

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

LEONARDO HAX DAMIANI

***SHPECK* - A Geochemical Speciation
Modelling Software**

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

Advisor: Prof. Dr. Carla Maria Dal Sasso Freitas
Coadvisor: Prof. Dr. Anthony J. Park

Porto Alegre
October 2015

CIP — CATALOGING-IN-PUBLICATION

Damiani, Leonardo Hax

SHPECK - A Geochemical Speciation Modelling Software /
Leonardo Hax Damiani. – Porto Alegre: PPGC da UFRGS, 2015.

79 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul.
Programa de Pós-Graduação em Computação, Porto Alegre, BR–
RS, 2015. Advisor: Carla Maria Dal Sasso Freitas; Coadvisor:
Anthony J. Park.

1. Geochemical Modelling. 2. Chemical Equilibrium. 3. Geo-
chemical Speciation. 4. Water-Rock interactions. 5. Multiphase
System. 6. Water-Rock simulation. 7. Software Engineering.
I. Dal Sasso Freitas, Carla Maria. II. Park, Anthony J.. III. Tít-
ulo.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitor de Pós-Graduação: Prof. Vladimir Pinheiro do Nascimento

Diretor do Instituto de Informática: Prof. Luis da Cunha Lamb

Coordenador do PPGC: Prof. Luigi Carro

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

SUMMARY

LIST OF ABBREVIATIONS AND ACRONYMS.....	5
LIST OF FIGURES	7
ABSTRACT	8
1 INTRODUCTION.....	9
1.1 Objectives of this work	10
1.2 Structure of this work.....	11
2 BASIC CONCEPTS FOR GEOCHEMICAL MODELING	12
2.1 Hydrogeochemistry Principles.....	12
2.1.1 Introduction to Chemical Reactions.....	12
2.1.1.1 Thermodynamic Equilibrium Reactions	13
2.1.1.2 Thermodynamic Equilibrium Constant.....	15
2.1.1.3 Activity of a solute	16
2.1.1.4 Ionic strength	16
2.1.1.5 Activity Coefficient.....	17
2.1.1.6 Saturation Index	18
2.1.1.7 Hydrogeochemistry common units	19
2.1.2 Hydrochemical processes.....	19
2.1.2.1 Acid-Base Reactions	19
2.1.2.2 Complexation	19
2.1.2.3 Oxidation-Reduction Reactions	20
2.1.2.4 Sorption.....	20
2.1.3 Geochemical Modeling.....	21
2.1.3.1 Geochemical Speciation Modeling	21
2.1.3.2 Other Types Of Geochemical Modeling	22
2.2 Summary.....	23
3 REVIEW OF AVAILABLE GEOCHEMICAL MODELING SOFTWARE	24
3.1 EQ3/6.....	24
3.1.1 Input/Output Options	25
3.1.2 User Interaction.....	26
3.1.3 File Formats	27
3.1.4 Software Environment and Installation Procedures	27
3.2 PHREEQC	28
3.2.1 Input/Output Options	28
3.2.2 User Interaction.....	29
3.2.3 File formats	32
3.2.4 Software Environment and Installation Procedures.....	32
3.3 MINTEQ.....	32
3.3.1 Input/Output Options	33
3.3.2 User Interaction.....	35
3.3.3 File formats	38
3.3.4 Software Environment and Installation Procedures	38
3.4 SOLMINEQ.88	38
3.4.1 Input/Output Options	39
3.4.2 User Interaction.....	40
3.4.3 File formats	40
3.4.4 Software Environment and Installation Procedures	41
3.5 Discussion	41
3.5.1 Existing Thermodynamic datasets	43

3.6 Summary.....	46
4 <i>SHPECK</i> - GEOCHEMICAL SPECIATION MODEL	48
4.1 Specification.....	49
4.2 Documentation	50
4.3 Intended users	50
4.4 Architecture	50
4.4.1 Technical Specification	51
4.5 Governing equations	52
4.6 Numerical Method	52
4.7 Algorithm.....	54
4.8 Graphical User Interface.....	56
4.8.1 <i>SHPECK</i> 's GUI	57
4.8.1.1 GUI Interaction	58
4.8.1.2 Chemical phases in the water.....	61
4.9 Database.....	62
4.9.1 Database Technologies.....	62
4.9.2 <i>LLNL</i> thermodynamic dataset parser	64
4.10 Summary.....	65
5 VERIFICATION AND EVALUATION	67
5.1 Case Study	67
5.1.1 Experimental validation of <i>Shpeck</i>	68
5.1.2 Computational comparative study	68
5.2 Database Evaluation	70
5.2.1 Configuration setup.....	72
5.2.2 Time analysis of fetching information	73
5.3 Summary.....	74
6 CONCLUSION	75
REFERENCES.....	77

LIST OF ABBREVIATIONS AND ACRONYMS

a	Activity
γ	Activity coefficient
ASCII	American Standard Code For Information Interchange
β_i	Stability Constant
DBH	Debye-Hueckel
E_a	Activation Energy
Eh	Redox Potencial
G	Gibbs Energy
GUI	Graphical User Interface
HCI	Human-Computer Interaction
I	Ionic Strength
IAP	Ion Activity Product
K	Equilibrium Constant
k_{diss}	Dissolution rate constant
k_0	Pre-exponential (Arrhenius) factor
LLNL	Lawrence Livermore National Laboratory
m	Molality
M	Molarity
MVC	Model-View-Controller
OS	Operating System
pH	Power of Hydrogen
R	Universal Gas Constant
SI	Saturation Index
T	Temperature

UI User Interface

USGS U.S. Geological Survey

LIST OF FIGURES

Figure 3.1	User interface of the PHREEQC for Windows (<i>PfW</i>).....	30
Figure 3.2	Example of the PHREEQC Notepad++ plugin	31
Figure 3.3	PHREEQC software called from the Notepad++ Plugin.....	31
Figure 3.4	PHREEQC software called from the Notepad++ Plugin.....	32
Figure 3.5	<i>MINTEQA2</i> initial menu options	36
Figure 3.6	<i>MINTEQA2</i> main menu	36
Figure 3.7	<i>MINTEQA2</i> Level 1 informations.....	36
Figure 3.8	<i>PRODEFA2</i> 's example of adding a specific aqueous species (Na^+)	37
Figure 4.1	<i>SHPECK</i> overview	49
Figure 4.2	Architecture of the software <i>SHPECK</i>	51
Figure 4.3	High-level algorithm of <i>SHPECK</i>	55
Figure 4.4	Algorithm of the Newton's Method Solver	56
Figure 4.5	Configuration tab	58
Figure 4.6	Chemical phases in the water tab.....	59
Figure 4.7	Results tab.....	60
Figure 4.8	Activity Coefficient Selection Method	60
Figure 4.9	Water composition detailed	61
Figure 4.10	ER Diagram of the database	63
Figure 4.11	Parser algorithm.....	64
Figure 5.1	Log activity ratio of Potassium to Sodium ions using the results from <i>SHPECK</i> ..	68
Figure 5.2	Na^+ log activity comparative study.....	69
Figure 5.3	Cl^- log activity comparative study.....	70
Figure 5.4	Mg^{+2} log activity comparative study	70
Figure 5.5	Ca^{+2} log activity comparative study.....	71
Figure 5.6	K^+ log activity comparative study	71
Figure 5.7	SO_4^{-2} log activity comparative study.....	72
Figure 5.8	Time elapsed in seconds X Reactions Accessed	73

ABSTRACT

A geochemical speciation modelling software is responsible for calculating the distribution of dissolved species between solutes and aqueous complexes, and also computes saturation indexes for different minerals. In this work we introduce *SHPECK*, a software program developed to model geochemical equilibrium systems using the mass-balance conditions based on the phase rule concept (GARRELS; CHRIST, 1965). *SHPECK* composes a system of mass-action equations coupled with equilibrium constraints and solve using *Newton-Raphson* method. Our software accepts any general combination of elements, species, and reactions, allowing the user to create different environments, simulations and, therefore, fully control any aspect and configuration of the model. It provides an interactive user interface as well as the support of a built-from-the-ground database structure that handles the management of the whole thermodynamic data used for the geochemical modeling. Also, we present the basic concepts for geochemical modeling followed by a computer science based review about the available geochemical modeling software. Finally, we validate *SHPECK* by modeling the diagenetic reactions observed in asiliciclastic reservoir and by performing a comparative study with other modelling software package. In addition to this, a database comparison was addressed and the results demonstrate a substantial improvement on the performance by the use of the *SHPECK*'s relational database comparing to the existent approaches.

Keywords: Geochemical Modelling. Chemical Equilibrium. Geochemical Speciation. Water-Rock interactions. Multiphase System. Water-Rock simulation. Software Engineering.

1 INTRODUCTION

Geochemical modelling describes the reactions that occur in a geological medium by representing the chemical reactions and composition of the water. The motivation of this study is to provide a better understanding of the diagenesis process of sediments during the burial process, especially the water-mineral chemical interaction. Applications of geochemical models are essential in several environmental problems, such as calculating the composition of natural waters, measuring flowing groundwater or surface water and the formation and dissolution of rocks and minerals in geologic formations. A geochemical speciation modelling software is responsible for calculating the distribution of dissolved species into solutes and aqueous complexes, and also for computing saturation indexes for different mineral species.

A mathematical model, as described in (SARKER R. A., 2008), requires three major components: decision variables (unknowns of the model); objective function (which needs to be optimized); and constraints (restrictions or limitation of the model). The decision variables depend on the type of the problem considered. The objective function represents the goal of the problem in term of decision variables. The constraints are the restrictions or limitations of the problem.

As stated in (DREVER, 2005), a chemical model is a theoretical construct that permits the calculation of chemical properties and processes using thermodynamic properties of reactions between solutes. Following this idea, a geochemical model is a chemical model developed for geological systems. A set of mathematical expressions represents these natural processes. Thermodynamic data used to establish the reactions and mimic nature are directly responsible for the accuracy of the geochemical system modeled; while the numerical method and computational algorithms determines the precision of results obtained through the modeling.

Prior the start of this work we experimented, reviewed and compared the available existing solutions. We present a thorough analysis of them, which exposes several gaps in the geochemical modeling solutions. This review makes clear the uniqueness of our proposed computational approach for the geochemical modeling problem.

In this work, a software system, named as *SHPECK*, was developed for calculating the chemical equilibrium of a geochemical system using the mass-balance conditions, a process known as *chemical speciation*. The software system receives as input any general combination of chemical elements, species, and reactions, allowing the user to create several different environments, simulations and, therefore, and to fully control any aspect and configuration of the model.

SHPECK provides an interactive user interface that allows the flexible setup of every parameter of the chemical simulation, as well as the support of a built-from-the-ground database structure that handles the management of the whole data used. The flexibility of parameter configuration offered by the visual interface further extend the possibilities of experimentation with the software, and this is a unique characteristic among commercial software products for geochemical speciation. Limitations on the database structure and input capability strongly restrict the potential use of the known geochemical packages. Therefore, *SHPECK* is unique by combining the modern computing technology with efficient numerical method.

The principles of chemical equilibrium calculation rely on the law of conservation of mass (also known as the principle of mass conservation), stated by Antoine Lavoisier, and chemical speciation, which was presented by Garrels in (GARRELS; CHRIST, 1965). The law of conservation of mass establishes that the total mass of an isolated system will remain constant and is independent of any chemical and physical changes taking place within the system. Therefore, the challenge of chemical equilibrium calculations is finding the number of moles that satisfies a system of equilibrium constraints at the moment where forward and reverse reactions rates are the same (equilibrium state). These constraints are organized in a form of linear conservation equations, which may be expressed in the form of either linear algebraic solute and charge balance equations or chemical equations (SMITH; MISSEN, 1983). For the sake of simplicity, in this work I will only deal with chemical equilibrium and not with chemical kinetics calculations since the first one requires only the solution of algebraic equation. The kinetic aspect is planned for integration in the future.

The system of equations drives and represents all the interactions between the components that make up the system, and is simulated by the program. By using the Newton-Raphson's solution method to that set of equations, it is possible to find a solution that satisfies the system of equations. One must note that the initial guess is generated automatically and used as a seed for the iterations. This method requires the usage of a Jacobian matrix and a residual vector during the algebraic calculations.

1.1 Objectives of this work

Soils and aquifers are heterogeneous, subsurface systems composed of a large number of components - dissolved salts, minerals, metals, gases, natural organics, and microorganisms. Geochemical modeling is an important tool for understanding processes occurring in such systems.

The main purpose of this work is to develop a geochemical speciation modelling software following a structured computational approach. The idea of our own geochemical speciation software has emerged as an application where it would be possible to add the benefits of smart user interface and efficient numerical method to the traditionally used method of speciation model development. The result is a compact and versatile tool with an intuitive and interactive user interface.

Therefore, I have made the necessary effort to understand and learn the complex aspects of a geochemical speciation model, and develop a software based on a solid knowledge in computer architecture, algorithms and software engineering.

A comparative study is presented to verify and evaluate *SHPECK*'s results. The software is systematically analyzed across two different approaches: experimentally and computationally. The benefits and the consistent flow (input and output) of *SHPECK* made it a singular tool.

1.2 Structure of this work

The rest of this report is structured as follows. In chapter 2, I present an overview of the basic concepts needed and technical concepts involved. Chapter 3 shows a thorough analysis and review of the commercial software available. Chapter 4 presents the *SHPECK* implementation with a detailed description of the whole system: design options; mathematical treatment; implementation and graphical user interface (*GUI*); algorithm validation and complexity; architecture and organization of the software as well as the database; data-flow; and iteration control. Chapter 5 presents a study case that validates *SHPECK*, and makes a broad observation of the approach used in this work. Chapter 6 brings the conclusion of this work.

2 BASIC CONCEPTS FOR GEOCHEMICAL MODELING

Applying Computer Science to solve problems and create solutions in different areas requires redefining obstacles outside normal boundaries and generating a new understanding of complex situations by thinking across two or more academic disciplines.

To develop this work, we had to delineate common goals for the different profiles that would take part on it, all of them with a clear view of their roles and with a noiseless communication in any direction. Although for the completeness of the text we should have included an introduction to the computer science aspects involved in building the geochemical model, we restrain ourselves to introduce the basic concepts of the application domain, i.e., geochemistry. The computational concepts and tools used in the development are addressed in Chapter 4

In the next section, we explain the essential geochemistry principles: an introduction to thermodynamics and geochemical processes. To finish, we focus on the geochemical modeling. If the reader feels comfortable with these topics, we recommend that you proceed to Chapter 3.

2.1 Hydrogeochemistry Principles

2.1.1 Introduction to Chemical Reactions

In thermodynamics, equilibrium is a state of dynamic balance where the ratio of the product and the reactant concentration is constant. There are three general approaches to calculating the composition of a solution at equilibrium (PETRUCCI, 2007).

1. Manipulation of equilibrium constants (K): The final concentrations are achieved by mathematical handling of the equilibrium constants; the idea is to express all the parts in terms of the measured equilibrium constant and initial conditions. Thermodynamics databases contain the values for the equilibrium constants obtained through experiments. Demonstration of this can be found in (KEHEW, 2000). The disadvantage of this method is that it may never converge when using this method for a huge number of reactions.
2. Gibbs Energy of the system: At equilibrium, the Gibbs Energy (G) is at a minimum. When the object of the study is a closed system - no mass neither entering nor leaving - the total number of atoms of each element will remain constant, therefore, achieving the minimum free energy. Due to the complexity in demonstrating how this method works, it will be suppressed here. An interesting algorithm for equilibrium calculation that uses

Gibbs energy is described in (LEAL et al., 2015). One of the disadvantages of this method lies in the effect of species that appear only in tiny quantities at equilibrium.

3. Manipulation of mass-balance: The total concentration of species that compose the system is the basis for this method. Smith (SMITH, 1980) explains this stoichiometric formulation approach. This method takes into account the stoichiometric approach among the species, which generates a system of non-linear mass-action equations. Mass-balance manipulation is the method chosen for this work, and the details are explained further in section 4.

Stoichiometric approaches have two general advantages over non-stoichiometric: in the case of real systems and for multiphase problems - in which singularities can occur in the linear equations (SMITH, 1980). It is important to remind that all the methods described above are equivalent, and can be verified in (ZEGGEREN; STOREY, 1970).

2.1.1.1 Thermodynamic Equilibrium Reactions

There are mainly two ways to describe chemical reactions: Equilibrium and Kinetic. Equilibrium is the moment where there is no more chemical energy to alter the distribution of mass between reactants and products in the system. The way to model a reaction depends on its rate: an equilibrium reaction is relatively fast on the mass transport process, while the kinetic reaction is slow. Therefore, when applying an equilibrium model to a reaction, it is assumed that the whole mass transfer happens at the same time when the reactant and product are put together, and this will configure an equilibrium situation. If the reaction rate is slow, it requires a kinetic description of the reaction. In this work, we will address equilibrium reactions (NORDSTROM D., 1986).

Assuming the independent equilibrium reactions:

$$0 \rightleftharpoons \sum_{i=1}^N v_{ji} \alpha_i \quad (j = 1, \dots, M) \quad (2.1)$$

where v_{ji} is the stoichiometric coefficient of the i -th species in the j -th reaction; and M represents the number of reactions and N the number of species, with $M < N$. The sign convention is to assign negative stoichiometric coefficients for reactants and positive for products. Assuming that all of the reactions in the system are in equilibrium, the chemical system must also

satisfy the mass-action equations:

$$K_j = \prod_{i=1}^N a_i^{v_{ij}} \quad (j = 1, \dots, M) \quad (2.2)$$

where K_j denotes the equilibrium constant of the j -th reaction; a denotes the activity of the i -th chemical species. The equilibrium constant depends on the temperature of the system; therefore, the equilibrium constant needs to be calculated according to the temperature of the system.

The driving force of a chemical reaction is related to the concentration of the reacting constituents and the concentrations of the reaction products. The law of mass-action states that any reaction will proceed to the right (dissolution) or to the left (precipitation) until the mass-action equilibrium is achieved. One must bear in mind that in nature it may take years or even thousands of years for that equilibrium to be achieved, and after a disturbance in the system, such as an addition of reactants, removal of products, changes in the temperature or pressure, the system will continue to proceed towards this new equilibrium (if the disturbances are frequent compared to the reaction rate, equilibrium may never be achieved) (FREEZE; CHERRY, 1979). Each of the dissolved species will have one representation of the non-ideal behavior of components in the solution, which is called *activity* and is presented in details on section 2.1.1.3.

Kinetic descriptions are applicable to any reaction but it is necessary to describe reactions that are slow in relation to mass transport. The following reaction has a k_1 and k_2 rates for the forward and reverse reactions, respectively



Each ion has a reaction rate related to the stoichiometry, and is expressed as

$$-\frac{r_A}{a} = -\frac{r_B}{b} = \frac{r_D}{d} = \frac{r_E}{e} \quad (2.4)$$

where a, b, d and e are stoichiometric coefficients of each one of the ions in the reaction. r_A, r_B, r_D and r_E are reaction rates, and they describe the time rate of change of concentration as function of rate constants and concentration. Each one of them express the rate of change at the chosen ion as the difference between the rate at which the component is being used in the

forward reaction and generated in the reverse reaction and is described as

$$r_A = -k_1(A)^{n_1}(B)^{n_2} + k_2(D)^{m_1}(E)^{m_2} \quad (2.5)$$

where n_1, n_2, m_1 and m_2 are empirical stoichiometric coefficients. When there are reactions in parallel or series the rate laws are even more complex. The dissolution rate constant (k_{diss}) of a chemical reaction depends on temperature. The relation between constant and temperature is given by the *Arrhenius equation*, described as

$$k_{\text{diss}} = A * e^{\left(\frac{-E_a}{R*T}\right)} \quad (2.6)$$

where k_0 is the pre-exponential (Arrhenius) factor, E_a is the activation energy, R is the universal gas constant, and T is the temperature in Kelvin. During the development of *SHPECK*, we will not deal with kinetic reactions.

2.1.1.2 Thermodynamic Equilibrium Constant

The *equilibrium constant* (K) is the value of the reaction quotient when the reaction has reached equilibrium. K depends only on the temperature and on the ionic strength of the solution. According to known reactions' equilibrium constant value, it is possible to express K as a temperature-dependent polynomial.

Equilibrium constants are determined by measurements of the relevant concentrations of the species under differing experimental conditions. Concentrations of species can be measured in multiple ways, and the use of these values in modeling requires adjustment to the conditions in the system being modeled. These adjustments, as well as the differences in conditions and different methods for determination, can lead to uncertainty in chemical speciation constants.

Several thermodynamics database are available. Each database includes reaction constants, reaction descriptions, solutes, species, enthalpy values, activity coefficient parameters, etc. During the development of *SHPECK* the *Lawrence Livermore National Laboratory (LLNL)* thermodynamic dataset was selected, it also contains the values of equilibrium constants at maximum of 8 temperatures. These data points were fitted to produce polynomials, which then allows our software to calculate the equilibrium constant to any temperature. Also another source of experimentally measured data is (PALANDRI L. JAMES, 2004).

The polynomial regression is specifically used to calculate the equilibrium constant of the compound at the desired temperature. Polynomial regression is one of several methods of

curve fitting, which is a process of constructing a curve that has the best fit to a series of data points. Polynomial regression is a statistic method that is a form of linear regression in which the relationship between the independent variable x and the dependent variable y is modeled as an n -th degree polynomial

2.1.1.3 Activity of a solute

Activity (a_i) is the "*thermodynamic concentration*" (or informally known as "*effective concentration*"). It is calculated as a product of activity coefficient and concentration (where i means the solute involved):

$$a_i = \gamma_i * m_i \quad (2.7)$$

where m_i is the molar concentration of the solute i and γ_i is the activity coefficient (explained in 2.1.1.5), which is a function of ionic strength (I).

2.1.1.4 Ionic strength

Ionic strength (I) is a measure of the concentration of ions in the solution. Mathematically the ionic strength of the solution is calculated according to

$$I = \frac{1}{2} \sum M_i z_i^2 \quad (2.8)$$

where M is the molar concentration of the specie i having a charge z . When I increases, activity coefficients decrease. In a very dilute solution, activity coefficient is closer to 1.0, and activity is nearly equal to the concentration. The decreasing trend is related to the "cage" of opposite charge particles around ions. There is a reversal of the trend in extremely concentrated solutions (brines), because beyond ionic strength of about $1\text{mol}/L$ there is an increase of activity coefficients with increasing ionic strength. This is related to decreasing amount of free water molecules, because most of water molecules are already bound around the solutes. For a matter of explanation, we will calculate the ionic strength of a CaCl_2 solution (composed by 0.5mol of

Ca^{+2} and 1mol Cl^{-1}):

$$I = \frac{1}{2}(z_{\text{Ca}}^2[\text{Ca}^{+2}] + \frac{1}{2}(z_{\text{Cl}}^2[\text{Cl}^{-1}]) \quad (2.9)$$

$$I = \frac{1}{2}(2_{\text{Ca}}^2[\text{Ca}^{+2}] + (-1)_{\text{Cl}}^2[\text{Cl}^{-1}]) \quad (2.10)$$

$$I = \frac{1}{2}(4 * 0.5 + 1 * 1) \quad (2.11)$$

$$I = 1.5\text{mol/L} \quad (2.12)$$

2.1.1.5 Activity Coefficient

There are different methods to calculate activity coefficient (γ) for ions:

- Debye-Hueckel: This method assumes that ions behave like spheres with charges located at their center points. The ions interact with each other by coulombic forces and the result of their analysis is as follows

$$\log\gamma_i = -Az_i^2\sqrt{I} \quad (2.13)$$

where A is a constant that is a function of temperature, z_i is the ion charge and I is the ionic strength of the solution.

- Davies equations: This is a variation of Debye-Hueckel equation that can be used when the ionic strength is relatively high. The equation is

$$\log\gamma_i = -Az_i^2\left(\frac{\sqrt{I}}{1 + \sqrt{I}} - 0.3I\right) \quad (2.14)$$

- B-dot: This model is presented as an activity model based on an equation similar to Davies and parameterized for solutions up to 3 molal ionic strength.

$$\log\gamma_i = -\frac{Az_i^2\sqrt{I}}{1 + a_iB\sqrt{I}} + \dot{B}I \quad (2.15)$$

where a is the ion size for each specie and A , B and \dot{B} are coefficients that vary with temperature and are present in the thermodynamic dataset used by geochemical modeling codes.

It is important to mention that there are other methods available for calculating activity coefficients, which are not going to be addressed here. Each one of the methods has its advantages and limitations. Debye-Hueckel equations are simple to apply, and is an extensible method for

including new species in the solution due to the fact that it requires a low number of (specific) arguments. Moreover, Debye-Hueckel can be applied to the most important temperatures in the field of aqueous geochemistry, but it works poorly for moderate or high ionic strength solutions. As to dissolution and precipitation, there is clearly a reaction happening during these processes, which means that some reactions are not in equilibrium. Also, pure solids have activity coefficients equal to one.

2.1.1.6 Saturation Index

Saturation index (*SI*) indicates the state of dis-equilibrium with respect to a given mineral. *SI* is expressed as

$$SI = \log(IAP/K) \quad (2.16)$$

When a mineral is in equilibrium within a solution, the *SI* is zero: a negative *SI* indicates undersaturation, and a positive *SI*, supersaturation. The Ion Activity Product (*IAP*) is calculated according to

$$IAP = \frac{[C]^c[D]^d}{[A]^a[B]^b} \quad (2.17)$$

where $[A]$, $[B]$, $[C]$ and $[D]$ are activities of the ions and a, b, c and d denote the respective stoichiometric values. The interpretation of *IAP* is the following:

- $IAP > K$: The reaction is progressing from left to right, producing more products. This state is also described to as supersaturated.
- $IAP = K$: The reaction is in equilibrium, and there is equal flow of reaction to the right and to the left
- $IAP < K$: The reaction is progressing from right to left, producing more reactants. This state is also described as undersaturated.

With the *SI* approach, it is possible to predict the reactive mineralogy from the water composition. If the *SI* for a mineral is less than zero, the aqueous solution is undersaturated with respect to that mineral - which corresponds to the state where the mineral may dissolve in order to reach equilibrium concentrations. If the *SI* is greater than zero, then more mineral may precipitate from the aqueous solution (oversaturated). To conclude, when the *SI* is close to zero (it is ok to consider a small range of values to be in equilibrium) , the water is in near- saturation with respect to that mineral (ALLEY, 1993).

2.1.1.7 Hydrogeochemistry common units

Molarity (M) is defined as mass in moles in 1 liter of solution and molality (m) is defined as mass in moles in 1 kilogram of solution. In dilute solutions, molarity is approximately equal to molality.

2.1.2 Hydrochemical processes

2.1.2.1 Acid-Base Reactions

The importance of acid-base reactions is clearly evident when one understands their influence on the pH. The pH scale represents how acidic or basic is the solution. The pH is described as

$$pH = -\log([H^+]) \quad (2.18)$$

where $[H^+]$ is the activity of the hydrogen ion. Solutions with pH greater than 7 are alkaline or basic, less than 7 are acidic and equals to 7 neutral solutions. The acid component has a tendency to loose protons, while a basic component has tendency to gain protons, and this interaction is called acid-base reactions and is described as



The reaction must be understood as that in the forward reaction, the proton lost by $Acid_1$ is gained by $Base_2$ and, in the reverse reaction, the proton lost by $Acid_2$ is gained by $Base_1$. The strength of an acid or base refers to the proportion of the protons that are lost or gained.

2.1.2.2 Complexation

Complexation is when an ion is formed by combining simpler cations, anions and sometimes molecules. Complexes are dissolved species that consist of two or more simples species. Considering the following reaction



A simple example of complexation for the reaction 2.20 is expressed in mass law form as:

$$K_{MnCl^+} = \frac{[MnCl^+]}{[Mn^{2+}][Cl^-]} \quad (2.21)$$

Each complex compound has a property called stability constant (β_i). It defines how the total concentrations of the components are distributed among the possible compounds (it may include other complexes) in the solution. For more details on this topic we refer to (KEHEW, 2000).

2.1.2.3 Oxidation-Reduction Reactions

Certain reactions involve the transfer of electrons between its components (reactants or products). As a result, there are changes in the oxidation states of one or more of the reactants and products. It is important to stress that the oxidation number is a hypothetical charge that an atom would have if the ion or molecule were to dissociate. This state can be different according to the solution.

In this work, *redox reactions* (as oxidation-reduction reactions are also known) are not going to be addressed. In order to get deeper understanding on this topic, we refer to (PETRUCCI, 2007)

2.1.2.4 Sorption

Sorption describes a process where compounds *stick* to surfaces of solids - which represents the removal of the solute from the solution onto mineral surface. There are three kinds of sorptions:

- Adsorption: The solute can held at the mineral surface as a hydrated species.
- Absorption: The solute can incorporate into the mineral structure at the surface.
- Ion Exchange: When an ion becomes sorbed to a surface by changing places with an ion previously residing on the mineral.

In this work, sorption are not addressed. For detailed information on this topic, we refer to (FREEZE; CHERRY, 1979)

2.1.3 Geochemical Modeling

The 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:** Based on the assumption of thermodynamic equilibrium reached in a relatively short time (no time factor is included in calculation). It takes into consideration only equilibrium reactions.
- **Geochemical Kinetic Models:** It also takes into account kinetic reactions and include the time factor.

As mentioned before, this work will focus on the first one - *geochemical equilibrium models*.

Geochemical equilibrium models can be further divided into three types: speciation models; inverse models (also called mass-balance models); and forward models (also called reaction-path models). Geochemical equilibrium models are considered *batch models* - which are basically closed vessels or reactors.

2.1.3.1 Geochemical Speciation Modeling

Speciation represents modeling based in the equilibrium of a system. A geochemical speciation modeling program calculates the distribution of dissolved species between free ions and aqueous complexes and also saturation indexes for different minerals. Sodium, for example, can be present in water as a free ion Na^+ , and also in the form of complexes with anions:

$$\text{Na}_{total}^+ = \text{NaCl}_{aq} + \text{NaOH} + \text{Na}^+ \quad (2.22)$$

where Na_{total}^+ is total sodium concentration from chemical analysis. Na_{total}^+ is a component (e.g., chemical formula unit used to describe a system) and Na^+ , NaCl_{aq} and NaOH are species (chemical entities which really exist in the system).

Information about the distribution of dissolved species is important, for example, for risk assessment of contamination by metals, because toxicity of metals depends on their speciation in the solution. Carbonate complexes of metals, for example, are less toxic than their free ions.

Field data necessary for input of speciation program are temperature, pH and results from laboratory chemical analysis (results obtained from a sample of the solution of interest).

Details of several available programs are presented and discussed in chapter 3. The development of *SHPECK*, which is our proposal of a geochemical speciation modeling software, is presented and thoroughly discussed in chapter 4.

2.1.3.2 Other Types Of Geochemical Modeling

There are other types of geochemical equilibrium models, as mentioned before. For completeness, we summarize their characteristics below.

- Inverse geochemical modeling This type of models, also known as mass-balance models, are used when chemistry of groundwater and solid phase composition are already known, and reactions that have already happened should be determined. It is used when we have access to 2 hydraulically connected points and composition of solid phase between these points. With these data in hand, it is possible to calculate and produce the reactions that will explain the changes of the water's chemistry. This approach leads to some uncertainties: stoichiometry of minerals in solid phase is not often well known; solution may be non-unique; and programs can produce several possible models for the same input. An interesting work about inverse geochemical modeling is (SHARIF, 2007).
- Forward geochemical modeling This type of models, also called reaction-path models, are used for prediction of water chemistry evolution along a flowline. The initial water chemistry is known and the aim of the program is to predict water chemistry at some point along the flow path. This kind of modeling introduces problems regarding kinetic and adsorption data, which are often missing and frequently limited.

2.2 Summary

- The importance of multidisciplinary problems: The power from the advances in computer science to adapt itself to other areas of knowledge is ever increasing. The best way to push the limits of our work is by redefining obstacles outside our normal boundaries and reach solutions based on new understanding of complex situations. On the other hand, understanding what is a software and how it connects the "*machine world*" to the "*real world*" can be non-trivial. I reinforce the idea that the extension of a software goes far beyond the lines of code and the user interface displayed on the screen. We often find the comparison that building software is somehow like building a house - this certainly is a helpful example to understand everything that is behind a software. Both house and software require: estimating costs, thinking about the requirements, plans, rules, standards, best practices, specifications timelines, reviews, milestones, testing, alterations, handover and warranty.
- Hydrogeochemistry principles: The groundwater's journey and what has happened to it can be discovered using hydrogeochemistry models. Studying water's chemical composition and properties is a complex and interesting task - many different factors take part in this environment. We can mention the equilibrium and kinetic reactions, the activity and the activity coefficient of the solute, the ionic strength of the solution, the saturation index, the equilibrium constant, etc.

3 REVIEW OF AVAILABLE GEOCHEMICAL MODELING SOFTWARE

The first geochemical models date back to the 70's (WESTALL et al., 1976), (WOLERY, 1979). Since then, these models are 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 model results depend on the methods used, and thermodynamic data and theoretical concepts applied. Therefore, it is crucial to verify the results and it is clear that there will be some differences among the results obtained by different software. Among the enormous variety of software available, some of them are developed for batch-type simulations only, while others have transport capabilities. Several of them do not incorporate graphical interfaces and are written in *FORTRAN*, while newer distributions are mainly written in C/C++ and, due to proprietary reasons, code is not distributed with the software. It is important to mention that, even in those who provide an integrated graphical user interface (*GUI*), it is often very tedious and time-consuming to generate input files for simulations.

We critically review and compare, in this article, the software packages that provide speciation modeling: *EQ3/6*, *PHREEQC*, *MINTEQ* and *SOLMINEQ*. We had given special attention to the geochemical modelling approaches and the computer-science related aspects of them, providing the comparison when the information is available.

3.1 *EQ3/6*

EQ3/6 consists of two programs: *EQ3* is a pure speciation code whose results *EQ6* subsequently process. It is a software package for geochemical modeling of aqueous systems written in *FORTRAN77*. *EQ3/6* includes a speciation-solubility solver, which is useful for analyzing groundwater chemistry data, calculating solubility limits and determining whether certain reactions are in states of partial equilibrium or disequilibrium. It also offers a reaction path calculation that models water/rock interaction or fluid mixture. *EQ3/6* supports several thermodynamic data files (these data files contain support for *Davies*, *B-dot*, *Debye-Huckel* equations, as well as support data for standard state and activity coefficient-related). It is developed to run under UNIX, and the full package distribution is not free (it requires a license). The whole work related to *EQ3/6* is described in (WOLERY, 1979), (WOLERY T. J., 1990) and (WOLERY, 1992).

3.1.1 Input/Output Options

The *datafilekey* and *inputfile* are files given to the program as arguments and they must be mutually consistent with the options and methods. For example, if they have different methods for calculating the activity coefficient there will be problems, and the results will be erroneous. In the similar manner, the *inputfile* must be using chemical data (for example, elements, species and compounds) that is known by the *datafilekey*.

Inside each file, there are a series of *blocks* that are combined to support the geochemical speciation. They are presented bellow:

- *Datafilekey*: Title; Miscellaneous parameters (temperature limit, activity coefficient parameters, pressure); chemical elements block; aqueous species block; pure minerals block; pure non-aqueous liquids block; gas species blocks; solid solutions blocks; references blocks
- *Inputfile*: Title; Special basis switches; temperature; pressure option; density; total dissolved salts (TDS) option; electrical balancing option; redox option; basis species constraints; ion exchanger creation flag; ion exchanger compositions; solid solutions compositions; alter/suppression options; iopt options; iopg options; iopr options; iodb options; numerical parameters; ordinary basis switches; saturation flag tolerance; aqueous phase scale factor.

EQ3/6 package produces different outputs depending on the version of software used. We will exemplify the output files in general by using the *EQ3NR* and *EQ6* output formats.

- *EQ3NR*: Two output files are generated: a *pickup* file and the normal output file. The *pickup* file can be used as input to *EQ6* software. The normal output file consists of six blocks: header section; input file echo; recap of input data; iterative calculations; principal results; and finally the end of *EQ3NR* run
- *EQ6*: This program generates three output files: a *tab* output file, a *pickup* output file and the normal output file. The *tab* file contains information that can be used to plot output results. The *pickup* file is the input to *EQ6*. The normal output file of *EQ6* is similar to the *EQ3NR*'s output file.

A small excerpt of an *EQ6* output file is shown in code 3.1. We can see that in this part of the output file, it specifies the components of the current problem with internal ID's for each component, specie, phase and so on.

Code 3.1 – Excerpt of *EQ6* output file

```

...
Entity Date Base Dimension Current Problem
Chemical Elements 81 81 6
Basis Species 201 259 7
Phases 1135 1159 29
Species 3031 3523 0
Aqueous Species 1769 1769 22
Pure Minerals 1120 1120 26
Pure Liquids 1 3 1
Gas Species 93 93 2
Solid Soutions 12 12 0
...

```

3.1.2 User Interaction

In *EQ3/6* the command prompt is used for all the user interaction. There are several functions inside *EQ3/6*; the appropriate command will trigger each one of them. From the existing software, there are *EQ3NR*, *EQ6* and *EQPT* just to name a few. The user must enter the command from the keyboard and must use this "command prompt". By pressing "CTRL+C" at any time the execution stops, literally "breaking" the process. For example, *EQ3* is run by commands of the form shown in code 3.2.

Code 3.2 – Running *EQ3* in *EQ3/6* package

```
>runeq3 datafilekey inputfile(s)
```

In this command, *datafilekey* and *inputfile* are arguments, being the former a three-character identification associated to which database should be used, while the latter is specifically the name of the input file, which can be more than one. Depending on which program from the package the user is using, it generates two or more output files (always in the *ASCII* format). As mentioned, the input file is entered in the program as an argument. Any regular text editor is sufficient to create or modify an input file (although it is not recommended that the user create an input file from scratch). There are several pre-existing input templates available and, if none of them matches the need of the user, *EQ3/6* recommends that the user generate a new one by copying existing blocks from any provided template.

The input files can contain instructions and parameters that try to recreate known user interactions method as shown in code 3.3.

Code 3.3 – Menu Option inside *EQ3/6* input files that mimics a "radio button"

```

ioprt(4) - Print a Table of Aqueous Species Concentrations, Activities, etc.:
[ ] (-3) Omit species with molalities < 1.e-8
[ ] (-2) Omit species with molalities < 1.e-12
[ ] (-1) Omit species with molalities < 1.e-20
[x] ( 0) Omit species with molalities < 1.e-100
[ ] ( 1) Include all species

```

3.1.3 File Formats

All the files discussed above are *ASCII* text files and, therefore, any regular text editor can be used to edit or build them. When the text file contains thermodynamic information needed by the program, the whole group of information is copied to the memory and, only after that, the software will be able to fetch information and continue the simulation's processing and natural flow. In memory, the storage of this information is not always optimized, specially because *ASCII* files are, in general, also not very well organized. Waste of memory here and there are expected and usual in this kind of text and memory management - scale this to a large amount of *ASCII* information - like in geochemical modeling systems - and it is easy to understand the risk that comes together with *ASCII* text files.

3.1.4 Software Environment and Installation Procedures

The *EQ3/6* package runs on Windows (95 and upper versions) and was developed in FORTRAN77. The support for UNIX computers has been discontinued. It has been developed and run at *Lawrence Livermore National Laboratory* on an *Alliant FX/80* and *Sun SPARCstations*.

The installation of *EQ3/6* is explained in details in (WOLERY, 1992). Due to the purpose of this work, details of the installation will be suppressed here. At the same time, it is interesting to mention that the whole installation process requires some level of experience with command prompt and *DOS*.

3.2 PHREEQC

PHREEQC stands for *PH* *RE*dox *EQ*uilibrium in *C* language, and is a widely used geochemical modeling software available from the USGS. Details can be found in (PARKHURST, 1995); It is available for download in versions for Windows and UNIX.

It was designed to perform a wide variety of low-temperature aqueous geochemical calculations based on an ion-association aqueous model, and has capabilities to:

- Speciation and Saturation Index calculations;
- Batch reaction and one-dimensional (1D) transport calculations involving reversible reactions (including aqueous, minerals, gas, solid-solution, surface-complexation, and ion-exchange equilibrium) and irreversible reactions (including specified mole transfer of reactants, kinetically controlled reactions, mixing of solutions and temperature changes);
- Inverse modeling, which finds sets of mineral and gas mole transfers that takes into account differences in composition between waters.

3.2.1 Input/Output Options

The input data for *PHREEQC* is arranged by *keyword data blocks*. Each block is organized with a keyword in the first line, followed by lines containing data related to the keyword. Keywords and their respective contexts are read from the database at the beginning of the run to define all the necessary parameters. After this database reading procedure, it continues reading the input file until it reaches the *END* keyword. As the input file is being read, the program will start putting the pieces together to perform the necessary calculations. An example of a *keyword data block* is shown in code 3.4. Among the possible keywords available in *PHREEQC* are EQUILIBRIUM PHASES, EXCHANGE, GAS PHASE, INVERSE MODELING, PHASES, REACTION, PRINT, SAVE, SOLUTION SPECIES, etc. Each one of these keywords contains specific arguments and parameters to be used. Certain keywords require some other specific keywords, and if something is missing from the input file, the results will be inconclusive. So building an input file is usually quite difficult and time-consuming task.

Code 3.4 – *PHREEQC* keyword data block example

```

EQUILIBRIUM_PHASES
Chalcedony  0.0      0.0
CO2(g)      -3.5     1.0
Gibbsite(c) 0.0      KAlSiO8  1.0
Calcite     1.0      Gypsum   1.0
pH_Fix      -5.0     HCl      10.0

```

The output file contains the result of the simulation defined in the input file, is also divided into *keyword blocks*. Among those are solution composition, description of the solution, redox couples, distribution of species (as can be seen in code 3.5), and saturation indices.

Code 3.5 – *PHREEQC*'s excerpt from the output file

```

...
-----Distribution of species-----

```

Species	Molality	Activity	Log Molality	Log Activity	Log Gamma	mole V cm3/mol
OH-	2.705e-006	1.647e-006	-5.568	-5.783	-0.215	-2.63
H+	7.983e-009	6.026e-009	-8.098	-8.220	-0.122	0.00
H2O	5.551e+001	9.806e-001	1.744	-0.009	0.000	18.07
C(4)	2.257e-003					
HCO3-	1.238e-003	8.359e-004	-2.907	-3.078	-0.170	27.87
NaHCO3	6.168e-004	7.205e-004	-3.210	-3.142	0.067	19.41
MgHCO3+	2.136e-004	1.343e-004	-3.670	-3.872	-0.201	5.82
MgCO3	7.301e-005	8.527e-005	-4.137	-4.069	0.067	-17.09
CaHCO3+	3.717e-005	2.572e-005	-4.430	-4.590	-0.160	9.96
CO3-2	3.128e-005	6.506e-006	-4.505	-5.187	-0.682	-0.34
CaCO3	2.256e-005	2.636e-005	-4.647	-4.579	0.067	-14.60
NaCO3-	1.477e-005	9.972e-006	-4.831	-5.001	-0.170	1.77

```

...

```

3.2.2 User Interaction

PHREEQC's distribution differs drastically according to the environment (*Windows* or *UNIX*). By this reason, the analysis of user interaction features will be done separately.

- *PHREEQC* for Windows: Due to the need of a geochemical modeling software and the lack of interface for *PHREEQC* in the first versions of the software, many efforts were made to create an interface to *PHREEQC*. The program *PhreeqcI* is a graphical user interface to *PHREEQC* that provides data entry screens for the keyword data blocks with a description of each input data item. It organizes the input file by using some *project tree* which facilitates viewing, selecting, editing and running the *PHREEQC* simulations.

It is critical to mention that *PhreeqcI* does not implement all the keyword blocks. Subsequently, *PHREEQC* version 2 with a graphical interface was launched, and the interface was extended to version 3. This graphical interface was named *PfW* (*Phreeqc for Windows*), but it has not been updated since 2011 (see Figure 3.1). The last effort in this sense was the development of an adaptation of the popular general-purpose text editor *Notepad++*. This modification comes with the following capabilities: syntax highlighting; autocompletion of keywords and identifiers; tips; colored numbers; parenthesis matching; commenting and uncommenting multiple lines at once; column editor; few shortcuts; and file recognition (see Figure 3.2). When running *PHREEQC* from the *Notepad++* adaptation, the command prompt is called and the simulation is executed, as in Figure 3.3. There are some options and shortcuts available from the *Notepad++*'s interface, which is shown in detail in Figure 3.4.

- *PHREEQC* for *UNIX*: Under *UNIX*'s distribution, *PHREEQC* runs from the command prompt. It can be launched by using the command shown in code 3.6

Code 3.6 – Command to run *UNIX*'s *PHREEQC*

```
phreeqc input output database screen_output
```

The "*input*" file contains the description of the simulation, the "*output*" file stores the result of the simulation, "*database*" specifies which database should be used, and finally the "*screen_output*" stores the information that will be shown on screen. If the user does not specify the names of "*output*", "*database*" or "*screen_output*" the geochemical modeling software will choose default values.

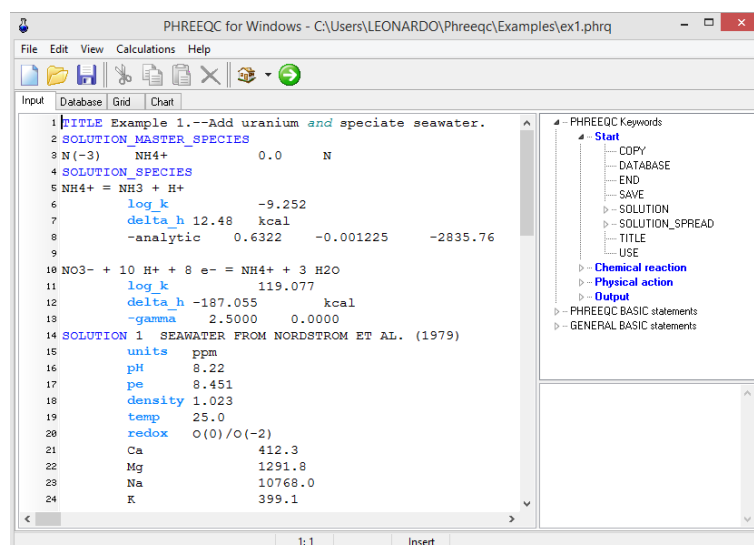
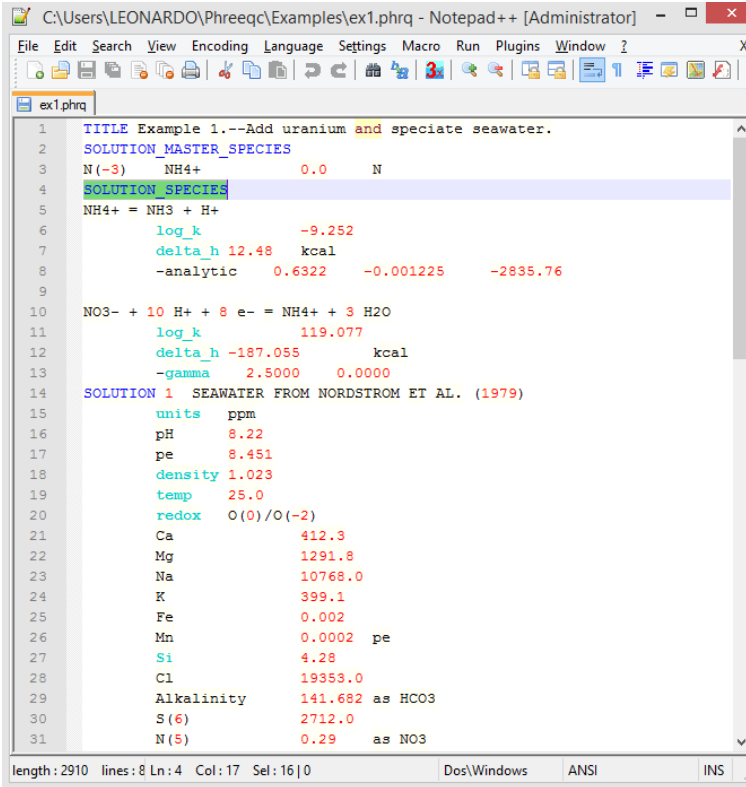


Figure 3.1 – User interface of the *PHREEQC* for Windows (*PfW*)

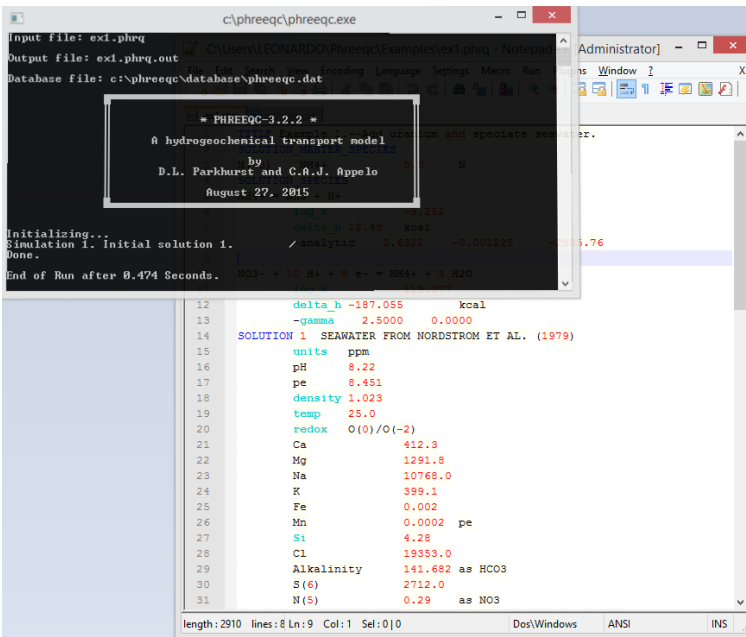


```

1  TITLE Example 1.--Add uranium and speciate seawater.
2  SOLUTION_MASTER_SPECIES
3  N(-3)  NH4+      0.0      N
4  SOLUTION_SPECIES
5  NH4+ = NH3 + H+
6      log_k      -9.252
7      delta_h 12.48      kcal
8      -analytic  0.6322  -0.001225  -2835.76
9
10 NO3- + 10 H+ + 8 e- = NH4+ + 3 H2O
11      log_k      119.077
12      delta_h -187.055      kcal
13      -gamma    2.5000      0.0000
14 SOLUTION 1 SEAWATER FROM NORDSTROM ET AL. (1979)
15      units      ppm
16      pH         8.22
17      pe         8.451
18      density    1.023
19      temp       25.0
20      redox      O(0)/O(-2)
21      Ca         412.3
22      Mg         1291.8
23      Na         10768.0
24      K          399.1
25      Fe         0.002
26      Mn         0.0002      pe
27      Si         4.28
28      Cl         19353.0
29      Alkalinity 141.682 as HCO3
30      S(6)       2712.0
31      N(5)       0.29      as NO3

```

Figure 3.2 – Example of the PHREEQC Notepad++ plugin



```

c:\phreeqc\phreeqc.exe
Input file: ex1.phrq
Output file: ex1.phrq.out
Database file: c:\phreeqc\database\phreeqc.dat

* PHREEQC-3.2.2 *
A hydrogeochemical transport model
by
D.L. Parkhurst and C.A.J. Appelo
August 27, 2015

Initializing...
Simulation 1. Initial solution 1.
Done.
End of Run after 0.474 Seconds.

12      delta_h -187.055      kcal
13      -gamma    2.5000      0.0000
14 SOLUTION 1 SEAWATER FROM NORDSTROM ET AL. (1979)
15      units      ppm
16      pH         8.22
17      pe         8.451
18      density    1.023
19      temp       25.0
20      redox      O(0)/O(-2)
21      Ca         412.3
22      Mg         1291.8
23      Na         10768.0
24      K          399.1
25      Fe         0.002
26      Mn         0.0002      pe
27      Si         4.28
28      Cl         19353.0
29      Alkalinity 141.682 as HCO3
30      S(6)       2712.0
31      N(5)       0.29      as NO3

```

Figure 3.3 – PHREEQC software called from the Notepad++ Plugin

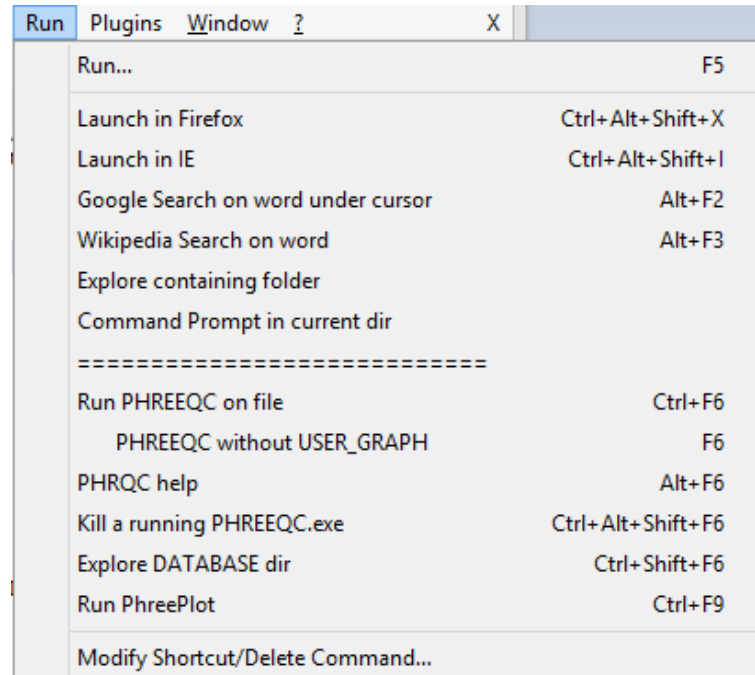


Figure 3.4 – PHREEQC software called from the Notepad++ Plugin

3.2.3 File formats

All of the files discussed above are in the *ASCII* text files format and, therefore, any regular text editor can be used. It is recommended that the editing of *PHREEQC* files be done by using the NotPhreeqcce or notepad++ adapted version (APELLO, 2011).

3.2.4 Software Environment and Installation Procedures

PHREEQC has support for Windows (32 and 64-bit), MacOS (OS 10.6+) and Linux. *PHREEQC* is currently on version 3 and with frequent updates, bug fixes and maintenance.

The *PHREEQC* version for Windows has a self-extracting file that is available for download from the USGS website and easily installed. The *UNIX* distribution comes with additional scripts and a makefile, and an instruction on how to compile and install the program.

3.3 *MINTEQ*

MINTEQ (FELMY et al., 1983) is a geochemical program to model aqueous solutions and the interactions of aqueous solutions with hypothesized assemblages of solid phases. It has a particular inclination to calculate equilibrium composition of dilute aqueous solutions.

The model is useful for calculating the equilibrium mass distribution among dissolved species, adsorbed species and multiple solid phases, although it has a much simpler treatment of the reactions.

It was originally developed in FORTRAN77 by the *Battelle Pacific Northwest Laboratory (PNL)* and continues to be maintained by the *Environmental Protection Agency (EPA)* to perform the necessary calculations regarding waste, sediments and ground water interaction. *MINTEQ* does not consider the kinetic reactions and works at fixed temperature (25 degrees Celcius). An extensive database adequate to a broad range of problems is part of the software, and there is no need for the user to change nor add anything (BROWN; ALLISON, 1987) (ALLISON J. D.; NOVO-GRADAC, 1991). The latest update on *MINTEQ* dates from 1990, and since then, there has been only some improvements especially on the usability and calculations. This version was named *MINTEQA2* and uses a well-developed thermodynamic database from the *USGS*. During this review we will address the *MINTEQA2* properties since it is the latest version and is clearly an improved version of the same program (*MINTEQ*).

3.3.1 Input/Output Options

The input files for *MINTEQA2* can be generated manually, but there is a supporting software called *PRODEFA2* that guides the user to accomplish this task. *PRODEFA2* is an interactive program used to create input files that will be addressed in details in section 3.3.2. Four files compose the input of *MINTEQA2*:

- Input file: The input file that contains the data input by the user. Typically, this file contains dissolved (i.e. Ca concentrations, pH, temperature) and solid phase (i.e. minerals, sorption sites) information for a water sample;
- Database file: This file contains the thermodynamic constants that govern the processes of interest (i.e. complexation constants, mineral solubilities, activity constants) which will be used to conduct calculations;
- Algorithm or executable file: These files contain the algorithms of the code, which solve the specified problem (usually using an iterative numerical approach) within the constraints imposed by the Database files and the information in the Input file.
- Output file: This file contains the results of the calculations performed by the Algorithm Files.

Among the input file's options, there are four levels of configuration. Each level controls some details of the simulation, and when these four levels are together, they compose a complete input file for *MINTEQA2*. It is important to mention that if the user does not want to specify all the details, there are default options that enable any non-experienced user to execute simple simulations.

1. Displays the current settings of system parameters such as temperature as well as program flag settings such as the number of iterations allowed;
2. Specify the chemistry of the system;
3. This level works as a "line editor" in displaying by category or TYPE those species that have been explicitly entered through level 2;
4. Deals with utility functions (output file details, for example);

If database, algorithm and output files are not specified, default options are used. Code 3.7 brings the *MINTEQA2*'s input file used in the study case discussed in chapter 5 and presented here as an example. The beginning of the file (first and second lines) contains a description of the simulation or input file; the third, fourth and fifth lines set configurations, such as temperature, unit chosen, eH, ionic strength, number of iterations, precipitation options. After that, the components that take part in the simulation, organized by internal id, concentration details and log of concentration are specified.

Code 3.7 – *MINTEQA2*'s input file

```
LHDAMIANI - STUDY CASE
Comparative study
25.00 MOLAL 0.000 0.000000E+00
0 0 1 0 1 0 0 0 1 1 0 0 0
0 0 0
330 6.026E-09 -8.22 y /H+1
410 1.045E-02 -1.98 y /K+1
500 4.793E-01 -0.32 y /Na+1
150 1.053E-02 -1.98 y /Ca+2
460 5.439E-02 -1.26 y /Mg+2
732 2.889E-02 -1.54 y /SO4-2
180 5.595E-01 -0.25 y /Cl-1
```

An excerpt of *MINTEQA2*'s output is presented in Code 3.8 - shows important parameters of the solution's components. The whole output file is divided into six parts:

1. Reproduction and interpretation of the input file;
2. Detailed listing of species read from the database files;
3. Iteration information and detailed information for each species;
4. Percentage distribution of components among dissolved and adsorbed species;

5. Provisional or equilibrated mass distribution, provisional or equilibrium ionic strength, equilibrium pH and pE, electrostatic surface potential and charge for electrostatic adsorption models;
6. Saturation indices for all database solids with respect to the solution;

Code 3.8 – *MINTEQA2*'s excerpt from the output file

```

...

PART 3 of OUTPUT FILE

MINTEQA2 v4.02 DATE OF CALCULATIONS: 5-JUN-2000 TIME: 14: 6:27

PARAMETERS OF THE COMPONENT MOST OUT OF BALANCE:

ITER      NAME      TOTAL mol/L    DIFF FXN    LOG ACTVITY    RESIDUAL
0   SO4-2      1.580E-03    6.594E-07    -2.91757      5.014E-07
1   SO4-2      1.580E-03    4.193E-04    -2.91775      4.192E-04
2   SO4-2      1.580E-03    8.125E-06    -3.01997      7.967E-06
3   SO4-2      1.580E-03    1.693E-07    -3.02220      1.135E-08

ID No  Name      Total Conc (M)    Conc (M)    log Activity    Diff fxn
410    K+1      7.700E-05      7.649E-05    -4.14759      5.093E-11
732    SO4-2    1.580E-03      1.266E-03    -3.02224      3.557E-09
2      H2O      0.000E+00      -1.049E-05    -0.00004      0.000E+00
330    H+1      0.000E+00      3.398E-03    -2.50000      0.000E+00
140    CO3-2    0.000E+00      2.916E-17    -16.66004      0.000E+00

-----
...

```

3.3.2 User Interaction

MINTEQA2 and *PRODEFA2* interactions are completely independent programs and *PRODEFA2* is used before *MINTEQA2* in order to generate the input file that will be consumed by the latter. Everything is done through the command prompt, as explained below. *PRODEFA2* provides a "walk-through" to generate an input file for *MINTEQA2*.

After opening the software and providing a valid name, it asks which part of the input file the user wants to create or edit (Figure 3.5). We will follow the suggested order and go through 4 levels, as previously discussed in this section. Figure 3.6 shows the main menu, which shows the organization of all the levels and works as a central hub of information. Figure 3.7 displays the necessary information about level 1. To change any of the entries on this screen, the user

must enter the number to the left of the entry and respond to the questions presented. All four levels carry out this type of interactions. Through these interactions, the user has access to the information in the database and can choose the data to be used for the model.

```

C:\Program Files (x86)\MINTEQA2\prodefa2.exe

Problem Definition Program for MINTEQA2 (PRODEFA2)
Version 4.03
May 2006
U.S. Environmental Protection Agency
Center for Exposure Assessment Modeling (CEAM)

Seed file (optional). Enter X to exit PRODEFA2.
Filename for MINTEQA2 input file to create (required).
Lhdaniani_example

-----
INPUT, OUTPUT, and DATABASE FILE SELECTION MENU
-----

1 = MINTEQA2 INPUT FILE to create (required):      Lhdaniani_ex
2 = SEED INPUT FILE (optional):
3 = COMPONENT DATABASE FILE (required):             comp.dbs
4 = MAIN THERMODYNAMIC DATABASE FILE (required):    thermo.unf
5 = GAUSSIAN DOM REACTION DATABASE FILE (required): gaussian.dbs
6 = SOLID SPECIES DATABASE FILE (required):          type6.unf

ENTER CHOICE TO CHANGE OR <enter> TO ACCEPT >

```

Figure 3.5 – MINTEQA2 initial menu options

```

C:\Users\LEONARDO\MINTEQA2\prodefa2.exe

MAIN MENU: SELECT OPTION PROB # 1

1 = EDIT LEVEL I <Change ionic strength, pH, Eh, temperature, adsorption
parameters, number of iterations, precipitation options, etc.>
2 = EDIT LEVEL II <Specify components, gas, redox, aqueous, and mineral
species, adsorption sites and reactions, add new species of all types>
3 = EDIT LEVEL III <Check, individually edit all entries>
4 = EDIT LEVEL IV <Sweep a range of pH, pE, or dissolved concentration;
Designate an auxiliary MINTEQA2 output file to receive equilibrated
output for spreadsheet import.>
M = MULTI-PROBLEM GENERATOR
X = EXIT <Write the current problem to the new MINTEQA2 input file
and EXIT PROGRAM>

ENTER CHOICE >

```

Figure 3.6 – MINTEQA2 main menu

```

C:\Users\LEONARDO\MINTEQA2\prodefa2.exe

5 = GAUSSIAN DOM REACTION DATABASE FILE (required): gaussian.dbs
6 = SOLID SPECIES DATABASE FILE (required):          type6.unf

ENTER CHOICE TO CHANGE OR <enter> TO ACCEPT >

EDIT LEVEL I PROB # 1

1 Title 1:
2 Title 2:
3 Temperature (Celsius): 25.00
4 Units of concentration: MOLAL
5 Ionic strength: TO BE COMPUTED
6 Inorganic carbon is not specified.
7 Terminate if charge imbalance exceeds 30% ? NO
8 Oversaturated solids ARE NOT ALLOWED to precipitate. EXCEPTIONS: Solids
listed in this file as TYPE -III (Infinite), -IV (Finite) or -U (Possible).
9 The maximum number of iterations is: 200
10 The method used to compute activity coefficients is: Davies equation
11 Level of output: INTERMEDIATE
12 The pH is: TO BE COMPUTED
13 The pe and Eh are: UNDEFINED
20 Return to filename and database selection menu

Enter selection to change any entry. Press ENTER to accept all settings.
ENTER CHOICE >

```

Figure 3.7 – MINTEQA2 Level 1 informations

For example, we chose to add a specific aqueous species to show the *PRODEFA2* interactions and present it step-by-step bellow:

1. Choose level 2 on main menu;
2. Choose option 1 (Specify AQUEOUS COMPONENTS: TOTAL CONCENTRATIONS or FIXED ACTIVITIES) inside the menu from level 2;
3. Choose option 1 (TOTAL DISSOLVED CONCENTRATION), it identifies how we want to enter this new component.
4. At this step, we are requested to enter the first letter for the component. Alternatives to this approach is to enter "-1" if you know the component id number or quit. We will add Na^+ to the system, and type the letter "N" and hit the enter button.
5. At this step, all the existing options available are shown. listed with an identifier. We are also requested to identify one of these to be added. This can be seen in 3.8.
6. After pressing Na^+ 's id (at this case was the number 4). We can finally enter the TOTAL DISSOLVED CONCENTRATION (MOLAL) of COMPONENT defined earlier on step 3. After adding the molal concentration for this species the system goes back to step 4 in case we want to continue adding other components.

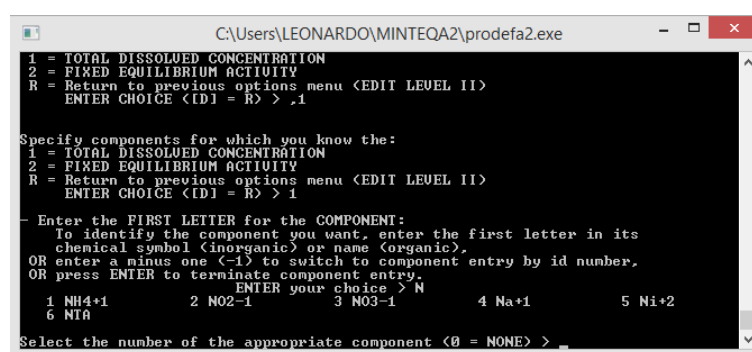


Figure 3.8 – *PRODEFA2*'s example of adding a specific aqueous species (Na^+)

There is also a software called *Visual Minteq* that tries to "humanize" *MINTEQA2* and was maintained by the KTH Royal Institute of Technology, located in Stockholm, Sweden. The latest release date of *Visual Minteq* latest is December 2013 at the version 3.0 and is available only for *Windows* operating systems since it was developed in Visual Basic.

3.3.3 File formats

All the files follow a regular text format (*ASCII*) and their purpose is defined in the extension of the file.

- Input files have the extension "*INP*";
- Test and help files have the extension "*HLP*";
- Output files have the extension "*LST*";
- Database files have the extension "*DBS*" or "*UNF*";
- Input file have the extension "*INP*";

3.3.4 Software Environment and Installation Procedures

It is currently at the version 4.03 (Windows only), and this release dates back to May 2006. The latest *UNIX* distribution is version 3.12 (which was also called BETA for UNIX) and dates back to August 1996. *MINTEQA2* is easily installed by a self-extractor installer that can be downloaded from *EPA*'s website (MINTEQ, 2006) and (MINTEQ, 1999). Included in the distribution package are also some important documentation, *PRODEFA2* software and several input/output template files. The UNIX version distribution comes with the source files and must be compiled and linked to run.

3.4 *SOLMINEQ.88*

SOLMINEQ.88 is a geochemical modeling program written in FORTRAN 77 and based on *SOLMNEQ* (KHARAKA; BARNES, 1973) with improved algorithms that resulted in a faster program execution and tighter convergence. The software has a database with the focus on organics aqueous species. It calculates the distribution of mass among aqueous species and complexes and calculates saturation indexes of minerals at different temperatures and pressures. It includes options as boiling, mixing of solutions, and partitioning of gases between water, oil and vapor phases. *SOLMINEQ.88* also contemplates mass transfer with the effects of dissolution and precipitation of minerals and options to calculate activity coefficients (KHARAKA Y. K.; DEBRAAL, 1988). The original version has no UI, but there were further studies with the intention of creating a user-friendly program that can be used to generate, edit and analyze

input and output files - *SOLINPUT*. *SOLMINEQ.88* model has not been developed any further nor improved since its first release.

3.4.1 Input/Output Options

The input of *SOLMINEQ.88* consists of two sets of data: fixed and variable. The first data contains the chemical composition of an aqueous fluid and options for processing these data; and the second data consists of the input required for using the first set of data.

The input file consists of six parts, and *SOLINPUT* guides the user through all of them:

1. Basic Parameters: enter the chemical and physical data for that sample;
2. Flags: controls how the software interprets, processes, and displays the data;
3. pH: controls the details of how the pH calculation is done;
4. Mass transfer: defines which mass transfer capabilities are used;
5. User Log K: makes temporary changes and extensions to the database;
6. Additional ions and minerals: temporarily adds user defined ions and minerals to a particular simulation;

The output file contains the results produced by *SOLMINEQ.88*, and consists of six parts: An input data echo that shows the values and options selected for each sample; A table listing the calculated tolerance factor for successive iterations on the anions; A list of input to *SOLMINEQ.88* including sample description, pH, Eh, temperature and so on; A table showing the distribution of species in solution; Ratios of a number of important cations and anions ; and the last one contains a table indicating the states of reactions for minerals considered:

Code 3.9 – *SOLMINEQ.88*'s excerpt from the output file

```

: Test Sample #1 for SOLMINEQ.88 - Modified Seawater at 25 C
TEMP HI TEMP DENS PRESS
0.2500E+02 0.0000E+00 0.1023E+01 0.0000E+00
PH EHM EHMC EMFZSC
0.8200E+01 0.5000E+00 0.9000E+01 0.9000E+01
CONCENTRATION UNITS : PPM
Na K Li Ca
0.1077E+05 0.3991E+03 0.1810E+00 0.4123E+03
SiO2 Cl S04 H2S
0.4280E+01 0.1935E+05 0.2712E+04 0.0000E+00
F P04 N03 NH3
0.1390E+01 0.6000E-01 0.2900E+00 0.3000E-01
Pb Zn Cu Mn
0.5000E-04 0.4900E-02 0.7000E-03 0.2000E-03
As U V
0.4000E-02 0.0000E+00 0.0000E+00
Acetate Oxalate Succinate CH4
0.1000E+00 0.1000E+00 0.1000E+00 0.1000E+00

```

3.4.2 User Interaction

The software that accompanies *SOLMINEQ.88* and handles the generation of input files and all the interactions is named *SOLINPUT* and described in (DEBRAAL; KHARAKA, 1989). All interactions are through command prompt input following displayed menus with several options. The user selects the option by entering the indication number and pressing enter (the indication number stays on the left of the option). Figure 3.10 shows this example of interaction.

Code 3.10 – *SOLMINEQ.88*'s example of user interaction

```

pH OPTIONS
1) Gas Addition Option
2) Gas-Water-Oil Distribution Option
3) Carbonate Mineral Saturation Option
4) CO2 Option
5) Tolerance factor for Mineral and CO2 Options
6) Return to Options Menu
Enter Choice (1-6)  ___

```

3.4.3 File formats

All the files that *SOLMINEQ.88* deals are regular *ASCII* text files and any text editor can be used to create, edit or view the files. The database files from *SOLMINEQ.88* have the

Table 3.1 – Geochemical comparison between speciation softwares

SOFTWARE	Aqueous Complexation	Precipitation and Dissolution	Reaction path	Kinetics	Multi-Activity Coefficient methods
EQ3/6	✓	✓	✓	✓	✓
PHREEQC	✓	✓	✓	✓	
MINTEQA2	✓	✓			
SOLMINEQ.88	✓	✓	✓	✓	✓

extension "TBL", while the input and output files, "IN" and "OUT", respectively. Some optional file with the extension "MIXFLE" can be used to specify mixture properties. Finally, if the user activates the restart option, also known as pickup option, the file "OUTIN" will be generated and it contains data from the current run.

3.4.4 Software Environment and Installation Procedures

As mentioned, *SOLMINEQ.88* had only one release and has been discontinued since then. It is available only for the *Windows* operating system.

SOLMINEQ.88 distribution requires knowledge of compiling and linking *FORTRAN77* programs. It also comes with the software *SOLINPUT*.

3.5 Discussion

In this section some important aspects and issues of each one of the software presented earlier are compared, analyzed, and discussed. Regarding geochemical features, table 3.1 brings certain aspects and features of the software.

From the perspective of computer science, we analysed and evaluated the software for the software engineering aspect taking into account all of the features described and extensively discussed in chapter 2.

Following points were taken into consideration:

- **The costs:** Costs are probably the most important thing that people look first when choosing software; however, it should not be the deciding factor. Different solutions use different pricing models and according to the purpose and utilization of the software.
- **Setup and versioning:** The installation of software is the act of making it ready for execution. Depending on the software different options are used to copy/generate files from the installation files to the local computer to be accessed by the operating system (*OS*). The *OS* also influences how this process is done. Each software has a different distribution package for different *OS*. Commonly software are distributed only to specific *OS*. Also common are software with disparate versions according to the *OS*, resulting in divergent features available on the same software defined by the *OS* types.
- **Customization and Integration:** Is the software a standard solution, and its supplier interested in making changes? This is the typical scenario where the user has great chances of finding problems ahead. Therefore, an interesting exercise is to think of how this software communicates with others. Options like “import” and “export” are vital when working with a large amount of data. With the advances in software, different software can be used to analyse the output and to be able to reach deeper insights about the information available.
- **Security and Control:** Security is one of the main issues one faces when considering a software. Data privacy is an important criteria. Ensuring that the software can guarantee no data loss or data leakage is important. The solutions that provide direct control to the database, details, and processes are most likely to require data privacy.
- **Infrastructure:** When choosing a software, the infrastructure it requires need to be carefully analyzed. Does it require Internet access? How much space on disk and memory does it use? Extra costs may be incurred if this is not addressed.
- **Core functionality:** This is one of the most important points to be analyzed. How good is the software focus on the needs of the user, and how good is the value the software brings to its users?
- **Graphical User Interface (*GUI*) and visualization:** This feature handles the interaction between the users and the software. When the software has a complex domain, such as

geochemical modeling, the *GUI* is even more important. It is responsible for allowing the user to consider all necessary possible options. The perfect *GUI* takes into consideration the necessary human behavior, senses, and how we interact with our world (from electronic devices to human relationships).

- **Support and Maintenance:** If the software, for some unknown reason, goes down, is the user able to reach someone for troubleshooting? Will the user be able to find a users' community to debate and share knowledge? If the software has a support team working to fix bugs, improve the performance, add new features and sharpen some of the old features, it means that the user will have a better infrastructure to work with.
- **Database:** All the data manipulated inside a software is stored and organized in a database. There are multiple ways of doing this. Many important things must be taken into account to decide which database fits best to the software. Since the 80's the relational database model represented by the *SQL* language has been the most popular. A conceptual database model is strongly recommended to produce a schema that considers all the structure and information needed by this software. Along the database schema, the security of this database must be addressed properly, for consistency and privacy. A good database design avoids redundant data (unnecessarily duplicated data). Poorly designed database generates inconsistent data (inaccurate data), which will lead to wrong decisions and, therefore, can result in failure of the software.

3.5.1 Existing Thermodynamic datasets

The contents of the databases are extracted from the *Lawrence Livermore National Laboratory (LLNL)* thermodynamic datasets that are used in *EQ3/6*, *PHREEQC* and others. The data sets are contributions from many authors that had measured thermodynamical and kinetic parameters of the minerals and reactions minerals over the years. Flat file databases' interactions are difficult and, if necessary, must be done carefully - the format of the file is composed of a series of rules and very often is not recommended that the user alters anything inside the flat file database due to the complexity and the errors that it might result. The following is a small excerpt that illustrates the structure that the information is organized in the *LLNL's* flat file:

- **Parameters:** Many used parameters are established on this section, among them are temperatures, pressures, debye huckel coefficients, bdot coefficients...

Code 3.11 – Excerpt of the section Parameters

```
* temperatures
      0.0000   25.0000   60.0000  100.0000
    150.0000  200.0000  250.0000  300.0000
* pressures
      1.0134    1.0134    1.0134    1.0134
      4.7600   15.5490   39.7760   85.9270
```

- **Elements:** This section is composed of information on pure elements. It also lists the mole weights of elements and the abbreviation:

Code 3.12 – Excerpt of the section Elements

```
Oxygen      (O )      mole wt.=   15.9994
Silver      (Ag)      mole wt.=  107.8680
Aluminum    (Al)      mole wt.=   26.9815
```

- **Basic Species:** This section lists atomic or molecular structural units for a mineral:

Code 3.13 – Excerpt of the section Basic Species

```
H2O
charge=  0.0      ion size=  0.0 A      mole wt.=   18.0152
2 elements in species
  1.000 O          2.000 H
```

- **Redox Couples:** This sections includes all chemical reactions in which molecules have their oxidation states changed. Redox reactions involve the transfer of electrons between species. The name comes from two concepts involved with electron transfer (reduction - loss of electrons - and oxidation - gain of electrons). Example:

Code 3.14 – Excerpt of the section Redox Couples

```
Cr++
charge=  2.0      ion size=  5.0 A      mole wt.=   51.9960 g
4 species in reaction
-1.000 H+          0.500 H2O          1.000 Cr+++
-0.250 O2(aq)
  33.6814   29.9291   25.6126   21.6721
  17.7896   14.7267   12.2289   10.1676
```

- **Aqueous Species:** This sections contains the water solutions. The word aqueous is applied to a solution or mixture in which water is the solvent. When a chemical species has been dissolved in water, this is denoted by writing (aq) after the chemical name. Example:

Code 3.15 – Excerpt of the section Aqueous Species

```

CO2 (aq)
  charge=  0.0      ion size=  4.0 A      mole wt.=  44.0098 g
  3 species in reaction
    -1.000 H2O      1.000 H+      1.000 HCO3-
    -6.5570    -6.3660    -6.3325    -6.4330
    -6.7420    -7.1880    -7.7630    -8.4650

```

- **Minerals:** This sections lists the physical properties and chemical formula of minerals:

Code 3.16 – Excerpt of the section Minerals

```

Calcite                                type= carbonate
  formula= CaCO3
  mole vol.=  36.934 cc      mole wt.=  100.0892 g
  3 species in reaction
    1.000 Ca++      1.000 HCO3-      -1.000 H+
    2.0683      1.7130      1.2133      0.6871
    0.0762      -0.5349      -1.2301      -2.2107

```

- **Gases:** This section lists individual atoms (e.g. noble gases or atomic gases), elemental molecules made from one type of atom (e.g. oxygen), or compounds made up of two or more molecules (e.g. carbon dioxide). Example:

Code 3.17 – Excerpt of the section Gases

```

CO2 (g)
  mole wt.=  44.0098 g
  3 species in reaction
    -1.000 H2O      1.000 H+      1.000 HCO3-
    -7.6827    -7.8184    -8.0628    -8.3849
    -8.8297    -9.3208    -9.8841    -10.6132

```

- **Oxides:** This sections contains the chemical compounds that consists of at least one oxygen atom and one other element in its chemical formula, and also does not include silica (Si):

Code 3.18 – Excerpt of the section Oxides

```

Al2O3
  mole wt.=  101.9616 g
  3 species in reaction
    -6.000 H+      2.000 Al+++      3.000 H2O

```

To achieve an applicable comparison we give grades from 1 to 5 to each aspect (where 1 is the lowest and 5 the highest possible grade). This "grading system" is done with the intention of comparing and normalizing the attributes of the software (see table 3.2¹).

¹Source:Author

Table 3.2 – Qualitative analysis of the Geochemical Speciation Software

SOFTWARE										
	Costs	Setup and versioning	Customization and Integration	Security and Control	Infrastructure	Core functionality	Graphical User Interface	Support and Maintenance	Database	Overall Average
EQ3/6	2	2	1	1	3	5	1	1	1	1.88
PHREEQC	4	4	2	2	3	4	2	3	1	2.77
MINTEQA2	4	2	1	1	2	2	1	1	1	1.66
SOLMINEQ.88	3	1	1	1	1	5	1	1	1	1.66

3.6 Summary

- **EQ3/6:** It represents a landmark in Geochemical Modeling. Unfortunately, it is inaccessible due to the elevated cost of licensing. The vast amount of information available online about *EQ3/6* makes it an excellent source of information. It used the computing tools and options that were available in 1970's and 80's, until the latest known release date in 1992. Since then, computing has clearly evolved, making it an obsolete and difficult to use. It has a large database with many pieces of information, but the contents are not transparent. It is hard to understand what exactly is the software doing; verifying if that is what the users wants is even more difficult. It is also important to mention that in *EQ3/6* documentation it recommends the user to use operating system commands (i.e. *ctrl+c*) to interact with the software - which is a highly risky procedure in modern computers.
- **PHREEQC:** It is the best option for users not experienced with software. It has a *GUI* and comes with a self-extractor installer. Important to mention here is that *PHREEQC*'s *GUI* is far from what a typical user of current era might expect. It is not clear in many aspects, and its usability is far from regular. In the geochemical modeling area, not all the users are familiar with tasks as compiling and linking computer programs. *PHREEQC* allows anyone with an interest to have a chance to perform a geochemical modeling simulation, even as many people have different ways of defining the problem. Database in *PHREEQC* seems to be a problem, as it uses a flat file database.
- **MINTEQA2:** From the geochemical point of view, *MINTEQA2* is the simpler of the four software analyzed in this work. The user interaction can be painful for anyone who are not familiar with command prompts. The complexity of the input file also makes it

difficult to use. Creating an input file without the subsidiary software *PRODEFA2* is a task close to impossible and learning how to use this subsidiary software is a very costly task. Taking into account that its last release dates back to 2006, it is difficult to justify using *MINTEQA2*.

- **SOLMINEQ.88:** This is another software that was a pioneer and laid the groundwork for many others to improve and progress the knowledge of geochemical modeling. *SOLMINEQ* used the computing tools that were accessible when it was developed in 1980's. There are less costly options easily attainable today, thus it is no longer a popular option for geochemical modeling.

4 *SHPECK* - GEOCHEMICAL SPECIATION MODEL

SHPECK is a geochemical speciation model based on the *Phase Rule* described in (GARRELS; CHRIST, 1965). This is also the principle that is used on the speciation models reviewed in the previous section.

As a computer simulation software, *SHPECK* is a complete integrated model that allows: the user to set up choosing; convert the chemical description to a system of equations; solving the system of equations; and present the results.

Figure 4.1 represents a briefing overview to make it clear what are the processes involved in *SHPECK*. In figure 4.1 (a), the user will interact with the interface to define the speciation settings and the chemical phases that compose the present water (necessary information will come from the database but is invisible to the user). In figure 4.1 (b), the processing core will take place and perform the several tasks involved in defining the speciation of a solution - described in details later in this chapter. In figure 4.1 (c), the relevant results are displayed in the *GUI* to the user.

This chapter will guide through the aspects of the development of *SHPECK* - a geochemical speciation modeling software. Software quality factors (either external and internal) were constantly addressed and studied. As external factors, we can mention:

- **Correctness:** The ability of performing the exact task as defined by the specification of a geochemical modeling software.
- **Robustness:** The ability of reacting appropriately to abnormal conditions that may be entered through the chemical composition by the user.
- **Extendibility:** The ease of adapting to changes of specification and reorganize the tasks.
- **Reusability:** The ability to serve for construction of many different applications. For example, adopting CO₂ pressure or including reactive transport to the solver.
- **Compatibility:** The ease of combining software elements with others. Working with new datasets and trying to connect different source of thermodynamic data can be complicated - if not impossible.

On the other side, we have internal factors:

- **Modularity:** By dividing the system into different parts is possible to have a better understanding. A set of different modules compose the whole system - each one responsible for a task.

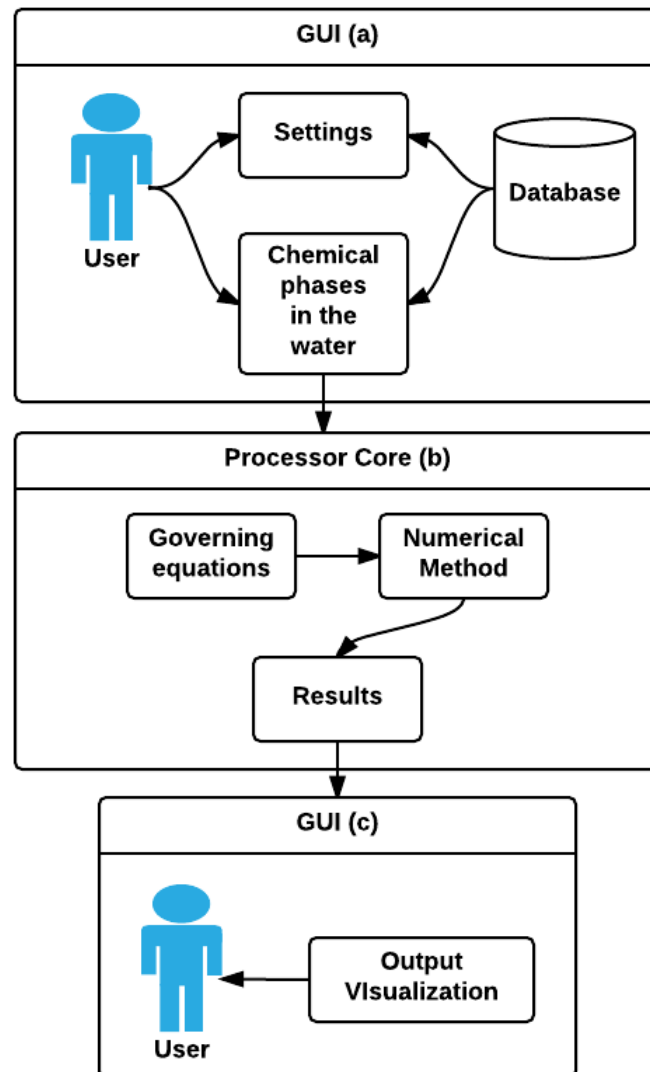


Figure 4.1 – *SHPECK* overview

- **Readability:** Adopting code conventions and good habits in programming are fundamental to the health of the software.

These quality factors are detailed in (MEYER, 2000).

4.1 Specification

SHPECK is a geochemical speciation modeling software designed to calculate the distribution of dissolved chemical elements and aqueous solutes and complexes, and it also calculates saturation indexes for different minerals.

4.2 Documentation

For internal documentation, *Doxygen* is used. *Doxygen* is an automated code documentation generating tool (HEESCH, 2013). This documentation will be available along with the distribution of *SHPECK* when possible.

4.3 Intended users

SHPECK is a software for geochemists, geologists, chemists, environmental engineers, forestry engineers, agronomists, laboratory analysts and others.

4.4 Architecture

The architecture of the software is responsible for assuring that all the internal and external stakeholders' concerns are preserved, addressed and satisfied. The design of the architecture must occur prior the start of the development - it avoids waste of resources and re-work in the future. *SHPECK*'s architectural design choices are done always to make sure that the software will bring value to the user.

Figure 4.2 shows the software architecture of *SHPECK*, which is modeled following the popular concept of *Model-View-Controller* (MVC) (GAMMA, 1994). *MVC* is an architectural pattern that divides the software into three interconnected parts:

- **Model:** It is an object representing data or even activity. For example, the algorithm and math behind calculating the activity coefficient by Debye-Hueckels' formula.
- **View:** It is a form of visualization of the state of the model. For example, which are the solutes that the user wants to add into the simulation?
- **Controller:** It offers facilities to change the state of the model. For example, define which algorithm for calculating the activity coefficient according to the user's choice or the value of the ionic strength (if the user did not specify which one to use).

In figure 4.2 any kind of information flow is indicated by the direction of the arrows going out or in from the components.

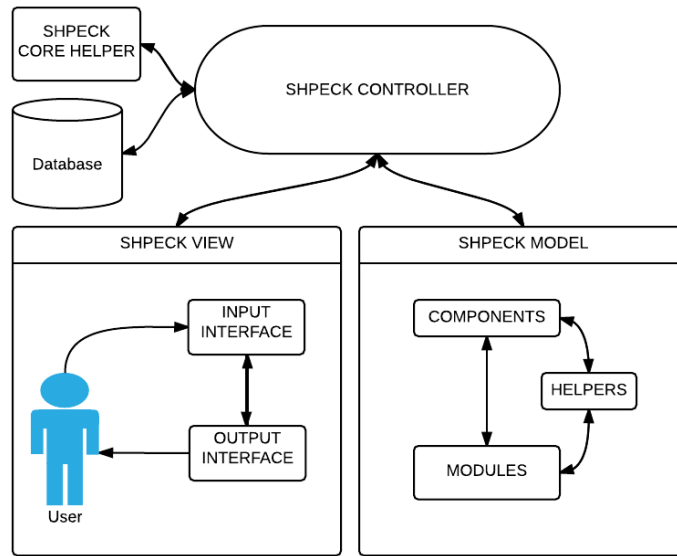


Figure 4.2 – Architecture of the software *SHPECK*

4.4.1 Technical Specification

SHPECK is a software developed using C++, which is a general-purpose programming language. C++ has a bias toward system programming that supports efficient low-level computation, data abstraction, object-oriented programming, and generic programming - as shown in (DALE, 2004) and (STROUSTRUP, 1997). It provides powerful and flexible mechanisms for abstraction. In other words, the language allows the programmer to introduce and use new types of objects that match the concepts of an application needed. Thus, C++ supports styles of programming that rely on relatively direct manipulation of hardware resources to deliver a high degree of efficiency. It can also address higher-level styles of programming that rely on user-defined types to provide a model of computation that is closer to human's view of the task being performed by a computer. Application libraries often support these higher level styles of programming. While developing *Shpeck*, two supporter tools were adopted:

- *Qt* : It is a cross-platform application and UI framework for developers using C++ (QT, 2014). *Qt* turns the development fast and easy. With extensive built-in library classes available, it provides a comprehensive range of functionality, as well as a convenient platform.
- *Armadillo* : It is a C++ linear algebra library that provides classes for vectors, matrices, cubes and a whole set of functions to operate on the classes (SANDERSON; CURTIN, 2014). Its syntax is similar to Matlab, and it has an up-to-date support with upgrades and new releases are available on monthly basis.

Important to mention that the system is being developed using *MAC* but since the beginning of the development process, the goal is to develop a multi-platform software to be distributed either in *Windows* and *LINUX-like* systems.

?? concepts and arguments are maintained close during the development of *SHPECK* to produce o high-class software.

4.5 Governing equations

SHPECK uses thermodynamic equilibrium reactions as equations for the calculation of multiphase systems in equilibrium. Details and treatments of these reactions are discussed in details in chapter 2. A set of mass-action equations (as in equation 2.2) compose the system, and the number of species and compounds that coexist in the system defines the number of equations. These equations model the geochemical speciation in a closed system taking into account the chemical properties of the solutes.

Aside from the mass-action equations there must be additional constraints to solve the equilibrium state of the system. In *SHPECK*, I use the concentration of the species to deal with the equilibrium state of the system. Therefore, I have the following configuration:

$$S_{aqueous} = N_{reactions} + N_{eqconstraints} \quad (4.1)$$

where $S_{aqueous}$ are the number of aqueous solutes in the solution, $N_{reactions}$ are the number of mass-action equations and $N_{eqconstraints}$ are the number of equilibrium constraints imposed by the user. In order to have a better efficiency on the method, we use equation 4.2 reformulated by applying natural logarithm in both sides, as expressed below:

$$\ln(K_j) = \ln\left(\prod_{i=1}^N a_i^{v_{ij}}\right) \quad (j = 1, \dots, M) \quad (4.2)$$

4.6 Numerical Method

In order to solve the system composed of the equilibrium state of the mass-action equations and the equilibrium constraints, *SHPECK* uses a numerical method that solves the complete set of equations simultaneously and find the value of the unknowns (solute concentrations).

The numerical method applied is a modification of the *Newton's method* (also known

as Newton-Raphson method), which is a method for finding successively better approximations to the roots of the given set of equations. It works with a derivative approach to the equations, which optimizes the time consumed to find the roots and makes the representation of the system easier.

SHPECK uses the *Newton-Raphson* method to solve a nonlinear system of equations which results in finding the roots of continuously differentiable equations.

The *Newton-Raphson* method is applied in order to achieve the best approximation possible to the solution that is being sought. It is an iterative algorithm where each step consists in minimizing the first-order approximation of the solution and its difference from regular *Newton's* method is that second derivatives are not required. *Newton-Raphson's* method is used to solve a system of coupled nonlinear equations. The first-order approximation of the function starts with an initial guess for the minimum values, the method proceeds by the iterations as shown in equations 4.3.

$$F(x + 1) = F(x) - J^{-1} * R \quad (4.3)$$

where F is function's result for the applied x , J^{-1} is the inverse of the Jacobian matrix, and R is the residual vector ((ISAACSON E., 1966)). The *Newton-Raphson's* method has a complexity $O(n^3)$ per iteration with a quadratic convergence.

The *Jacobian Matrix* is the matrix of all first-order partial derivatives of the equations and it is defined as

$$J_{mn} = \frac{\partial y^m}{\partial x^n} \quad (4.4)$$

where the y^i 's are a new coordinate system defined in terms of the original coordinate system, the x^i 's. In differential equation theory, the Jacobian matrix plays a key role in defining the stability of solutions.

The residual vector (R) is defined as a vector containing the resulting values for each equation.

$$R = \begin{pmatrix} F(x_1) \\ \vdots \\ F(x_m) \end{pmatrix} \quad (4.5)$$

where m is the number of unknowns (or mass-action equations plus equilibrium constraints).

The algorithm consists of iteratively calculating new approximations for the unknown

values, through the matrix equation:

$$[J]_{iteration}^{-1} * \alpha[U]_{iteration+1} = [R]_{iteration} \quad (4.6)$$

where J is the Jacobian Matrix; $\alpha[U]_{iteration+1}$ is the unknown composition at the next iteration; iteration is the iteration number and R is the residual vector. With this, it is possible to state the $[U]_{iteration+1}$ value with:

$$[U]_{iteration+1} = [U]_{iteration} + \alpha[U]_{iteration+1} \quad (4.7)$$

The initial guess of the solutions is an approximation that the user provides to the *Newton-Raphson* method. This method needs a *seed* to start the calculations (usually this guess is used for $F(0)$). If the guess is close to the real root value the number of iterations necessary to obtain the solution is small. If the guess is far from the real solution, more iterations are needed to find the correct solution.

Specifically, *SHPECK*'s equations can be described as: $F_1(x_1, \dots, x_n), \dots, F_m(x_1, \dots, x_n)$. The partial derivatives of all these equations with respect to the variables x_1, \dots, x_n can be organized in a m-by-n matrix, the Jacobian matrix, as bellow:

$$J = \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \dots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1} & \dots & \frac{\partial F_m}{\partial x_n} \end{pmatrix} \quad (4.8)$$

In our case, $m = n$ and the *jacobian matrix* is a square matrix generated once and updated every iteration. With the equations selected and organized, the derivatives of each reaction towards each variable are calculated and the *jacobian matrix* is modeled and stored.

It is important to realize that the main complication of using the *Newton-Raphson* method to solve a system of nonlinear equations is defining the equations (and their derivatives) included in the *Jacobian Matrix*. As the number of equations and unknowns increases (n), so does the number of elements in the *Jacobian* (n^2).

4.7 Algorithm

SHPECK's algorithm takes as its input a specification of the system's state. For example, this consists of the values for the concentrations of the solutes in the solution, the temperature of the system, the method used to calculate activity coefficient, etc. It then calculates the system's

state by executing the numerical iterative process of achieving a solution. Figure 4.3 presents the high-level algorithm in details. It is important to understand that each of the boxes in this algorithm represents a set of instructions, calculations, and conditional clauses. Due to this work's scope we only specify the algorithm inside the box called "Newton's Method Solver". The algorithm of how the governing equations control the numerical methods is presented in figure 4.4.

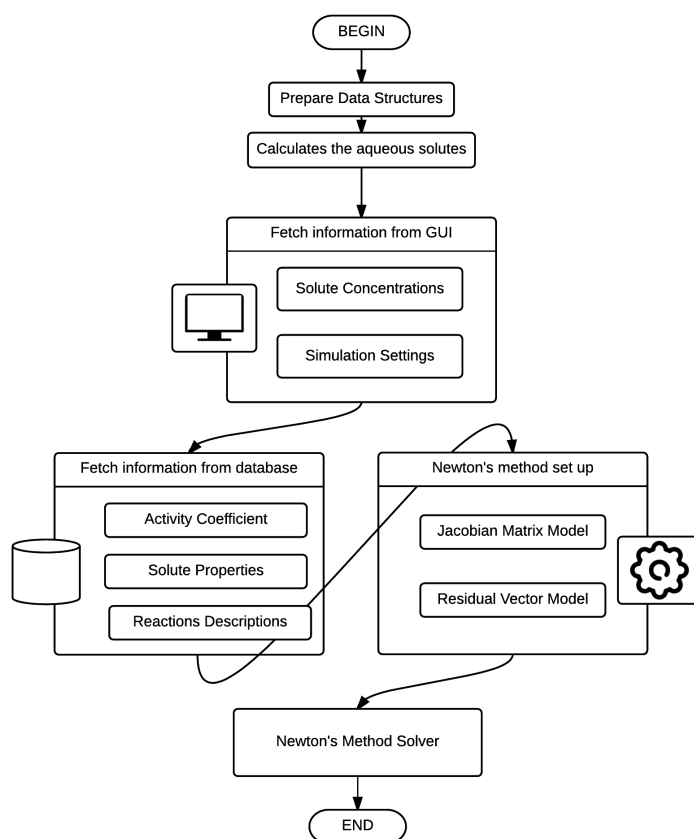


Figure 4.3 – High-level algorithm of *SHPECK*

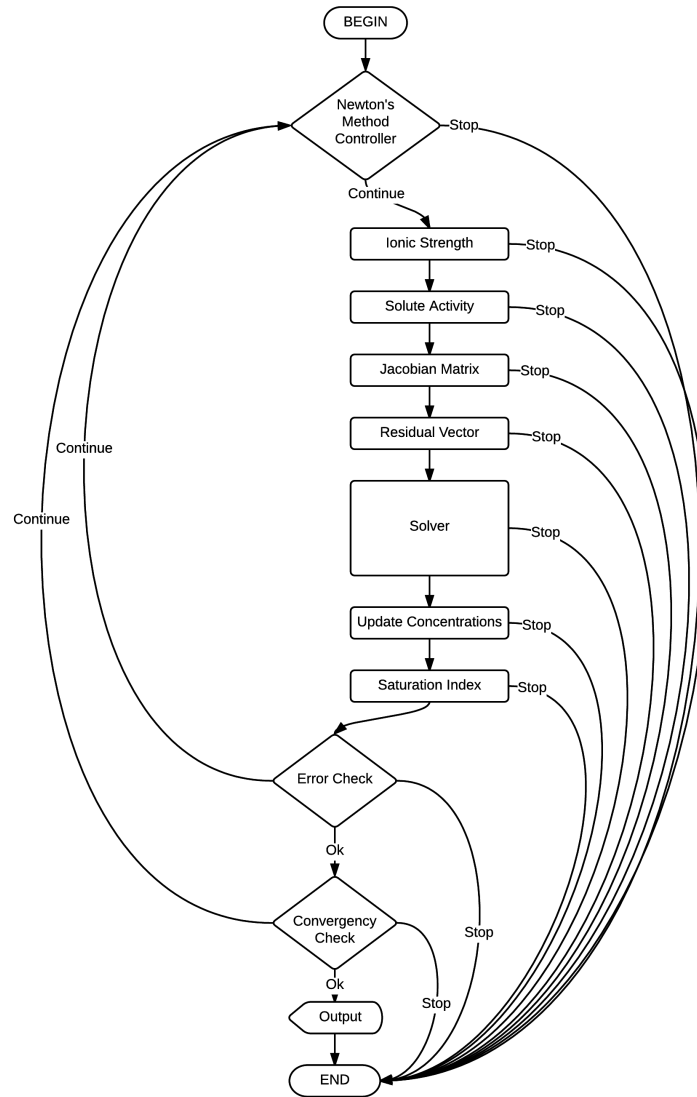


Figure 4.4 – Algorithm of the Newton's Method Solver

4.8 Graphical User Interface

Due to the enormous amount of options connected to the nature of geochemical modeling, it is necessary to develop the software as an intuitive and user-friendly utility. The development of the GUI introduces several stages of processes related to it:

- **Planning :** This moment is where the developer needs to conceptualize and evaluate the options toward constructing the core of the program that is intuitive and friendly to the user. The planning is an essential process. The content, features, and details of the software need to be strictly defined and organized. There is a small gap from a useful software from one that is not useful at all. Especially, guessing the role of a user is a difficult task, simply because the developer is not the final user for its applications. It is important also to find out if there are comparable software already in the public domain.

- **Building** : The planning for the implementation of the features of the software is important to be well defined at this point. To build the tool *Qt* - already presented above - was chosen due to its many advantages, since it is easy to customize, no coding was necessary, templates are available, provides a simple drag and drop *GUI* builder, and is secure and reliable;
- **Ensuring Usability** : The challenge in the field of Human-Computer Interaction (HCI) investigates the way in which people use computers. Techniques and tools are used to find relevant standards. These include tests with real users, evaluation by experts, gathering user feedback, and usage logging;

4.8.1 *SHPECK*'s GUI

In existing geochemical modeling software systems, the *GUI* are either poorly implemented, or not implemented at all. The *GUI* of *SHPECK* enables the user to streamline and conveniently construct a geochemical environment for modeling. *SHPECK*'s *GUI* works based on tabs, where each tab is responsible for viewing a specific point of the software. It is presented them in detail below:

- **Configurations tab**: this allows users to view and manipulate basic system settings and to control temperature, activity coefficient calculation method, the number of iterations, solver options, and error and convergence criteria. It is presented in figure 4.5.
- **Chemical phases in the water tab**: it allows users to create and edit the composition of the water that will be used in the geochemical speciation model. It has the complete catalog of species available from the database. The species in the tab that have concentrations other than zero compose the water. This tab is presented in figure 4.6.
- **Tab of results**: Allows the user to see relevant input information and the outputs of the geochemical speciation model, such as, temperature, ionic strength, pH of the solution, final concentration for the species, saturation indexes, etc. It is presented figure 4.7.

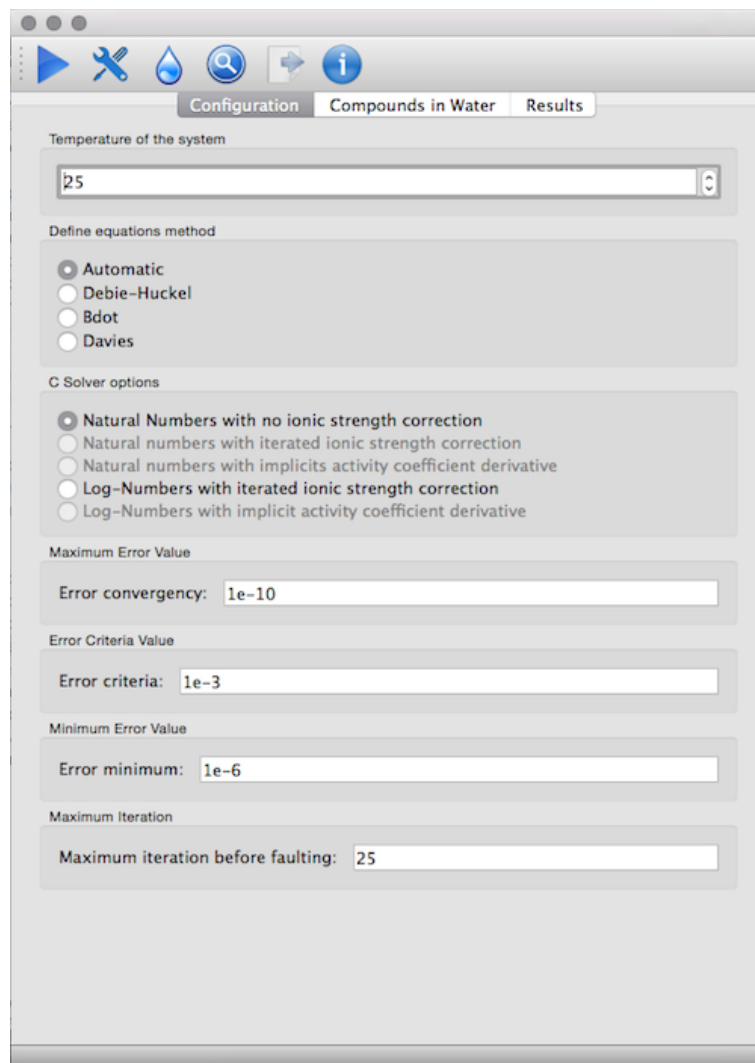


Figure 4.5 – Configuration tab

4.8.1.1 GUI Interaction

To demonstrate how the user interaction works in *SHPECK*, the selection of which method used to calculate the activity coefficient will be detailed. Figure 4.8 shows the available options (displayed to the user by radio buttons). There are four: automated, *Debye Hueckel*, *Bdot* and *Davies*. The automated option will choose among the available methods which one fits best to the modeled system. The methods were described in 2.1.1.5. Inside the logic of the program, the system identifies which one should be used (code 4.1) and calls the referring method (code 4.2).

Configuration Compounds in Water Results

Water Composition

	concentration ▼	name
1	1	H2O
2	0,5595	Cl-
3	0,47932	Na+
4	6,0256e-09	H+
5	0	Ag+
6	0	Al+++
7	0	Al(OH)3(aq)
8	0	Am+++
9	0	As(OH)4-
10	0	Au+
11	0	B(OH)3
12	0	Ba++
13	0	Br-
14	0	Ca++
15	0	Co++
16	0	CO3--
17	0	CO2(aq)
18	0	CH3COO-
19	0	Cr+++
20	0	Cs+

Revert Submit

Figure 4.6 – Chemical phases in the water tab

Code 4.1 – Activity coefficient calculation method implementation

```

if ((iStrength < I_STRENGTH_DBH_MAX && option == 0) || (option == 1)){
    // Debye-Huckel Equations
    qDebug() << MODEL_OF_EQUATIONS_DBH;
    if (option == 0)
        selectedEqMethod = "Automatic_(Debye-Huckel)";
    actCoeffList = ActCoeffDBH::actCoeffDBH(actCoeffDatabase.getADH(),
        actCoeffDatabase.getBDH(), iStrength,
        variables, charges, ionSizes);
}

```

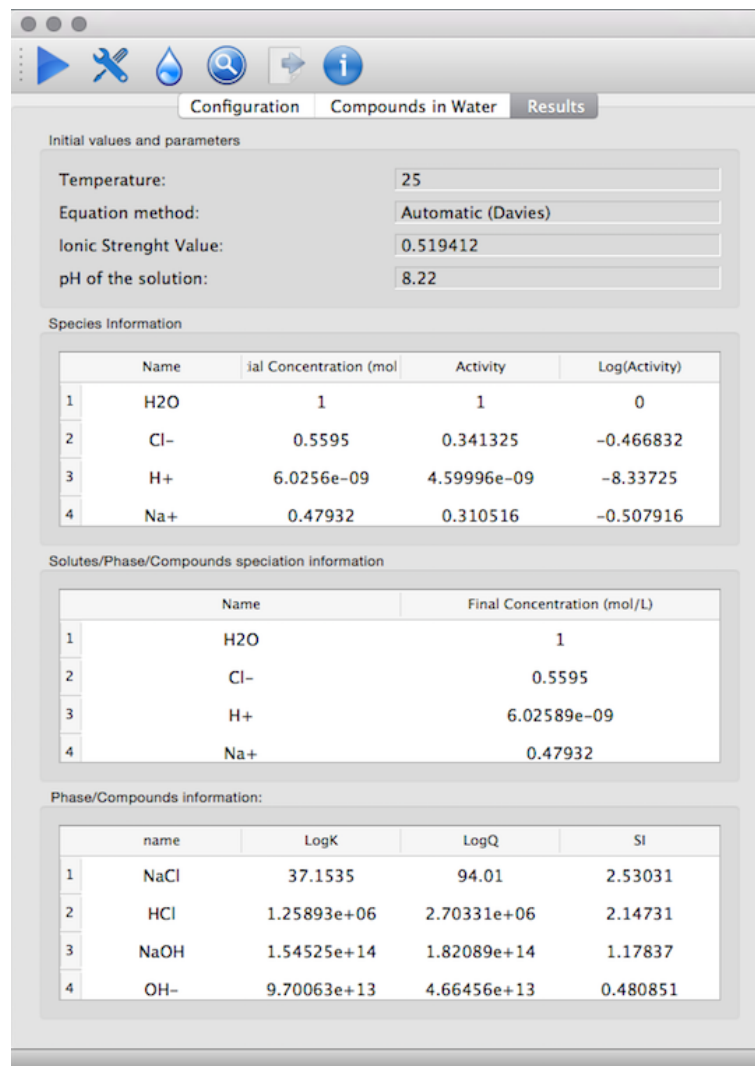


Figure 4.7 – Results tab

The screenshot shows a dialog box titled 'Activity Coefficient Calculation Method'. It contains four radio buttons for selecting the calculation method:

- ☒ Automatic
- ☐ Debye-Huckel
- ☐ Bdot
- ☐ Davies

Figure 4.8 – Activity Coefficient Selection Method

Code 4.2 – Debye Hueckel method implementation

```

QList<double> ActCoeffDBH::calcActCoeff(double adh,
                                       double bdh,
                                       double ionicStrength,
                                       QStringList variables,
                                       QStringList charges,
                                       QStringList ionSize)
{
    QList<double> actCoeff;
    for (int i = 0; i < variables.size(); i++){
        double value = (adh * qPow(charges.at(i).toDouble(),2.0) * qPow(
            ionicStrength,0.5) * -1)/
            (1 + ionSize.at(i).toDouble()* bdh * qPow(ionicStrength
                ,0.5));
        value = qPow(10.0,value);
        actCoeff.push_back(value);
    }
}

```

4.8.1.2 Chemical phases in the water

The chemical composition displayed to the user in the 'compounds tab' comes from the database. It shows the chemical species that are present in *SHPECK*'s database. Among these, the user can define the concentration for each one and this is corresponding to the water composition that will be simulated and speciated. A simple water composition is shown in details in figure 4.9. When the *SHPECK* runs, this water is fetched from the database (code 4.3), each query is treated as one component of the water (code 4.4) and each component is added to the water composition of the system (code 4.5).

Water Composition		
	concentration ▼	name
1	1	H2O
2	0,5595	Cl-
3	0,47932	Na+
4	6,0256e-09	H+

Figure 4.9 – Water composition detailed

Code 4.3 – Fetching the water composition from the database method

```
// Get water composition information from database
status = DatabaseHelper::getWaterComposition(&waterShpeck, dataBase, &pHValue
);
```

Code 4.4 – Excerpt of the water composition saving code

```
if (query.exec(sql))
{
    while(query.next())
    {
        // Set values inside the class WaterComposition
        // from the query returned from the database
        waterShpeck->addComponent (&query);
    }
}
```

Code 4.5 – Detailed composition of the water

```
Composition* composition = new Composition();
composition->setConcentration(query->value(0).toFloat());
composition->setName(QString(query->value(1).toString()));
composition->setCharge(query->value(2).toFloat());
composition->setIonSize(query->value(3).toFloat());
composition->setMoleWt(query->value(4).toFloat());
composition->setNumberOfElements(query->value(5).toInt());
componentsList.push_back(*composition);
```

4.9 Database

As can be seen in this chapter, the algorithm of *SHPECK* is not trivial and requires many interactions between many different entities. The database is responsible for providing the data that will flow through *SHPECK*. Chapter 3 makes clear that using a flat file database poses difficulties to users. Potential issues of using flat file databases are: duplication of the information; non-unique records; difficulty in updating and maintenance; inherently inefficient; rigid (difficult data format); and insecure ???. In *SHPECK*, we use a relational database, which is a model of a database that prevents the problems faced in flat file database previously discussed. A relational database works based on a *query*, which is an information request from the database. Besides that, there are three important advantages:

- Memory saving, since the information is not fully load in main memory during run time. The software fetches the information when required by each module. The data is permanently stored and relationally structured according to the requirements of the application.
- Direct access to relevant information: *SQL* queries are built and fetched when required by the software modules, avoiding sequential search over text files. Only the chemical phases that are part of the simulated reaction are load to memory to be processed by the software. Complex queries and concatenation of queries result in a faster and more efficient use of the available resources. The semantic connection between the relational data allows richer queries semantically when fetching data.

The database architecture, presented in details in figure 4.10, was defined after studying the algorithm and determining what would be the structure and the information requested. The relational database embraces all the important data requested for a geochemical modelling software and the structure was defined in order to compact, organize, structure and make the information access efficient by the modules.

4.9.1 Database Technologies

The implementation was done with *SQLite* database (HIPPI, 2015), which is a software-library that implements a self-contained, transactional *SQL* open source database engine. Currently it is the most widely deployed *SQL* database engine in the world. Some of its advantages are presented below:

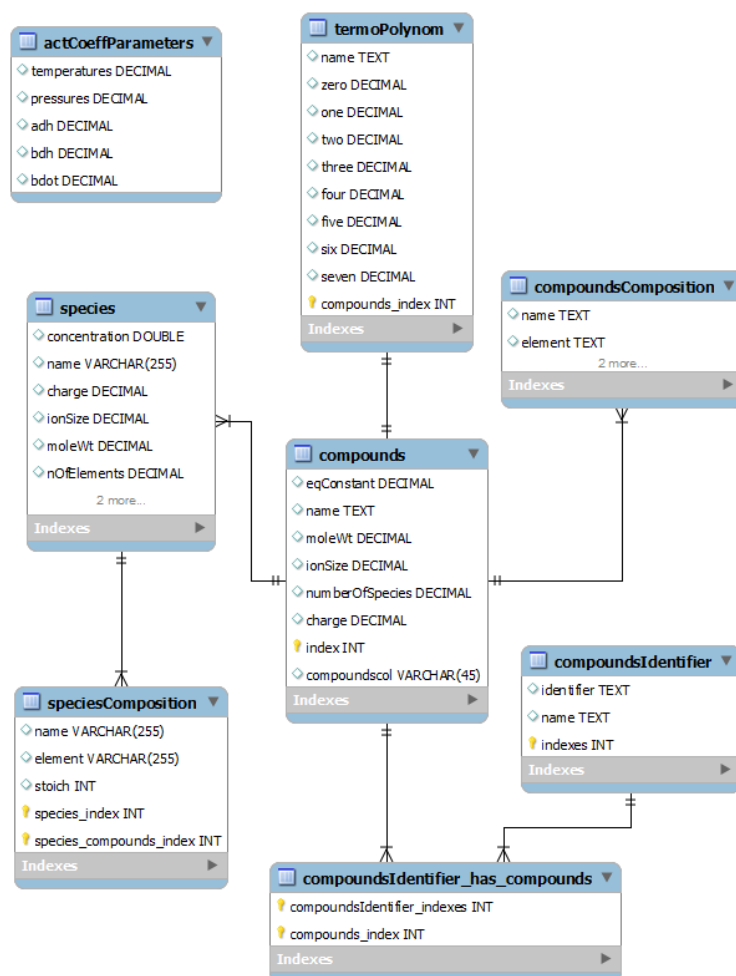


Figure 4.10 – ER Diagram of the database

- **Zero-Configuration:** *SQLite* does not need to be installed before it is used, there is no setup procedure;
- **Serverless:** The process that wants to access the database reads and writes directly from the database files on disk. There is no intermediary server process (nor interprocess communication using *TCP/IP*);
- **Single Database File:** An *SQLite* database is a single file located in the directory hierarchy. *SQLite* database can be easily copied onto a USB memory stick or emailed for sharing;
- **Stable Cross-Platform Database File:** A database file written on one machine can be copied and used on a different machine with a different architecture. Furthermore, *SQLite* is backwards compatible (newer versions can read and write older database files);

4.9.2 LLNL thermodynamic dataset parser

Once the structure and the technology of our database were defined, a parser was prepared to extract the information from the flat file database. The parser was generated carefully in order to ensure that all of the data and delimiters were identified and treated properly. This task was extremely time-consuming because of irregularities found in the flat files (i.e. wrong tab or space formatting, different encoding of the files, etc). Figure 4.11 describes in details the parser.

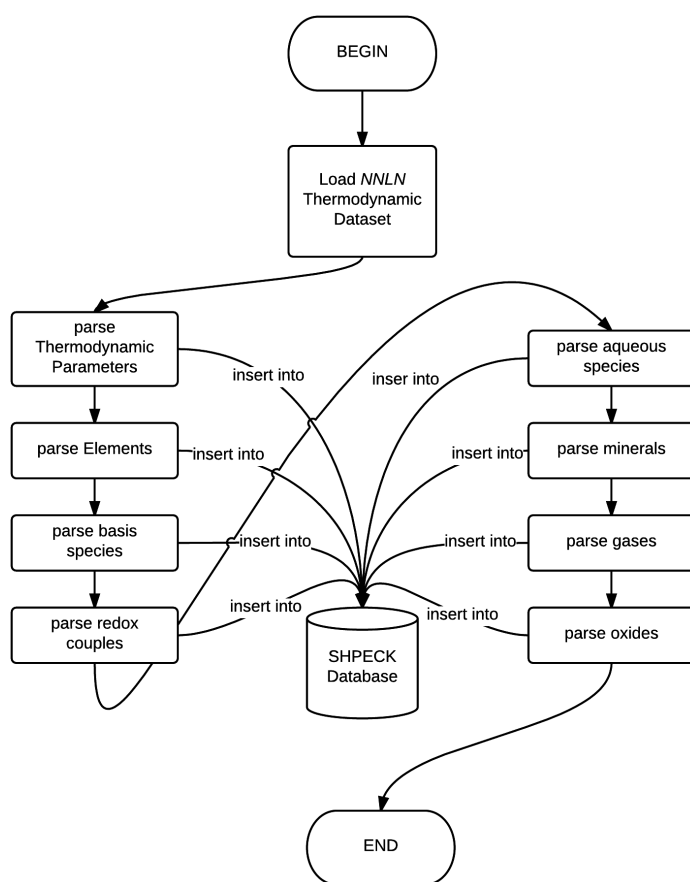


Figure 4.11 – Parser algorithm

4.10 Summary

- **Architecture:** *SHPECK* follows the architectural pattern called *MVC*; Its benefits are the complete separation of responsibilities and concerns among the parts of the software. Thus, there is no mixing of codes between them. Another advantage is its flexibility that allows the software to grow and develop further. Figure 4.2 displays the *MVC* pattern existing in *SHPECK*.
- **Governing Equations:** *SHPECK* is a geochemical speciation modeling software that drives the behavior of the aqueous system based on a set of mass-action equations combined with equilibrium constraints. A mass-action equation is described in equation 2.2 and reinforced here:

$$K_j = \prod_{i=1}^N a_i^{v_{ij}} \quad (j = 1, \dots, M) \quad (4.9)$$

where K_j denotes the equilibrium constant of the j -th reaction; a denotes the activity of the i -th chemical species.

- **Numerical Method:** *SHPECK* applies the *Newton-Raphson* method to solve the nonlinear system of equations. The concept of the method is described in equation 4.3 and reinforced here:

$$F(x + 1) = F(x) - J^{-1} * R \quad (4.10)$$

where F is function's result for the applied x , J^{-1} is the inverse of the Jacobian matrix, and R is the residual vector. The quadratic rate of convergence of this method compensate the expensive calculation inherent to it.

- **Graphical User Interface:** The *GUI* mission is to enable the user to use the full potential *SHPECK* has to offer. It is intuitive and user-friendly to allow the user to focus on essential duties that are imperative to modeling a geochemical environment. We use the popular approach of *tab* panels, which are separated according to their purposes: configurations and settings, water composition, and results visualization.
- **Database:** A geochemical modeling software is dependent on the information provided in the database. *SHPECK* structures its data in a *SQLite* relational database, which is unique approach among the comparable software available. *SHPECK*'s database is composed of the information on elements, species, compounds, reactions and thermodynamic

Table 4.1 – Updated version of table 3.2

SOFTWARE	Costs	Setup and versioning	Customization and Integration	Security and Control	Infrastructure	Core functionality	Graphical User Interface	Support and Maintenance	Database	Overall Average
EQ3/6	2	2	1	1	3	5	1	1	1	1.88
PHREEQC	4	4	2	2	3	4	2	3	1	2.77
MINTEQA2	4	2	1	1	2	2	1	1	1	1.66
SOLMINEQ.88	3	1	1	1	1	5	1	1	1	1.66
SHPECK	4	4	2	3	3	5	5	5	5	4

constants found in the *LLNL* thermodynamic dataset. A parser for *LLNL* flat file database was created to fetch this information.

- *SHPECK*'s software engineering: A set of techniques provided in ?? were carefully followed along every step in the development of *SHPECK*.
- The updated version of table 3.2, including *SHPECK*, is demonstrated in table 4.1.

5 VERIFICATION AND EVALUATION

In order to test and evaluate the simulator, we model the diagenetic reactions observed in Snorre Field reservoir sandstones of Norwegian North Sea. The main reservoir horizons of the field are the fluvial sandstones in the upper member of the Upper Triassic Lunde Formation and the Upper Triassic to Lower Jurassic Statfjord Formation (??). The sandstones sampled for this study, according to Morad (MORAD. S., 1990), belong to the upper member of the Lunde Formation. The sandstones are dominantly fine- to medium-grained and arkosic with framework constituents of quartz (40-80%), K-feldspar (5-12%), plagioclase (15-45%), muscovite, biotite and clay minerals that include smectite, mixed-layer clay minerals, chlorite and subordinate amounts of kaolinite and illite. Subordinate rock fragments include intraformational mudstone and carbonate clasts and extrabasinal grains of quartz-feldspar-mica aggregates that probably represent granitic rocks and/or schist or gneisses. Mica and detrital clay minerals seldom make up more than 2% of the total mineral content in the sandstones. Diagenetic clay minerals include pore-filling kaolinite and pore-lining smectite, mixed-layer chlorite-smectite, and chlorite. Other cements include the carbonates (0.0-25%) which play a significant role in porosity reduction in some of the sandstones. Authigenic overgrowths are primarily quartz, anatase and minor albite, pyrite and barite.

5.1 Case Study

For the model's validation, the diagenetic reactions observed in Snorre Field reservoir sandstones, Norwegian North Sea was simulated. The modeling set up and results given in Morad (MORAD. S., 1990) describes the texture, origin, chemistry of the sandstones reservoirs in terms of the water composition and temperature. Against this *SHPECK*'s results were examined.

The descriptions presented in (MORAD. S., 1990) of the diagenetic reactions that take place in the Snorre Field allows us to generate a computationally comparative study. We model and compare the same diagenetic environment using *SHPECK*, *PHREEQC* and *MINTEQA2*. The water composition is detailed in (NORDSTROM, 1979).

We approach the environment described above following two methodologies: experimental and computational. The first one analyzes and compare the behavior of *SHPECK* with petrographic analysis through thin sections (available in (MORAD. S., 1990)). The second treats the direct comparison of *SHPECK*'s result side-by-side with other software's.

5.1.1 Experimental validation of Shpeck

As stated in (MORAD. S., 1990), the model presented in (EGEBERG; AAGAARD, 1988) calculates activities of the various ions of formations waters using ion association model (originally described in (WIGLEY, 1977)). The thermodynamics data used are given in (HELGESON; KIRKHAM, 1974b), (HELGESON; KIRKHAM, 1974a), (HELGESON; KIRKHAM, 1976), (WALTRER; HELGESON, 1977), (HELGESON H. C.; BIRD, 1978) and (HELGESON H. C.; FLOWERS, 1981).

The activity diagram generated for a known temperature and log activity ratio of Potassium to Sodium ions of (AAGAARD, 1990) is provided as input to *SHPECK* model (Figure 5.1). The results show a consistent pattern: as the temperature rises the potassium activity gets higher over sodium's, which means that the phases associated to the ion potassium (i.e. K-feldspar, Illite, etc) are dissolving.

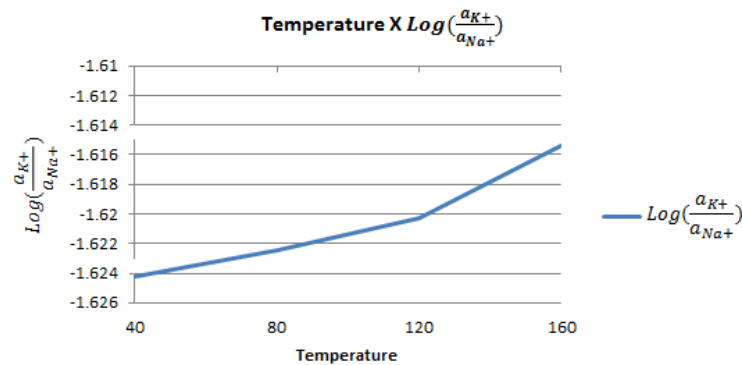


Figure 5.1 – Log activity ratio of Potassium to Sodium ions using the results from *SHPECK*

5.1.2 Computational comparative study

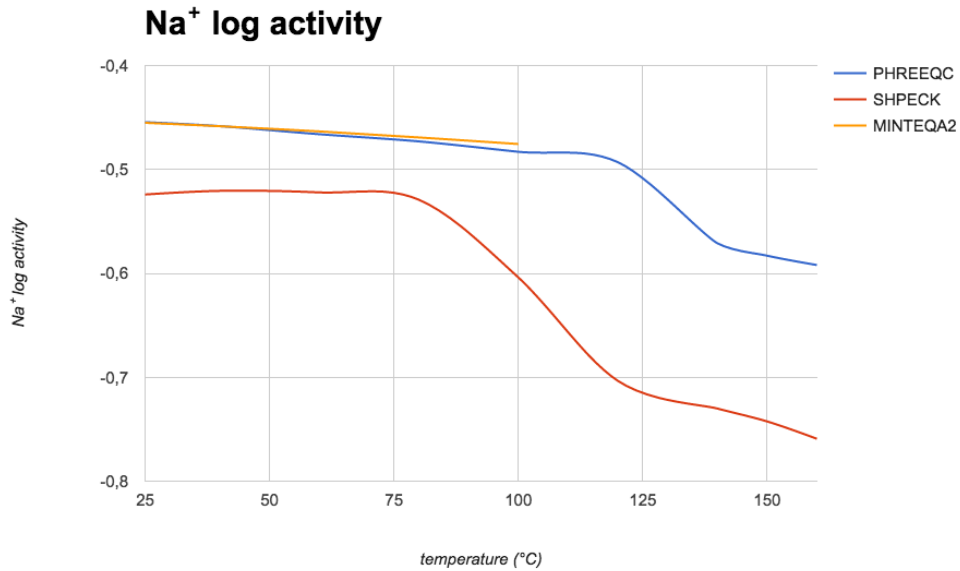
By modeling the same environment using three distinct software packages, we attain a relevant comparison among the numerical methods and algorithms. Chemical composition of the water adopted in the models is taken from (NORDSTROM, 1979) which provides the chemical composition of the seawater (table 5.1). The comparative study tested temperatures that have varied from 100°C to 100°C. In *MINTEQA2*, due to limitations of its thermodynamics equilibrium database, the maximum temperature available is 100°C.

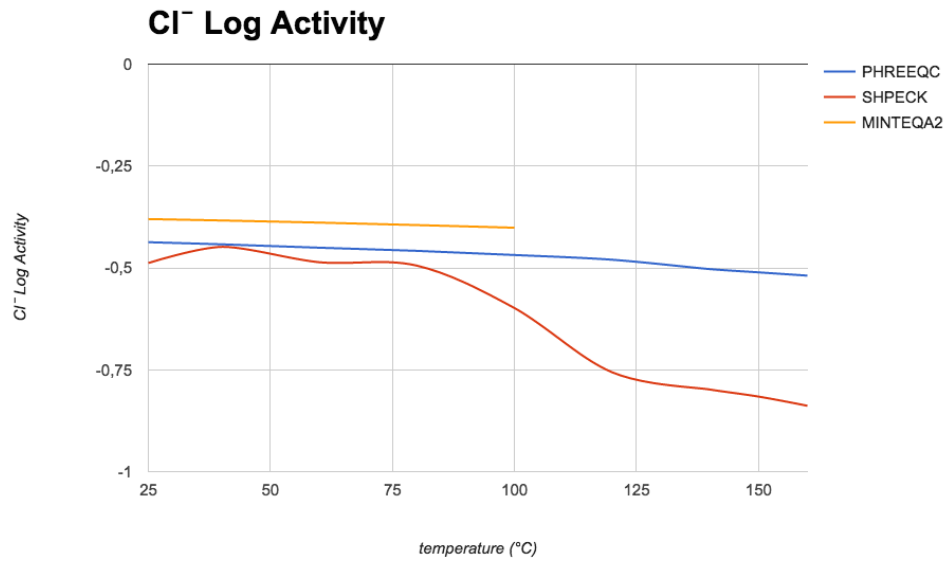
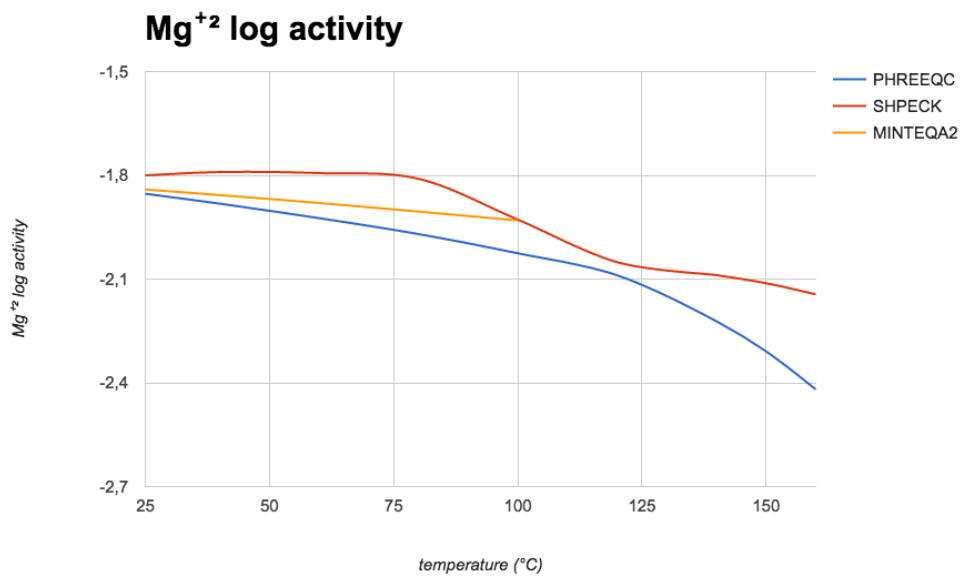
Table 5.1 – Chemical composition of the solution in the sea water at 25° in *mM/LC*

Al^{3+}	K^+	Na^+	Ca^{2+}	Mg^{2+}	Fe^{2+}	SiO_2	SO_4^{2-}	Cl^-	pH
7.59e-5	10.45	479.32	10.53	54.39	3.66e-5	0.073	28.893	559.5	8.22

We aim to model the diagenetic processes that best represent the behavior of ions in the water-rock interactions. Figures 5.2, 5.3, 5.4 and 5.5 present the most representative ions of the solution. It is possible to see that the behavior of *SHPECK* is similar to both *PHREEQC* and *MINTEQA2* in most of the cases, especially in temperatures under 100°C. We observe a more dissimilar behavior between *SHPECK* and *PHREEQC* in temperatures higher than 100°C, but the results were never completely opposite. This can be explained by the temperature intervals where the equilibrium constant K is not completely defined.

This is a known issue from *LLNL* thermodynamic dataset: sometimes equilibrium constants have no measures in literature and are treated as unknown. *SHPECK* adopts the nearest equilibrium constant known value. Unfortunately we do not have access to the whole software's details in order to describe how they treat this issue.

Figure 5.2 – Na^+ log activity comparative study

Figure 5.3 – Cl⁻ log activity comparative studyFigure 5.4 – Mg⁺² log activity comparative study

5.2 Database Evaluation

The geochemical models used for comparison studies use text files as the database. The goal of this section is to make clear the difference and - more importantly - the benefits of *SHPECK*'s relational database. Most of the information inside a geochemical database is related to each other (i.e. a mineral phase is described by a reaction, a reaction is composed by solute species, and a solute species is composed of chemical elements). *SQLite* databases are naturally

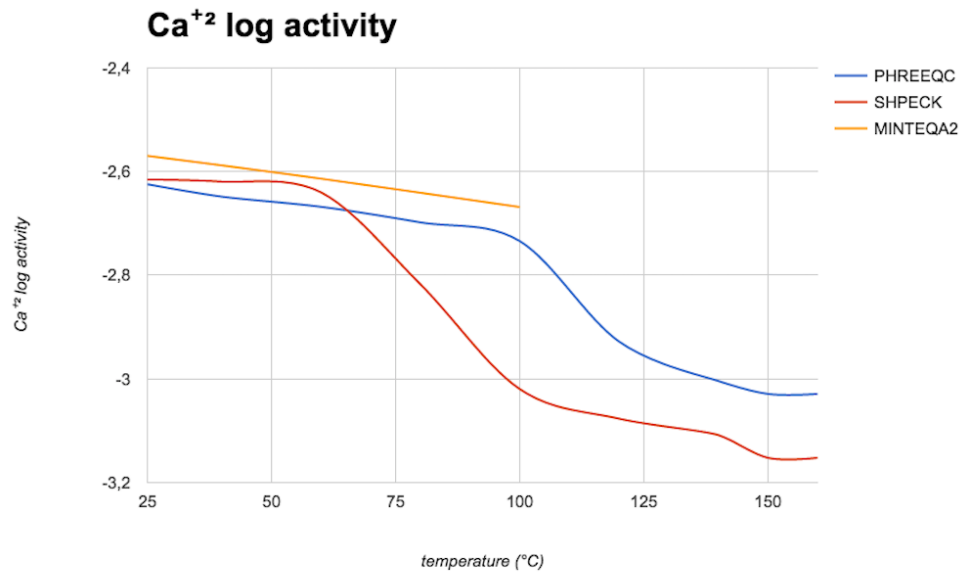


Figure 5.5 – Ca²⁺ log activity comparative study

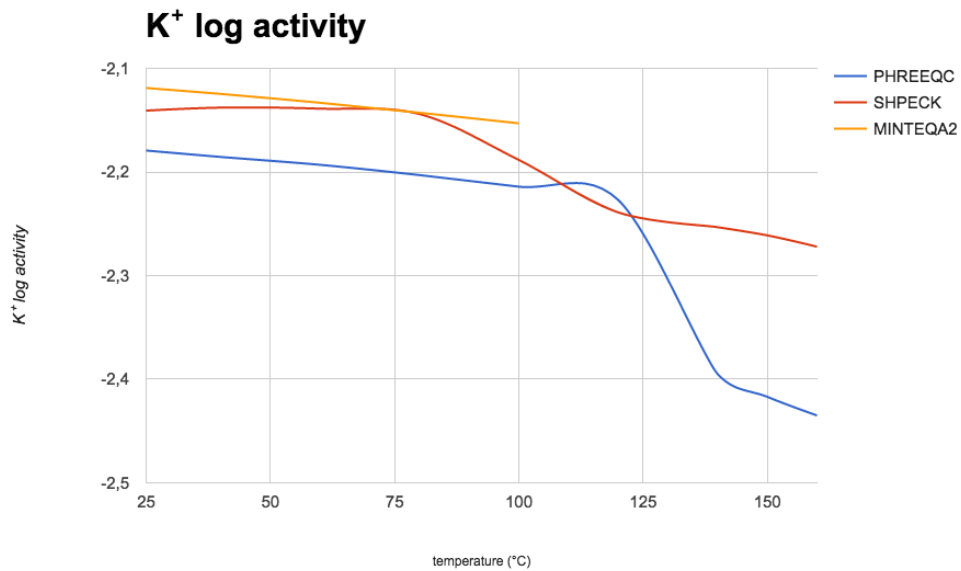


Figure 5.6 – K⁺ log activity comparative study

a structure where the data can be related to each other, and this significantly improves the performance and robustness of the application. On *SQLite* databases, the data can be accessed using *SQL* queries that reduce the complexity and increase the speed on information retrieval - code 5.1 shows a typical query issued from *SHPECK*'s *SQLite* query. Text database files only allow sequential access to the information that it contains. In relation databases, the *SQL* language allows random access and it is used to selectively locate needed data.

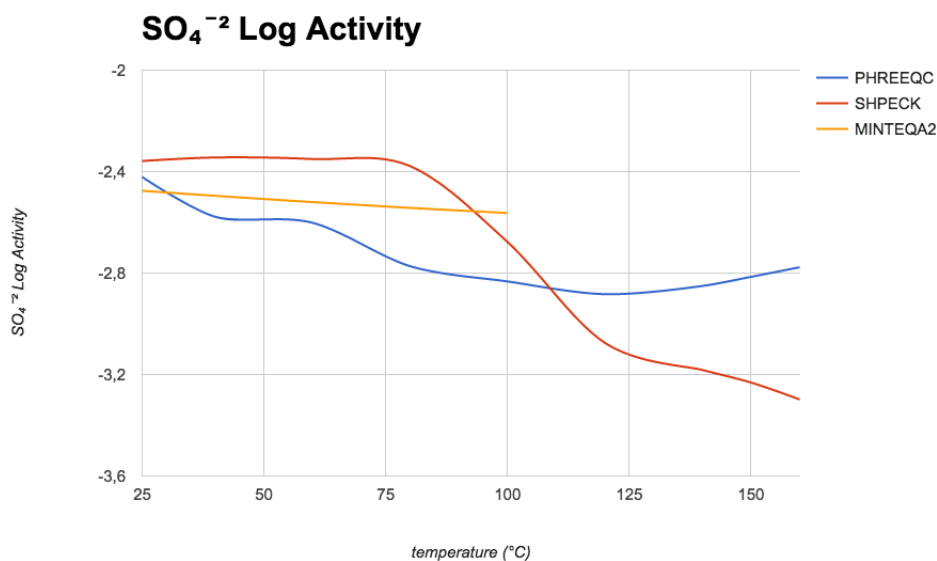


Figure 5.7 – SO_4^{-2} log activity comparative study

Code 5.1 – *SHPECK*'s *SQLite* example query

```
SELECT compoundsComposition.element, compoundsComposition.name,
       compoundsComposition.stoich, compounds.eqConstant, termoPolynom.zero,
       termoPolynom.one, termoPolynom.two, termoPolynom.three, termoPolynom.
       four, termoPolynom.five, termoPolynom.six, termoPolynom.seven FROM
       termoPolynom, compoundsComposition, compounds WHERE
       compoundsComposition.name = compounds.name AND termoPolynom.name =
       compounds.name AND compounds.name = XXXXX;
```

Important to note that at each simulation the value "XXXX" in code 1 is updated to the simulation's solute name. Also interesting to point that the information fetched with this query comes from three different tables and, in only one query, many relevant information for the simulation is fetched.

5.2.1 Configuration setup

Our experiments' methodology was first defining what was the information necessary from the database to generate the set of reactions. After that, we define the order that this information was necessary and how to fetch it. At the end, the measurement metric is to verify how long the program was idle waiting for this information and when it finally was available. In both the relational database and flat file experiments, we used the *NNLN* thermodynamic dataset. The tests were executed on a MacBook air, i7 processor, 1.7GHz, 8GB RAM running OS X 10.9.3.

5.2.2 Time analysis of fetching information

The objective of the time analysis is to measure how long it takes to fetch the same information from different types of database. The response time is considered as the sum of the processing time and the time waiting for the availability of the resource. It is necessary to understand that until the software has received the information requested from the database it is inactive and in a standby mode. In order to analyze the response time within a geochemical analysis point of view, we discuss not only the access time but also the implications of that information.

When fetching any information from a thermodynamic dataset, it is important to take into consideration additional allied data will also have to be retrieved. For example, when fetching a reaction (as expressed in equation 2.1) data, the basic information consists of the compounds that take part on this reaction and the related stoichiometric values. Behind this action, the database must also provide information about the compounds itself (i.e. charge, ion size, mole weight, elements in that specie, formula, mole volume) as well as the reaction (i.e., thermodynamic equilibrium constant coefficients, etc).

Figure 5.8 indicates the time elapsed (in seconds) that it takes to retrieve the necessary information related to a chemical reaction from the database. In this example, we simulated from 20 to 580 randomly selected reactions in the database. It is possible to see that *SHPECK*'s database has improved approximately 40% in the average time elapsed to fetch the information if compared to the regular text file database used by other simulation packages.

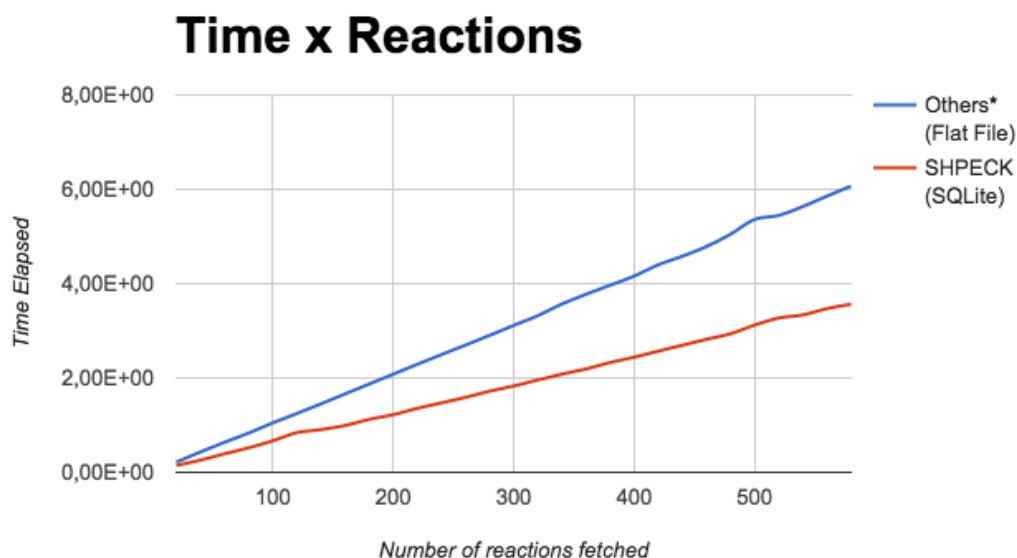


Figure 5.8 – Time elapsed in seconds X Reactions Accessed

5.3 Summary

- **Diagenesis:** This term refers to chemical and physical changes taking place in a rock due to chemical reactions. In this study we reproduce the diagenetic reactions observed in Snorre Field reservoir sandstones, Norwegian North Sea. The environment modeled is described in (MORAD. S., 1990) and the chemical composition of the water in (NORDSTROM, 1979).
- **Comparative study:** We performed a comparative study of *SHPECK* and other available geochemical speciation software. The results prove that *SHPECK* produce comparable results. The discrepancies are minimal and the differences can be justified by the software-specific implementations of the mathematical and computational treatment to the set of equations and other parameters.
- **Database evaluation:** The lack of a relational database in the geochemical modeling software makes it clear that none of the existing options were developed with emphasis on a computer science emphasis. *SHPECK* uses a *SQLite* database specially developed to support a geochemical speciation modeling software. This software design option enables the software to achieve efficient performance, and handle complex data queries and retrieval. A comparative graph highlights the advantages in time elapsed by the number of reactions requested.

6 CONCLUSION

This work provides details of *SHPECK*, a modeling software package intended to raise the quality of available geochemical speciation. The program innovates by including a visual user interface that allows to parametrize the reaction simulation, as well as a relational database architecture component. The user interface implements a dynamic tool that facilitates the simulation case preparation. *SHPECK* accepts any general combination of chemical elements, mineral species and chemical reactions, along with definitions of boundary contour conditions of temperature, pH, pressure, etc.

SHPECK is built with considerations for computer science's priorities and theories. As would be expected, the first part of this work was to understand the topic of geochemical modeling. Subsequently, available alternative options of geochemical models were evaluated. Based on the findings, the program *SHPECK* was designed to address the weakness of the existing alternative models.

The first challenge was to design the database that would fit the information of the geochemical speciation modeling software. This information was parsed from a well-established source and organized in a *SQLite* relational database. Once *SHPECK*'s database was ready, the speciation methodology was coded.

Speciation is the calculation of the distribution of dissolved solutes in the water and saturation indices of various minerals. *SHPECK* calculates the speciation of a solution based on a set of mass-balance equations, which are solved iteratively using the Newton-Raphson numerical method.

Once the implementation of both database and mathematical treatment was completed, the user interface was developed. *SHPECK*'s *GUI* allows significantly simpler method of setting up a geochemical model compared to the available methods.

The completed software was tested and its results compared with the results obtained through the alternative softwares. The results produced by *SHPECK* were found to be consistent with those of the alternative models.

SHPECK's results accuracy is equivalent and inside the expect range to the study case developed in this work. The results show that all the three software compared have nearly the same results with the same inputs. Slight differences are found when simulating with temperatures higher than 100°C, which appear to arise from discrepancies in the thermodynamic properties of the equilibrium constant.

Due to reasons, like time and purpose, there are still some points along this work where

there are enhancements to be done. For instance, add the kinetic reactions to *SHPECK*'s solver - it is a delicate topic since adding kinetic reactions implies not only working with mass balance equations but also with chemical elemental mass evolution. The elemental mass evolution describes mass change through mass transfer and kinetic reactions of solids and solute-solute interaction - described in details in (PARK, 2014). Moreover, we also plan to create a distribution platform where users can register, interact, collaborate in forums, indicate bugs, and download *SHPECK*.

REFERENCES

- AAGAARD, P. E. A. Diagenetic albitization of detrital k-feldspars in jurassic, lower cretaceous and tertiary elastic reservoir rocks from offshore norway. if. formation water chemistry and kinetic considerations. **Jour. Sed. Petrology**, 1990.
- ALLEY, W. M. **Regional Ground-Water Quality**. [S.l.]: Wiley New York, 1993. 634 p.
- ALLISON J. D., B. D. S.; NOVO-GRADAC, K. J. Minteqa2/prodefa2, a geochemical assessment model for environmental systems: version 3.0. **Environmental Research Laboratory, Office of research and development, U.S. Environmental Protection Agency**, 1991.
- APELLO, C. **NotPHREEQC - Notepad++ with PHREEQC Plugin**. 2011. Disponível em: <www.hydrochemistry.eu/ph3>.
- BROWN, D. S.; ALLISON, J. D. **MINTEQA1, an equilibrium metal speciation model: user's manual**. [S.l.]: Environmental Research Laboratory, Office of research and development, U.S. Environmental Protection Agency, 1987.
- DALE, N. B. **Programming and problem solving with C++**. [S.l.]: Jones Bartlett Publishers, 2004.
- DEBRAAL, J. D.; KHARAKA, Y. K. Solinput: A computer code to create and modify input files for the geochemical program solmineq.88. **U.S. Geological Survey Books and Open-File Reports**, 1989.
- DREVER, J. I. **Surface and Ground Water, Weathering, and Soils**. [S.l.]: Elsevier, 2005.
- EGEBERG, P. K.; AAGAARD, P. Formation water chemistry in relation to the stability of detrital and authigenic minerals in elastic reservoirs from offshore norway of oslo. **University of OSLO**, 1988.
- FELMY, A. R. et al. Minteq: A computer program for calculating aqueous geochemical equilibria. **Environmental Research Laboratory, Office of research and development, U.S. Environmental Protection Agency**, 1983.
- FREEZE, R. A.; CHERRY, J. A. **Groundwater**. [S.l.]: Prentice Hall, 1979.
- GAMMA, E. E. A. **Design Patterns. Elements of Reusable Object-Oriented Software**. [S.l.: s.n.], 1994.
- GARRELS, R. M.; CHRIST, C. M. **Solutions, Minerals, and Equilibria**. [S.l.]: Harpers' Geoscience Series, 1965.
- HEESCH, D. van. **Doxygen**. 2013. Disponível em: <<http://www.stack.nl/~dimitri/doxygen/>>.
- HELGESON, H. C.; KIRKHAM, D. H. Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures: II. debye-hückel parameters for activity coefficients and relative partial molal properties. **Am. Jour. Science**, 1974.
- HELGESON, H. C.; KIRKHAM, D. H. Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures: 1. summary of the thermodynamic/electrostatic properties of the solvent. **Am. Jour. Science**, 1974.

HELGESON, H. C.; KIRKHAM, D. H. Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures: Iii. equations of state for aqueous species at infinite dilution. **Am. Jour. Science**, 1976.

HELGESON H. C., D. W. J. M. N. H. W.; BIRD, D. K. Summary and critique of the thermodynamic properties of rock-forming minerals. **Am. Jour. Science**, 1978.

HELGESON H. C., K. D. H.; FLOWERS, G. C. Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures: Iv. calculation of activity coefficients, osmotic coefficients, and apparent molal and standard and relative partial molal properties science. **Am. Jour. Science**, 1981.

HIPP, D. R. **SQLite**. 2015. Disponível em: <<https://www.sqlite.org>>.

ISAACSON E., K. H. B. **Analysis of Numerical Methods**. [S.l.]: Wiley New York, 1966.

KEHEW, A. **Applied Chemical Hidrogeology**. [S.l.]: Prentice Hall, 2000. 368 p.

KHARAKA, Y. K.; BARNES, I. Solmneq: Solution-mineral equilibrium computations. **NTIS Tech Report**, 1973.

KHARAKA Y. K., G. W. D. A. P. K. P. E. H.; DEBRAAL, J. D. Solmineq.88; a computer program for geochemical modeling of water-rock interactions: Water-resources investigations report. **U.S. Geological Survey Books and Open-File Reports**, 1988.

LEAL, A. M. et al. A chemical kinetics algorithm for geochemical modelling. **Applied Geochemistry**, v. 55, p. 46–61, 2015.

MEYER, B. **Object-Oriented Software Construction**. [S.l.]: Prentice Hall, 2000.

MINTEQ. 1999. Disponível em: <http://eng.odu.edu/cee/resources/model/minteqa_unix.shtml>.

MINTEQ. 2006. Disponível em: <<http://www2.epa.gov/exposure-assessment-models/minteqa2>>.

MORAD. S., E. A. Albitization of detrital plagioclase in triassic reservoir sandstones from the snorre field, norwegian north sea. j. sediment. **U.S. Geological Survey Books and Open-File Reports**, 1990.

NORDSTROM, D. K. e. a. A comparison of computerized chemical models for equilibrium calculations in aqueous systems. **American Chemical Society**, 1979.

NORDSTROM D., M. K. **Geochemical Thermodynamics**. [S.l.]: Prentice Hall, 1986. 477 p.

PALANDRI L. JAMES, K. K. Y. A compilation of rate parameters of water-mineral interaction kinetics for application to geochemical modeling. **U.S. Geological Survey**, 2004.

PARK, A. J. Water-rock interaction and reactive-transport modeling using elemental mass-balance approach: I. the methodology. **American Journal of Science**, 2014.

PARKHURST, D. L. User's guide to phreeqc - a computer program for speciation, reaction-path, advective-transport, and inverse geochemical calculations. **U.S. Geological Survey Water-Resources Investigations Report**, 1995.

PETRUCCI, R. H. **General Chemistry: Principles & Modern Applications**. [S.l.]: New Jersey : Prentice-Hall, 2007.

QT. **Qt Application Framework**. 2014. Disponível em: <<http://www.qt.io>>.

SANDERSON, C.; CURTIN, R. **Armadillo C++ linear algebra library**. 2014. Disponível em: <<http://www.arma.sourceforge.net>>.

SARKER R. A., N. C. S. **Optimization Modelling - A Practical Approach**. [S.l.]: CRC Press, 2008.

SHARIF, M. U. E. A. Inverse geochemical modeling of groundwater evolution with emphasis on arsenic in the mississippi river valey alluvial aquifer, arkansas (usa). **Journal of Hidrology**, 2007.

SMITH, W. R. The computation of chemical equilibria in complex systems. **American Chemical Society**, 1980.

SMITH, W. R.; MISSEN, R. W. **Chemical reaction equilibrium analysis: Theory and algorithms**. [S.l.]: John Wiley, 1983.

STROUSTRUP, B. **The C++ Programming Language (Third ed.)**. [S.l.: s.n.], 1997.

WALTRER, T. E.; HELGESON, H. C. Calculations of the thermodynamic properties of aqueous silica and the solubility of quartz and its polymorphs at high pressures and temperatures. **American Journal of Science**, 1977.

WESTALL, J. et al. Mineql, a computer program for the calculation of chemical equilibrium composition of aqueous systems. **American Chemical Society**, 1976.

WIGLEY, T. M. L. A computer program for determining the equilibrium speciation of aqueous solutions. **Technical Bulletin No. 20, Climatic Research Unit, Univ. East Anglia, U.K**, 1977.

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.

WOLERY, T. J. Eq3/6, a software package for geochemical modeling of aqueous systems: package overview and installation guide (version 7.0). **American Chemical Society - Symposium series, No 416**, 1992.

WOLERY T. J., E. A. Current status of the eq3/6 software package for geochemical modeling in chemical modeling of aqueous system. **American Chemical Society - Symposium series, No 416**, p. 104–116, 1990.

ZEGGEREN, F. V.; STOREY, S. H. **The Computation of Chemical Equilibria**. [S.l.]: Cambridge University Press, 1970.