Gustavo Jandt Feller. – Porto Alegre: PPGC da UFRGS, 2018.

72 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2018. Advisor: Carla Maria Dal Sasso Freitas.

1. Visualization, Ensemble, Geochemical Simulations. I. Dal Sasso Freitas, Carla Maria. II. Título.

AGRADECIMENTOS

Gostaria de agradecer primeiramente a minha família, porque sem eles para me dar suporte durante o Mestrado, não seria possível finalizá-lo.

Quero agradecer também a ajuda e paciência da minha orientadora, Carla Maria Dal Sasso Freitas, teve desde minha Graduação até agora.

Também quero agradecer a meus amigos e meus colegas de laboratório, por toda a ajuda com dúvidas técnicas e por estarem disponíveis para espairecer ou simplesmente para conversar.

Finalizando, quero agradecer ao Instituto de Informática da UFRGS por toda a estrutura fornecida para que esse trabalho fosse possível ser realizado.

ABSTRACT

Scientists use simulations to help understanding complex phenomena and processes when they are expensive, difficult or even impossible to reproduce as they occur in the real world. With the of increase of the computational power along the years, scientists were able to simulate more complex and longer phenomena, resulting in increasing volumes of data being produced. Then, they face larger and/or highly complex data sets to analyze. In some situations, scientists want to understand the behavior of their domain of interest in different conditions, and to do so they run multiple simulations with varying parameters. These simulations results constitute ensemble data sets, and analyzing these data sets demands both overview and detailed visual representations, as well as interactive and numerical features for exploration purposes. A specific problem domain that can use simulations ensembles is geochemistry, where scientists often want to study the interaction between water and rocks, which can give an understanding about the conditions oil reservoirs developed along millions of years. The goal of this work is to help geologists and geochemists to explore these complex data through an interactive visualization interface, so they can get insights to answer their questions about the simulated phenomena. We developed Geochemical Simulation Ensembles Visualization (GEVIs), a visualization tool, considering experts' main tasks for exploring these data. The system was evaluated with a case study and experimental use by experts. Results from both evaluations suggest that GEVIs comply with the basic requirements users have regarding visualization of ensembles data sets.

Keywords: Visualization, Ensemble, Geochemical Simulations.

Visualização de Múltiplas Execuções de Simulações Geoquímicas

RESUMO

Cientistas usam simulações para entender fenômenos e processos complexos quando eles são caros, difíceis e mesmo impossíveis de reproduzir da forma como ocorrem no mundo real. Com o aumento dos recursos computacionais ao longo dos anos para realizar computação de alto desempenho, simulações se tornaram mais complexas e longas, resultando em aumento do volume de dados produzidos. Cientistas se deparam com grandes e/ou complexos conjuntos de dados para analisar. Em algumas situações os cientistas querem entender o comportamento de seu domínio de interesse em diferentes condições, e para isso eles executam múltiplas simulações variando parâmetros. Esses resultados de simulações geram múltiplas instâncias de dados, e analisar esses dados demandam tanto representações visuais gerais como detalhadas, assim como facilidades de interação e numéricas de modo que eles possam responder as questões de interesse. Um domínio científico específico que pode se utilizar de múltiplas execuções de simulações é a geoquímica, onde cientistas geralmente estudam a interação entre água e rocha, dando uma compreensão sobre as condições que reservatórios de óleo e gás se desenvolveram durante milhões de anos. O objetivo desse trabalho é ajudar geólogos e geoquímicos através de uma interface de visualização interativa explorar esses dados complexos e, dessa maneira, chegarem à resolução de seus problemas. Foi desenvolvida uma ferramenta denominada Geochemical Simulation Ensembles Visualization (GEVIs) considerando as principais tarefas que usuários especialistas realizam para explorar esses dados. O sistema foi avaliado através de um estudo de caso e de sessões de experimentação livre por usuários especialistas. Os resultados de ambas as avaliações sugerem que GEVIs atende os requisitos básicos de visualização que usuários têm em relação a dados de múltiplas simulações.

Palavras-chave: Visualização, Múltiplas Simulações, Simulações Geoquímicas.

LIST OF ABBREVIATIONS AND ACRONYMS

GEVis	Geochemical Simulation Ensembles Visualization
GUI	Graphical User Interface
nD	n-dimensional
SLR	Systematic Literature Review
SQL	Structured Query Language
SUS	System Usability Scale

LIST OF FIGURES

Figure 2.1 The framework execvuted by a geologist when describing its geochemical model in a computational system. 17

Figure 3.1 DiagenViz UI: left part is the variable selection panel and right part is the visualization panel. 21
Figure 3.2 Example of a three-dimensional plot, showing Albite volume fraction by time and distance. 22
Figure 3.3 Publications per year 28
Figure 3.4 Quantity of papers per research area 29
Figure 3.5 Quantity of papers employing each visualization technique 31
Figure 3.6 Visualization techniques usage in each research area 33
Figure 3.7 Quantity of papers employing each visualization task 35
Figure 3.8 Quantity of papers with each analysis technique 37

Figure 4.1 Header file example 40
Figure 4.2 Data file showing data values from two time steps 41
Figure 4.3 A graphical vision of GEVIs architecture with how the user acts in the system and the processes occuring in the system 42
Figure 4.4 Relational schema for simulation ensembles 43
Figure 4.5 NoSQL (MongoDB) schema for simulation ensembles 44
Figure 4.6 An overview of the window concept of GEVIs 45
Figure 4.7 Temporal and spatial domain visualization techniques. 47
Figure 4.8 Scatter plot matrix showing the mineralization rate of anhydrite, calcite and dolomite of 25 simulations 48

Figure 5.1 Volume fraction of each solid (represented by one chart each) through time for all simulations of the first ensemble 52
Figure 5.2 Volume fraction of each solid (represented by one chart each) through time for all simulations of the first ensemble selecting 25°C and 50°C 52
Figure 5.3 Volume fraction of each solid (represented by one chart each) through time for all simulations of the first ensemble selecting 75°C to 125°C 53
Figure 5.4 Volume fraction of each solid (represented by one chart each) through time for all simulations of the second ensemble 53
Figure 5.5 Comparing the saturation of each solid (represented by one chart each) through time of both ensembles 54

Figure 6.1 Chart showing the answers of all users in the SUS questionnaire 57
Figure 6.2 Chart showing the SUS score of each subject 57
Figure 6.3 Chart showing the answers of all users in the system specific questionnaire 59

LIST OF TABLES

Table 3.1 Research Questions 24
Table 3.2 Data fields extracted 27
Table 3.3 Research areas and specific problems addressed in the papers 28
Table 3.4 Visualization techniques categories 30
Table 3.5 Visualization task categories 34
Table 3.6 Analysis techniques categories 36

Table 5.1 Description of the mineralogy used for all simulations 50
Table 5.2 Water composition used for all simulations of the first ensemble 50
Table 5.3 Water composition used for all simulations of the second ensemble, changing the pH 51

Table 6.1 SUS Score for the Adjective Ratings as in (BANGOR; KORTUM; MILLER, 2009) 58

Table A.1 Selected Studies from the SLR 70
Table A.2 Selected Studies from the SLR 71

Table B.1 SUS questions 72
Table B.2 System specific questions for the user evaluation experiment 72

CONTENTS

1 INTRODUCTION 12
1.1 Motivation 12
1.2 Objectives and Contribution 13
1.3 Organization of the Dissertation 14
2 BACKGROUND 15
2.1 Diagenetic processes 15
2.2 Geochemical Modeling 16
3 RELATED WORKS 19
3.1 Visualizations of Geochemical Simulations 19
3.1.1 DiagenViz 20
3.2 Visualization of Ensemble Simulations 23
3.2.1 Systematic Review 23
3.2.2 Systematic Literature Review Protocol 24
3.2.2.1 Research Questions 24
3.2.2.2 Search method and data sources 25
3.2.2.3 Inclusion and exclusion criteria 26
3.2.2.4 Data Extraction 26
3.2.2.5 Study selection 27
3.2.3 SLR Results 27
3.2.3.1 Evolution of the amount of published articles 28
3.2.3.2 Research areas 28
3.2.3.3 Visualization techniques 30
3.2.3.4 Visualization tasks 33
3.2.3.5 Analysis techniques 35
3.2.3.6 Coordinated multiple views 37
3.2.4 Further Comments 37
4 GEOCHEMICAL SIMULATION ENSEMBLES VISUALIZATION 39
4.1 Data Description and Users' Tasks 39
4.1.1 Data Description 39
4.1.2 Motivating Users' Tasks 41
4.2 GEVIs Architecture 42
4.2.1 Service Layer 42
4.2.2 Application Layer 45
4.2.2.1 Temporal Visualization 46
4.2.2.2 Spatial Domain Visualization 46
4.2.2.3 Multivariate Visualization 48
4.3 Remarks 48
5 CASE STUDY 50
5.1 Hypothetical Ensembles 50
5.2 First Task: Parameter Influence 51
5.3 Second Task: Verify Similarity with Nowadays Conditions 51
5.4 Third Task: Comparing Ensembles 54
6 EVALUATION WITH EXPERT USERS 55
6.1 Evaluation Process Design 55
6.2 Evaluation Results 56
6.2.1 SUS Results 56
6.2.2 System Specific Questionnaire 58
6.3 Final Remarks 59

7 CONCLUSIONS AND FUTURE WORKS 61
REFERENCES 62
APPENDIX A — SELECTED STUDIES FOR THE SLR 70
APPENDIX B — GEVIS EVALUATION QUESTIONNAIRES 72

1 INTRODUCTION

Scientists use simulations to help understanding complex phenomena and processes when they are expensive, difficult or even impossible to reproduce as they occur in the real world. Physical models for experimentation may be too expensive or difficult to reproduce, for example, when they require the construction of large and/or complex prototypes or buildings, or even landscapes. Such models might be even impossible to build due to several conditions: for example, in nature, a geological process takes thousands or millions of years to evolve, which of course is impossible to reproduce in a life-time.

Simulations are among the first scientific applications of computers, and also among the first motivations for the establishment of visualization as an important research area(MCCORMICK; DEFANTI; BROWN, 1987).

With the of increase of computational power along the years, simulations have become more complex and longer, resulting in increasing volumes of data being produced. The analysis of results of many simulation runs is a difficult task and has been motivating research on visualization throughout the years.

A recent scenario related to visualization of simulation results is being targeted by some researchers: the visualization of simulation ensembles. Simulation ensembles are sets of simulations results, each simulation varying from each other in parameters settings, simulation models, or even different algorithms or numerical methods. The complexity of such ensembles is due to the fact they are: (i) multidimensional, (ii) multivariate, (iii) time evolutive and (iv) multivalued (a variable in a cell in a certain time is represented by many values, each value represented by its respective simulation) (WILSON; POTTER, 2009).

1.1 Motivation

In Geochemistry, scientists and researchers often want to study the interaction between water (with solutes) and rocks, but realistic physical experiments might no be possible in a feasible time, because these interactions take hundreds to million years to occur.

Such scenario is typical in the study of diagenetic processes. Diagenesis is defined as the set of chemical, physical and biological changes through which the

sediments pass since their deposition, during and after lithification, and before the metamorphic conditions. Diagenetic processes are controlled by factors such as temperature, pressure, minerals, activity of the ions dissolved in water and organic systems (ROS, 1996).

In order to perform such studies, they can use simulators that help them to understand what happened in the past to have the conditions a field presents in the current days. However, they still lack the understanding of how the processes behave in different conditions, depending on temperature, pressure, and other boundary conditions.

Another problem geochemists face is that different simulators provide different answers to their questions, because of different approaches and different numerical models they may use to give an answer. Thus, comparing these results also would help scientists to better understand the natural processes they are simulating.

This scenario lead to the need of tools for the analysis of simulation ensembles, where a scientist could explore sets of simulation results, comparing results obtained under different conditions. Our research question is then: "Would a set of interactive visualization techniques help researchers in understanding different results in simulation ensembles?"

1.2 Objectives and Contribution

The main objective of this work is to propose and evaluate an interactive visualization solution to help geochemists to analyze an ensemble or compare different ensembles. Although the solution is devised to geochemical simulations, the overall approach can be used for other scenarios, such as climate and weather simulations and heat diffusion simulations, for example. We based our proposal in a systematic literature review of visualization of ensembles data sets.

The main contribution of this work is the visualization solution itself, which allows a user to:

- Customize the visualization of different aspects of the data to help users in their cognitive process of understanding the results of a simulation run, and
- Build connected visualizations through a network of different views to allow the comparison of different simulation ensembles.

A secondary contribution is the systematic literature review we performed prior to the design and development of our tool.

1.3 Organization of the Dissertation

This dissertation is organized as follows:

- Chapter 2 gives a short introduction to the problem domain that motivated our work and where we applied the proof-of-concept prototype.
- Chapter 3 reviews the visualization solutions adopted by the most used simulators, and presents a systematic literature review on ensembles visualization.
- Chapter 4 presents the rationale, design and implementation details about the solution we propose for visualizing geochemical simulations ensembles.
- Chapter 5 describes an hypothetical usage of our tool for analyzing results from simulations performed with hypothetical data by means of the provided visualizations and associated interactive features.
- Chapter 6 presents evaluation sessions conducted with 4 expert users, and discuss our findings from this evalution.
- Chapter 7 concludes our work by discussing contribution and limitations, and draws comments about future work.

2 BACKGROUND

In this section we shortly introduce diagenesis and geochemical modelling, just to provide the context that motivated the work, and where we applied our proof-of-concept prototype.

2.1 Diagenetic processes

As mentioned before, diagenesis is the set of chemical, physical and biological changes through which sediments pass until metamorphic rocks are formed. So, it implies changes that occur since deposition, during and after a process known as lithification, and before the metamorphic conditions.

Diagenetic processes are active, and the sedimentary minerals react to restore equilibrium in an environment where pressure, temperature and chemical composition are changing. The reactions in the system can increase or decrease permeability and porosity (WORDEN; BURLEY, 2003). All these processes correspond to the formation of the present rocks, and they occurred along millions of years. A geologist studying diagenesis usually wants to understand the processes that have occurred during that time, as well as factors that may have influenced the oil quality of a determined region. So, simulations are run to test hypotheses about how an oil reservoir formed in the past, and ultimately, to determine its quality.

Usually, geologists classify the stages of diagenesis as: Eodiagenesis, Mesodiagenesis and Telodiagenesis. During Eodiagenesis, the depositional processes are affected by the proximity of the surface, and then they occur in low temperatures and depths. Mesodiagenesis is the stage in which the sediments and rocks are buried at depths that are not influenced by surface conditions, and thus occur in higher temperatures. At last, Telodiagenesis is the stage where rocks are affected by processes associated to erosion and uplifts (ALI et al., 2010).

Since diagenetic processes are influenced by temperature, pressure, minerals that are present in the environment, activity of the ions dissolved in water and organic systems (ROS, 1996), they can be modelled as geochemical processes.

2.2 Geochemical Modeling

Geochemical modeling typically refers to the process of describing the distribution and reactivity of solutes in a given solution. Geochemical models can be divided into two groups:

- Geochemical Equilibrium Models, which are used under the assumption that thermodynamic equilibrium is reached in a relatively short time, so no time factor is included in the calculation. They take into consideration only equilibrium reactions, and are considered batch models, which are basically closed vessels or reactors.

- Geochemical Kinetic Models, that take into account kinetic reactions (besides equilibrium reactions) and include the time factor.

Geochemical modelling is only useful as a forecasting tool if there is the possibility of validating the results. In real life, this is the goal that most often becomes non-achievable, because of the complexity of natural systems, insufficient field data and uncertainties related to how a system would change along time. A model must be treated as a simplification of reality, and its precision is dependent on how it is capable of estimating the probability of a forecast to be true or false (NORDSTROM, 1992).

The first geochemical models date back to the 70's (WESTALL, 1976)(WOLERY, 1979). Since then, these models have been used to solve complex geochemical problems, such as speciation; determination of minerals' saturation indexes; mixing of different waters; calculation of stoichiometric reactions; interaction between solids, fluids and gaseous phases; calculation of equilibrium/kinetic controlled reactions; reactive transport; and mass-law calculations.

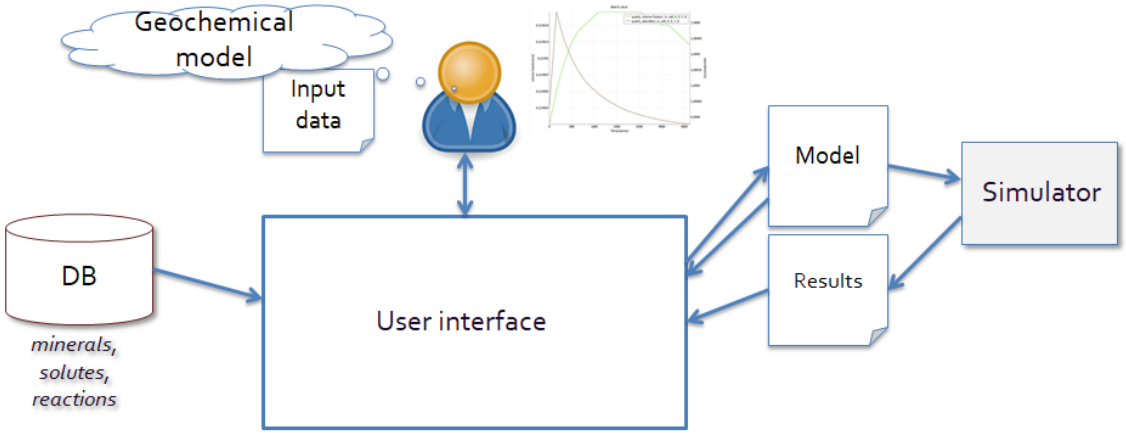
The quality of the results obtained from such a model depend on the methods used, and the thermodynamic data and theoretical concepts applied. Therefore, verifying the results is essential, and it is clear that there will be some differences between the results obtained from different software. Among the enormous variety of software available, some of them are developed for batch-type simulations only, while others have transport capabilities. Batch-type models are those where the spatial domain is modelled as a single cell, where one can optionally have water on top of the sediment. The existence of water allows modelling a *surface* condition, while a

cell with sediment only corresponds to a *subsurface* condition. On the other hand, models with transport conditions are used for simulating one-dimensional (1D), two-dimensional (2D) or three-dimensional (3D) spatial domains. In these models, water flows through the domain cells contributing differently for the water-rock interaction processes.

A general geochemical simulation process is roughly divided into 3 major stages: data input, simulation core and data output.

Data input consists of collecting information related to the geological medium of interest, through chemical analysis made in laboratory and through stratigraphic studies on the sedimentary basin. These data are (i) water composition, (ii) mineral composition, (iii) kinetics and thermodynamics reactions, (iv) burial history (depth of rock formation, estimated time of the occurrence of lithology transformations, pressure, temperature) and (v) spatial domain (batch in an one-cell domain, one-, two-, and three-dimensional domain). Data can be input through a script or filling in a form in a graphical user interface (GUI). Fig. 2.1 has a graphical description of the user interacting with a simulation system.

Figure 2.1: The framework execvuted by a geologist when describing its geochemical model in a computational system.



Data is entered to the simulation core, which starts the simulation execution steps. In this stage, numerical methods are used to solve geochemical equations of fluid-rock interaction in the geological medium defined in the input data. As the simulation process runs, system state is updated for each simulation step, and partial simulation results are generated. This process goes on until the system reaches a steady state or a user-defined maximum simulation time.

Data output is occurs at the end of each simulation step. The data generated by the simulation execution is stored in a file, usually text. Each simulator has its

own standard for input and output files.

Next chapter revises visualizations provided by current, widely used geochemical simulation software.

3 RELATED WORKS

3.1 Visualizations of Geochemical Simulations

We researched in the literature to find works which have visualization of geochemical simulations, but there was none. So we started to research the commercial softwares, and found two well-known simulators for geochemical modeling: Geochemist's Workbench (GWB)[1] and Toughreact [2].

In GWB, the user sets an initial geochemical system to be taken to thermodynamic equilibrium. The software automatically inserts a known volume of water in the system (1 kg). Then, the user sets the amounts of solutes present in that water. GWB starts the calculations and the necessary iterations that lead to a speciation model. [3]

When GWB finishes the simulation, output data is generated. Data contained in the output file are temperature, pressure, pH, ionic strength, water activity, mass of solvent, dissolved solids, solution density and mass of the rock. A list of aqueous species is also output with all solutes present in the model. An important indicator is the "Saturation Index – SI" of the fluid, which informs if: (i) mineral and solution are in equilibrium; (ii) solution is super-saturated; or (iii) solution is under-saturated.

Toughreact can be used in one-, two-, or three-dimensional geological domains in heterogeneous physical and chemical environments, i.e., a wide range of conditions. Input files are provided through a GUI (PetraSim). Firstly, the user selects the solutes that will compose the aqueous phase, and then selects the lithology of interest describing the geological environment. Kinetics and thermodynamic parameters are adjusted after the user builds the interaction model. Once all requirements are satisfied, the software starts the simulation.

Toughreact output data is generated basically to provide plots of the quantity of solute and volume variation versus simulation time. If the user wants to visualize saturation index, Toughreact generates text files that need to be exported as spreadsheets, like EXCEL.

As for visualization, GWB provides the tool named Gtplot that allows users

[1] <http://www.gwb.com/>
[2] <http://www.thunderheadeng.com/petrasim/>
[3] A geochemical speciation modeling software calculates the distribution of dissolved species between free ions and aqueous complexes and also saturation indexes for different minerals.

to display simulation results with 2D visualization techniques, such as line plots, pie charts, color maps, contour plots, vector plots and star plots. However, regarding diagenesis, only line plots are used.

Regarding Thoughreact output, PetraSim provides more visualization techniques like line plots, 3D iso-surface visualization, vector and 3D contour plots.

Both GWB and Thorughreact, however, do not support visualization of more than one simulation run results. If the user wants to compare simulations output, he/she must manually combine the plots generated by each simulation.

3.1.1 DiagenViz

In a previous work, we developed a visualization tool called DiagenViz, for displaying results within the context of a diagenetic process modeling project (FELLER, 2014) (FELLER; KLUNK; FREITAS, 2015).

DiagenViz is implemented in C++, using the Qt Framework [4] for the GUI and two external libraries for visualization: QCustomPlot [5] for the 2D plots, and QwtPlot3D [6] for 3D plots. The GUI is divided into two panels: the visualization panel and the variable selection panel (see Fig. 3.1).

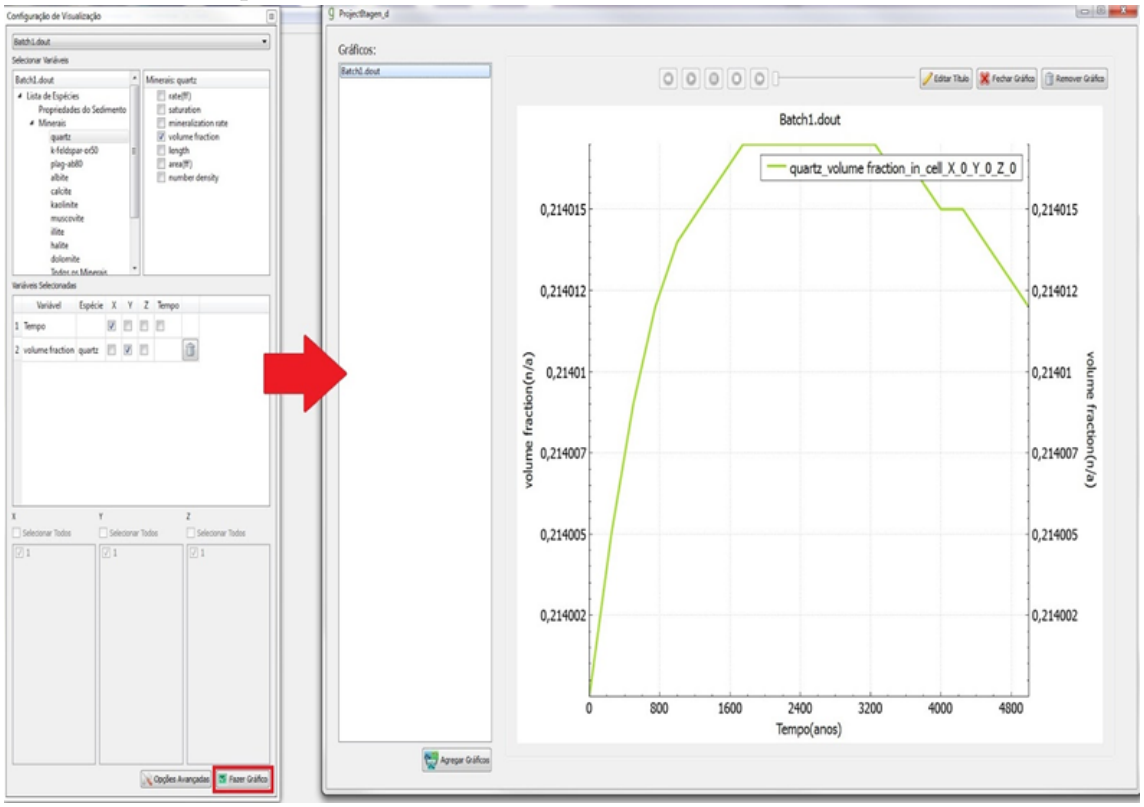
The variable selection panel is divided in three main parts : (A) variable selection itself, (B) axis selection and (C) time and/or cell selection.

In (A), the user selects the variable that he wants to analyze. A tree widget is used to select a species or a variable to be analyzed, depending on what the user wants to focus. In (B), the user selects the variable to be represented in each axis. By default, the plot will have at least one variable (time), but it may also have distances, depending of the dimension of the domain. Also, when the user selects one of the default variables for one of the axis, its list will fade out from (B), because it will be plotted. In (C), the user defines which time step or cells in the domain she/he wants to visualize, and this combines all selected cells and time steps the user has selected.

As an option, after setting (C), the user can define if the data on a specific axis is in linear scale or logarithmic scale. Also, the user can filter the data he/she

[4] <https://www.qt.io/>
[5] <http://www.qcustomplot.com/>
[6] <https://github.com/sintegrial/qwtplot3d>

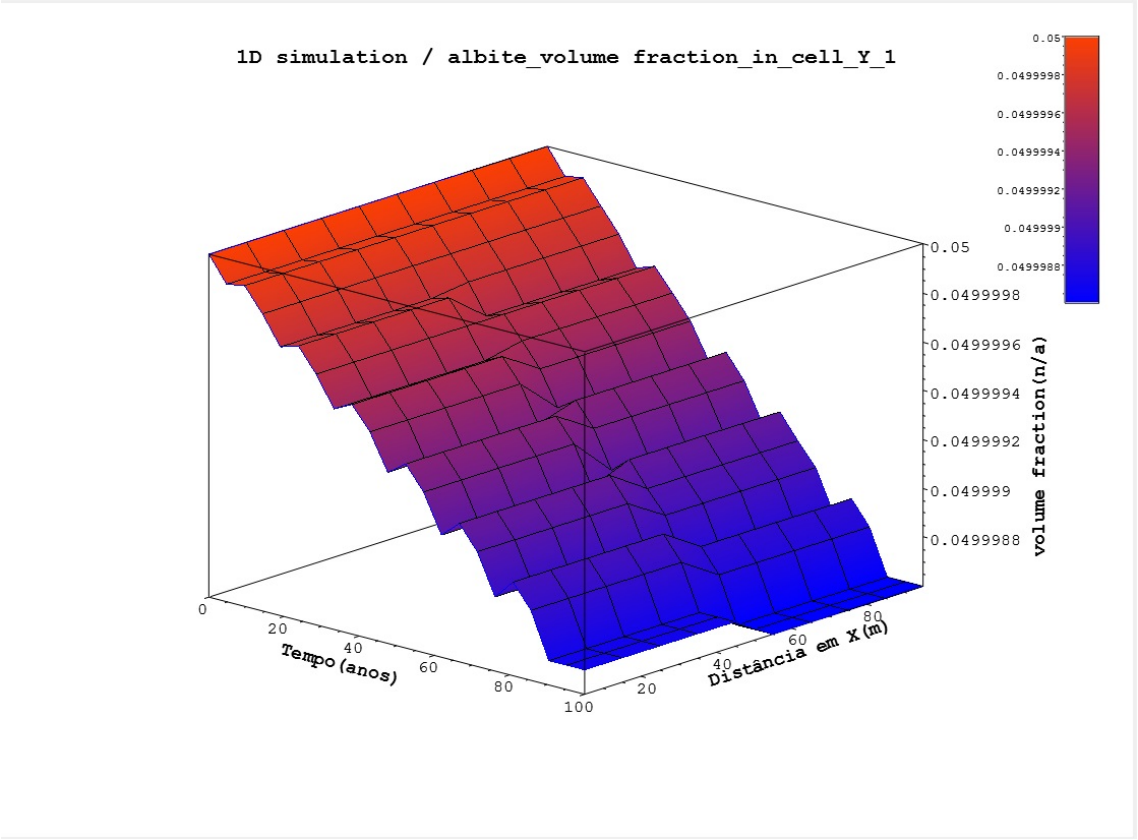
Figure 3.1: DiagenViz UI: left part is the variable selection panel and right part is the visualization panel.



wants to visualize, selecting the interval of values to be plotted.

The visualization panel plots the variables chosen by the user. The visualization techniques provided by the tool were chosen based on user preference. The users of DiagenViz were the geologists that interacted constantly with the simulator development team. The techniques are quite simple in terms of visualization, but they are based on the techniques that geologists were used to: (i) line plots, (ii) line plots with two Y axes, (iii) scatterplots and (iv) 3D surface plots. Line plots are used when one of the variables is continuous, i.e time or distance. Line plots with two Y axes are used when the user selects two different types of variables for the Y axis, e.g saturation and volume fraction, and X axis depicts a continuous variable. Scatterplots are used when the user wants to compare two non-continuous variables, to analyze their relation, e.g concentration of Ca^{++} and Calcite saturation. 3D surface plots (Fig. 3.2) are used when the user wants to analyze the variation of one variable in relation to two others, as for example, Quartz saturation per time and distance. For all the techniques, we also implemented animation to allow display of the plots along the simulation time (the animation controls are at the top of each plot).

Figure 3.2: Example of a three-dimensional plot, showing Albite volume fraction by time and distance.



Two users evaluated DiagenViz based on the comparison between GWB, Petrasim and DiagenViz. One of them (User1) being a young specialist in geochemistry that frequently uses GWB and Petrasim/Toughreact for his research, and the other (User2) a senior specialist in geochemistry that does not use any of these tools, since he has developed his own simulator.

In general, PetraSim got the worst results in most of the aspects. Both users complained about the limitation of the plots in PetraSim, because they could only visualize solute concentration and mineral volume fraction, but not saturation index, which is one of the most important variables for diagenesis studies. Two good points noticed by User2 about PetraSim are the speed and simplicity in plotting the available variables, even it is not complete. Another drawback was catched by both users: PetraSim does not give to users ways to explore data. This may result from the lack of variables to be shown.

In comparing GWB and DiagenViz, both users did not show a consensus as in PetraSim. We noticed that the young specialist preferred DiagenViz rather than GWB, but as for the senior specialist, DiagenViz and GWB were similar, but GWB

was found better than DiagenViz in some concepts.

From this previous work, we kept the idea of allowing the user to interactively choose which variables were to be mapped to which visual dimension, mainly because the simulator for which we developed the proof-of-concept provides hundreds of variables for each simulation step, and is up to the user to choose which one is important to use as a comparison among the members f a ensemble simulation.

3.2 Visualization of Ensemble Simulations

In this section, we review related work on visualization of ensemble simulations. We followed a systematic review protocol, with selected research questions, and obtained an overview of selected articles, which allowed for design decisions in our work.

3.2.1 Systematic Review

A systematic review was conducted with the following objectives:

- To identify which visual representations are used in different scientific domains for the visualization of ensemble data sets.

- To verify which interactive visualization tasks are important.

- To identify which scientific domains use visualization techniques for presenting results from multiple simulation runs.

We followed the guidelines for systematic literature review (SLR) provided by Kitchenham and Charters (KITCHENHAM; CHARTERS, 2007). Although they were published as guidelines for reviews in software engineering, they are general enough for our purposes. We also based our review in the results published by other SLR in visualization for different problems (SHAHIN; LIANG; BABAR, 2014)(NOVAIS et al., 2013)(YUSOFF; SALIM, 2015)(CARROLL et al., 2014).

3.2.2 Systematic Literature Review Protocol

The protocol of SLR dictates five aspects to be defined before searching for articles: (i) research questions, (ii) search method, (iii) inclusion and exclusion criteria, (iv) filtering options and (v) data to be extracted from each article.

3.2.2.1 Research Questions

We have formulated 6 questions, which are shown in Table 3.1 along the motivation for each one.

Table 3.1: Research Questions

Research Question	Motivation
RQ1: What are the visualization techniques implemented for ensemble data sets visualization?	To understand what matters to scientists when they visualize ensembles of simulation results and to get an overview of the visual metaphors used.
RQ2: Which research areas use visualization of ensemble data sets?	To identify which research areas deals mostly with simulation ensembles.
RQ3: What is the relation between visualization techniques and research domains?	To understand what visualization techniques and visual metaphors are used by each research area.
RQ4: What are the visualization tasks used in ensemble data sets visualization?	To investigate how scientists interact with visualization techniques.
RQ5: What kind of analyses are used in ensemble data sets visualization?	To investigate what kinds of algorithms are used in the analysis of an ensemble.
RQ6: Are coordinated multiple views (CMV) used in most of the works?	Observe the potential use of such views in a future implementation of a system for visualizing ensemble data sets.

Our main objective can be achieved by answering three questions (RQ1, RQ2 and RQ3). RQ1 is directed towards knowing the visualization techniques used in all the articles, and provides an overview of what is important to scientists of all research areas. RQ2 looks for the research areas addressed by the articles found, so it can help to identify which areas can be explored in future research. Answers to RQ3 allow to correlate the visualization techniques with the research area to have a more complete analysis of how visualization techniques are used and who use them.

RQ4 can help in understanding how the scientists interact with a visualization system, if they prefer to interact with the visualization itself or through widgets in a graphical user interface. This may eventually help designers of visualization techniques for simulation systems.

RQ5, like RQ4, is a question focused on the user, but it focuses more on the analytical process, i.e., more on the perception side, while RQ4 is more focused on the action side (SACHA et al., 2016). Discovering how a user behaves trying to understand the data, and what algorithms are used in the analysis are also important for designing a visualization tool.

RQ6 also will address how users prefer to observe and interact their results: all data summarized in one visualization technique or through different visualization techniques.

The answers for all these questions will help designers to start understanding how scientists prefer to visualize their simulation results, and how they prefer to interact with the visualization techniques.

3.2.2.2 Search method and data sources

At first, we performed a manual search to collect some of the most relevant works, which allowed us to elaborate the review protocol. Then, an automatic search strategy was used in the following data sources: (i) IEEE Xplore[7], (ii) ACM Digital Library[8], (iii) ScienceDirect[9], (iv) SpringerLink[10] and (v) Wiley Online Library[11]. In all of them, except for SpringerLink, we matched the search terms with title, keywords or abstract, while in SpringerLink we had to match with the full-text. We did not include Google Scholar because, besides it produces low precision (and many irrelevant) results, it has a considerable overlap with the used data sources, creating an unnecessary effort. Different for other sources, in Springer, we could not filter through meta-data, so our query was used in full-text only for this case.

We used the strategies listed in Kitchenham guide to formulate the search query (KITCHENHAM; CHARTERS, 2007). As major terms of the search we identified "visualization", "ensemble" and "simulation". The resulting query was the

[7] <http://ieeexplore.ieee.org/Xplore/home.jsp>
[8] <http://dl.acm.org>
[9] <http://www.sciencedirect.com>
[10] <http://link.springer.com>
[11] <http://onlinelibrary.wiley.com>

following:

*(visualization **OR** visualisation **OR** visual) **AND** ensemble **AND** (simulation **OR** simulations)*

We searched the data sources looking for the literature published from 2009 until January 2017. We started from 2009 because it was in that year that Wilson and Potter study of ensemble data sets defined visualization of multiple runs of simulations as a research problem (WILSON; POTTER, 2009).

3.2.2.3 Inclusion and exclusion criteria

The main purpose of inclusion and exclusion criteria is to select relevant works that help answer questions in a systematic literature review. In our SLR, the inclusion criteria are:

1. Article revised and available in text format.
2. Article introduces a visualization technique or a visualization system to visualize ensemble data sets.

The exclusion criteria are:

1. Editorials, abstracts, posters and tutorials.
2. Not written in English.
3. Visualization of simulations that do not produce ensemble data sets.
4. Duplicated works.

3.2.2.4 Data Extraction

For each paper in a SLR study one needs to extract relevant data. Table 3.2 shows the data items extracted for each article along with their description. D1 to D4 are items for an overview study and are recommended by Kitchenham and Charters (KITCHENHAM; CHARTERS, 2007). D5 to D9 are related to our research questions, with D5 to D8 being categorical data and D9, a Boolean answer, "yes" or "no". D5 was extracted to answer RQ1 and to help answering RQ3; D6 is related to RQ4, D7 with RQ2 and, with D5, we covered RQ3, while D8 is related to RQ5 and D9 to RQ6.

Table 3.2: Data fields extracted

Data item	Description
D1: Author	The author(s) of the paper
D2: Year	The year of paper publication
D3: Title	The title of the paper
D4: Event	The venue or event where the paper was published
D5: Visualization technique(s)	The visualization technique(s) used in the paper
D6: Interaction technique(s)	The interaction feature(s) present in the visualization technique(s)
D7: Research area(s)	The research area(s) where the visualization technique(s) was(were) appliedied
D8: Analysis technique(s)	The automated analysis technique(s) used to help in user analysis
D9: Use of CMV	If the work uses coordinated multiple views

3.2.2.5 Study selection

We divided the study selection in stages. In the first stage using the query in all data sources we found a total of 3,771 articles. In a second stage, filtering by reading the publication title and keywords, we selected 69 articles. The third stage of filtering was based on reading the abstract: we selected 52 articles. In the fourth stage we read the articles, and we selected a total of 39 articles. With these 39 works, we took their references and applied a similar process to extrapolate, trying to reach works we could not get with the query. Applying the same process, we selected 12 more articles, and then we ended up with a total of 51 papers.

3.2.3 SLR Results

In this section we present the results of the analysis of the 51 selected papers. All the selected papers are cited Tables A.1 and A.2.

At first, we analyze data about the evolution of the publications of ensemble data sets visualization along the years. After, we report answers to our research questions (listed in section 3.2.2.1), starting with answers for RQ2 in Section 3.2.3.2. Then, RQ1 and RQ3 will be discussed in Section 3.2.3.3. In Section 3.2.3.4, we discussed results regarding RQ4, while in Section 3.2.3.5 we address RQ5. Finally, Section 3.2.3.6 reports our results regarding RQ6.

3.2.3.1 Evolution of the amount of published articles

Figure 3.3: Publications per year

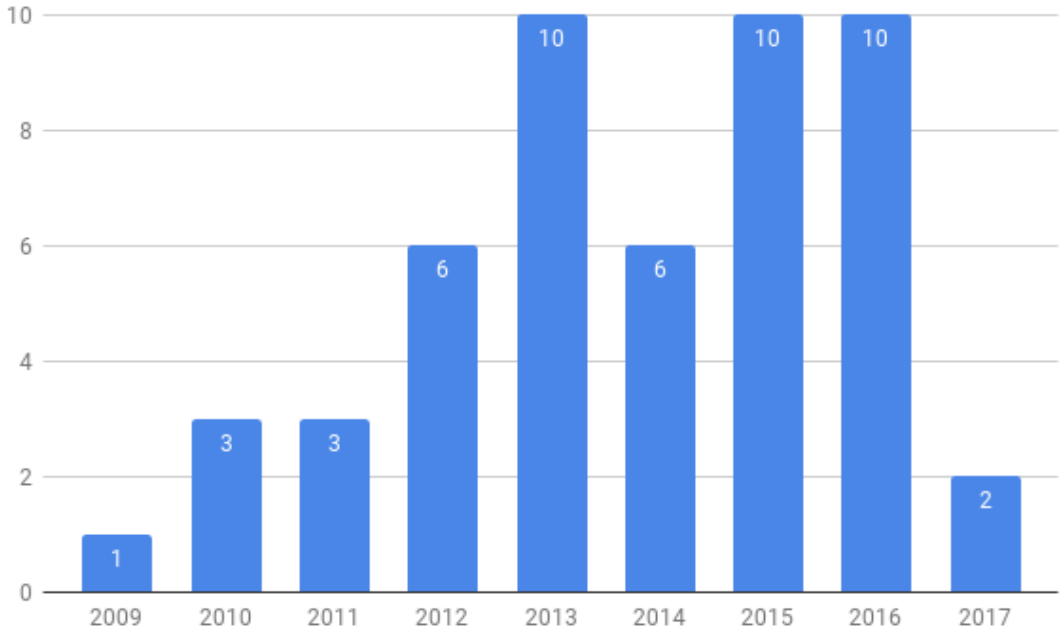


Figure 3.3 shows the quantity of published works per year. We can notice a steady evolution of published works from 2011 to 2013, and it remained stable from 2013 to 2016, with 2014 being an exception. Since we stopped our survey in August 2017, it is likely that the number of published papers remain stable, confirming that the subject still has open questions to be addressed.

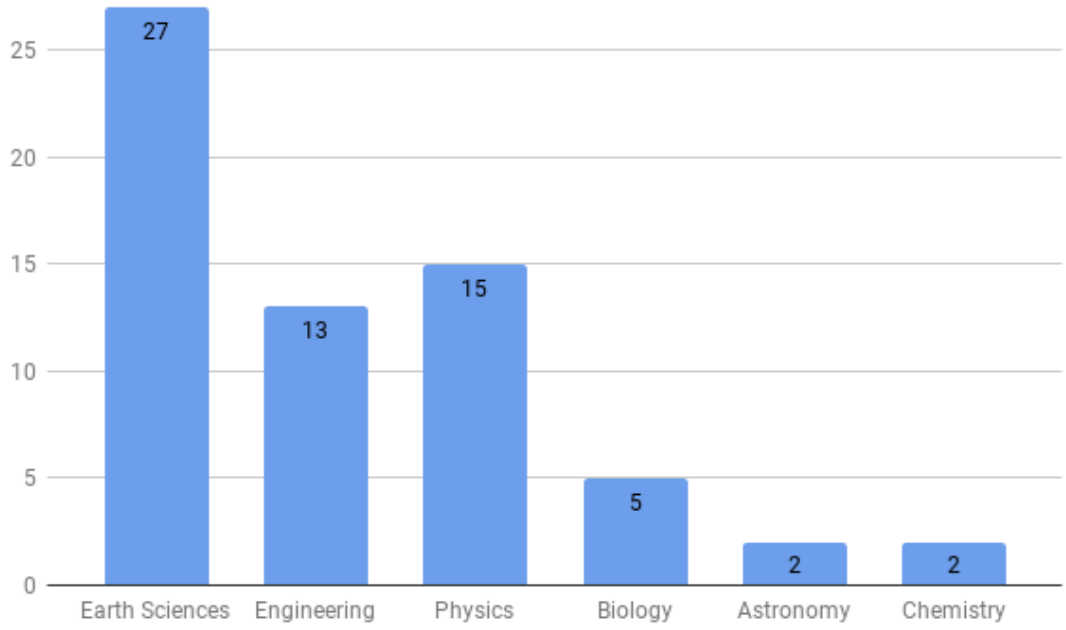
3.2.3.2 Research areas

Table 3.3: Research areas and specific problems addressed in the papers

Research area	Problem/main subject of papers
Astronomy	Study of celestial objects and their processes
Biology	Study of living organisms
Chemistry	Study of matter composition, structure, properties and changes
Engineering	Application of scientific principles to design and/or develop structures, machines and apparatus
Earth Sciences	Study of Earth and its composition, structure, physical properties and processes
Physics	Study of nature and its phenomena

To answer RQ2 ("Which research areas use visualization of ensemble data sets?") and to help answering RQ3 ("What is the relation between visualization

Figure 3.4: Quantity of papers per research area



techniques and research domains?"), we first need to classify the selected studies according to the research area and the visualization techniques they use.

We observed that many of the selected papers aimed at helping researchers from different scientific areas. Then, the best approach was to classify them based on academic disciplines. We chose the academic disciplines as enumerated by Bates (BATES, 2007), and we excluded the categories that do not deal with numerical simulations, since this is the context of our work.

In Table 3.3 we list the research areas categories and their definition that helped us to classify the selected articles. There is a paper that belongs to more than one category because the use case is from an intersection of areas (XIAO et al., 2015). Another consideration is that some papers have multiple use cases for validating their visualizations (MIRZARGAR; WHITAKER; KIRBY, 2014). The results of this classification is shown in Fig. 3.4.

Earth Sciences is the research area where we found most application of visualization techniques for ensemble data sets. Within this domain, the research field that shows the majority of use cases is climate and weather simulations, followed by ocean simulations. One aspect we noticed in the articles about climate and weather is that they choose mainly two sources of data: (i) Weather Research and Forecasting (WRF) Model[12] and (ii) Goddard Earth Observing System, Version 5 (GEOS-5)

[12] <http://www.wrf-model.org>

[13].

The second main research area is Physics. Fluid dynamics is the most used example, followed by heavy ion collisions. Different from Earth Sciences data, fluid dynamics simulation data do not come from known data sources, so they are likely to originate from self-developed simulators or from partners.

The third significant research area we found is Engineering, which of course is a wide discipline, encompassing several research fields. Most of the works are visualization of simulations of car engines, followed by flooding in cities.

Summarizing the findings related to RQ2, the research areas with most usage of visualization of ensemble data sets are Earth Sciences, with 54.4% of the selected works, Physics, 31,6%, and Engineering, 22,8%.

3.2.3.3 Visualization techniques

Table 3.4: Visualization techniques categories

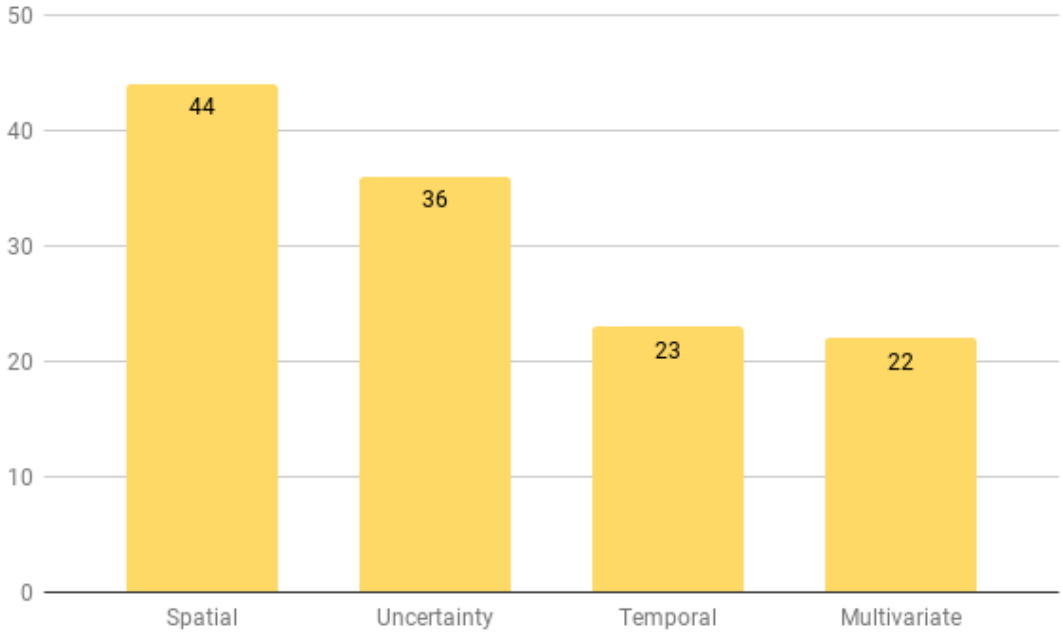
Visualization technique	Definition
Spatial	Techniques for visualizing data associated to a spatial domain often in real world
Temporal	Techniques for visualizing the evolution of data along time
Multivariate	Techniques for visualizing values of two or more different data features
Uncertainty	Techniques using a visual metaphor for representing an aggregation of values

Classifying visualization techniques is a hard task, because there are many aspects to take in consideration. Techniques are usually classified according to the data it aims to present (CHI, 2000)(SHNEIDERMAN, 1996)(TORY; MOLLER, 2004)(MUNZNER, 2014). This is the most common way of classifying techniques because researchers want to understand the whole structure of the data to make their assumptions. This type of classification is said to be a low-level taxonomy (TORY; MOLLER, 2004), but the problem is that it just considers the researcher side, and users have different ideas about visualization (CHENGZHI; CHENGHU; TAO, 2003).

Other way to define taxonomies for visualization techniques is to create specific classifications (PRICE; BAECKER; SMALL, 1993)(WENZEL; BERNHARD; JESSEN, 2003)(LIP[Pleaseinsertintopreamble]A et al., 2012)(KEHRER; HAUSER,

[13]<https://gmao.gsfc.nasa.gov/GEOS/>

Figure 3.5: Quantity of papers employing each visualization technique



2013). It is easier to solve a taxonomy problem for a specific case than building a more general taxonomy. Moreover, since it limits problems, such taxonomy may consider other aspects, e.g., visual metaphors, dimensions of the graphical representation, analysis techniques, etc. (WENZEL; BERNHARD; JESSEN, 2003)(KEHRER; HAUSER, 2013). The disadvantage of specific taxonomies is the difficulty to extrapolate from a domain to another, since this other domain will have its specific problems, which can not be dealt with using taxonomies from other domains.

To classify the selected studies we used a mixture of both approaches. The categories were based on data because all the visualization techniques we are dealing with are for scientific research, so the users, i.e., researchers, are more concerned about understanding the results of their simulations, so they know how the data is structured and what they want to visualize. Another consideration is that the definition of ensemble data sets is already associated to data types (i.e., multidimensional, time evolutive, multivariable and multivalued), so we brought this definition for our categories. Table 3.4 describes the categories of visualization techniques we used and how they should be interpreted.

In Fig. 3.5 one can observe the dominance of spatial visualization techniques, as 2-dimensional maps and 3-dimensional volumes, being reported in 89.5% of the studies. The second most used visualization techniques category is uncertainty visualization techniques, in 73.7% of the studies. Color is the most used visual attribute,

being employed to represent statistical values, followed by different kinds of box plots (WHITAKER; MIRZARGAR; KIRBY, 2013)(MIRZARGAR; WHITAKER; KIRBY, 2014). Then, we observed temporal visualization techniques, mainly represented by line charts, in 45.6% of the studies. This is a natural finding because simulation produces data along time. At last, we observe several works employing multivariate techniques (38.6% of the studies), mainly scatter plots and parallel coordinates. The results showed the importance of visualizing data associated to the simulation domain in an aggregated way, so the users can obtain at first an overview of the behavior of certain variables over the domain space.

We have also analyzed the visualization techniques of the three most important research areas to answer RQ3 ("What is the relation among visualization techniques and research domains?"). Then, we combined the classification of visualization techniques with research areas, and the results are shown in Fig.3.6.

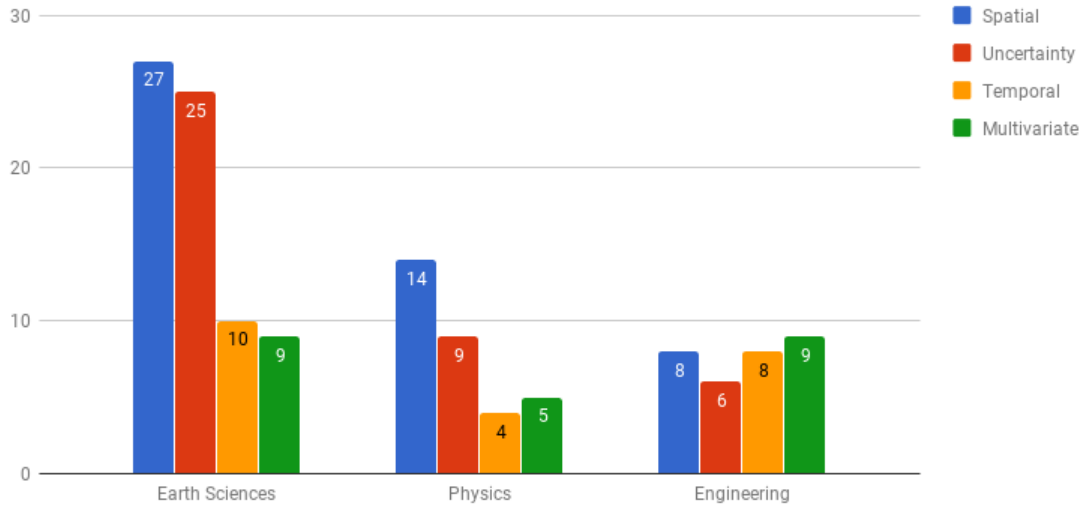
One can notice that in Earth Sciences all the studies use some real world space reference, and it is mainly due to the definition of the research area itself. All the studies use some spatial visualization technique, and we found that 2-dimensional maps are the most frequently used visual representation in these studies. Another consideration is the large amount of studies that employ some uncertainty technique (92.6% of the works), mainly due to the need of predicting some phenomenon. We were not expecting so rare use of temporal visualization techniques, only 37% of the studies employed some technique of this category.

Regarding Physics, it is almost the same scenario as for Earth Sciences. 93.3% of the studies use some spatial visualization technique, mainly due to the fact that in the most used cases (fluid dynamics) it is necessary to visualize how the matter behaves in real world. The main difference, when compared to Earth Sciences, is the use of uncertainty visualization techniques in 60% of the studies.

Engineering is a different scenario, where we found a more distributed use of different visualization techniques. 69.2% of the articles report the use of multivariate visualization techniques, like scatter plots and parallel coordinates. In 61.5% of the studies, spatial and temporal visualization techniques are used, and uncertainty visualization being least used, in 50% of the articles. As mentioned before, different engineering research fields imply simulations for different problems, involving diverse data sets, rising the need for different techniques.

Finally, we can comment on results for answering RQ1 and RQ3. As for RQ1

Figure 3.6: Visualization techniques usage in each research area



("What are the visualization techniques implemented for ensemble data sets visualization?"), usually spatial and uncertainty visualization techniques are the most important, but since it is always necessary to understand the needs of researchers, it is likely to assume that the other two techniques are necessary as well. To answer RQ3 ("What is the relation between visualization techniques and research domains?"), we found that in Earth Sciences and Physics the most used visualization techniques are spatial and uncertainty techniques, while in Engineering all four visualization kinds of visualization techniques are used.

3.2.3.4 Visualization tasks

The study of how users interact with visualizations is important to improve the usability of the visualization techniques and helping users to increase their capacity of creating more hypothesis (SACHA et al., 2016) and investigating them.

There is a large number of studies classifying user tasks (SHNEIDERMAN, 1996)(KEHRER; HAUSER, 2013)(YI et al., 2007)(KEIM, 2002). We can describe the way users interact with a visualization using the well-known mantra, "Overview first, zoom and filter, then details-on-demand" (SHNEIDERMAN, 1996). This mantra introduced a user interface visualization strategy that researchers call *Overview + detail*. Another visualization strategy is *Focus + context*, where the overview (context) and the details (focus) are viewed simultaneously (CARD; MACKINLAY; SHNEIDERMAN, 1999), with the user changing the focus in some interactive way, e.g. fish-eye (FURNAS, 1986). Both of these interaction approaches can be used in

the visualization of ensemble data sets.

To classify the selected studies by the tasks the user performs within a visualization system, we used Brehmer and Munzner typology (BREHMER; MUNZNER, 2013), because, besides being a classification based on previous ones, they use it to help in describing tasks of known visualization tools. Since we want to classify user interactions within a visualization system, we are using most of the *"how?"* techniques observed by them.

The list of techniques we used to classify the selected studies by their interaction features is shown in Table 3.5. The other categories described by Brehmer and Munzner were not found in the selected studies. We should also state that we classified the techniques by what was explicitly described in the articles, implicit interactions were not considered. For example, zooming and rotation, which are part of the "navigate" category, can be implicit for a 3-dimensional visualization, but they were only considered if they are referred to in the text.

Table 3.5: Visualization task categories

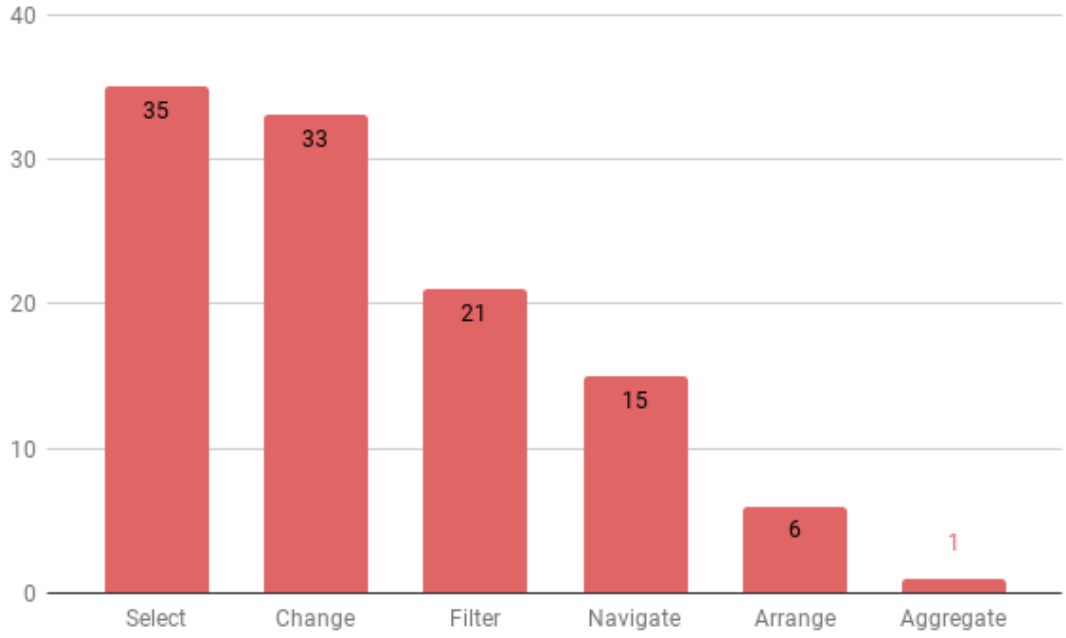
Visualization interactive task	Definition
Select	Selection of elements
Navigate	Alter user viewpoint
Arrange	Organizing elements
Change	Alter visual encoding
Filter	Adjust exclusion and inclusion criteria
Aggregate	Change granularity

In 36 studies, representing 70.6%, we found interactive techniques thus showing the importance of interaction in visualization systems.

In Fig. 3.7 we present the quantity of studies addressing each interaction technique: 97.2% of the studies which used some interaction used "select" operations to "select ensemble members" or some point or area in a map. In 91.7% of the papers, authors report the changing of the visual encode, which is used most to highlight selected items. In 58.3% of the articles, filtering is used, being brushing the most used filtering technique. Interaction techniques for navigation were used in 41.7% of the works, being zooming techniques used in some 2-dimensional graphics and necessary in 3-dimensional visualizations. Arrange is one of the less used technique due to the few studies using clustering and aggregate (DEMIR; DICK; WESTERMANN, 2014). The user orders to change the granularity of the domain using a semantic zoom, the other do not change granularity, they just do a real

zoom.

Figure 3.7: Quantity of papers employing each visualization task



Regarding RQ4, *selecting* is important because the user usually needs to select the ensemble members to perform some other tasks; *change* allows to modify the visual encoding to highlight the selected members so the users are kept oriented within the ensemble; *filtering* is used the lower the data items the scientist wants to visualize, by brushing or by some other parameter definition, and *navigate* is important for exploring 3-dimensional domains.

3.2.3.5 Analysis techniques

When dealing with ensemble data sets, users often face the need of analyzing multivalued data (WILSON; POTTER, 2009), which is difficult to analyze for each ensemble member at the same time. To overcome this problem some techniques aggregate data so fewer values or relations among data values are easier to represent and understand.

Table 3.6 presents the categories we use for describing the different analysis techniques we found being employed by the surveyed studies. There are other classifications for analysis techniques (KEHRER; HAUSER, 2013)(FEW, 2009), but we needed a broader way to classify the selected studies, since we wanted to consider analysis techniques ranging from descriptive statistics and Principal Component

Analysis (PCA), for example, to clustering and neural networks.

Table 3.6: Analysis techniques categories

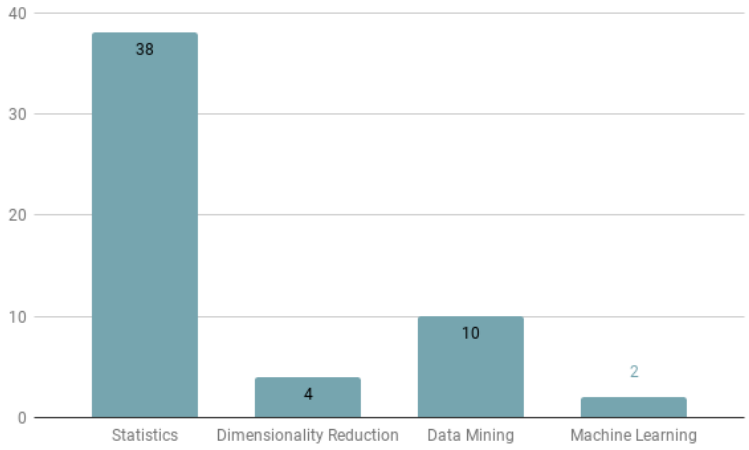
Analysis technique	Definition
Statistics	Inputs a list of values, outputs one or more values resulting from a function or procedure
Dimensionality reduction	Inputs a multidimensional data set, outputs a reduced-dimension data set that can be plotted in 2D or 3D
Data Mining	Inputs a data set, outputs relations resulting from a deterministic function
Machine Learning	Inputs a data set, outputs a result from a non-deterministic function

We considered a study as belonging to the category of statistics if the study uses a method which receives a list of values, executes a function that calculates one or more values, and the formulas employed in the calculation do not change. An example of such statistics is the quartile trend chart (POTTER et al., 2009). Regarding dimensionality reduction techniques we considered them as methods that receive a list of data comprising many attributes for each data item, execute some function which, in the end, will help to understand correlation and similarity of the data items. Common examples of dimensionality reduction techniques are Principal Components Analysis (PCA) and Multidimensional Scaling (MDS). As for the data mining category, given a set of data items, the technique returns some more complex relations between the data items. Clustering is the most used example of data mining technique in visualization. Finally, regarding machine learning techniques the input is some data (it can be one item or a list of items), and the output is some other data (it can be a complex relation, as data mining) and this data can change how the function evaluates further input data, with the possibility of processing previous data yielding different results.

Fig. 3.8 is the distribution of the analysis techniques among the 51 selected studies. Eleven studies did not fit in any of our categories, because they were based just on observations. Considering the 40 studies that fitted in the categories, 95% use statistics and the other techniques are used in 25% or less the cases.

So, answering RQ5 ("What kind of analyses are used in ensemble data sets visualization?"), the use of statistics in the visualization seems the most common approach used to help researchers in their analysis. This may change with the more often use of visual analytics techniques with ensemble data sets.

Figure 3.8: Quantity of papers with each analysis technique



3.2.3.6 Coordinated multiple views

Coordinated Multiple Views (CMV) is an approach for exploratory visualization used to investigate data (usually a large volume of data) by means of different, coordinated representations (BALDONADO; WOODRUFF; KUCHINSKY, 2000)(ROBERTS, 2007). The advantages of using CMV are the discovery of relationships that can be hard or impossible to visualize with a single visualization. It is said to improve user cognition in finding the data he/she wants, since each view intends to improve the understanding of different aspects of the data. However, it is still rarely used in commercial systems, being employed more often in academic research (ANDRIENKO; ANDRIENKO, 2007).

We used the CMV definition to classify the selected studies: they were classified as "Yes" if they use CMV, or "No" if they did not. In 64.7% of the cases, CMV is used, while in 35.3% it is not.

So, related to RQ6 ("Is coordinated multiple views (CMV) used in most of the works?"), yes, CMV is used in most of the studies, proving its importance to users. Many of the studies that do not use CMV, have as objective to visualize spatial data or a specific aspect in one visualization.

3.2.4 Further Comments

We investigated how researchers design systems to explore this type of data, considering what is important to visualize for the most used research areas, how the

users interact with the system and how they visualize the relations or values in the visualization. It is aimed at helping researchers to design systems and to develop other ways for exploring ensemble data sets.

Unfortunately, this study has also some threats to validity (in a SLR, we have to point out the threats to validity of the employed method and results). We think the human decision bias is the biggest threat, because it comes in the every stage of this study, from selecting the studies to classifying the remaining ones. To try to mitigate the problem in the selection of the studies we extrapolated using the references sections of articles selected in the automatic research. However, it still brings bias which can interfere in the result, but it was lowered. Other biases we are aware of were the fatigue in selecting the studies from the automatic search due to the large quantity of papers found (3,771).

We observed that using multidimensional visualization techniques is most important in many areas, but the use of other techniques can be important as well. So, before designing new techniques, the researcher still needs to consult the users about what they want to explore in the data. We also identified the intense use of interaction techniques and CMV in the visualization of different aspects of the data as an important solution. And how researchers use analysis techniques was also presented, with the majority using statistics, but we think the other techniques can be more explored to help the users to understand more aspects of the data.

4 GEOCHEMICAL SIMULATION ENSEMBLES VISUALIZATION

In this chapter we present the design and implementation of GEVis, a system which objective is to help geologists in exploring data from ensembles of geochemical simulations. Although our approach aims at being general for any simulation ensembles, the proof-of-concept prototype we built considers data output from a geochemical simulator used in a current project at UFRGS.

At a certain extent we follow the "What-Why-How" framework used by (MUNZNER, 2014). Firstly, we describe the data coming from the simulator our users adopted, and then we present the users' tasks that we targeted in our work. Then we describe the GEVis architecture and the visualization techniques we provided along implementation details.

Also we are using our experience in designing visualization systems for geologists with what was found in the SLR (in Section 3.2.1) to design our system. We tried to use all visualization techniques described in the SLR (Section 3.2.3.3) to give the user different views about the data and used all interactive tasks, but *Aggregate* (Section 3.2.3.4) to give the user alternatives to interact with its data (even some of the techniques are not used much).

4.1 Data Description and Users' Tasks

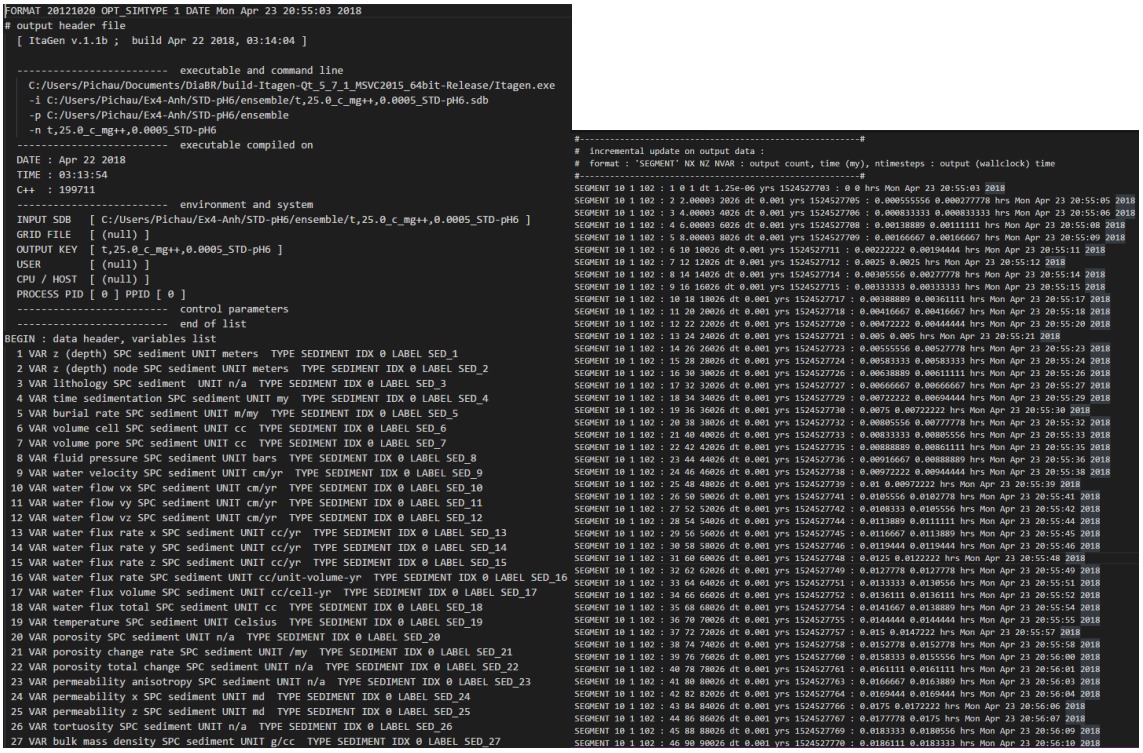
Before start the work, we need to understand how the data is disposed and what the users want to visualize in their desired visualization system. So first we describe the data format provided by the simulator specified in section 2.2, then we describe the user desired tasks in the system.

4.1.1 Data Description

The data our tool takes for visualization is the output from geochemical simulator. It comes as two files for each simulation: a *header file* that contains the description of the variables in the output, also containing the summary of each time step, (Fig. 4.1 and a *data file* that contains the values of the variables defined in the header file for each time step and each cell (Fig. 4.2. Each variable in the header

file is defined as a composition of textual information (Fig. 4.1a).

Figure 4.1: Header file example



```
FORMAT 20121020 OPT_SIMTYPE 1 DATE Mon Apr 23 20:55:03 2018
# output header file
  [ ItaGen v.1.1b ;  build Apr 22 2018, 03:14:04 ]

  ------------------------  executable and command line
    C:/Users/Pichau/Documents/Dia8R/build-Itagen-Qt_5_7_1_MSVC2015_64bit-Release/Itagen.exe
    -i C:/Users/Pichau/Ex4-Anh/STD-pH6/ensemble/t,25.0_c_mg++,0.0005_STD-pH6.sdb
    -p C:/Users/Pichau/Ex4-Anh/STD-pH6/ensemble
    -n t,25.0_c_mg++,0.0005_STD-pH6
  ------------------------  executable compiled on
  DATE : Apr 22 2018
  TIME : 03:13:54
  C++  : 199711
  ------------------------  environment and system
  INPUT SDB   [ C:/Users/Pichau/Ex4-Anh/STD-pH6/ensemble/t,25.0_c_mg++,0.0005_STD-pH6 ]
  GRID FILE   [ (null) ]
  OUTPUT KEY  [ t,25.0_c_mg++,0.0005_STD-pH6 ]
  USER        [ (null) ]
  CPU / HOST  [ (null) ]
  PROCESS PID [ 0 ] PPID [ 0 ]
  ------------------------  control parameters
  ------------------------  end of list
BEGIN : data header, variables list
  1 VAR z (depth) SPC sediment UNIT meters  TYPE SEDIMENT IDX 0 LABEL SED_1
  2 VAR z (depth) node SPC sediment UNIT meters  TYPE SEDIMENT IDX 0 LABEL SED_2
  3 VAR lithology SPC sediment  UNIT n/a  TYPE SEDIMENT IDX 0 LABEL SED_3
  4 VAR time sedimentation SPC sediment UNIT my  TYPE SEDIMENT IDX 0 LABEL SED_4
  5 VAR burial rate SPC sediment UNIT m/my  TYPE SEDIMENT IDX 0 LABEL SED_5
  6 VAR volume cell SPC sediment UNIT cc  TYPE SEDIMENT IDX 0 LABEL SED_6
  7 VAR volume pore SPC sediment UNIT cc  TYPE SEDIMENT IDX 0 LABEL SED_7
  8 VAR fluid pressure SPC sediment UNIT bars  TYPE SEDIMENT IDX 0 LABEL SED_8
  9 VAR water velocity SPC sediment UNIT cm/yr  TYPE SEDIMENT IDX 0 LABEL SED_9
 10 VAR water flow vx SPC sediment UNIT cm/yr  TYPE SEDIMENT IDX 0 LABEL SED_10
 11 VAR water flow vy SPC sediment UNIT cm/yr  TYPE SEDIMENT IDX 0 LABEL SED_11
 12 VAR water flow vz SPC sediment UNIT cm/yr  TYPE SEDIMENT IDX 0 LABEL SED_12
 13 VAR water flux rate x SPC sediment UNIT cc/yr  TYPE SEDIMENT IDX 0 LABEL SED_13
 14 VAR water flux rate y SPC sediment UNIT cc/yr  TYPE SEDIMENT IDX 0 LABEL SED_14
 15 VAR water flux rate z SPC sediment UNIT cc/yr  TYPE SEDIMENT IDX 0 LABEL SED_15
 16 VAR water flux rate SPC sediment UNIT cc/unit-volume-yr  TYPE SEDIMENT IDX 0 LABEL SED_16
 17 VAR water flux volume SPC sediment UNIT cc/cell-yr  TYPE SEDIMENT IDX 0 LABEL SED_17
 18 VAR water flux total SPC sediment UNIT cc  TYPE SEDIMENT IDX 0 LABEL SED_18
 19 VAR temperature SPC sediment UNIT Celsius  TYPE SEDIMENT IDX 0 LABEL SED_19
 20 VAR porosity SPC sediment UNIT n/a  TYPE SEDIMENT IDX 0 LABEL SED_20
 21 VAR porosity change rate SPC sediment UNIT /my  TYPE SEDIMENT IDX 0 LABEL SED_21
 22 VAR porosity total change SPC sediment UNIT n/a  TYPE SEDIMENT IDX 0 LABEL SED_22
 23 VAR permeability anisotropy SPC sediment UNIT n/a  TYPE SEDIMENT IDX 0 LABEL SED_23
 24 VAR permeability x SPC sediment UNIT md  TYPE SEDIMENT IDX 0 LABEL SED_24
 25 VAR permeability z SPC sediment UNIT md  TYPE SEDIMENT IDX 0 LABEL SED_25
 26 VAR tortuosity SPC sediment UNIT n/a  TYPE SEDIMENT IDX 0 LABEL SED_26
 27 VAR bulk mass density SPC sediment UNIT g/cc  TYPE SEDIMENT IDX 0 LABEL SED_27
```

```
#-------------------------------------------------------#
#  incremental update on output data :
#  format : 'SEGMENT' NX NZ NVAR : output count, time (my), ntimesteps : output (wallclock) time
#-------------------------------------------------------#
SEGMENT 10 1 102 : 1 0 1 dt 1.25e-06 yrs 1524527703 : 0 0 hrs Mon Apr 23 20:55:03 2018
SEGMENT 10 1 102 : 2 2.00003 2026 dt 0.001 yrs 1524527705 : 0.000555556 0.000277778 hrs Mon Apr 23 20:55:05 2018
SEGMENT 10 1 102 : 3 4.00003 4026 dt 0.001 yrs 1524527706 : 0.000833333 0.000833333 hrs Mon Apr 23 20:55:06 2018
SEGMENT 10 1 102 : 4 6.00003 6026 dt 0.001 yrs 1524527708 : 0.00138889 0.00111111 hrs Mon Apr 23 20:55:08 2018
SEGMENT 10 1 102 : 5 8.00003 8026 dt 0.001 yrs 1524527709 : 0.00166667 0.00166667 hrs Mon Apr 23 20:55:09 2018
SEGMENT 10 1 102 : 6 10 10026 dt 0.001 yrs 1524527711 : 0.00222222 0.00194444 hrs Mon Apr 23 20:55:11 2018
SEGMENT 10 1 102 : 7 12 12026 dt 0.001 yrs 1524527712 : 0.0025 0.0025 hrs Mon Apr 23 20:55:12 2018
SEGMENT 10 1 102 : 8 14 14026 dt 0.001 yrs 1524527714 : 0.00305556 0.00277778 hrs Mon Apr 23 20:55:14 2018
SEGMENT 10 1 102 : 9 16 16026 dt 0.001 yrs 1524527715 : 0.00333333 0.00333333 hrs Mon Apr 23 20:55:15 2018
SEGMENT 10 1 102 : 10 18 18026 dt 0.001 yrs 1524527717 : 0.00388889 0.00361111 hrs Mon Apr 23 20:55:17 2018
SEGMENT 10 1 102 : 11 20 20026 dt 0.001 yrs 1524527718 : 0.00416667 0.00416667 hrs Mon Apr 23 20:55:18 2018
SEGMENT 10 1 102 : 12 22 22026 dt 0.001 yrs 1524527720 : 0.00472222 0.00444444 hrs Mon Apr 23 20:55:20 2018
SEGMENT 10 1 102 : 13 24 24026 dt 0.001 yrs 1524527721 : 0.005 0.005 hrs Mon Apr 23 20:55:21 2018
SEGMENT 10 1 102 : 14 26 26026 dt 0.001 yrs 1524527723 : 0.00555556 0.00527778 hrs Mon Apr 23 20:55:23 2018
SEGMENT 10 1 102 : 15 28 28026 dt 0.001 yrs 1524527724 : 0.00583333 0.00583333 hrs Mon Apr 23 20:55:24 2018
SEGMENT 10 1 102 : 16 30 30026 dt 0.001 yrs 1524527726 : 0.00638889 0.00611111 hrs Mon Apr 23 20:55:26 2018
SEGMENT 10 1 102 : 17 32 32026 dt 0.001 yrs 1524527727 : 0.00666667 0.00666667 hrs Mon Apr 23 20:55:27 2018
SEGMENT 10 1 102 : 18 34 34026 dt 0.001 yrs 1524527729 : 0.00722222 0.00694444 hrs Mon Apr 23 20:55:29 2018
SEGMENT 10 1 102 : 19 36 36026 dt 0.001 yrs 1524527730 : 0.0075 0.00722222 hrs Mon Apr 23 20:55:30 2018
SEGMENT 10 1 102 : 20 38 38026 dt 0.001 yrs 1524527732 : 0.00805556 0.00777778 hrs Mon Apr 23 20:55:32 2018
SEGMENT 10 1 102 : 21 40 40026 dt 0.001 yrs 1524527733 : 0.00833333 0.00805556 hrs Mon Apr 23 20:55:33 2018
SEGMENT 10 1 102 : 22 42 42026 dt 0.001 yrs 1524527735 : 0.00888889 0.00861111 hrs Mon Apr 23 20:55:35 2018
SEGMENT 10 1 102 : 23 44 44026 dt 0.001 yrs 1524527736 : 0.00916667 0.00888889 hrs Mon Apr 23 20:55:36 2018
SEGMENT 10 1 102 : 24 46 46026 dt 0.001 yrs 1524527738 : 0.00972222 0.00944444 hrs Mon Apr 23 20:55:38 2018
SEGMENT 10 1 102 : 25 48 48026 dt 0.001 yrs 1524527739 : 0.01 0.00972222 hrs Mon Apr 23 20:55:39 2018
SEGMENT 10 1 102 : 26 50 50026 dt 0.001 yrs 1524527741 : 0.0105556 0.0102778 hrs Mon Apr 23 20:55:41 2018
SEGMENT 10 1 102 : 27 52 52026 dt 0.001 yrs 1524527742 : 0.0108333 0.0105556 hrs Mon Apr 23 20:55:42 2018
SEGMENT 10 1 102 : 28 54 54026 dt 0.001 yrs 1524527744 : 0.0113889 0.0111111 hrs Mon Apr 23 20:55:44 2018
SEGMENT 10 1 102 : 29 56 56026 dt 0.001 yrs 1524527745 : 0.0116667 0.0113889 hrs Mon Apr 23 20:55:45 2018
SEGMENT 10 1 102 : 30 58 58026 dt 0.001 yrs 1524527746 : 0.0119444 0.0119444 hrs Mon Apr 23 20:55:46 2018
SEGMENT 10 1 102 : 31 60 60026 dt 0.001 yrs 1524527748 : 0.0125 0.0122222 hrs Mon Apr 23 20:55:48 2018
SEGMENT 10 1 102 : 32 62 62026 dt 0.001 yrs 1524527749 : 0.0127778 0.0127778 hrs Mon Apr 23 20:55:49 2018
SEGMENT 10 1 102 : 33 64 64026 dt 0.001 yrs 1524527751 : 0.0133333 0.0130556 hrs Mon Apr 23 20:55:51 2018
SEGMENT 10 1 102 : 34 66 66026 dt 0.001 yrs 1524527752 : 0.0136111 0.0136111 hrs Mon Apr 23 20:55:52 2018
SEGMENT 10 1 102 : 35 68 68026 dt 0.001 yrs 1524527754 : 0.0141667 0.0138889 hrs Mon Apr 23 20:55:54 2018
SEGMENT 10 1 102 : 36 70 70026 dt 0.001 yrs 1524527755 : 0.0144444 0.0144444 hrs Mon Apr 23 20:55:55 2018
SEGMENT 10 1 102 : 37 72 72026 dt 0.001 yrs 1524527757 : 0.015 0.0147222 hrs Mon Apr 23 20:55:57 2018
SEGMENT 10 1 102 : 38 74 74026 dt 0.001 yrs 1524527758 : 0.0152778 0.0152778 hrs Mon Apr 23 20:55:58 2018
SEGMENT 10 1 102 : 39 76 76026 dt 0.001 yrs 1524527760 : 0.0158333 0.0155556 hrs Mon Apr 23 20:56:00 2018
SEGMENT 10 1 102 : 40 78 78026 dt 0.001 yrs 1524527761 : 0.0161111 0.0161111 hrs Mon Apr 23 20:56:01 2018
SEGMENT 10 1 102 : 41 80 80026 dt 0.001 yrs 1524527763 : 0.0166667 0.0163889 hrs Mon Apr 23 20:56:03 2018
SEGMENT 10 1 102 : 42 82 82026 dt 0.001 yrs 1524527764 : 0.0169444 0.0169444 hrs Mon Apr 23 20:56:04 2018
SEGMENT 10 1 102 : 43 84 84026 dt 0.001 yrs 1524527766 : 0.0175 0.0172222 hrs Mon Apr 23 20:56:06 2018
SEGMENT 10 1 102 : 44 86 86026 dt 0.001 yrs 1524527767 : 0.0177778 0.0175 hrs Mon Apr 23 20:56:07 2018
SEGMENT 10 1 102 : 45 88 88026 dt 0.001 yrs 1524527769 : 0.0183333 0.0180556 hrs Mon Apr 23 20:56:09 2018
SEGMENT 10 1 102 : 46 90 90026 dt 0.001 yrs 1524527770 : 0.0186111 0.0183333 hrs Mon Apr 23 20:56:10 2018
```

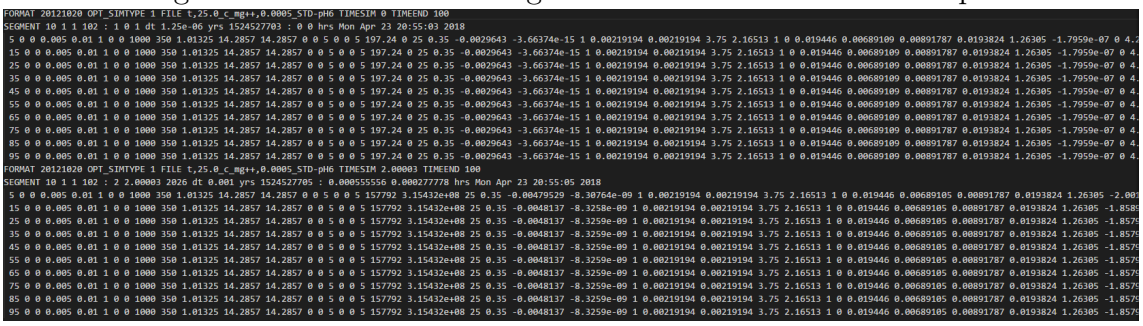
(a) Variable information in the header file. (b) Time steps information in the header file.

There are five possible types of variables depending on to what entities they refer to: (i) sediment, (ii) water column, (iii) element, (iv) solute and (v) solid. The *Sediment* type variables are related to the sediment itself as porosity, temperature, water velocity, and so on, being the only species of the type. *Water column* variables are only available if the simulation uses evaporation methods. *Element* variables are about some information of the quantity of each chemical element present in the related cell. *Solute* variables represent solute concentration in the water and activity in reactions, and each solute present in the system is defined in the data input, as H^+ (representing pH), HCO_3^-, etc. *Solid* variables are related to minerals(e.g. Quartz and Calcite), they present precipitation and dissolution, saturation, volume fraction and mineralization rate for each mineral.

Also in the header file there is a short description of all time steps, telling the quantity of cells in each dimension, the number of variables, the time in years of the time step and clock information the time step was calculated (Fig. 4.1b). This helps in the code allocate the data structures to save the values.

The actual data is in the data file (Fig. 4.2), where each time step is separated by a tag *FORMAT*, with the data of the current time of the time step and the time

Figure 4.2: Data file showing data values from two time steps.



the simulation is going to end. The next row begins with a tag *SEGMENT* which is equal to the rows in the header file describing the summary of the time step. The following rows until the next *FORMAT* tag is the data of each variable described in the header file for each domain cell. The first three values represent the distance of the cell from the domain origin, and the next values represent the data of the variables, described in the order they appear in the header file.

Using Munzner's classification (MUNZNER, 2014), the output data is a *Field*, because the variables (described in the header file) are associated to each cell at each time step. Variables associated with distance or time are considered *continuous* data, and variables associated to another variable is *non-continuous*.

4.1.2 Motivating Users' Tasks

To define what typical users would want to visualize from sets of geochemical simulation results, and the tasks they would need to perform with these data, we interviewed two geologists. Both are users of the geochemical simulator mentioned in the previous section. We conducted informal interviews with these users to find out what would be their intentions to use a visualization system to visualize ensemble data set, which is new to them. As a result of the interviews, we compiled the following short list of tasks:

1. Discover what and how parameters influence simulations within an ensemble.

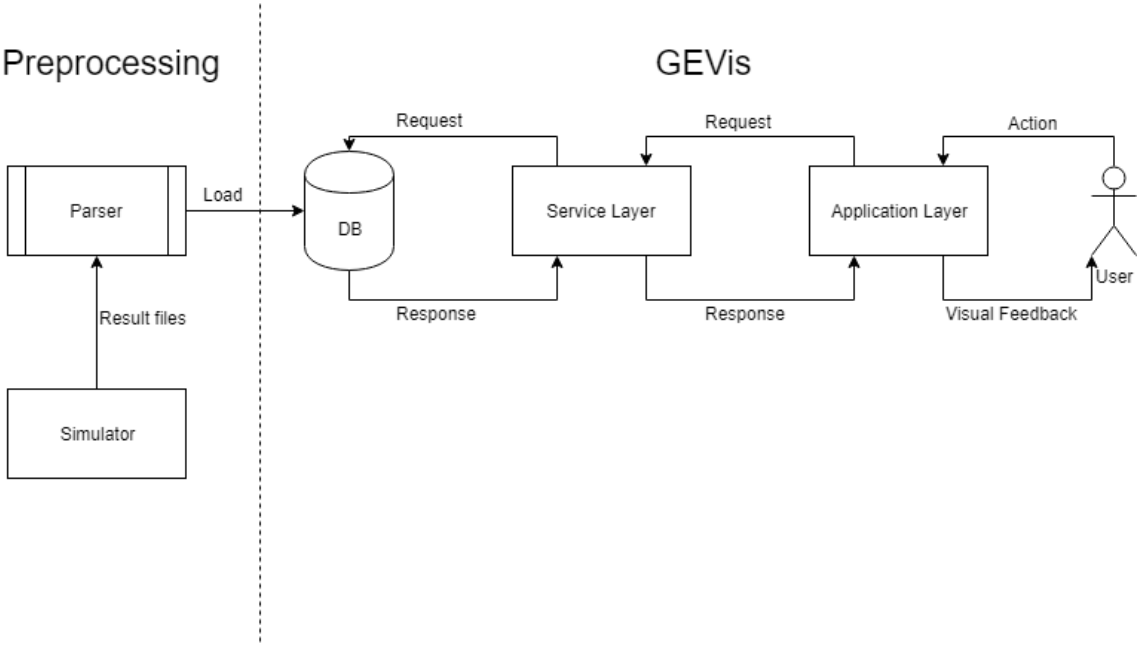
2. Verify which simulation conditions approximate from the nowadays conditions (from a geological point of view).

3. Compare ensembles from different simulators.

4.2 GEVIs Architecture

The architecture of the system follows a common Web system architecture, where there are two layers: (i) a service layer and (ii) an application/view layer. The service layer is responsible fro retrieving data from the data base, process it if needed, and then send it to the requester. The application/view layer is the graphical user interface, which runs in the user machine. This section gives detail about these two layers. Fig. 4.3 has a graphical vision which clarifies the GEVIs architecture.

Figure 4.3: A graphical vision of GEVIs architecture with how the user acts in the system and the processes occuring in the system.



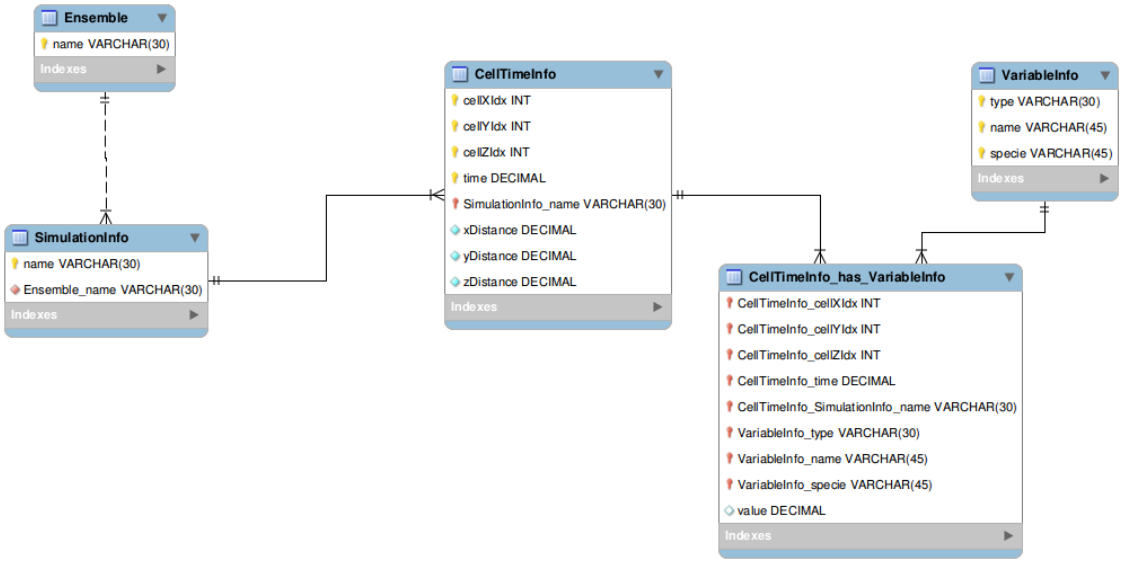
4.2.1 Service Layer

The service layer is characterized as a data provider. It can be described as having to parts: (i) the database and (ii) the data provider server.

At first, we modelled the simulation ensembles as a schema for a relational data base to understand the relationships we would have within the data sets provided by the simulator. The result of this modelling is in Fig. 4.4. The *Ensemble* table has the list of ensembles available for visualization by the GEVis user. *SimulationInfo* is the list of all the simulations, and it tells to which ensemble a simulation

belongs to. The *CellTimeInfo* table describes each cell in each time step for each simulation. *VariableInfo* lists all variables that can be used by the simulations. Finally, *CellTimeInfo has VariableInfo* represents the relation of *CellTimeInfo* and *VariableInfo*, in this case representing the value of a variable in a cell in a certain time step.

Figure 4.4: Relational schema for simulation ensembles



After this first relational model, we modelled our data set as a NoSQL schema because we adopted MongoDB[1] documents[2] as data base. This NoSQm odeLl is presented in Fig. 4.5. In this model, an ensemble is a list of simulations and is represented as a document. Each variable is also treated as a document. The relationship between variables and a cell in a time step of a simulation, which we modeled in Fig.4.4 as the table *CellTimeInfo has VariableInfo*, we defined as a list of variable values in a MongoDB document.

Our choice for a NoSQL schema is justified by the fact tha comparing both models, it was noticed a large amount of redundant data needed in the relational data model for retrieving data, while in MongoDB, since it stores documents using a similar concept of JSON files, there is less redundancy. We developed a parser to read the data from simulations results and store it in the MongoDB database.

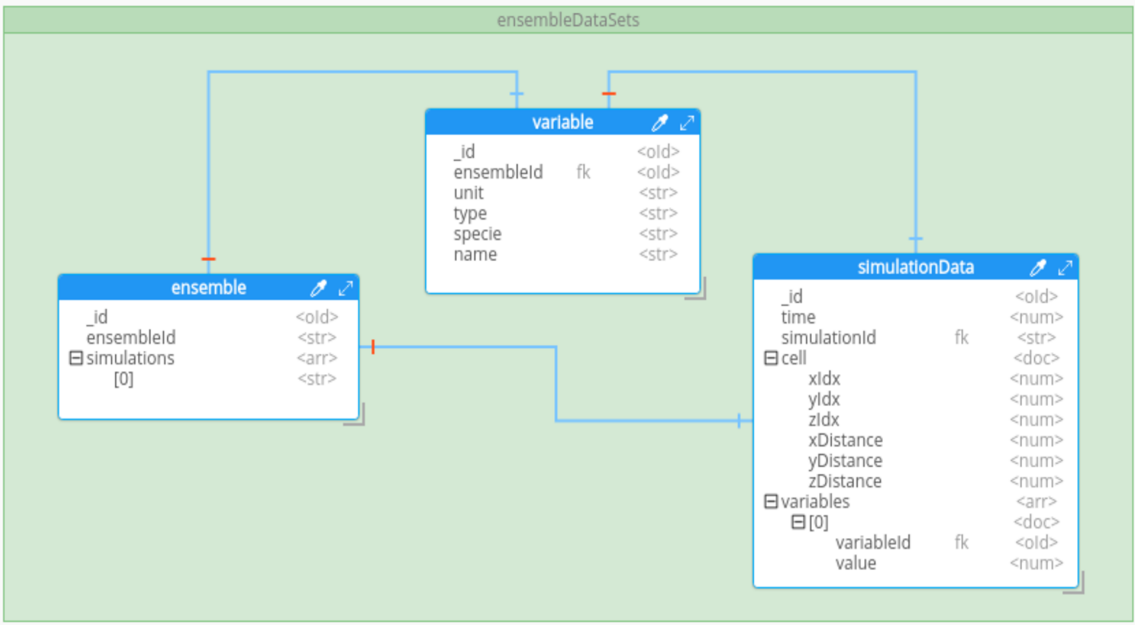
For the data provider server, we used the concept of Web Services[3] to communicate the data to the application layer. It passes to the application layer:

[1] <https://www.mongodb.com/>

[2] MongoDB does not used the concept of table, but uses the concept of documents (see <https://docs.mongodb.com/>).

[3] "A Web service is a software system designed to support interoperable machine-to-machine interaction over a network."(W3C, 2004)

Figure 4.5: NoSQL (MongoDB) schema for simulation ensembles



- Data about the ensembles
- Variables in an ensemble
- Cell quantity of the domain
- Data values of a variable in a cell for all time steps
- Data values of a variable in a certain time instant for each simulation in the whole domain,
- Data values of a set of variables in a certain cell and for a certain time instant.

The data provider server was developed using NodeJS[4].

In the cases where the server needs to retrieve data from a specific time instant, there is a problem to be solved. The time steps can be different between different simulations, so a simulation might not have data values for a specific time instant. Then, to obtain data related to a time instant from a simulation that does output data with that time stamp, the server needs to provide interpolated data. We adopted cubic spline interpolation method (MCKINLEY; LEVINE, 1998), instead of other methods (Lagrange or Newton) because with higher order polynomial function we would introduce unwanted oscillations in data values, which would likely be producing inconsistencies in the values.

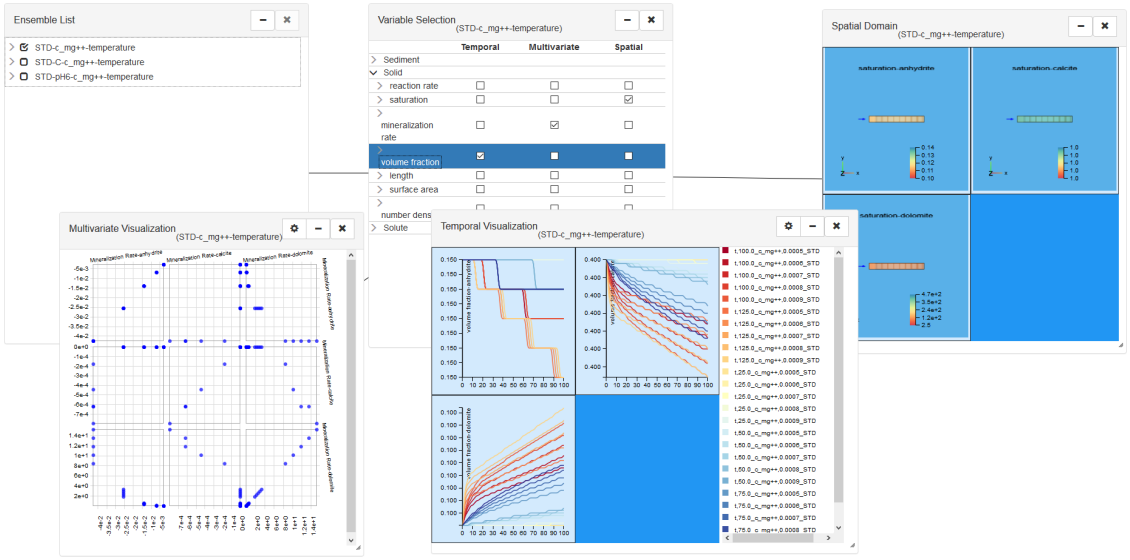
[4] <https://nodejs.org/en/>

4.2.2 Application Layer

The application layer corresponds to the software layer through which the user interacts with the system. Our application, as a visualization application, has the goal of helping users to understand phenomena or models by analyzing visual metaphors representing the data. The 2D visualization techniques were implemented using D3[5], the 3D visualization techniques was implemented using three.js[6] and for the window system it was used jQuery[7]. Our code is available online in GitHub[8] repository in the url <https://github.com/GJFeller/GEVIs>.

The application user interface (Fig.4.6) uses the concept of floating windows for displaying the visual representations, so the user has a flexible way to place the visualizations side by side, and "physically" linked by lines, and also uses the CMV concept (Section 3.2.3.6) when the user interacts with a visualization technique, using *brushing* and *selection*, it is reflected in the other views. This approach is based on (DUNNE et al., 2012) and (CAVA, 2017). This windowed design is most helpful in comparing different ensembles, and its flexibility also helps users to resize and relocate their visualizations depending on the course of data exploration.

Figure 4.6: An overview of the window concept of GEVIs



GEVIs first displays the ensembles available for visualization (upper left part of Fig.4.6), giving information about which simulations compose them. When an

[5] <https://d3js.org/>
[6] <https://threejs.org/>
[7] <https://jquery.com/>
[8] <https://github.com/>

ensemble is selected from this list, the user can interact with its variables to select the features he/she wants to analyze (window named "Variables Selection").

An important concept of GEVIs is the division of the possible visualizations into three groups: (i) temporal visualization, (ii) spatial domain visualization and (iii) multivariate visualization. Examples of these 3 visualizations are shown in Fig. 4.6. The user selects these features to have different perspectives of the data. Each visualization technique has some interactive features, and the interaction with a window is reflected in the other windows showing the same ensemble. In the following sections, we give details about these coordinated visualization techniques.

4.2.2.1 Temporal Visualization

In the temporal visualization, the user can understand how the variables behave along the simulated time, and interacting with the other visualizations, mainly the multivariate visualization, she/he can understand how the parameters affect the simulation in a certain cell.

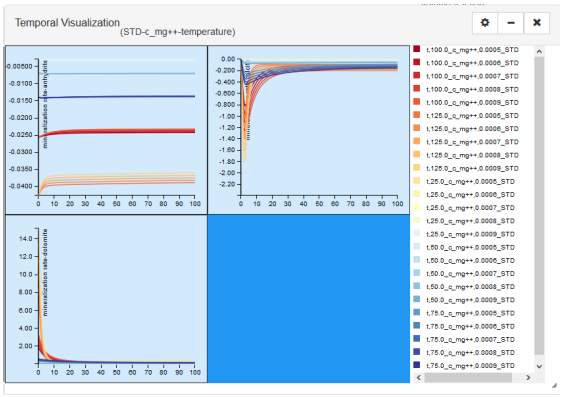
To visualize how a variable behaves along time in different simulation runs, the user can choose to represent them in a multiple series chart that can help users observe how each simulation evolves. Each line in a chart corresponds to a simulation run. If the user selects more than one variable, the charts are placed following a grid layout (Fig. 4.7a. It was used because is what the users are used to, but in the future we want to test other techniques with them as an improvement of the system.

In the temporal visualization, the user can select a time instant, and other existing views are updated with the values of variables at that time instant. If no time instant is selected, values at time instant 0 are shown in the spatial domain visualization, for example.

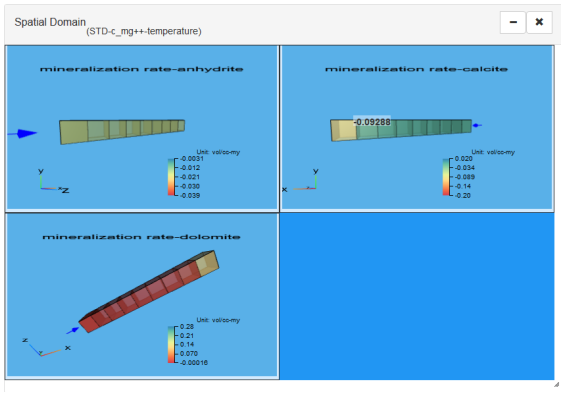
4.2.2.2 Spatial Domain Visualization

The spatial domain of the simulation is an important, almost mandatory information to be displayed, because it represents how different lithotypes are distributed across space, which may provide hints about how they affect the simulations behavior. We adopted a 3D domain visualization even with 1D simulation domains. The idea of using 3D to visualize such spatial domains is because geologists are used to interact with 3D representations of geological models.

Figure 4.7: Temporal and spatial domain visualization techniques.



(a) Temporal visualization: multiple series chart showing the mineralization rate of anhydrite, calcite and dolomite varying along time



(b) Spatial domain visualization showing a 1D domain, color coded with the mineralization rate of anhydrite, calcite and dolomite

In the spatial domain view, we first show the division in cells, even the user has not selected any variable to visualize. When the user selects a variable, the cells are displayed using colors that map the mean value of that variable for all the simulations in a certain instant of time. Although displaying such approximate value does not have a meaning, we aimed to provide some hint about the behaviour of the variable along the whole simulation.

By default, when the user does not interact with the spatial domain view, the system considers the first cell as the selected cell (the cell marked by the arrow (see Fig. 4.7b). As the user interacts and selects multiple cells, the other views display the mean value for the variables in the selected cells.

In terms of visualization techniques, we used a simple grid visualization with color mapping the average values of all selected simulations for each cell because in the simulator GUI development we used bar chart to represent the percentage of solids in a mineralogy, but it was claimed that it can confuse geologist because it can think about spatial domain, so if we used a spatial-temporal techniques (e.g. separate the cells in time instants and give a color to each time instant representing the average of the variable in the time instant for that cell) it may confuse them, so we prefer to keep simple for the user.

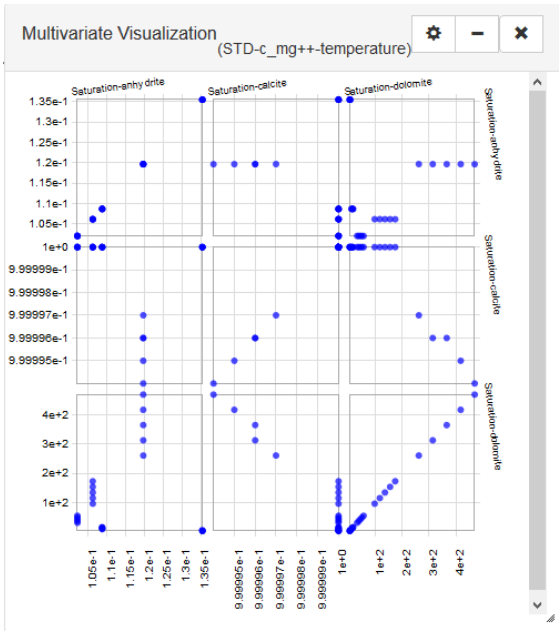
In our system we have a technical limitation from the simulator which is treating only rectilinear grids and only the 1D simulations are well concept, so in all the examples used we are going to use only rectilinear grids in 1D simulations, but in the future we expect to improve with different.

4.2.2.3 Multivariate Visualization

GEVis provides a multivariate visualization in the form of a scatter plot matrix (Fig. 4.8) or parallel coordinates. Scatter plot matrix is useful for guessing about possible correlation between variables. In the scatter plot matrix, each marker is a different simulation, while in parallel coordinates simulations are represented by lines. By default, it is shown to the user the scatter plot matrix, but using the configuration button in the window, they can change to parallel coordinates plot.

By default, all simulations are shown in this view, but by using brushing, the user can select a subset of simulation runs, and the other views are update accordingly, showing only data values from the selected simulations.

Figure 4.8: Scatter plot matrix showing the mineralization rate of anhydrite, calcite and dolomite of 25 simulations



4.3 Remarks

The proof-of-concept prototype of GEVis provides visualization techniques for supporting the three main tasks our users pointed out in the interviews.

To discover what and how parameters influence simulations within an ensemble, the user can observe the temporal visualization comparing the variables behavior along time for the different runs. Each simulation run may have different values for input parameters and the user can observe the behavior of the output

values of the selected variables.

The comparison of different ensembles can be performed in the same way. The user just have to select simulation runs from different ensembles in the GEVis configuration panel.

Finally, to verify which simulation conditions approximate the nowadays conditions, from a geological point of view, the user has to select the appropriate variables related to minerals and visualize the last time instant.

5 CASE STUDY

Although we provided light evidences that the users can perform their elicited high levels tasks with GEVis, in this chapter, we use synthetic simulations trying to provide stronger evidences that our design is able to support users' tasks.

5.1 Hypothetical Ensembles

A problem we encountered in this work was to obtain real reservoir data that would be necessary to build simulation cases. Usually, oil companies do not publicize reservoir data because they are of private interest, and finding (and compiling) such data in the literature is a hard task and demands expert knowledge. To avoid any problems, we used hypothetical data which were used to help geologists to understand how changes in concentration affect simulations.

Our hypothetical data set is composed by two simulations, and we considered each of these simulations candidates for being ensembles. Moreover, since they were batch (0D) simulations, we also converted them to 1D by artificially extending the spatial domain.

In all the ensembles we varied the temperature between 25°C and 125°C and varied concentration of solute Mg^{++} from 0.0005 mol/L to 0.0009 mol/L. We considered a domain of 10 cells, each one with 10 m of thickness, in a depth of 1500 m and a water inlet flux of 5cm/year, entering the same water composition. The mineralogy we used was the same for all simulations (Table 5.1).

Table 5.1: Description of the mineralogy used for all simulations

Solid	Volume Fraction (%)	Grain Diameter (mm)
Anhydrite	15	0.01
Calcite	40	0.01
Dolomite	10	0.01

Table 5.2: Water composition used for all simulations of the first ensemble

Solute	Concentration (mol/L)
pH	7
Total C	0.01
SO_4--	0.05
Mg^{++}	0.0005 to 0.0009

Table 5.3: Water composition used for all simulations of the second ensemble, changing the pH

Solute	Concentration (mol/L)
pH	6
Total C	0.01
SO_4--	0.05
Mg^{++}	0.0005 to 0.0009

5.2 First Task: Parameter Influence

The temperature has a great influence in the simulation results (as stated in Section 3.1.1), but the typical user does not know how the variation of Mg^{++} may influence the simulation. So, he can use the first ensemble for exploring this case.

In Fig. 5.1, one can visualize for Anhydrite, the greater the temperature, faster is its dissolution. The same behavior we notice in calcite, but for dolomite it precipitates more in higher temperature. To analyze better how the variation of concentration of Mg^{++} influences the simulation, one can select different temperature intervals in the multivariate visualization and analyze each plot (Fig. 5.2 and Fig. 5.3). The user notices that for Anhydrite and for calcite in higher concentration of Mg^{++}, faster is their dissolution, and for dolomite as higher is this concentration, higher it precipitates. This can be explained as follows: since dolomite reaction has Mg^{++} to form it, a higher concentration of Mg^{++} accelerates dolomite formation, and it also uses the other solids in this process.

5.3 Second Task: Verify Similarity with Nowadays Conditions

Our typical geologist receives (from the lab) data about a probable condition in the past regarding solid composition and water condition. Since this is a guess, the user needs to test it under different conditions to try to find stronger evidences about how was the most probable conditions in the past for describing the rock formation in a certain area.

Since our case is hypothetical, we can not actually use it to verify real conditions. For example, consider the geologist is using these simulations as tests for a more precise simulation. He simulates only 100 years, instead of thousands or million years. He expects the Anhydrite to precipitate instead of dissolve: he observes that the trend in the first ensemble is wrong, so he needs to change another

Figure 5.1: Volume fraction of each solid (represented by one chart each) through time for all simulations of the first ensemble

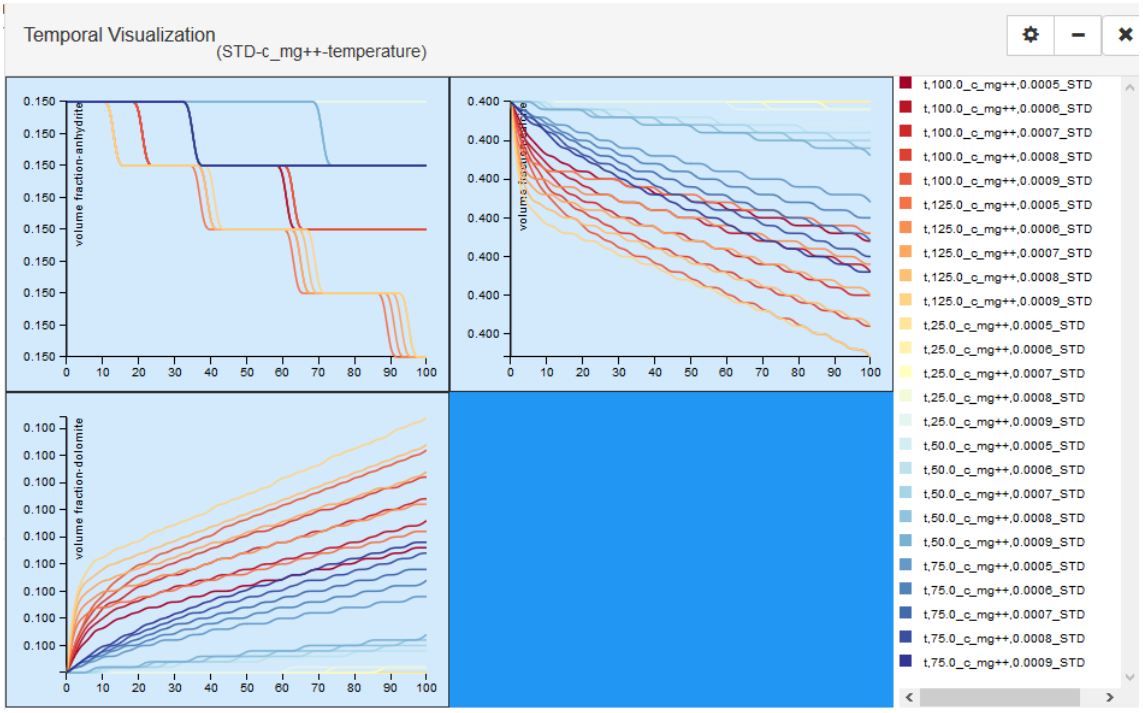
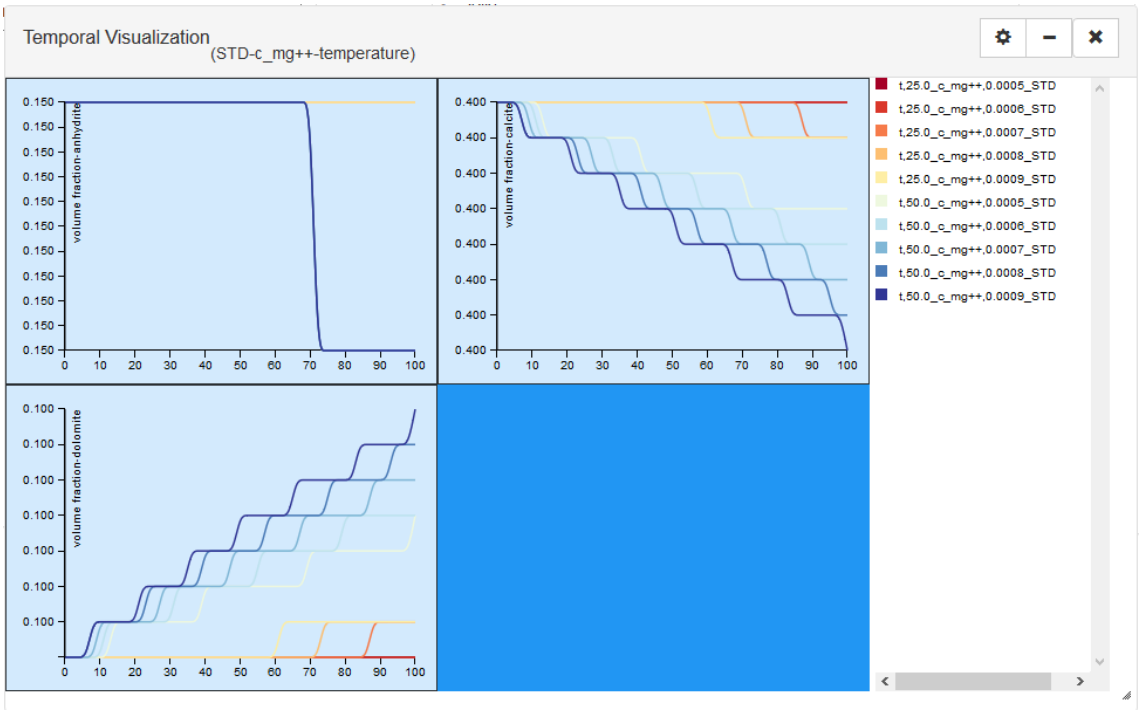


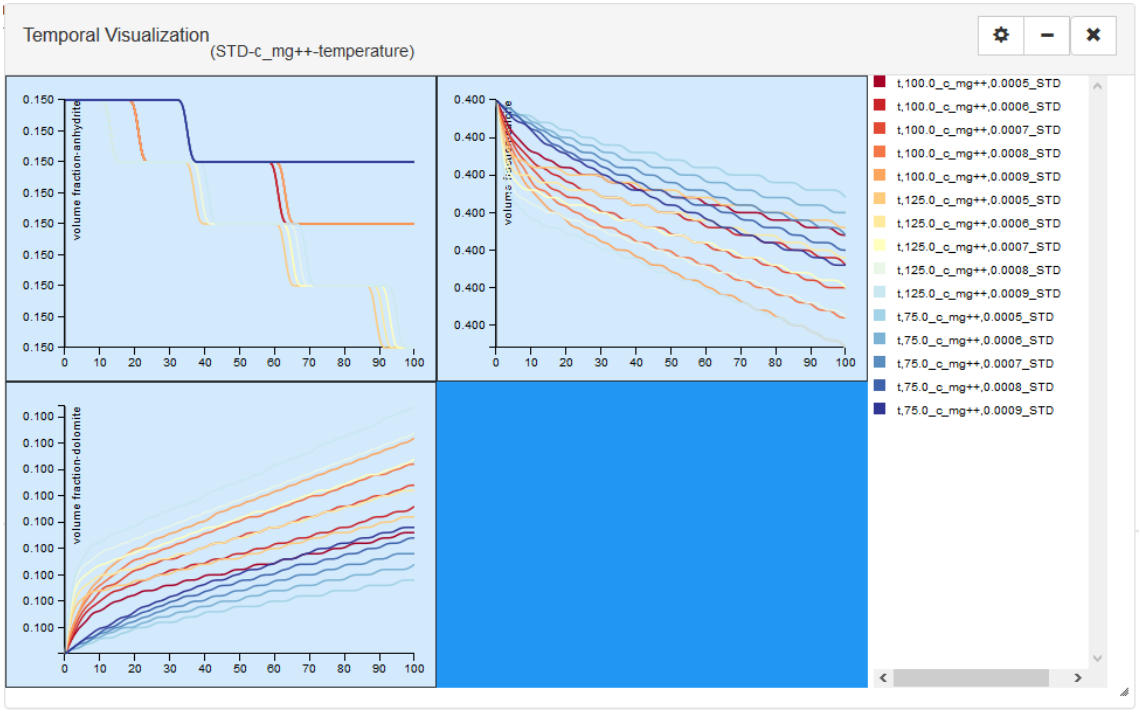
Figure 5.2: Volume fraction of each solid (represented by one chart each) through time for all simulations of the first ensemble selecting 25°C and 50°C



parameter to try to reach the expected condition.

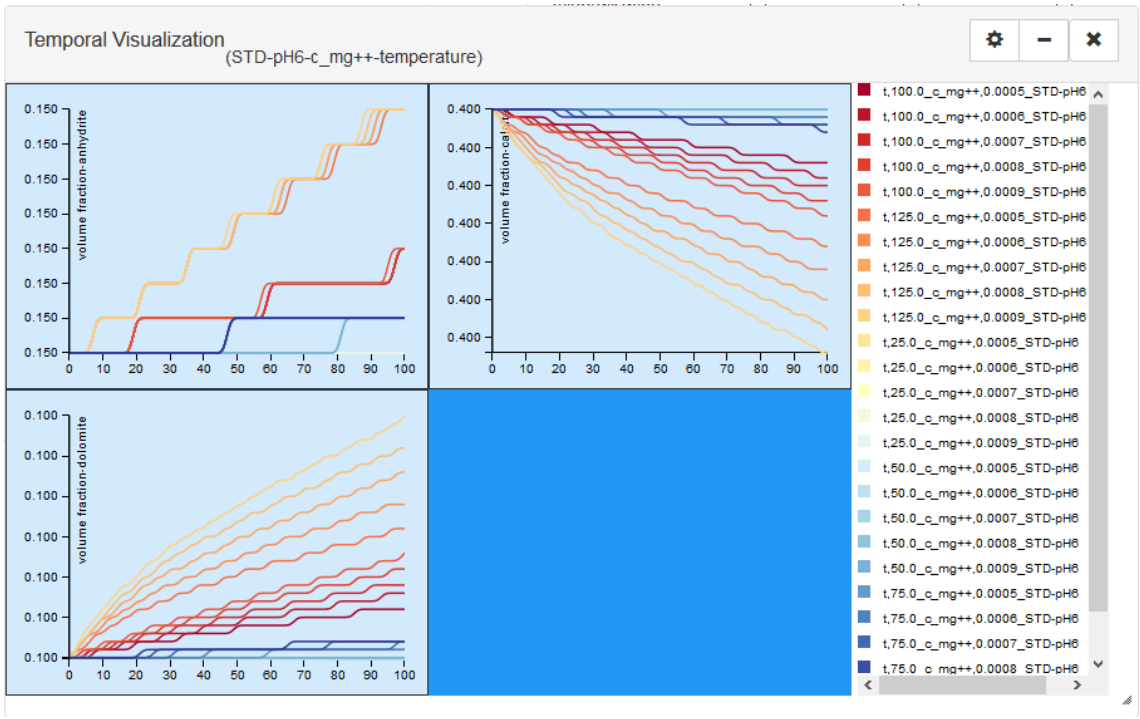
The user then analyzes the second ensemble which was run using different pH values. Analyzing Fig. 5.4, he observes that precipitation of Anhydrite is different from the first ensemble. He can assume that, in this case, the second ensemble is a

Figure 5.3: Volume fraction of each solid (represented by one chart each) through time for all simulations of the first ensemble selecting 75°C to 125°C



strong candidate for describing the conditions needed for a more precise simulation.

Figure 5.4: Volume fraction of each solid (represented by one chart each) through time for all simulations of the second ensemble

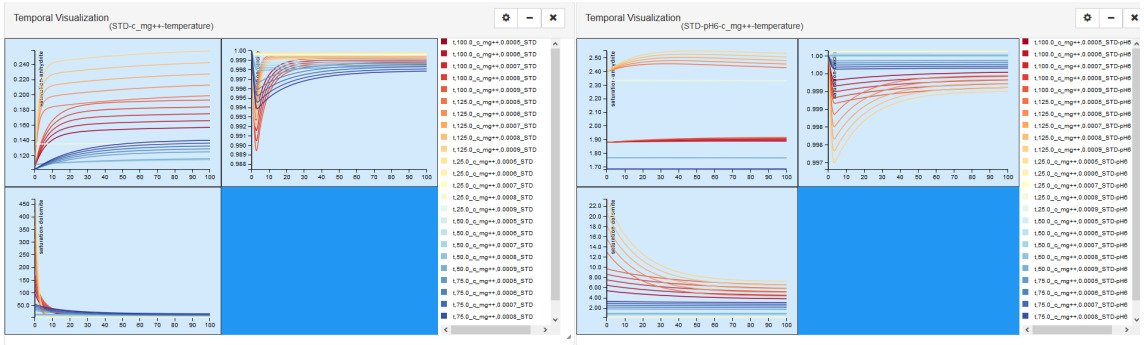


5.4 Third Task: Comparing Ensembles

Now, we consider that the user wants to compare both ensembles to understand, for example, why the pH 6 simulation runs had the different behavior presented in Section 5.3. The user may want to compare them side by side. The windowed design of the application allows it in a easy way (as in Fig. 5.5). The user can visualize one run at a time, but this requires a large cognitive effort and more steps to finally compare them (for example, printing the window).

In this example, the user can analyze that the lower pH value increases the saturation of Anhydrite and decreases the saturation of dolomite. So, this way the user can say Anhydrite is using some solutes dolomite was expected to use, slowing dolomite precipitation (in reactions, they both share solutes based on Carbon).

Figure 5.5: Comparing the saturation of each solid (represented by one chart each) through time of both ensembles



6 EVALUATION WITH EXPERT USERS

While in the previous chapter we developed a case study to show the use of our tool in scenarios related to the users' tasks that we gathered from interviews with users, in this chapter we present an evaluation we conducted with four expert users. Two of them were interviewed in the beginning of the work.

6.1 Evaluation Process Design

We invited four subjects for using GEVIs in typical analysis tasks. Three of them have a B.Sc. degree in Geology: User1 and User2 have PhD degrees and User3 has a MSc deegree. User1 is also an expert in Geochemistry, with experience in developing geochemical simulators. User2 and User3 have experience in using geological computational tools. User4 has a BSC degree in Chemistry and a PhD in Geology, and has experience in using geological tools and geochemical simulators. All users experimented GEVIs with the ensembles explained in section 5.1.

All users, except User4, performed the evaluation in the same room and with the same computer, one at a time. User4 did the experiment remotely since GEVIs is a web-based application.

Each user was invited to first read about how the ensembles were created and their mineralogy and water composition. After, they were allowed to use the system freely to evaluate also if the tasks provided through the informal interview (as explicited in section 4.1.2) are really important to them and if they can be achieved. Since currently they have different activity profiles, they may want to visualize different aspects of the data. For example, geologists concentrate more on understanding solid data, while geochemists concentrate also in understanding the solutes relation with solid precipitation and dissolution). In average, the users spent 30 minutes using the system. We did not recorded their interactions with the system, because we thought using something to record their actions would increase their stress using the system, and it could bias the evaluation.

After the use of GEVIs, the expert users answered two questionnaires: (i) a System Usability Scale (SUS) questionnaire (BROOKE et al., 1996) (Table B.1) and (ii) a questionnaire evaluating the tasks they can perform with the system and the visualization techniques (Table B.2). In all questionnaires, expect Q6 and Q7

in the second questionnaire, use Likert scale (LIKERT, 1932) for the answers, with 1 being "Strongly disagree", and 5 "Strongly agree". Questions Q6 and Q7 asked for a textual answer. We used the SUS questionnaire as a metric to define the usability of the system and the specific questionnaire to evaluate the tasks the users could perform in the system and also evaluate the visualization techniques used. These two questionnaires can be related to help to understand in which tasks we can improve usability to make the task more accessible to the user.

6.2 Evaluation Results

In this section we are going to discuss the results of the evaluation for each questionnaire. As the order the subjects answered in the evaluation questionnaire, first we are going to discuss the SUS questionnaire results and then the system specific questionnaire results.

6.2.1 SUS Results

To obtain the SUS score, we calculated the individual score for each one of the 4 subjects, and after we calculated the average of all subjects. The answers of all subjects for each SUS question is shown on Fig. 6.1, and the SUS score of all the subjects is in Fig. 6.2.

To evaluate the quality of system, we adopted Bangor's ratings (BANGOR; KORTUM; MILLER, 2009). The average SUS score of all subject was 70,625, while the minimum score was 57.5 and the maximum score was 90. According to Table 6.1, this average is in the *Good* category (considering we are above the standard deviation of *Ok* and in the standard deviation of *Good*), but below the average of *Good* by little. We interpret this result as a good result of our work, because data to be visualized is complex and the users had very few time to get acquainted with the system. Another aspect we can consider positive is that the minimum score we obtained was *Ok*, and the maximum score was an *Excellent*.

Analyzing individually the results of each question, we observe answers for SUS2 ("I found the system unnecessarily complex.") SUS3 ("I thought the system was easy to use"), which are related to the complexity and the easy of use of the

Figure 6.1: Chart showing the answers of all users in the SUS questionnaire

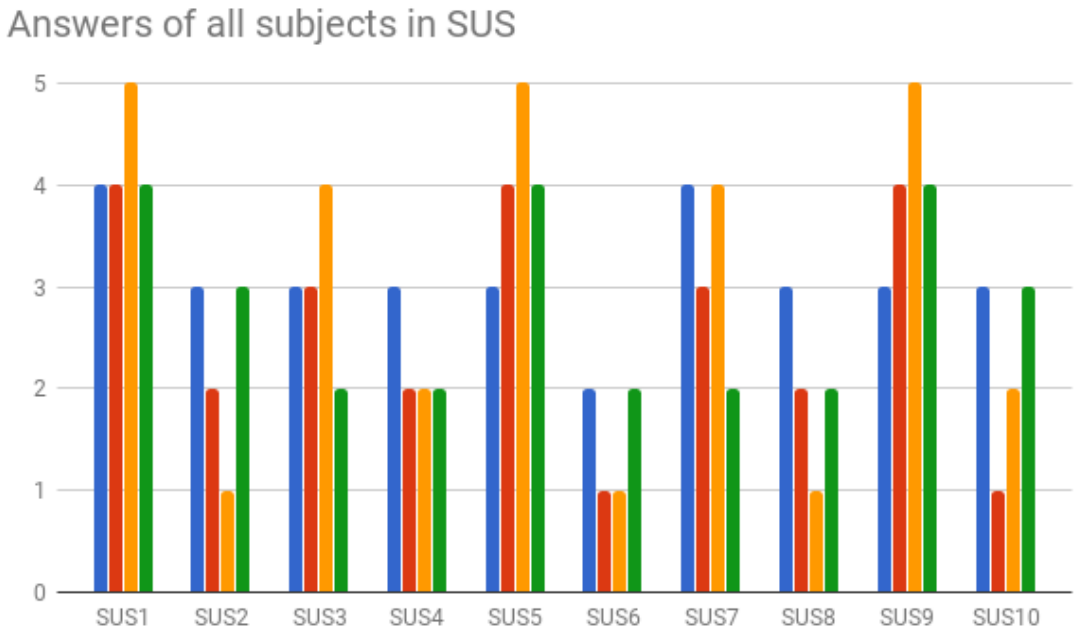
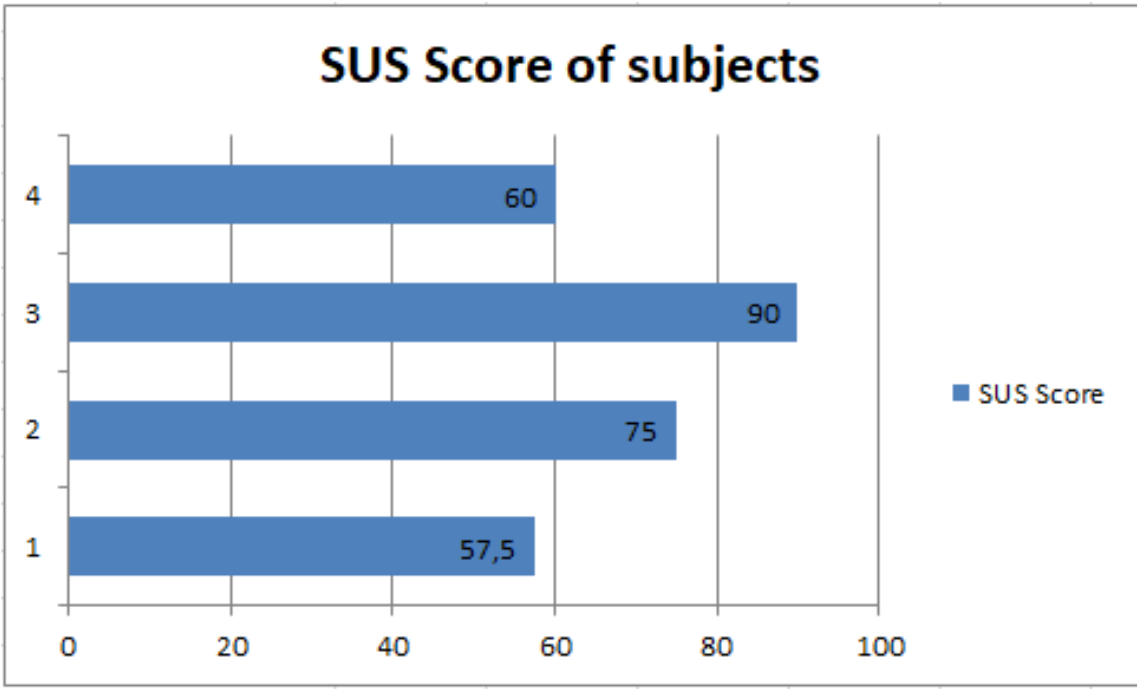


Figure 6.2: Chart showing the SUS score of each subject



system, respectively. There are some negative aspects shown by the answers, which need further evaluation to verify if the difficulties of the users were from the data itself or from some design mistake. For questions SUS7 ("I would imagine that most people would learn to use this system very quickly") and SUS10 ("I needed to learn a lot of things before I could get going with this system"), regarding learnability, GEVIs obtained a *Good* result, but there is also a need to evaluate if these answers

Table 6.1: SUS Score for the Adjective Ratings as in (BANGOR; KORTUM; MILLER, 2009)

Adjective	Mean SUS Score	Standard Deviation
Worst Imaginable	12.5	13.1
Awful	20.3	11.3
Poor	35.7	12.6
OK	50.9	13.8
Good	71.4	11.6
Excellent	85.5	10.4
Best Imaginable	90.9	13.4

come from what a user need to learn to understand the data or if the user interface does not help the users to use it, even though their are experts in the application domain.

6.2.2 System Specific Questionnaire

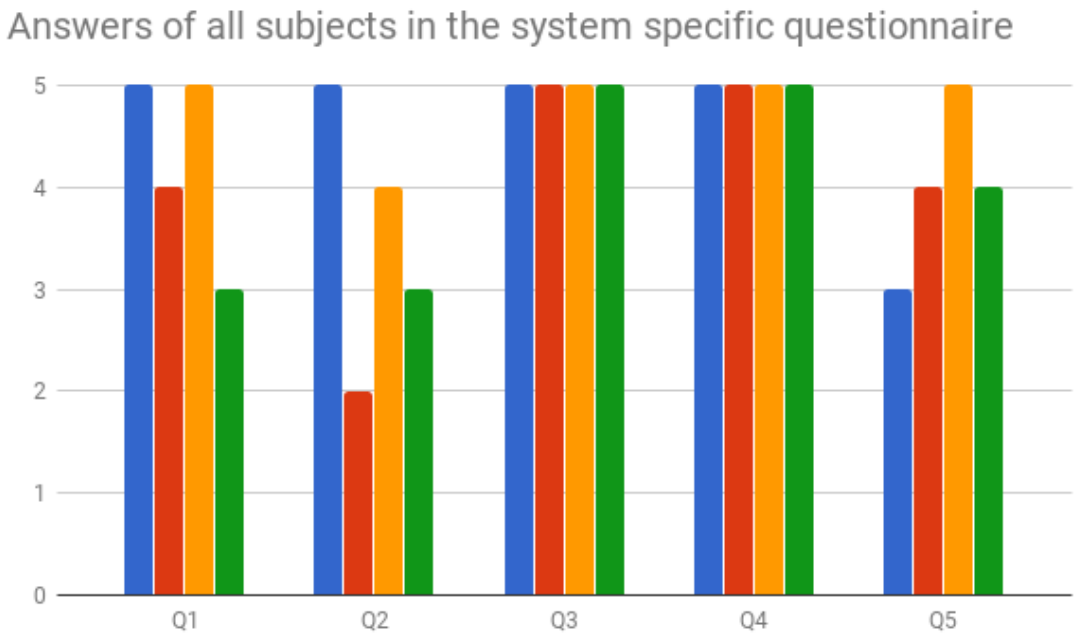
The main purpose of this questionnaire was to get a feedback from users regarding if they think they would be able to perform their tasks of analyzing the ensembles with the system. We also aimed at evaluating if the visualization techniques helped them to understand the simulations. The questions of this questionnaire are shown in Table B.2.

Analyzing the answers (Fig. 6.3), we observe a huge disparity in Q2 ("I think I can find out what are the simulations that approximate the expected results"). This can be a consequence of not using real data, so we do not have expected results, but expected behavior. We need to study how we can, visually, help the users to find the expected results or behavior (e.g. creating a filter of expected values or if a solid precipitates or dissolves) in order to reduce this probable problem.

In Q1 ("I think I can find out the influence of some characteristics in the simulations"), we have some disparity in answers: it was commented that the system is very good to visualize how varying one parameter can influence the simulations, varying two parameters is Ok, but varying three may be difficult, so we need to improve this kind of visualization. One possibility is separating the simulations in groups by the variable value and visualizing the aggregation of simulations using this value.

Q5 ("I think the proposed visualization techniques are enough to visualize

Figure 6.3: Chart showing the answers of all users in the system specific questionnaire



the results") reflects the situation that some visualization techniques are missed by users in the system. They commented two aspects to improve in the visualizations:

1. Comparing the same variable (e.g. saturation) for different minerals in the same plot and

2. Finding some visual metaphor to understand the mineralogy as all in time and space.

For the first aspect, we can use the idea of merging plots by dragging one into another. For second one, this was suggested as future work.

6.3 Final Remarks

This chapter described the evaluation of GEVis by expert users. Although a previous tool (DiagenViz) was developed considering suggestions from expert users, and this work was built based on that experience, GEVis aims at helping the analysis and visualization of simulation ensembles, where the data sets are larger and more complex than in previous work. Along the case study and the experiments performed by users, we collected several suggestions and ideas for improvements, which we will discuss in the next section.

Since User4 tested remotely and we do not put an analyzer of user actions in the system, we do not have information about its use, just the results. But User1 tested the system for about 1 hour, User2 tested for about 15 minutes and User3 tested for about 30 minutes. Other thing to consider is talking with them after the test and their answer to the questionnaire, they said the system was good, has some minor problems (bugs and some features they thought interesting but it was available to use) but it has an huge potential to be used in the future.

7 CONCLUSIONS AND FUTURE WORKS

Visualization of simulations ensembles is a complex problem, due mainly to the complexity of data, considering that they are multidimensional, time evolutive, multivariate and multivalued. Providing methods to help users to explore such data, and make it easier the understanding of the several facets data might have, is still a big challenge. We addressed the specific problem of ensembles of geochemical simulations results.

First we researched about how other geochemical simulation systems deal with visualization and we developed, as a previous work, an user interface to visualize single geochemical simulation. Then we researched about visualization of multiple simulation results to figure how we could solve the problem of visualizing multiple geochemical simulations. We used SLR method to review the works in the area.

We designed a system to visualize these ensembles, based on a windowed design to make it flexible for the user to place the visualizations in the workspace and interact with them, using what we found in the SLR we made with self experience from previous work. We evaluated our system based on two methods: a case study, where we analyzed user tasks and discussed how a user would accomplish such tasks with our tool, and an experiment with four expert users. In this experiment, they found the system adequate for use, but pointed out that it still needs some refinements for achieving the status of a good or excellent system.

As a suggestion, GEVis can be used in two situations: (i) to learn about geochemistry, like how some solute concentration influences solid saturation, and (ii) to find out the conditions in which reservoirs are formed along the years.

Regarding immediate future work, we need to analyze again the results from the experiments performed by the expert users for improving the current prototype. Next, we need to test other visualization and analysis techniques, like dimensional reduction techniques. These would help users to understand in a faster way how variables in an ensemble correlates to each other, and would also help them to understand common behaviors across different simulations.

As a subsequent work, we would like to apply machine learning methods to facilitate user tasks in terms of visual feedback and data processing.

REFERENCES

ALABI, O. S. et al. Comparative Visualization of Ensembles Using Ensemble Surface Slicing. **Visualization and Data Analytics**, v. 8294, p. 1–12, 2012. ISSN 1996-756X. Disponível em: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1345538>.

ALI, S. A. et al. Diagenesis and reservoir quality. **Oilfield Review**, v. 22, n. 2, p. 14–27, 2010.

ANDRIENKO, G.; ANDRIENKO, N. Coordinated multiple views: a critical view. In: **Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV '07. Fifth International Conference on**. [S.l.: s.n.], 2007. p. 72–74.

BALDONADO, M. Q. W.; WOODRUFF, A.; KUCHINSKY, A. Guidelines for using multiple views in information visualization. In: **Proceedings of the Working Conference on Advanced Visual Interfaces**. New York, NY, USA: ACM, 2000. (AVI '00), p. 110–119. ISBN 1-58113-252-2. Disponível em: <http://doi.acm.org/10.1145/345513.345271>.

BANGOR, A.; KORTUM, P.; MILLER, J. Determining what individual sus scores mean: Adding an adjective rating scale. **Journal of usability studies**, Usability Professionals' Association, v. 4, n. 3, p. 114–123, 2009.

BATES, M. J. Defining the information disciplines in encyclopedia development. **Information Research**, v. 12, n. 4, p. 4–12, 2007.

BENSEMA, K. et al. Modality-driven classification and visualization of ensemble variance. **IEEE Transactions on Visualization and Computer Graphics**, v. 22, n. 10, p. 2289–2299, Oct 2016. ISSN 1077-2626.

BISWAS, A. et al. Visualization of time-varying weather ensembles across multiple resolutions. **IEEE Transactions on Visualization and Computer Graphics**, v. 23, n. 1, p. 841–850, Jan 2017. ISSN 1077-2626.

BOCK, A. et al. Visual verification of space weather ensemble simulations. In: **2015 IEEE Scientific Visualization Conference (SciVis)**. [S.l.: s.n.], 2015. p. 17–24.

BREHMER, M.; MUNZNER, T. A multi-level typology of abstract visualization tasks. **IEEE Transactions on Visualization and Computer Graphics**, v. 19, n. 12, p. 2376–2385, Dec 2013. ISSN 1077-2626.

BROOKE, J. et al. Sus-a quick and dirty usability scale. **Usability evaluation in industry**, London–, v. 189, n. 194, p. 4–7, 1996.

BRUCKNER, S.; MÖLLER, T. Result-Driven Exploration of Simulation Parameter Spaces for Visual Effects Design. **IEEE Transactions on Visualization and Computer Graphics**, v. 16, n. 6, p. 1468–1476, nov 2010. ISSN 1077-2626. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5613488>.

BUTNARU, D. et al. Fast insight into high-dimensional parametrized simulation data. In: **2012 11th International Conference on Machine Learning and Applications**. [S.l.: s.n.], 2012. v. 2, p. 265–270.

CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. **Readings in information visualization: using vision to think**. [S.l.]: Morgan Kaufmann, 1999.

CARROLL, L. N. et al. Visualization and analytics tools for infectious disease epidemiology: A systematic review. **Journal of Biomedical Informatics**, v. 51, p. 287–298, 2014. ISSN 1532-0464. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1532046414000914>.

CAVA, R. A. **Abordagens heterogêneas para a exploração interativa de grafos multivariados**. Tese (Doutorado) — Universidade Federal do Rio Grande do Sul, 2017. Disponível em: <http://hdl.handle.net/10183/157522>.

CHEN, H. et al. Uncertainty-aware multidimensional ensemble data visualization and exploration. **IEEE Transactions on Visualization and Computer Graphics**, v. 21, n. 9, p. 1072–1086, Sept 2015. ISSN 1077-2626.

CHENGZHI, Q.; CHENGHU, Z.; TAO, P. Taxonomy of visualization techniques and systems–concerns between users and developers are different. In: **Proceedings of Asia GIS Conference**. [S.l.: s.n.], 2003.

CHI, E. H. A taxonomy of visualization techniques using the data state reference model. In: **Proceedings of the IEEE Symposium on Information Vizualization 2000**. Washington, DC, USA: IEEE Computer Society, 2000. (INFOVIS '00), p. 69–75. ISBN 0-7695-0804-9. Disponível em: <http://dl.acm.org/citation.cfm?id=857190.857691>.

COFFEY, D. et al. Design by dragging: An interface for creative forward and inverse design with simulation ensembles. **IEEE Transactions on Visualization and Computer Graphics**, v. 19, n. 12, p. 2783–2791, Dec 2013. ISSN 1077-2626.

DEMIR, I.; DICK, C.; WESTERMANN, R. Multi-charts for comparative 3d ensemble visualization. **IEEE Transactions on Visualization and Computer Graphics**, v. 20, n. 12, p. 2694–2703, Dec 2014. ISSN 1077-2626.

DUNNE, C. et al. Graphtrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history. In: ACM. **Proceedings of the SIGCHI conference on human factors in computing systems**. [S.l.], 2012. p. 1663–1672.

FELLER, G. J. **Visualização de resultados de simulação de processos diagenéticos**. Bachelor's Thesis — Universidade Federal do Rio Grande do Sul, 2014.

FELLER, G. J.; KLUNK, M. A.; FREITAS, C. M. D. S. Diagenviz: Interactive analysis of simulation results. In: . [S.l.: s.n.], 2015.

FERSTL, F.; BüRGER, K.; WESTERMANN, R. Streamline variability plots for characterizing the uncertainty in vector field ensembles. **IEEE Transactions on Visualization and Computer Graphics**, v. 22, n. 1, p. 767–776, Jan 2016. ISSN 1077-2626.

FEW, S. **Now You See It: Simple Visualization Techniques for Quantitative Analysis**. 1st. ed. USA: Analytics Press, 2009. ISBN 0970601980, 9780970601988.

FOFONOV, A.; MOLCHANOV, V.; LINSEN, L. Visual analysis of multi-run spatio-temporal simulations using isocontour similarity for projected views. **IEEE Transactions on Visualization and Computer Graphics**, v. 22, n. 8, p. 2037–2050, Aug 2016. ISSN 1077-2626.

FURNAS, G. W. Generalized fisheye views. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 1986. (CHI '86), p. 16–23. ISBN 0-89791-180-6. Disponível em: <http://doi.acm.org/10.1145/22627.22342>.

GOSINK, L. et al. Characterizing and visualizing predictive uncertainty in numerical ensembles through bayesian model averaging. **IEEE Transactions on Visualization and Computer Graphics**, v. 19, n. 12, p. 2703–2712, Dec 2013. ISSN 1077-2626.

GUO, H. et al. Coupled ensemble flow line advection and analysis. **IEEE Transactions on Visualization and Computer Graphics**, v. 19, n. 12, p. 2733–2742, Dec 2013. ISSN 1077-2626.

HAO, L.; HEALEY, C. G.; BASS, S. A. Effective visualization of temporal ensembles. **IEEE Transactions on Visualization and Computer Graphics**, v. 22, n. 1, p. 787–796, Jan 2016. ISSN 1077-2626.

HAZARIKA, S.; DUTTA, S.; SHEN, H. W. Visualizing the variations of ensemble of isosurfaces. In: **2016 IEEE Pacific Visualization Symposium (PacificVis)**. [S.l.: s.n.], 2016. p. 209–213.

HÖLLT, T. et al. Visualizing uncertainties in a storm surge ensemble data assimilation and forecasting system. **Natural Hazards**, v. 77, n. 1, p. 317–336, 2015. ISSN 1573-0840. Disponível em: <http://dx.doi.org/10.1007/s11069-015-1596-y>.

HUMMEL, M. et al. Comparative visual analysis of lagrangian transport in cfd ensembles. **IEEE Transactions on Visualization and Computer Graphics**, v. 19, n. 12, p. 2743–2752, Dec 2013. ISSN 1077-2626.

HöLLT, T. et al. Extraction and visual analysis of seismic horizon ensembles. In: OTADUY, M.-A.; SORKINE, O. (Ed.). **Eurographics 2013 - Short Papers**. [S.l.]: The Eurographics Association, 2013. p. 69–72. ISSN 1017-4656.

HöLLT, T. et al. Visual analysis of uncertainties in ocean forecasts for planning and operation of off-shore structures. In: **2013 IEEE Pacific Visualization Symposium (PacificVis)**. [S.l.: s.n.], 2013. p. 185–192. ISSN 2165-8765.

HöLLT, T. et al. Ovis: A framework for visual analysisof ocean forecast ensembles. **IEEE Transactions on Visualization and Computer Graphics**, v. 20, n. 8, p. 1114–1126, Aug 2014. ISSN 1077-2626.

JAREMA, M. et al. Comparative visual analysis of vector field ensembles. In: **2015 IEEE Conference on Visual Analytics Science and Technology (VAST)**. [S.l.: s.n.], 2015. p. 81–88.

KEHRER, J.; HAUSER, H. Visualization and visual analysis of multifaceted scientific data: A survey. **IEEE Transactions on Visualization and Computer Graphics**, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 19, n. 3, p. 495–513, mar. 2013. ISSN 1077-2626. Disponível em: <http://dx.doi.org/10.1109/TVCG.2012.110>.

KEIM, D. A. Information visualization and visual data mining. **IEEE Transactions on Visualization and Computer Graphics**, v. 8, n. 1, p. 1–8, Jan 2002. ISSN 1077-2626.

KITCHENHAM, B. A.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. In: **Technical report, Ver. 2.3 EBSE Technical Report. EBSE**. [S.l.: s.n.], 2007.

KONYHA, Z. et al. Interactive visual analysis of families of curves using data aggregation and derivation. In: **Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies**. New York, NY, USA: ACM, 2012. (i-KNOW '12), p. 1–8. ISBN 978-1-4503-1242-4. Disponível em: <http://doi.acm.org/10.1145/2362456.2362487>.

LIKERT, R. A technique for the measurement of attitudes. **Archives of psychology**, v. 22, n. 140, p. 5–53, 1932.

LIPŞA, D. R. et al. Visualization for the physical sciences. **Comput. Graph. Forum**, The Eurographs Association & John Wiley & Sons, Ltd., Chichester, UK, v. 31, n. 8, p. 2317–2347, dez. 2012. ISSN 0167-7055. Disponível em: <http://dx.doi.org/10.1111/j.1467-8659.2012.03184.x>.

LIU, R.; GUO, H.; YUAN, X. A bottom-up scheme for user-defined feature comparison in ensemble data. In: **SIGGRAPH Asia 2015 Visualization in High Performance Computing**. New York, NY, USA: ACM, 2015. (SA '15), p. 1–4. ISBN 978-1-4503-3929-2. Disponível em: <http://doi.acm.org/10.1145/2818517.2818531>.

LIU, R.; GUO, H.; YUAN, X. User-defined feature comparison for vector field ensembles. **Journal of Visualization**, p. 1–13, 2016. ISSN 1875-8975. Disponível em: <http://dx.doi.org/10.1007/s12650-016-0388-0>.

LIU, R. et al. Comparative visualization of vector field ensembles based on longest common subsequence. In: **2016 IEEE Pacific Visualization Symposium (PacificVis)**. [S.l.: s.n.], 2016. p. 96–103.

LIU, S. et al. Gaussian mixture model based volume visualization. In: **IEEE Symposium on Large Data Analysis and Visualization (LDAV)**. [S.l.: s.n.], 2012. p. 73–77.

MAHFOUD, E.; LU, A. Gaze-directed immersive visualization of scientific ensembles. In: **Proceedings of the 2016 ACM Companion on Interactive Surfaces and Spaces**. New York, NY, USA: ACM, 2016. (ISS Companion '16), p. 77–82. ISBN 978-1-4503-4530-9. Disponível em: <http://doi.acm.org/10.1145/3009939.3009952>.

MATKOVIC, K. et al. Visual analytics for complex engineering systems: Hybrid visual steering of simulation ensembles. **IEEE Transactions on Visualization & Computer Graphics**, IEEE Computer Society, Los Alamitos, CA, USA, v. 20, n. 12, p. 1803–1812, 2014. ISSN 1077-2626.

MATKOVIć, K. et al. Interactive visual analysis of multiple simulation runs using the simulation model view: Understanding and tuning of an electronic unit injector. **IEEE Transactions on Visualization and Computer Graphics**, v. 16, n. 6, p. 1449–1457, 2010. ISSN 10772626.

MATKOVIć, K. et al. Interactive visual analysis of large simulation ensembles. In: **2015 Winter Simulation Conference (WSC)**. [S.l.: s.n.], 2015. p. 517–528.

MCCORMICK, B.; DEFANTI, T.; BROWN, M. Visualization in scientific computing. **Computer Graphics**, v. 21, n. 6, p. 247–307, Nov 1987.

MCKINLEY, S.; LEVINE, M. Cubic spline interpolation. **College of the Redwoods**, v. 45, n. 1, p. 1049–1060, 1998.

MIRZARGAR, M.; WHITAKER, R. T.; KIRBY, R. M. Curve boxplot: Generalization of boxplot for ensembles of curves. **IEEE Transactions on Visualization & Computer Graphics**, IEEE Computer Society, Los Alamitos, CA, USA, v. 20, n. 12, p. 2654–2663, 2014. ISSN 1077-2626.

MüLLER, C. et al. On the utility of large high-resolution displays for comparative scientific visualisation. In: **Proceedings of the 8th International Symposium on Visual Information Communication and Interaction**. New York, NY, USA: ACM, 2015. (VINCI '15), p. 131–136. ISBN 978-1-4503-3482-2. Disponível em: <http://doi.acm.org/10.1145/2801040.2801045>.

MUNZNER, T. **Visualization Analysis and Design**. [S.l.]: CRC Press, 2014. ISBN 9781466508910.

NORDSTROM, D. K. On the evaluation and application of geochemical models. In: **Proceedings of the 5th CEC Natural Analogue Working Group and Alligator Rivers Analogue Project**. [S.l.: s.n.], 1992. p. 375–385.

NOVAIS, R. L. et al. Software evolution visualization: A systematic mapping study. **Information and Software Technology**, v. 55, n. 11, p. 1860–1883, 2013. ISSN 0950-5849. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0950584913001298>.

OBERMAIER, H.; BENSEMA, K.; JOY, K. I. Visual trends analysis in time-varying ensembles. **IEEE Transactions on Visualization and Computer Graphics**, v. 22, n. 10, p. 2331–2342, Oct 2016. ISSN 1077-2626.

PHADKE, M. N. et al. Exploring ensemble visualization. **Proc. SPIE**, v. 8294, p. 130–134, 2012. Disponível em: <http://dx.doi.org/10.1117/12.912419>.

PIRINGER, H. et al. Comparative visual analysis of 2d function ensembles. **Computer Graphics Forum**, v. 31, n. 3.3, p. 1195–1204, 2012. ISSN 01677055. Disponível em: <http://doi.wiley.com/10.1111/j.1467-8659.2012.03112.x>.

PÖTHKOW, K.; WEBER, B.; HEGE, H. C. Probabilistic marching cubes. **Computer Graphics Forum**, v. 30, n. 3, p. 931–940, 2011. ISSN 01677055.

POTTER, K. et al. Ensemble-vis: A framework for the statistical visualization of ensemble data. **ICDM Workshops 2009 - IEEE International Conference on Data Mining**, p. 233–240, 2009.

PRICE, B. A.; BAECKER, R. M.; SMALL, I. S. A principled taxonomy of software visualization. **Journal of Visual Languages & Computing**, v. 4, n. 3, p. 211–266, 1993. ISSN 1045-926X. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1045926X83710153>.

RIBIčIć, H. et al. Visual analysis and steering of flooding simulations. **IEEE Transactions on Visualization and Computer Graphics**, v. 19, n. 6, p. 1062–1075, June 2013. ISSN 1077-2626.

ROBERTS, J. C. State of the art: Coordinated multiple views in exploratory visualization. In: **Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV '07. Fifth International Conference on**. [S.l.: s.n.], 2007. p. 61–71.

ROS, L. F. D. Composition controls on sandstones diagenesis: compr. summ. **Uppsala Diss. Facul. Sci. Tech.**, v. 198, p. 1–24, 1996.

SACHA, D. et al. The role of uncertainty, awareness, and trust in visual analytics. **IEEE Transactions on Visualization and Computer Graphics**, v. 22, n. 1, p. 240–249, Jan 2016. ISSN 1077-2626.

SANYAL, J. et al. Noodles : A Tool for Visualization of Numerical Weather Model Ensemble Uncertainty. **Ieee Transactions on Visualization and Computer Graphics**, v. 16, n. 6, p. 1421–1430, 2010.

SCHARNOWSKI, K. et al. Comparative visualization of molecular surfaces using deformable models. **Computer Graphics Forum**, v. 33, n. 3, p. 191–200, 2014. ISSN 1467-8659. Disponível em: <http://dx.doi.org/10.1111/cgf.12375>.

SHAHIN, M.; LIANG, P.; BABAR, M. A. A systematic review of software architecture visualization techniques. **Journal of Systems and Software**, v. 94, p. 161–185, 2014. ISSN 0164-1212. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0164121214000831>.

SHNEIDERMAN, B. The eyes have it: A task by data type taxonomy for information visualizations. In: **Proceedings of the 1996 IEEE Symposium on Visual Languages**. Washington, DC, USA: IEEE Computer Society, 1996. (VL '96), p. 336–343. ISBN 0-8186-7508-X. Disponível em: <http://dl.acm.org/citation.cfm?id=832277.834354>.

SHU, Q. et al. Ensemblegraph: Interactive visual analysis of spatiotemporal behaviors in ensemble simulation data. In: **2016 IEEE Pacific Visualization Symposium (PacificVis)**. [S.l.: s.n.], 2016. p. 56–63.

SISNEROS, R. et al. Contrasting climate ensembles: A model-based visualization approach for analyzing extreme events. **Procedia Computer Science**, v. 18, p. 2347–2356, 2013. ISSN 1877-0509. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1877050913005498>.

SPLECHTNA, R. et al. Interactive interaction plot. **The Visual Computer**, v. 31, n. 6, p. 1055–1065, 2015. ISSN 1432-2315. Disponível em: <http://dx.doi.org/10.1007/s00371-015-1095-x>.

SPLECHTNA, R. et al. Interactive visual steering of hierarchical simulation ensembles. In: **2015 IEEE Conference on Visual Analytics Science and Technology (VAST)**. [S.l.: s.n.], 2015. p. 89–96.

STEED, C. A. et al. Big data visual analytics for exploratory earth system simulation analysis. **Computers & Geosciences**, v. 61, p. 71–82, 2013. ISSN 0098-3004. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0098300413002215>.

THOMPSON, D. et al. Analysis of large-scale scalar data using hixels. **1st IEEE Symposium on Large-Scale Data Analysis and Visualization 2011, LDAV 2011 - Proceedings**, p. 23–30, 2011. ISSN 2197666X.

TORY, M.; MOLLER, T. Rethinking visualization: A high-level taxonomy. In: **Proceedings of the IEEE Symposium on Information Visualization**. Washington, DC, USA: IEEE Computer Society, 2004. (INFOVIS '04), p. 151–158. ISBN 0-7803-8779-3. Disponível em: <http://dx.doi.org/10.1109/INFOVIS.2004.59>.

W3C. **Web Services Glossary**. 2004. Disponível em: <https://www.w3.org/TR/2004/NOTE-ws-gloss-20040211>.

WANG, J. et al. Multi-resolution climate ensemble parameter analysis with nested parallel coordinates plots. **IEEE Transactions on Visualization and Computer Graphics**, v. 23, n. 1, p. 81–90, Jan 2017. ISSN 1077-2626.

WASER, J. et al. Many plans: Multidimensional ensembles for visual decision support in flood management. **Computer Graphics Forum**, v. 33, n. 3, p. 281–290, 2014. ISSN 1467-8659. Disponível em: <http://dx.doi.org/10.1111/cgf.12384>.

WASER, J. et al. Nodes on Ropes: A Comprehensive Data and Control Flow for Steering Ensemble Simulations. **IEEE Transactions on Visualization and Computer Graphics**, v. 17, n. 12, p. 1872–1881, dec 2011. ISSN 1077-2626. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6064950>.

WENZEL, S.; BERNHARD, J.; JESSEN, U. Visualization for modeling and simulation: A taxonomy of visualization techniques for simulation in production and logistics. In: **Proceedings of the 35th Conference on Winter Simulation: Driving Innovation**. Winter Simulation Conference, 2003. (WSC '03), p. 729–736. ISBN 0-7803-8132-7. Disponível em: <http://dl.acm.org/citation.cfm?id=1030818.1030915>.

WESTALL, J. **MINEQL: A computer program for the calculation of chemical equilibrium composition of aqueous systems**. [S.l.]: Massachusetts Institute of Technology, Water Qaulity Laboratory, 1976.

WHITAKER, R. T.; MIRZARGAR, M.; KIRBY, R. M. Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles. **IEEE Transactions on Visualization and Computer Graphics**, v. 19, n. 12, p. 2713–2722, Dec 2013. ISSN 1077-2626.

WILSON, A. T.; POTTER, K. C. Toward visual analysis of ensemble data sets. In: **Proceedings of the 2009 Workshop on Ultrascale Visualization - UltraVis '09**. New York, New York, USA: ACM Press, 2009. p. 48–53. ISBN 9781605588971. Disponível em: <http://portal.acm.org/citation.cfm?doid=1838544.1838551>.

WOLERY, T. J. Calculation of chemical equilibrium between aqueous solution and minerals: the eq3/6 software package. **Lawrence Livermore National Laboratory, Livermore CA, U.S.A.**, 1979.

WORDEN, R.; BURLEY, S. Sandstone diagenesis: the evolution of sand to stone. **Sandstone Diagenesis: Recent and Ancient**, International Association of Sedimentologists, v. 4, p. 3–44, 2003.

XIAO, J. et al. Hierarchical visual analysis and steering framework for astrophysical simulations. **Transactions of Tianjin University**, v. 21, n. 6, p. 507–514, 2015. ISSN 1995-8196. Disponível em: <http://dx.doi.org/10.1007/s12209-015-2605-7>.

YI, J. S. et al. Toward a deeper understanding of the role of interaction in information visualization. **IEEE Transactions on Visualization and Computer Graphics**, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 13, n. 6, p. 1224–1231, nov. 2007. ISSN 1077-2626. Disponível em: <http://dx.doi.org/10.1109/TVCG.2007.70515>.

YUSOFF, N. M.; SALIM, S. S. A systematic review of shared visualisation to achieve common ground. **Journal of Visual Languages & Computing**, v. 28, p. 83–99, 2015. ISSN 1045-926X. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1045926X1400158X>.

APPENDIX A — SELECTED STUDIES FOR THE SLR

Table A.1: Selected Studies from the SLR

Study Id	Citation
S1	(POTTER et al., 2009)
S2	(MATKOVIć et al., 2010)
S3	(SANYAL et al., 2010)
S4	(BRUCKNER; MÖLLER, 2010)
S5	(THOMPSON et al., 2011)
S6	(WASER et al., 2011)
S7	(PÖTHKOW; WEBER; HEGE, 2011)
S8	(PIRINGER et al., 2012)
S9	(ALABI et al., 2012)
S10	(PHADKE et al., 2012)
S11	(BUTNARU et al., 2012)
S12	(LIU et al., 2012)
S13	(KONYHA et al., 2012)
S14	(STEED et al., 2013)
S15	(GOSINK et al., 2013)
S16	(HUMMEL et al., 2013)
S17	(WHITAKER; MIRZARGAR; KIRBY, 2013)
S18	(SISNEROS et al., 2013)
S19	(GUO et al., 2013)
S20	(COFFEY et al., 2013)
S21	(HöLLT et al., 2013a)
S22	(RIBIčIć et al., 2013)
S23	(HöLLT et al., 2013b)
S24	(SCHARNOWSKI et al., 2014)
S25	(MIRZARGAR; WHITAKER; KIRBY, 2014)
S26	(WASER et al., 2014)
S27	(DEMIR; DICK; WESTERMANN, 2014)
S28	(HöLLT et al., 2014)
S29	(MATKOVIC et al., 2014)

Table A.2: Selected Studies from the SLR

Study Id	Citation
S30	(LIU; GUO; YUAN, 2015)
S31	(JAREMA et al., 2015)
S32	(XIAO et al., 2015)
S33	(SPLECHTNA et al., 2015a)
S34	(MATKOVIć et al., 2015)
S35	(SPLECHTNA et al., 2015b)
S36	(MüLLER et al., 2015)
S37	(CHEN et al., 2015)
S38	(BOCK et al., 2015)
S39	(HÖLLT et al., 2015)
S40	(LIU et al., 2016)
S41	(HAO; HEALEY; BASS, 2016)
S42	(SHU et al., 2016)
S43	(MAHFOUD; LU, 2016)
S44	(BENSEMA et al., 2016)
S45	(FERSTL; BüRGER; WESTERMANN, 2016)
S46	(LIU; GUO; YUAN, 2016)
S47	(FOFONOV; MOLCHANOV; LINSEN, 2016)
S48	(OBERMAIER; BENSEMA; JOY, 2016)
S49	(HAZARIKA; DUTTA; SHEN, 2016)
S50	(WANG et al., 2017)
S51	(BISWAS et al., 2017)

APPENDIX B — GEVIS EVALUATION QUESTIONNAIRES

Table B.1: SUS questions

ID	Question
SUS1	I think that I would like to use this system frequently
SUS2	I found the system unnecessarily complex
SUS3	I thought the system was easy to use
SUS4	I think that I would need the support of a technical person to be able to use this system
SUS5	I found the various functions in this system were well integrated
SUS6	I thought there was too much inconsistency in this system
SUS7	I would imagine that most people would learn to use this system very quickly
SUS8	I found the system very cumbersome to use
SUS9	I felt very confident using the system
SUS10	I needed to learn a lot of things before I could get going with this system

Table B.2: System specific questions for the user evaluation experiment

ID	Question
Q1	I think I can find out the influence of some characteristics in the simulations
Q2	I think I can find out what are the simulations that approximate the expected results
Q3	I think the feature of comparing simulation sets can be useful for my studies
Q4	I think the visualization techniques help me to understand the simulations
Q5	I think the proposed visualization techniques are enough to visualize the results
Q6	If you feel some visualization is missing in the tool, please describe it
Q7	Please, write down any other observations you might have about the system