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

SUMMARY

LIST OF ABBREVIATIONS AND ACRONYMS

a   Activity

ASCII   American Standard Code For Information Interchange

CRUD   Create / Read / Update / Delete

CPU   Central Processing Unit

DBH   Debie-Hueckel

E_a Activation Energy

Eh   Redox Potencial

FLOPS   Floating-Point Operations per Second

G   Gibbs Energy

GUI   Graphical User Interface

GWB   The Geochemist's Workbench

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
γ	Activity coefficient
β_i	Stability Constant

LIST OF FIGURES

ABSTRACT

Chemical modelling is essential to comprehend many environmental problems. We present a flexible and general computer software, called *SHPECK*, to calculate the geochemical speciation modelling dynamically and efficiently. Geochemical modelling applications require an extremely high level of computations in a single simulation. The method uses the stoichiometric approach (also known as law of mass-action approach) coupled with mass-action equations and a system of equilibrium constants solved using Newton's method. An adaptive control scheme of internal factors of the simulation is adopted to guarantee that chemical and physical processes are respected and follow the literature. The software operates with customised options as capabilities to specify pH (activity of Hydrogen ions) of an aqueous solution, concentrations of the species, convergence criteria and maximal number of iterations. The proposed algorithm developed is described carefully from a software development point of view. We present a comparison study about the existing geochemical modelling solvers. Finally, a study case using different solvers applied to the same system: evaporites (halite and sylvite) in an aqueous solution containing sodium (Na^+), chlorine (Cl^-) and potassium (K^+). An application of this geological behaviour is the influence of the salt dome in turbidite reservoirs of oil and gas.

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

1 INTRODUCTION

Geochemical modelling corresponds to the design of the reactions that occur in a geological structure through the usage of chemical properties (either thermodynamics and kinetics) to describe it. The need to understand the Earth's interior (both at high-temperature - magma - and low-temperature - aqueous solutions near the surface) motivates the effort in this area of study with the development of models and simulations. The 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 between free ions and aqueous complexes and also saturation indexes for different minerals.

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, such as thermodynamics. Following this idea, a geochemical model is a chemical model developed for geologic systems. Geochemical models incorporate chemical models. A set of mathematical expressions represents these natural processes and handle modelling the system. The thermodynamics and kinetics data used to establish the reactions and mimic nature are directly responsible for the accuracy and precision of the geochemical model.

In this work, we develop a software that through the stoichiometry formulation calculates the chemical equilibrium of a geochemical system using the approach of imposing mass-balance conditions according to the species of the system. This process is known as chemical speciation, and the software was named as *SHPECK*. It 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. Also in this work, we show a thorough analysis of the available existing solutions, and we made clear the uniqueness of our computational approach to the geochemical modelling problem.

Using a high-level and object-oriented programming language, we could implement an

efficient solution that models geochemical speciation. *SHPECK* provides an interactive and intuitive user interface - unique among geochemical speciation software - as well as the support of a built-from-the-ground database structure that handles the management of the whole data used by *SHPECK*. These two contributions are presented as the result of an extensive study about the available software normally in use to perform geochemical speciation simulations. Their flow of information (input and output) are old, complexes and prone to error. It is also important to mention that these software are working with flat file databases. Both of these characteristics are responsible for frequent errors, problems and wrong interpretations. Therefore, the we could clearly notice a lack of a software that would either bring computer science and geochemical modelling side-by-side.

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 atom and charge balance equations or chemical equations (SMITH; MISSEN, 1983). For the sake of simplicity, in this work we will only deal with chemical equilibrium and not with chemical kinetics calculations since the first one requires only the solution of algebraic equation. It is planned to integrate kinetics reactions in the future.

The system of equations drives and represents all the interactions between the components of the simulation. By applying newton-Raphson's method to that set of equations, is possible to find a suitable solution that satisfies the system constraints and the convergence criteria. This method requires the usage of a Jacobian matrix and a residual vector during the algebraic calculations. Geochemical modelling speciation has an important application in processes that occur in turbidite reservoirs. The process of the water coming from a salt dome contains a high concentration of salts as sodium (Na^+), chlorine (Cl^-) and potassium (K^+). Compactation, cementation, dissolution or recrystalization can be observed inside turbidites when this process happens. These processes might change drastically, for example, the porosity of the rock and, therefore, the storage capacity of oil and gas.

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, microorganisms, animals and plants. The subsurface is one of the most complex systems studied by scientists and engineers today. Because of this, geochemical modelling has gained importance and is being accepted as a useful tool to interpret subsurface geochemical processes. Geochemical speciation is based on thermodynamics concepts and the assumption of chemical equilibrium in geochemical reactions. The idea of our own geochemical speciation software has emerged as an application where it would be possible to apply all the physical, chemical aqueous, geochemistry and linear algebra concepts, and develop a useful tool with an intuitive and interactive user interface. The most usual approach found in the area of geochemical modelling is a geochemical expert that develops a solution to solve his particular problem and generates a specific code or algorithm - a solution that most of the times is not very reliable and has no scalability. In this work, the approach is that the computer science expert made the necessary efforts to understand and learn all the complex aspects of a geochemical speciation model and develop a software based on a solid knowledge in computer architecture, algorithms and software engineering.

The main purpose of this work is to develop a geochemical speciation modelling software following a structured computational approach.

The rest of this work is structured as follow. In chapter 2, we present an overview of the basic concepts needed and technical concepts involved in this work. Chapter 3 shows a thoroughly 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 user interface (UI); algorithm validation and complexity; architecture and organization of the software as well as the database; data-flow; and iteration control. In Chapter 5, it is presented a study case with an interesting and relevant scenario; the results that validate *SHPECK* and a broad comparison between solutions previously addressed in this work. Chapter 6 brings the conclusion of this work. Finally, this work contains an Appendix A, which is a presentation and an analysis of a linear algebra library used for the development of *SHPECK* called *Armadillo C++* and appendix B, which is an automated analysis of *SHPECK*'s database.

2 BASIC CONCEPTS FOR GEOCHEMICAL MODELLING

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 modeller, 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 hydrogeochemistry principles: an introduction to thermodynamics; and hydrochemical processes; And to finish, we focus on the geochemical modelling with a special section for it. 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 Thermodynamics

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 close system - no particles 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).

It is also important to mention that any analysis resulting from a water sample must be carefully taken. Any geochemical investigation is useless if the integrity of the water or the solid phase is compromised. Results of interpretation and modelling might be incorrect if the sampling was not done properly. A main objective is to obtain a water sample with the same chemical composition as that of the water in its original environment, for example, an aquifer or a surface water (DEUTSCH, 1997).

2.1.1.1 Thermodynamic Equilibrium Reactions

There are mainly two ways to describe thermodynamic equilibrium reactions: Equilibrium and Kinetic. Both of them formulate a closed system and describe the position of the maximum thermodynamic equilibrium. 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:

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

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 the stoichiometric coefficient negative for reactants and positive for products. Assuming that all the reactions in the system are in equilibrium, the chemical system must also satisfy the mass-action equations:

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

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.

It has been known that the driving force of a chemical reaction is related to the concentration of the constituents that are reacting and the concentrations of the products of the reaction. 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 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 will never be achieved) (FREEZE; CHERRY, 1979). Each of the dissolved species will have one representation of the nonideal 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

$$aA + bB \underset{k_1}{\overset{k_2}{\rightleftharpoons}} dD + eE \qquad (2.3)$$

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

$$-\frac{r_A}{a} \;=\; -\frac{r_B}{b} = \frac{r_D}{d} = \frac{r_E}{e} \tag{2.4}$$

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 follow

$$r_A \;=\; -k_1(A)^{n1}(B)^{n2} + k_2(D)^{m1}(E)^{m2} \tag{2.5}$$

where $n1, n2, m1$ and $m2$ 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 temperatue. The relation between constant and temperature is given by the *Arrhenius equation*, described as

$$k_diss = A * exp(\frac{-E_a}{R * T}) \tag{2.6}$$

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*), also known as *stability constant*, 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 determine the value for at any temperature by a polynomial fitting technique or polynomial regression.

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 modelling requires adjustment to the conditions in the system being modelled. 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 nowadays. They include reaction con-

stants, reaction descriptions, solutes, species, enthalpy values, activity coefficient parameters, etc. During the development of *SHPECK* we selected the *Lawrence Livermore National Laboratory* (*LLNL*) thermodynamic dataset - it contains also the values of a 8th degree polynomial, which allows our software to calculate the equilibrium constant to any temperature and is used often by several geochemical modelling software. Also another source of experimentally measured data is (PALANDRI L. JAMES, 2004).

In geochemical modelling, 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 modelled as an *n-th* degree polynomial. In our case, the polynomial regression is necessary in order to obtain the equilibrium constant for compounds found in the solution system.

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):

$$a_i = \gamma_i * m_i \tag{2.7}$$

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

2.1.1.4 Ionic strength

Mathematically the ionic strength of the solution is calculated according to

$$I = 0.5 \sum M_i z_i^2 \tag{2.8}$$

where *M* is the molar concentration of the specie *i* having a charge *z*.When *I* increases, activity coefficients decrease. In very diluted solutions, activity coefficient is equals to *1.0*, and activity is equal to concentration. The decreasing trend is related to the "cage" of opposite charge particles around ions. There is reversal of the trend in extremely concentrated solutions (brines),

because beyond ionic strenght of about $1 mol/L$ there is an increase of activity coefficients with increasing ionic strength. This is related to decreasing amount of free water because most of water is already bound around dissolved species. For a matter of explanation, we will calculate the ionic strength of a $CaCl_2$ solution (composed by $0.5 mol$ of Ca^{+2} and $1 mol$ Cl^{-1}):

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

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

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

$$I = 1.5 mol/L \tag{2.12}$$

2.1.1.5 Activity Coefficient

There are different methods to calculate γ for ions:

- Debie-Hueckel: They assumed 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

$$log\gamma_i = -Az_i^2\sqrt{I} \tag{2.13}$$

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: Is a variation of Debie-Hueckel that can be used when the ionic strength is relatively high. The equation is as follow

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

- 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.

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

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

18

Important to mention that there are other methods available for calculating activity coefficients, which are not going to be addressed here. Pure solids have activity coefficient equal to one. 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 regarding moderate or high ionic strength. As to dissolution and precipitation, there is clearly a reaction happening during these processes, which means that some reactions are not in equilibrium.

2.1.1.6 Saturation Index

The saturation index (*SI*) indicates the degree of saturation with respect to a given mineral; in other words, it defines if a reaction will be in equilibrium or not. *SI* is expressed as

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

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

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

where [A], [B], [C] and [D] are activitys of each ion. The interpretation of *IAP* is the following:

- IAP > K : The reaction is progressing from right to left, producing more products. In a ground water solution, the water is supersaturated.
- IAP = K : The reaction is in equilibrium, there is no flow neither to the right nor to the left. In a ground water solution, the water and the mineral are in equilibrium.
- IAP < K : The reaction is progressing from left to right, producing more reactants. In a ground water solution, the water is undersaturated.

With the *SI* approach, it is possible to predict the reactive mineralogy of the subsurface from the ground-water data without collecting samples of the solid phase and analyzing the mineralogy. If the *SI* for a mineral is less than zero, the aqueous solution is undersaturated with respect to that mineral - which corresponds to the fact that the mineral will not precipitate and may dissolve in order to reach equilibrium concentrations. If the *SI* is greater than zero,

then the mineral is not reactive and the 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) , it means that the water is saturated with respect to that mineral (ALLEY, 1993). From the mentioned before, it is possible to state the following:

- SI < 0 : Mineral is undersaturated;
- SI = 0 : Mineral is in equilibrium with the solution;
- SI > 0 : Mineral is oversaturated;

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. Concentration in miliequivalents per liter is concentration in milimoles per liter multiplied by charge of an ion.

2.1.2 Hydrochemical processes

2.1.2.1 Acid-Base Reactions

The importance of acid-base reactions is cleary evident when one understands their influence on the pH. The pH is a master variable in charge of controlling chemical systems, and it is described as

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

where $[H^+]$ is the activity of the hydrogen ion. The interpretation of the values is as follows:

- pH < 7 : acid solution;
- pH = 7 : neutral solution;
- pH > 7 : basic solution;

The acid substance has a tendency to loose protons, while a basic substance has tendency to gain protons, and the interation between acids and bases is called acid-base reactions and is described as

$$Acid_1 + Base_2 = Acid_2 + Base_1$$ (2.19)

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 and Speciation

Complexation is when an ion is formed by combining simpler cations, anions and sometimes molecules. This process facilitates the transport of potentially toxic substances and form what is called a complex. Due to the impact of this process in contamination problems, it has acquired a huge importance in practical and commercial fields. A simple example of complexation is the following

$$Mn^{2+} + Cl^- \quad = \quad MnCl^+ \tag{2.20}$$

The calculation of the distribution of metals among complexes (*speciation*) involves the solution of a series of mass-law transport equations. The mass-law equation of the reaction 2.20 is described below

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

Each complex compound has a property called stability constant (β_i). It defines how the total concentration of the components are distributed among the possible compounds (it may include other complexes) in the solution.

2.1.2.3 Oxidation-Reduction Reactions

Groundwater environment's reactions involve the transfer of electrons between its components (gaseous, dissolved or solid constituents). As a result, there are changes in the oxidation states 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 Adsorption and ion exchange

Adsorption systems treat water by adding a substance, such as activated carbon or aluminia, to the water supply. Adsorbents attract contaminants by chemical and physical processes that cause them to *stick* to their surfaces for later disposal. This mechanism is often used to remove contaminants like *arsenic* or *fluoride* (mostly organic contaminants) from water reservoirs. Ion exchange works simillarly but it is focused on inorganic contaminants in a particle-free water. Ion exchange is most often used to remove hardness or nitrate (mostly inorganic soluble molecules).

In this work, adsorption and ion exchange are not going to be addressed. In order to get deeper understanding on this topic, we refer to (FREEZE; CHERRY, 1979)

2.1.3 Geochemical Modelling

The geochemical modelling is the design of the geochemical reactions responsible for the migration of dissolved species. 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 takes into account also kinetic reactions and include the time factor. As kinetic data is measured experimentally, there still is a lack of kinetic data available for many geochemical processes.

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

Inside geochemical equilibrium models we can mention three divisions: speciation models; inverse models (also called mass-balance models); forward models (also called reaction-path models); and reactive transport (coupled) models. Regarding the relation to spatial coordinates, geochemical equilibrium models are considered *batch models* - which are basically closed vessels or reactors.

2.1.3.1 Geochemical Speciation Modelling

Speciation represents modelling based in the equilibrium of a system. A geochemical speciation modelling 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:

$$\text{Na}^+_{total} = \text{NaCl}_{aq} + \text{NaOH} + \text{Na}^+ \tag{2.22}$$

where Na$^+_{\text{total}}$ is total sodium concentration from chemical analysis. Na$^+_{\text{total}}$ is a component (e.g., chemical formula unit used to describe a system) and Na$^+$, NaCl$_{\text{aq}}$ 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.

Saturation index (SI) is used to determine the direction of geochemical processes. When *SI > 0* the mineral precipitates from the water, and when *SI < 0*, the mineral dissolves in contact with the water, if it is present in solid phase. 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).

Common problems solved using speciation programs are:

- There is a sample with high concentration of dissolved sodium and we need to know the distribution of sodium between Na$^+$ and different complexes (for example, $NaCl_{\text{aq}}$ or NaOH) because different forms of sodium have different characteristics;

- There are ground water samples that had been in contact with granitic masses and we want to verify the possibility of precipitation of minerals like *Albite* (a plagioclase feldspar mineral whose formula is *NaAlSi$_3$O$_8$*).

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

2.1.3.2 Other Types Of Geochemical Modelling

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

- Inverse geochemical modelling 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 modelling is (SHARIF, 2007).

- Forward geochemical modelling 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 modelling introduces problems regarding kinetic and adsorption data, which are often missing and frequently limited.

2.2 Summary

- 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"* is something non-trivial. We reinforce the idea that the extension of a software goes way beyond the lines of code and the interface showing up 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: Thermodynamics is one of the substantial basis for the history of physics, chemistry and the science in general as we know it. It is crucial to understand the role that it has in the natural aspects of the world that we live. The comprehension of thermodynamics goes through the analysis, awareness and interrelations of the several factors that compose this complex *maze*. From the wide range of factors, 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, etc.

3 REVIEW OF AVAILABLE GEOCHEMICAL MODELLING SOFTWARE

The importance of geochemical models has increased lately due to the variety of applications, but the first models date back to the 70's as can be observed (WESTALL et al., 1976), (WOLERY, 1979). Since then, these models are used to solve problems as speciation; determination of minerals' saturation indexes; adjustment of equilibrium for minerals; mixing of different waters; calculation of stoichiometric reactions; mixing of solids, fluids and gaseous phases; calculation of equilibrium/kinetic controlled reactions; reactive transport; and mass-law calculations.

The quality of the chemical analysis depend on the methods used, thermodynamic data and theoretical concepts applied. Therefore, it is crucial to verify the results and it is clear that there will be some differences in the results according to the software used. 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 ts 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 groundwater simulation.

The goal of this chapter is to review other programs that perform modelling of aqueous geochemical systems. It is not possible nor the purpose of this work to present all the existing software but to critically review and compare some aspects of them.

In this work, only programs that provide speciation modelling are reviewed. They are the following: *EQ3/6*; *PHREEQC*; *MINTEQ*; and *SOLMINEQ*;

We will present relevant detail about the geochemical point of view and then, we will critically analyze and discuss each software from the computer science point of view. It's worth mentioning that sometimes is not possible to analyse the same aspects of different softwares due to lack of information.

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-Hueckel 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) (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 meaningless. Another point to be taken into consideration and that follows the same idea, 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; qqueous 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 software that is 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 consists of six *blocks*: header; input echo; input recap; iterative calculations; principal results; and the end of *EQ6* run.

A small excerpt of an *EQ6* outpuf file is shown in code 3.1. We can see that in this part of the output file, it is specified 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

28

from the package the user is using, it generates from two to several output files (always in the *ASCII* format). As mentioned, the input file will be 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"

```
iopr(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 informations 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 modelling 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 modelling 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 modelling, 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 will continue reading the input file until the 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 and wrong. So building an input file is usually quite difficult.

30

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 will contain the results of the simulation defined in the input file also divided into *keyword blocks*. Among those, we can mention 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--------------------------

                                  Log      Log      Log    mole V
   Species       Molality   Activity Molality Activity   Gamma   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 modelling software and the lack of interface for *PHREEQC* in the first versions of the software, many efforts were done 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. After this, another effort was done and *PHREEQC* version 2 with a graphical interface was launched, the interface being kept later in version 3. This graphical interface was baptized *PfW* (Phreeqc for Windows) but it has no updates since 2011. It can be seen in 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; This can be seen in 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 will store the results of the simulation, *"database"* specifies which database should be used, and *"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 modelling software will choose default values.

3.2.3 File formats

All 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).
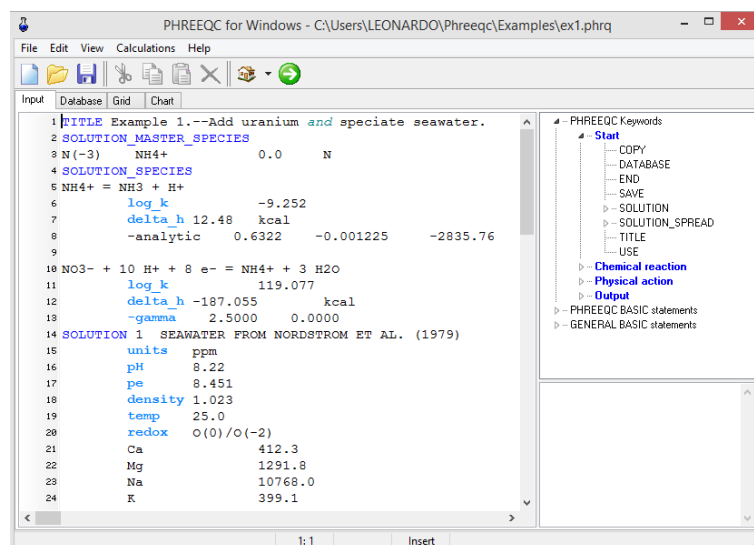
32

Figure 3.1 – User interface of the PHREEQC for Windows (*PfW*)
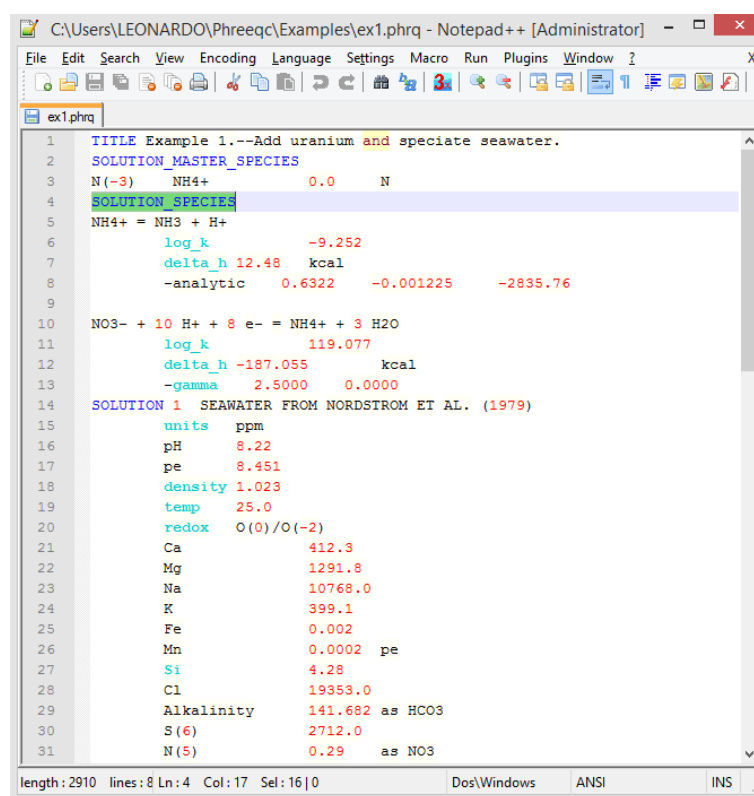
Figure 3.2 – Example of the PHREEQC Notepad++ plugin

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 can is available for download from the USGS website and easily installed. The *UNIX* distribution comes with additional scripts and a makefile and the user should follow some steps to compile and install

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

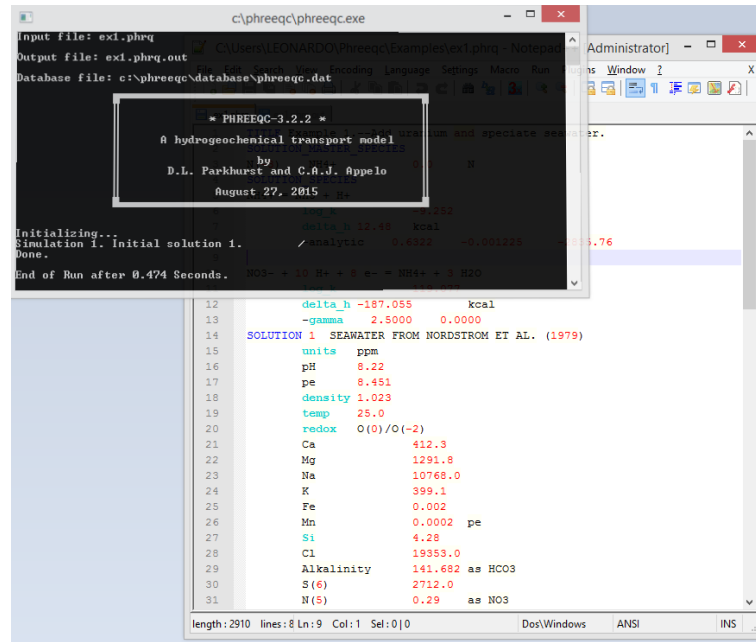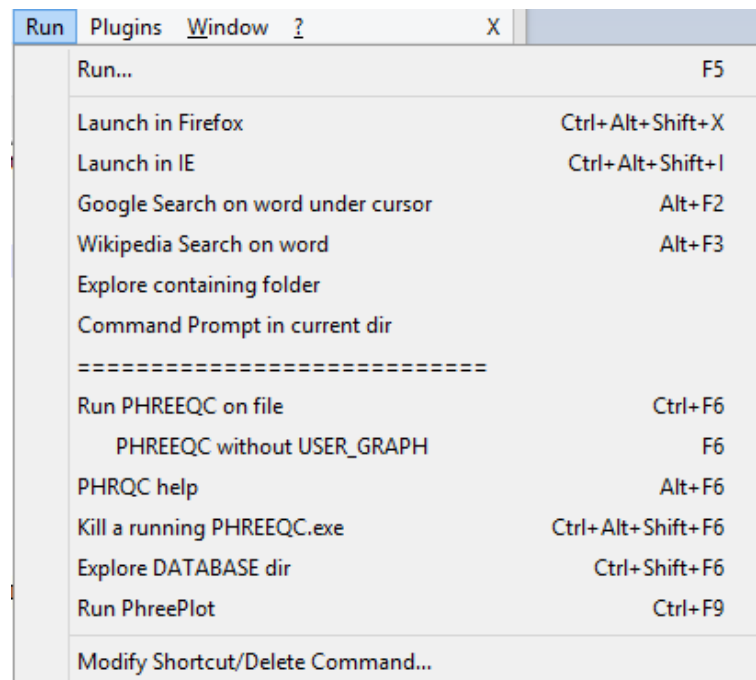Figure 3.4 – PHREEQC software called from the Notepad++ Plugin

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 Battelle Pacific Northwest Laboratory (*PNL*) and continued by the Environmental Protection Agency (*EPA*) to perform the necessary calculations regarding waste, sediments and ground water. *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 another 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 parts compose the input file:

- 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*. Important to mention that if the user does not want to specify all the details for every level, 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 example. The beginning of the file (first and second lines) contains a description of the simulation or input file; the third, fourth and fifth line brings settings configuration: details as temperature, unit chosen, eH, ionic strength, number of iterations, precipitation options. After that, comes the components that take part in the simulation, organized by internal id, concentration details and log of concentration;

Code 3.7 – *MINTEQA2*'s input file

```
LHDAMIANI - STUDY CASE
Comparative study
25.00 MOLAL  0.000  0.00000E+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 - is possible to see 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 specie;

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 potencial 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 ACTVTY    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 and following we will detail *PRODEFA2*'s interaction in an example step-by-step. *PRODEFA2* provides a "walk-through" to generate an input file for *MINTEQA2*.

After opening the software and providing a valid name, it will ask which part of the input file the user wants to create or edit as shown in 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, it is the organization of all the levels and works as a central hub of information. Figure

3.7 displays the necessary information about level 1, in order 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 the four levels do this kind of interactions. Through these interactions, the user has access to all the information in the database and can choose specifically about what is the model.
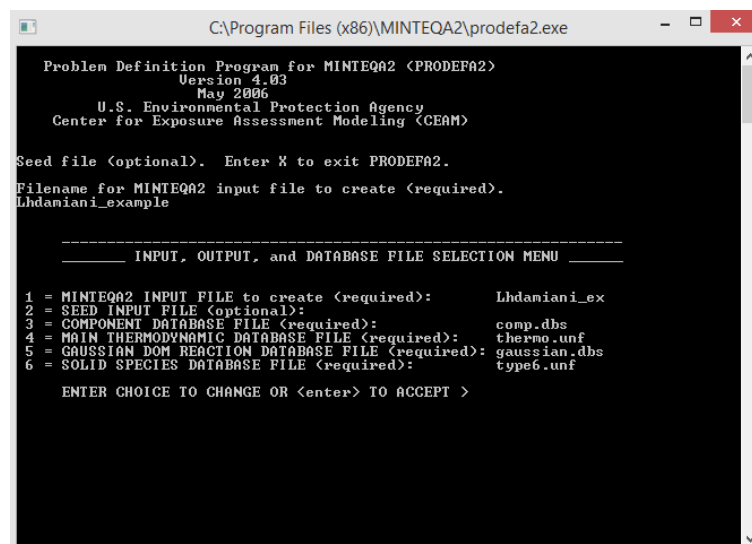

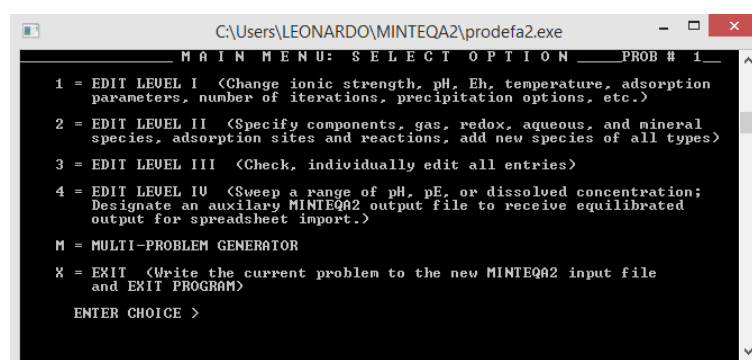Figure 3.5 – *MINTEQA2* initial menu options

Figure 3.6 – *MINTEQA2* main menu

During this work, for sake of illustration, we chose to add a specific aqueous specie 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;

Figure 3.7 – *MINTEQA2* Level 1 informations

3. Choose option 1 (TOTAL DISSOLVED CONCENTRATION), it identifies how we want to entre 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 components id number or quit. We will add Na^+ to the system. So, we type the letter "N" and hit enter button.

5. At this step, all the existing options available are shown listed with an identifier and we are requested to identify which one of these we want to add. 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 specie 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 aquoues specie (Na^+)

There is also a software called *Visual Minteq* that tries to "humanize" *MINTEQ* and was maintained by the KTH Royal Institute of Technology, located in Stockholm, Sweden. *Visual Minteq* latest release date 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 on 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"*;

3.3.3.1 Software Environment and Installation Procedures

It is currently at the version 4.03 (Windows only), interesting to mention that this release date back to May 2006. The latest *UNIX* distribution is version 3.12 (which was also called BETA for UNIX) and date 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 an FORTRAN 77 written geochemical modeling program based on SOLMNEQ (KHARAKA; BARNES, 1973) with improved algorithms that result 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 a vapour phase. *SOLMINEQ.88* also contemplates mass transfer with the effects of dissolution and precipitation of minerals and options to calculate activity coefficient model (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 contains the chemical composition of an aqueous fluid and options for processing these data; and the last Consists of input data required for using them.

The input file is composed by six parts, *SOLINPUT* guides the user throught all of them:

1. Basic Parameters: enter the chemical and physical data for that sample;
2. Flags: controls how the software interprets, process and display 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: make 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 of the computations by *SOLMINEQ.88* and it 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 os on; A table showing the distribution of species in solution; Ratios of a number of cations and anions of importance in geochemical processes; and th e 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 O.OOOOE+00 0.1023E+01 O.OOOOE+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 SO4 H2S
0.4280E+01 0.1935E+05 0.2712E+04 O.OOOOE+00
F PO4 NO3 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 O.OOOOE+00 O.OOOOE+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 accompany *SOLMINEQ.88* and handles the generation of input files and all the interactions is named *SOLINPU* and described in (DEBRAAL; KHARAKA, 1989). All the interactions use the command prompt - the menus with several options are generated and displayed. 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

The input files are regular *ASCII* text files and any regular text editor is good for creating or editing.The database files from *SOLMINEQ.88* have the extension "TBL" and the output files

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 in compiling and linking FORTRAN77 programs. It also comes with the software *SOLINPUT* that makes the user interactions, as explained before, easier.

Interesting to point that along the installation manual, they recommend that curious compiling options are activated, for example *"do not check for array bounds"*, and *"do not check subprogram interfaces"*.

3.5 Discussion

On this section we compare, analyze and discuss some important aspects and issues of each one of the software presented earlier in this work. Regarding geochemical features, table 3.1 brings certain aspects and features of the software.

Regarding the computer science point of view, we analyse and evaluate the software always thinking how good is the software engineering behind it. Taking into account all the aspects described and extensively discussed in chapter 2 is essential.

The points that we discuss and take into consideration are the following:

- The costs: Costs are probably the most 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 of the solution it will be decided.

- Setup and versioning: The installation of software is the act of making it ready for execution. According to the software, a particular installation process is done - which involves copying/generating 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

Table 3.1 – Geochemical comparison between speciation softwares

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

has a different distribution package to each *OS* - or not. Not hard to find software with distribution only to specifics *OS*, meaning that if the user uses other *OS* it is impossible to use that software. Also not hard to find same software with disparate versions according to the *OS* - resulting in divergent features available on the same software defined by the *OS* that is being used.

- Customization and Integration: The software is a standard solution, and its supplier is not 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 communicate with others? Options like "import" and "export" are vital when working with a large amount of data. With the advances in software, you might want to use a different software to analyse the output and to be able to reach deeper insights about the information available. These insights might be unique and lead to incredible breakthroughs.

- Security and Control: Security is one of the main issues one face when considering solutions. Nobody wants to share private data and details with others. 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 process is most likely to have a private required by some users.

- Infrastructure: When choosing a software, the infrastrucutre that it requires need to be carefully analyzed to verify if it matches the disponibility. Does it requires internet access? How many space on disk and memory it uses? Extra costs may occur if this is not

thought earlier.

- Core functionality: This is one of the most important points to be analyzed. How good the software focus on the needs of the user and how good is the value that the software brings to its users by performing the core activity successfully.

- Graphical User Interface (*GUI*) and visualization: Handles the interactions between the users and the electronic devices. 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 think all the possible options and to take all the advantages of the software. The perfect *GUI* takes into consideration all the human behaviour, 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 to question about the issue or discuss what happened? 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.

- 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 consider all the structure and information needed by this software. Along the database schema, the security of this database must be addressed properly - either for consistency and privacy reasons. 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

All the content of the databases are extracted from the *Lawrence Livermore National Laboratory (LLNL)* thermodynamic datasets - which are used in *EQ3/6*, *PHREEQC* and others. The data itself, is contribution from many authors that had measured thermodynamical and kinetic parameters of several minerals over the years. In order to illustrate the structure that the

information is organized in the *LLNL*'s flat file databases, we present some of the content on it:

- Parameters: Many used parameters are stablished 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 by all the existing pure elements. It also brings the mole weight of the element and the abreviation. Example:

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 contains the chemically identical atomic or molecular structural units in a solid array. Example:

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 atoms have their oxidation stated 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 specie has been dissolved in water, this is denoted by writing (aq) after the chemical name. Example:

46

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 contains the naturally occurring substances that can be solid and inorganic representable by a chemical formula and has an ordered atomic structure. Example:

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 contains individual atoms (e.g. noble gases or atomic gases), elemental molecules made from one type of atom (e.g. oxygen) or compound molecules made from a variety of atoms (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. Example:

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 worst and 5 the highest and best possible grade). This "grading system" is done with the intention to obtain a normalization towards differents aspects an interesting output for the comparison and in table 3.2[1] we can in details.

Table 3.2 – Qualitative analysis of the Geochemical Speciation Softwares

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 Modelling. 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 knowledge encourages and pushes the understanding of geochemical modelling further. It used all the computing tools and options that were available until the release date. Since then, computing has clearly evolved, making it an obsolete and hard to use the software. It has a large database with many pieces of information on it, but it is not exactly clear to the user. It is hard to understand what exactly is the software doing; verifying if that is what the users wants is even more difficult. Also important to mention that in *EQ3/6* documentation it recommends to use operating system commands (i.e. *ctrl+c*) to interact with the software - which is highly innapropriate and dangerous to be done from the computer science point of view.

- 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 that *PHREEQC*'s *GUI* is far from what a regular user might want - it is not clear in many aspects, and its usability is far from regular. In the geochemical modelling area, not all the users have intimacy with tasks as compiling and linking computer programs. *PHREEQC* allows anyone with an interest to have a chance to perform a geochemical modelling simulation even though many people have different ways of defining the problem. Database in *PHREEQC* seems to be a problem - it uses a flat file database.

- MINTEQA2: From the geochemical point of view, *MINTEQA2* is the simpler from the

---

[1]Source:Author

four software analyzed in this work. The user interaction can be painful for anyone that does not know how to use the command prompt, besides that, the complexity of the input file makes any way a hard way. Creating the 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 date back to 2006, it is hard to motivate and try to give *MINTEQA2* any consideration nowadays.

- SOLMINEQ.88: An interesting software that was also a pioneer and layed the ground for many others to improve and progress the knowledge in the geochemical modelling field. *SOLMINEQ* used the computing tools that were accessible when it was developed - nowadays there is no tendency that someone will start using it and learning its complex input and output. There are less costly options easily attainable, and that produce analogous results. One must bear in mind that using a software that recommends to use compiling options such as *"do not check for array bounds"* is something weird - in other words: risky.

4 *SHPECK* - GEOCHEMICAL SPECIATION MODEL

SHPECK is a geochemical speciation model equation-based simulator, there is a governing theory that can guide the construction of mathematical models based on a set of equations. The "equation-based" term refers to simulations based on the kinds of global equations we associate with physical theories - many of these theories present in the foundation concepts of *SHPECK* are originally from the *Phase Rule*. Phase rule allows us to predict the number of stable phases that may exist in equilibrium for a particular system and is describe fully in (GARRELS; CHRIST, 1965) and is the base principle used in chemical speciation.

As a computer simulation software, *SHPECK* represents the entire process. This process includes choosing a model; finding a way of implementing that model in a form that a computer can run; calculating the output of the algorithm; visualizing and studying the resultant data.

This chapter will guide through all the aspects of the development of *SHPECK* - a geochemical speciation modelling software.

4.1 Specification

SHPECK is a geochemical speciation modelling software responsible for calculating the distribution of dissolved species between free ions and aqueous complexes and also saturation indexes for different minerals.

4.2 Architecture

The architecture of a software is responsible for assuring that all the internal and external stakeholders' concerns are preserved, addressed and satisfied. To not develop something that will not need restructuration and code refactoring later, it is necessary to see the big picture of the whole *SHPECK*'s sofware. This drives and is present during the entire development, starting with the technical decisions until the satisfaction and the value brought to the user.

Figure 4.1 brings the software architecture of *SHPECK*, which is modelled 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-Huckels' formula.

- View: It is some 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).
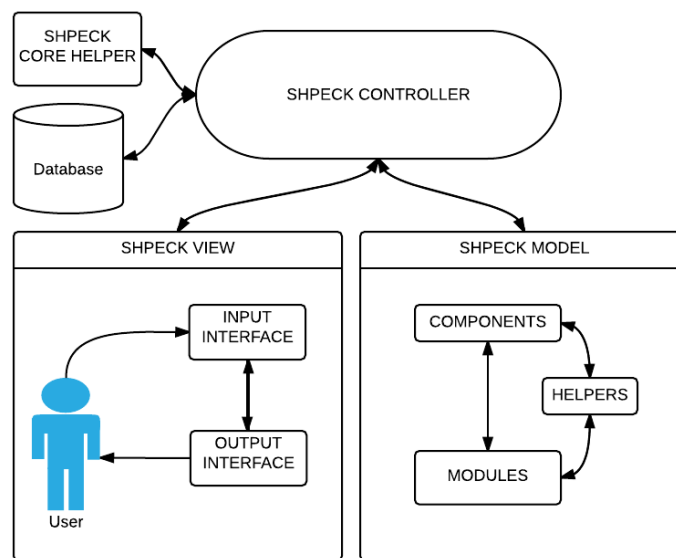

Figure 4.1 – Architecture of the *SHPECK* software

4.2.1 Technical Specification

SHPECK is a software developed using C++, which is a general-purpose programming language. C++ has a bias towards system programming that supports efficient low-level computation, data abstraction, object-oriented programming, and generic programming (DALE, 2004) (STROUSTRUP, 1997). It provides powerful and flexible mechanisms for abstraction; that is, language constructs that allow 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 lots of inbuilt library classes it provides a comprehensive range of functionality which also turns the debugging and testing way more productive and effective.

- *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). With a syntax similar to Matlab and an up-to-date support with upgrades and new releases almost every month.

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 *UNIX-like* systems.

4.3 Governing equations

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

Aside with the mass-action equations we must have other constraints to solve the equilibrium state of the system. In *SHPECK*, we use the concentration of the species (other possible constraints are the solute activity, the number of moles, total volume, etc.) to deal with the equilibrium state of the system. Therefore, we have the following configuration:

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

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:

$$ln(K_j) = ln(\prod_{i=1}^{N} a_i^{v_{ij}}) \qquad (j = 1, ..., M) \qquad (4.2)$$

4.4 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 methodology that computes simultaneously and find the value of the unknowns.

The numerical method applied 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 a 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*'s method to solve a nonlinear system of equations which results in finding the roots of continuously differentiable equations.

The *Newton-Raphson*'s method is applied in order to achieve the best approximation possible to the solution that is being sought. It since it is an iterative algorithm where each step consists in minimizing the first-order approximation of the solution. *Newton-Raphson*'s method is a modification of *Newton*'s method for finding a minimum of a function. 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.

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

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

$$R = \begin{pmatrix} F(x_1) \\ \vdots \\ F(x_m) \end{pmatrix} \qquad (4.4)$$

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:

$$[J]^{-1}_{iteration} * \alpha[U]_{iteration+1} = [R]_{iteration} \tag{4.5}$$

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 matrix. With this, is possible to state the $[U]_{iteration+1}$ value with:

$$[U]_{iteration+1} = [U]_{iteration} + \alpha[U]_{iteration+1} \tag{4.6}$$

The initial guess of the solutions is an approximation that the user provides to the *Newton-Raphson*'s 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 than the number of iterations necessary to obtain the solution will be small. If the guess is something completely nonsense, and far from the real solution, more iterations are going to be needed to find the correct solution.

The *Jacobian Matrix* receives the equations and generates a model that does it automatically The *Jacobian Matrix* is the matrix of all first-order partial derivatives of the equation and it is defined as

$$J_{mn} = \frac{\partial y^m}{\partial x^n} \tag{4.7}$$

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.

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

$$J = \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_n} \end{pmatrix} \tag{4.8}$$

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

Important to realize that the main complication with using a the *Newton-Raphson*'s method to solve a system of nonlinear equations is defining all the functions included in the

54

Jacobian Matrix. As the number of equations and unknowns increases (n), so does the number of elements in the *Jacobian* (n^2).

4.5 Algorithm

SHPECK's algorithm takes as its input a specification of the system's state. For example, the values for the concentrations of the species in the solution, the temperature of the system, the method to calculate activity coefficient, etc. It then calculates the system's state by executing all the processes in the necessary order that the model was conceptualized. Figure 4.2 presents the high-level algorithm in details; 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", understanding how the governing equations controls the numerical methods is a fundamental piece. Newton's Method Solver internal algorithm is presented in figure 4.3.
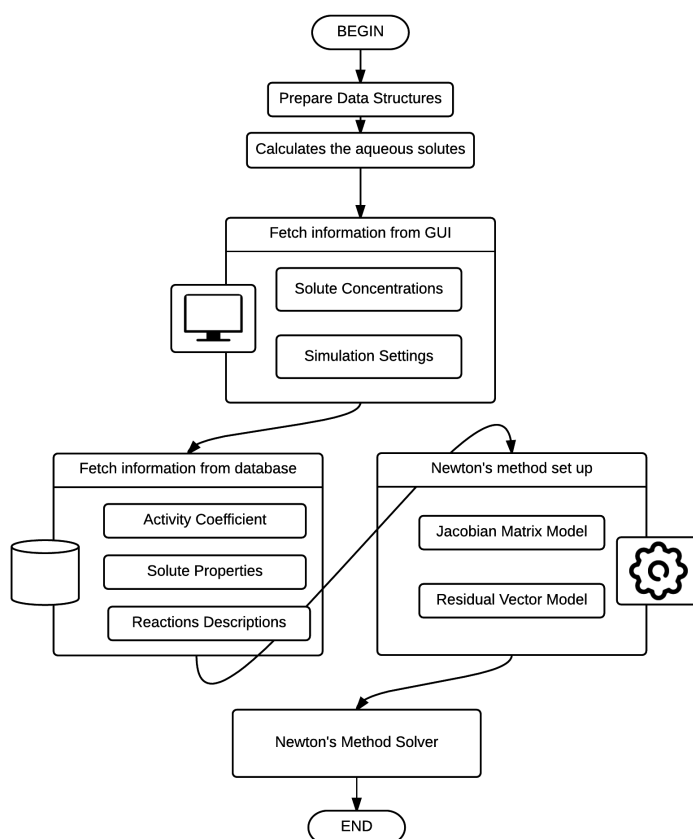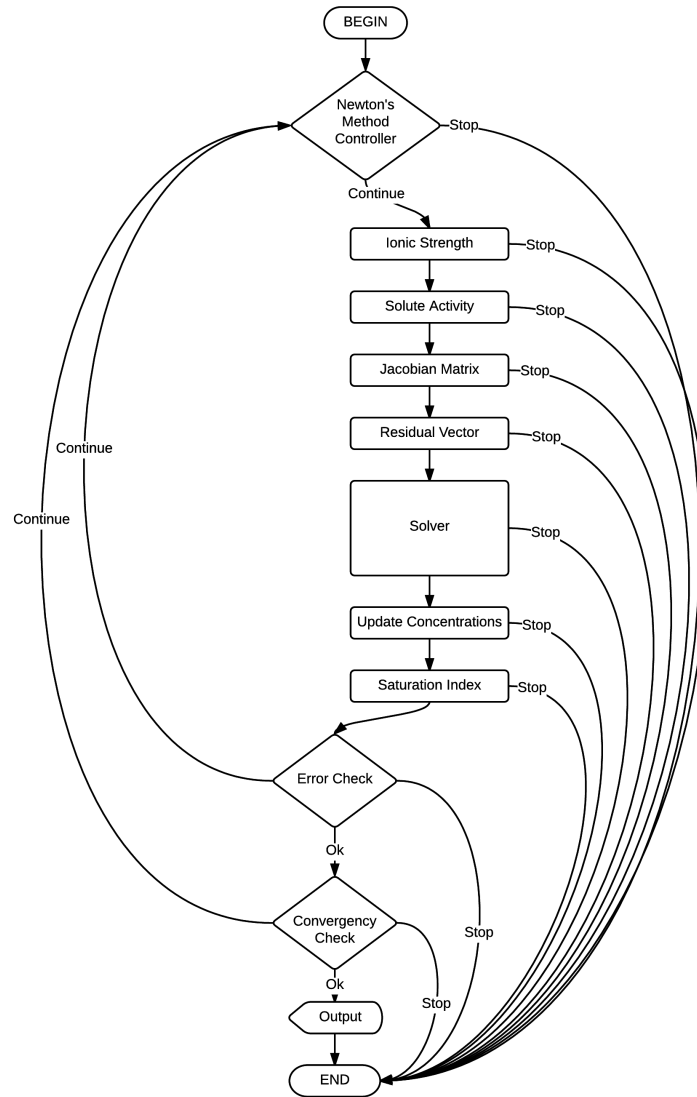

Figure 4.2 – High-level algorithm of *SHPECK*

Figure 4.3 – Algorithm of the Newton's Method Solver

4.5.1 Complexity of the algorithm

The *Newton-Raphson*'s method has a complexity $O(n^3)$ per iteration with a quadratic convergence.

4.6 Graphical User Interface

Due to the enormous amount of options connected to the nature of geochemical modeling, the development of a GUI is a challenge itself. It is necessary to develop a software that is intuitive and user-friendly and has an accurate and efficient approach to all of the options available in nature and the system. The development of the GUI introduces several stages of processes related to it:

- Planning : This moment is where the developer needs to think a way to contemplate all the options from the core of the program and make them easily, intuitive and friendly to the final user. The planning is an essential process and must be correct. All the content, features, details of the software need to be strictly defined and organized in order not to break anything afterwards nor forget it (which will result in a not used software). A slight path divides something really useful from something not useful at all - assumptions and the challenge of the developer to think as a user is sometimes tough because many of the developers are not the final user for its applications. Is important also to find out what others have done or are doing - during the study of others software their GUI were analyzed carefully.

- Building : All the planning done for needs to be implemented and the main features of the software are important to be well defined at this point. To build the tool *Qt* - already presented above - was chosen due to its many advantages as easy to customize; no coding; templates; simple drag and drop *GUI* builder; secure and reliable;

- Ensuring Usability : The challenge in the field of Human-Computer Interaction (HCI) investigates way people use computers is huge. Techniques and tools are used to find some relevant standards, and some measurements are available. Just to name a few, tests with real users; evaluation by experts; gathering user feedback; Usage logging;
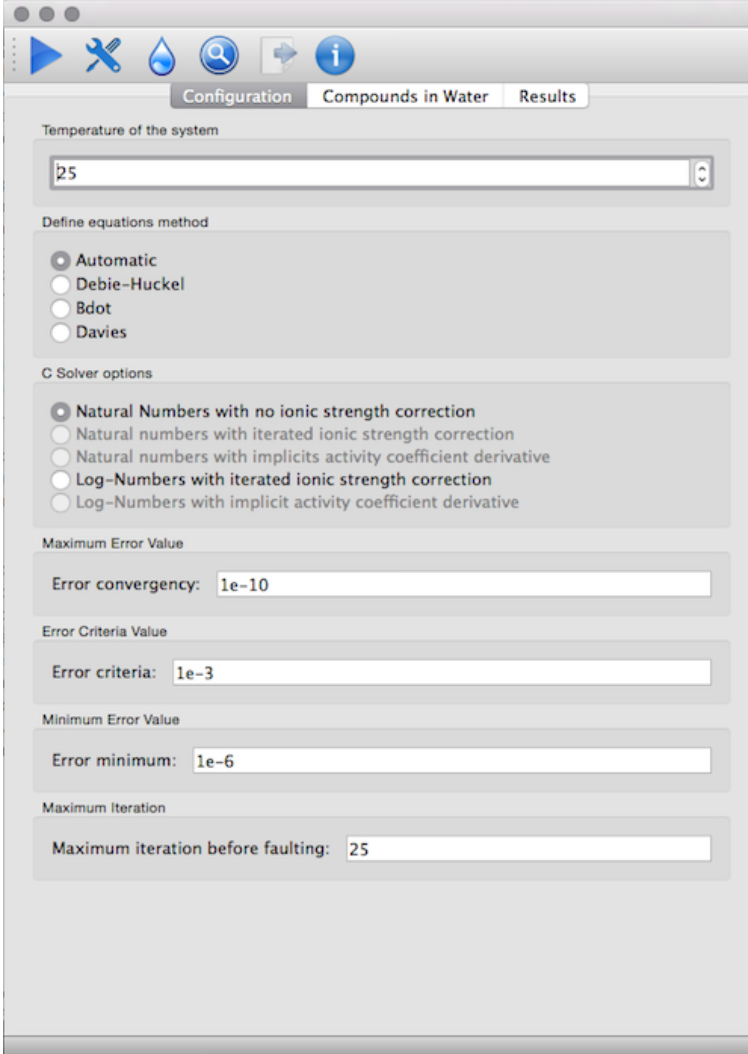
4.6.1 *SHPECK*'s GUI

Among geochemical modelling software, the GUI was either not or poorly implemented. We see*SHPECK* as geochemical speciation modelling software with a potential to broaden its influence. The *GUI* enables the user to focus on essential duties and not struggle with the mental power wasted on things that do not matter when modelling a geochemical environment. We think of the GUI with a bright idea that the less skill it requires, the better it is. *SHPECK*'s GUI works based on tabs, each tab is responsible for viewing an specific point of the software. We present them in details bellow:

- Configurations tab: Allows users to view and manipulate basic system settings and controls such as temperature, activity coefficient calculation method, the number of iterations, solver options, maximal error and convergence criteria. It is presented in 4.4.

- Compounds in the water tab: Allows users to create and edit the composition of the water that will be used in the geochemical speciation model. It has all the species available

from the database and its concentration. The species in the tab that have concentration different than zero compose the water. This tab is presented in 4.5.

• Results tab: Allows the user to see relevant input information and the outputs of the geochemical speciation model such temperature, ionic strength, pH of the solution, final concentration for the species, saturation indexes, etc. It is presented in 4.6.
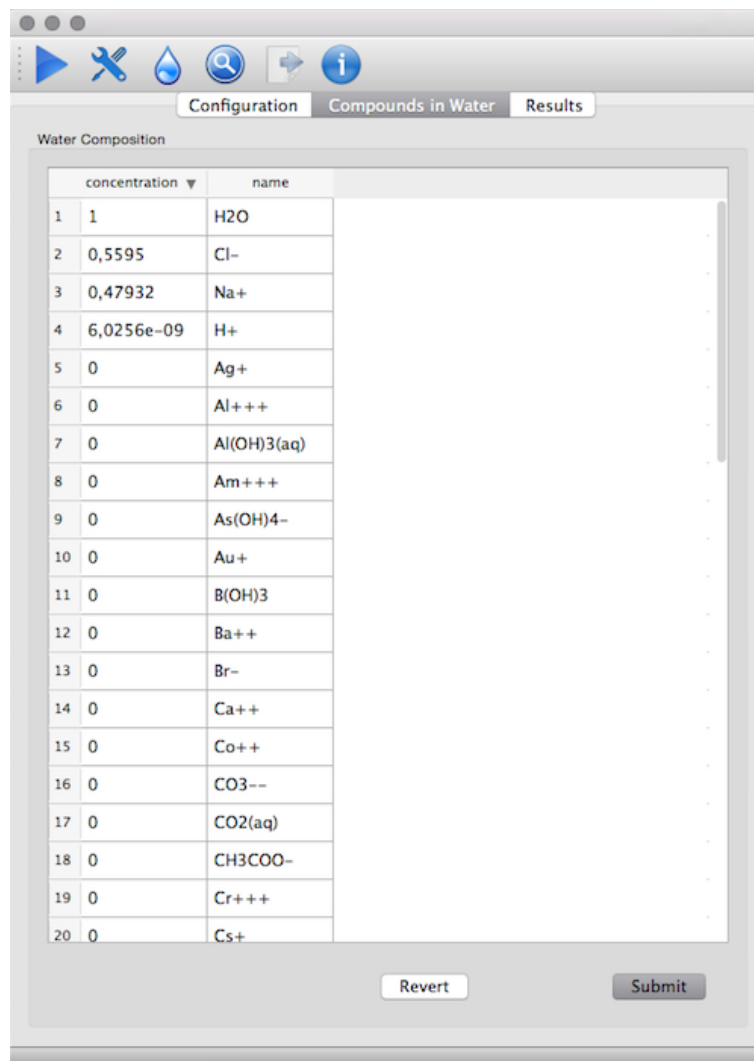

Figure 4.4 – Configuration tab

Figure 4.5 – Water composition tab

4.7 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 all the data that will flow through and works as the *seed* provider. Chapter 3 makes clear that using a flat file database is something harmful to the software. Potential issues of using flat file databases are duplication of the information; non-unique records; hard to update; inherently inefficient; harder to change data format; poor at complex queries; no security; In *SHPECK*, we use a relational database, which is a model of a database that prevents all the problems faced in flat file database previously discussed and enable to take full advantage of the information on it. Relational databases work based on a *query*, which is an information request from the database. Besides that, there are three important advantages:

- A relational database has the advantage of not necessarily be in the memory of the com-
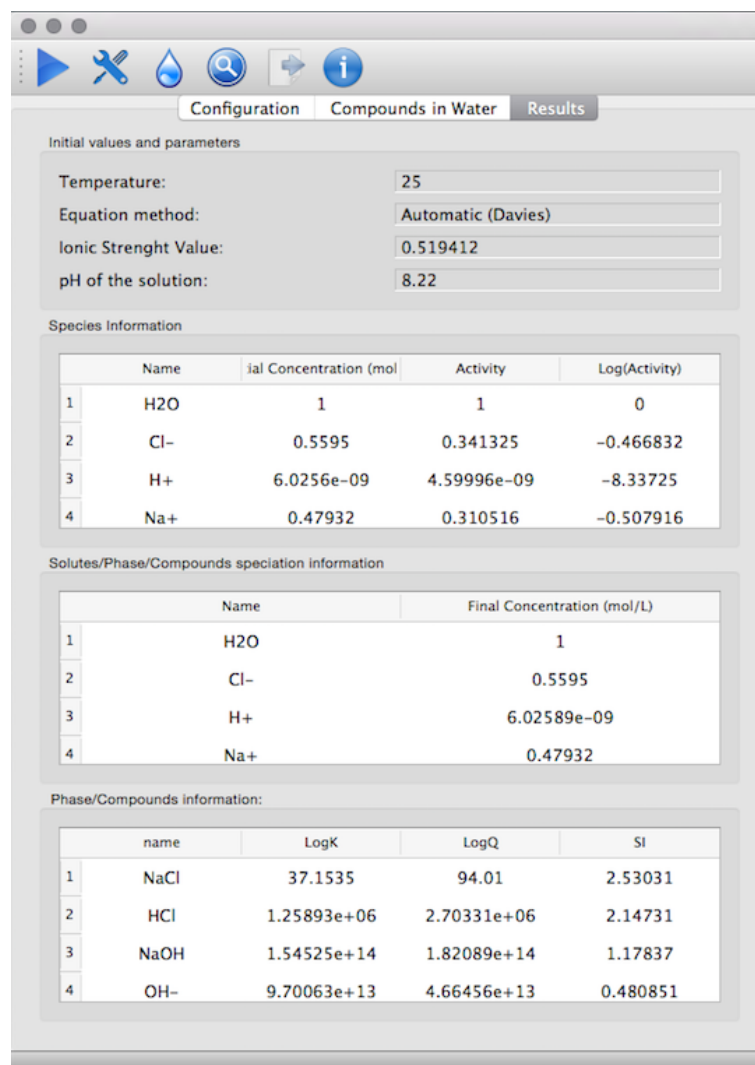
Figure 4.6 – Results tab

puter - often these databases are larger than the computer's memory. What happens is that the software fetches the information on-the-go.

- Complex queries enable the versatility and efficiency when fetching relevant information from the database. Instead of scanning the whole flat file, the query can pre-define the categories of information - even from multiple tables - which will be sought. Also important to mention, a query can be composed and concatenated on runtime execution - meaning that *SHPECK* only fetches from the database information relevant to the specific simulation that is being executed (species, compounds, reactions, etc.). Complex queries and concatenation of queries result in a faster and more efficient use of the available resources.

Instead of being something to worry about, the database is one of the most critical and significant infrastructure parts of our software. The database architecture was defined after studying the algorithm and determining what would be the structure and the information needed - it is presented in details in figure 4.7. We developed a relational database that embrace all the

important data already in use from other software but with a newer approach and structured in tables - naturally some changes in the structure were necessary in order to compact, organize, structure and make it easier to use.
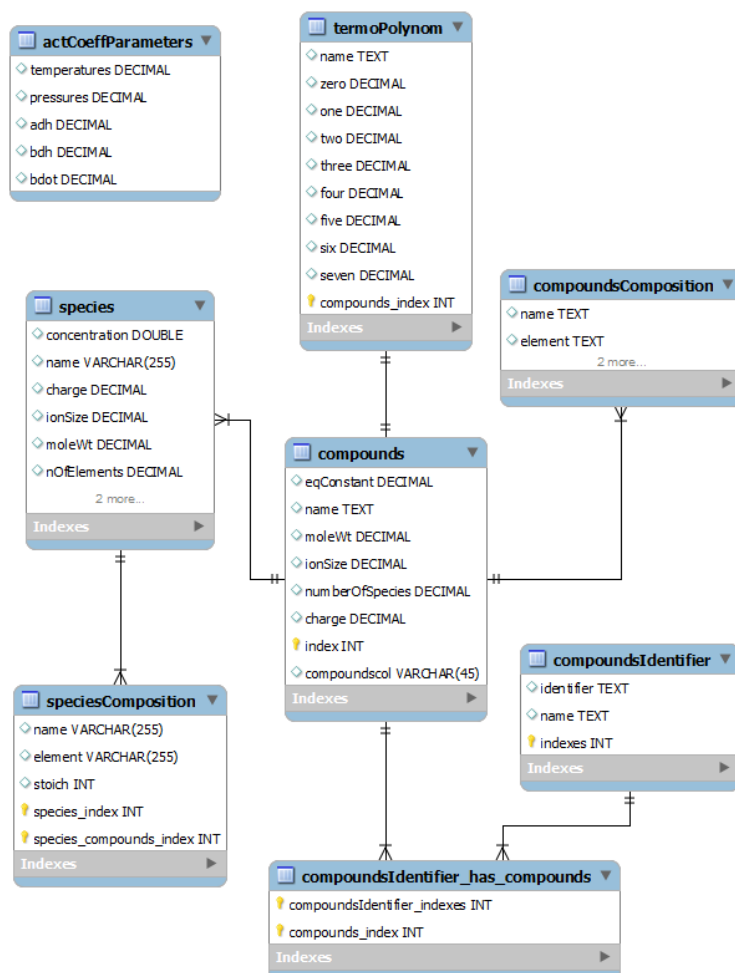
Figure 4.7 – ER Diagram of the database

4.7.1 Database Technologies

We use the SQLite database (HIPP, 2015), which is a software library that implements a self-contained, transactional *SQL* database engine, open source and currently is the most widely deployed *SQL* database engine in the world. Some of its advantages are explained bellow:

- Zero-Configuration: *SQLite* does not need to be installed before it is used, there is no setup procedure;

- Serverless: The process that wants to acess 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);

4.7.2 *LLNL* thermodynamic dataset parser

Once we defined what would be the structure and the technology of our database, we need to find the information to populate it - the information that would feed *SHPECK*. We create a parser for the flat file database that extracted the information and arranged it following our structure. The parser was carefully generated in order not to misunderstand any structure nor miss a delimiter - this task was extremely time-consuming and proportionally important for the success of *SHPECK*.

4.8 Summary

- Architecture: *SHPECK* follows the architectural pattern called *MVC*; It's benefits are the complete separation of responsibilities and concerns among the parts of the software; there is no mixing of codes between them. Another advantage that worth mentioning is the flexibility for the software to grow and develop itself. Figure 4.1 displays the *MVC* pattern existing in *SHPECK*.

- Governing Equations: *SHPECK* is a geochemical speciation modelling software that drives the behaviour 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:

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

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's* method to solve the non-linear system of equations. The concept of the method is described in equation 4.3 and reinforced here:

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

Where F is function's result for the applied x, J^{-1} is the inverse of the Jacobian matrix, and R is the residual matrix. 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 fully the potential *SHPECK* has to offer. It is intuitive and user-friendly to allow the user to focus on essential duties that are imperative to model a geochemical environment. We use the popular approach of *tab* panels, which are separated according the purpose of it: configurations and settings, water composition and results visualization.

- Database: A geochemical modelling software is severely dependent on the information inside it. *SHPECK* structure all of its data in a *SQLite* relational database, unique among the commercial software available. *SHPECK*'s database is composed by the information about elements, species, compounds, reactions and thermodynamic constants from the

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

5 VERIFICATION AND VALIDATION

During the development of this work, we have always been in contact with geochemists to present them the partial results of our geochemical speciation modelling software. We collected feedbacks regarding how to achieve a better software, as well as checking if our solution is fulfilling its purpose with consistent results.

As final evaluation of this work, we selected an application relevant to petroleum systems. Many physical-chemical reactions happens during the generation, migration and storage of oil. *Diagenesis* is the definition of the several processes that are involved and it is driven by multiple factors as temperature, pressure, mineral composition, water composition, activity of the solutes, pH, etc. The *diagenesis* is responsible for compaction and precipitation of minerals (TUCKER, 2001) and therefore, porosity, solubility and permeability of these reservoirs. The study of diagenesis is important because it allows to understand the geologic history of rocks, specially sedimentary rocks. In sedimentary rocks, the deposition of sediments are compacted in different layers and cemented by minerals that precipitate from reactions in a chemically very active environment. The *diagenesis* reactions happens because the components are always trying to reach equilibrium, and therefore, they tend to interact with each others (BURLEY S. D.; WAUGH, 1985). Using geochemical modelling softwares is a powerful tool to understand the diagenetic processes and the natural conditions that occur in this natural environment. The goal is to numerically model this environment and analyse the results of the diagenetic reactions with a petrographic analysis of the modeled reservoirs.

5.1 Case Study

We reproduce the diagenetic reactions observed in Snorre Field reservoir sandstones, Norwegian North Sea. Morad (MORAD. S., 1990) describes the texture, origin, chemistry of the sandstones reservoirs in terms of the water composition and temperature. With the help and the results from this study, we can verify *SHPECK*'s results from two perspectives:

- Experimentally: Morad examined two hundred representative core samples with standard optical microscopy and petrographically described the reservoirs. This analysis confirm that the our geochemical model suits the natural environment's reality;

- Computationally: The descriptions presented in (MORAD. S., 1990) of the diagenetic reactions that take place in the Snorre Field allows us to generate a comparative study. We

model and compare the same environment using *SHPECK*, *PHREEQC* and *MINTEQA2*. The water composition is detailed in (NORDSTROM, 1979);

5.1.1 Experimentally 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 (HELGE-SON; KIRKHAM, 1974b), (HELGESON; KIRKHAM, 1974a), (HELGESON; KIRKHAM, 1976), (WALTRER; HELGESON, 1977), (HELGESON H. C.; BIRD, 1978) and (HELGESON H. C.; FLOWERS, 1981).

After studying the composition of formations waters, the goal is to demonstrate what characteristics do the components of that water adopt. This is possible by generating and analysing the graph, in this case, expressed as a function of temperature and the log activity ratio of Potassium to Sodium ions - this can be verified in details in (AAGAARD, 1990).

Figure 5.1 brings the graph using the results from *SHPECK*. It is consistent with (AA-GAARD, 1990) and analysed in (MORAD. S., 1990). The pattern detected is as the temperature rises and greater burial depths, the potassium activity gets higher over sodium's.
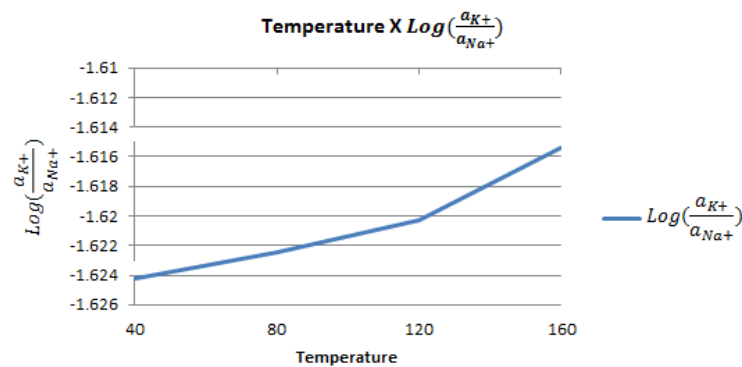

Figure 5.1 – Log activity ratio of Potassium to Sodium ions using the results from *SHPECK*

66

Table 5.1 – Chemical composition of the solution in the sea water at 25^oC

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

5.1.2 Computationally comparative

By modelling the same environment using three different software, we achieve a relevant comparison among the numerical methods and algorithms.

The environment modelled is described in (MORAD. S., 1990) and the chemical composition of the water in (NORDSTROM, 1979). The latter provides the chemical composition of the initial solution for the sea water. It can be verified the major components in table 5.1 - in mM/L. The temperature on this comparative study interpolate from 25^oC to 160^oC. In MINTEQA2, due restrictions to its thermodynamics equilibrium database, the maximal temperature available is 100^oC.

The diagenetic processes modelled by the water-rock interactions are responsible for defining how of ions present the solution are going to behave. Figures 5.2, 5.3, 5.4, 5.5, 5.6 and 5.7 present the most representative ions in the solution. It is possible to see that the behavior of *SHPECK* accompany both *PHREEQC* and *MINTEQA2* in most of the cases, specially in temperatures under 100^oC. We see a more distant behavior among *SHPECK* and *PHREEQC* in temperatures higher than 100^oC but never completely opposites. This is explained due to temperatures where the equilibrium constant *K* is not completely defined. This is a known issue from *LLNL* thermodynamic dataset: sometimes equilibrium constants that are not measured are misterpreeted. *SHPECK* uses the following approach: if there is one unknown equilibrium constant needed, it uses the closest known value. Unfortunately we do not have access to all the informations about each software; therefore, we can not describe exactly what other softwares do with this issue.
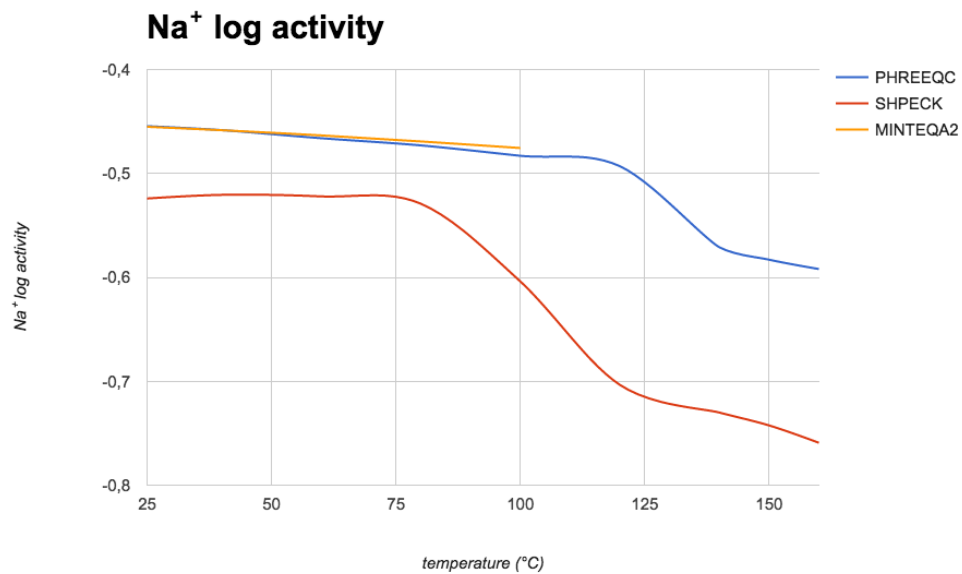
Figure 5.2 – Na⁺ log activity comparative study

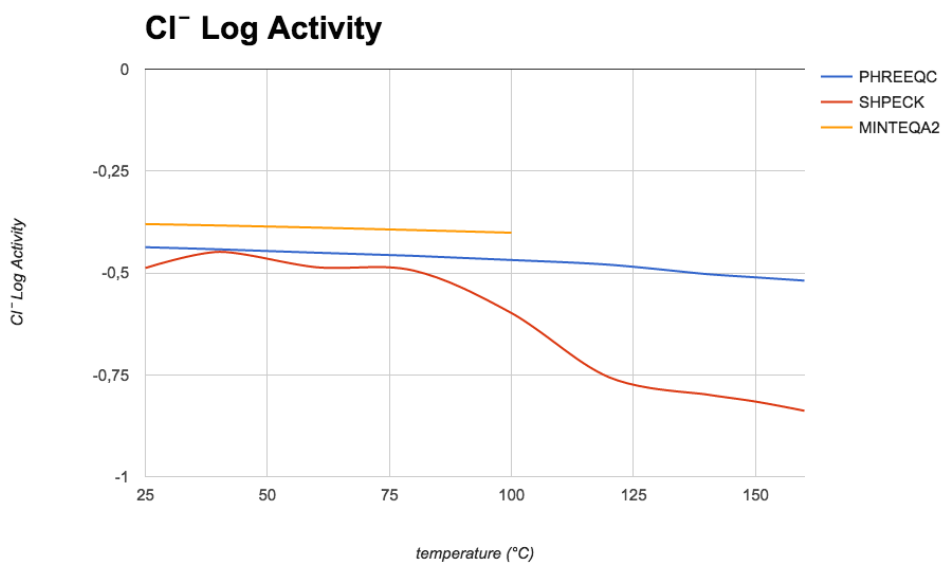Figure 5.3 – Cl⁻ log activity comparative study

5.2 Database Evaluation

The database is the source of all the data and thermodynamic information inside a geochemical modelling software. The goal of this section is to make clear the difference and - more important - the benefits of *SHPECK*'s relational database if compared to others. In geochemical modelling software the common approach is to use flat file databases; therefore, we're going to describe a time, space and expressiveness analysis between the *LLNL* thermodynamic dataset
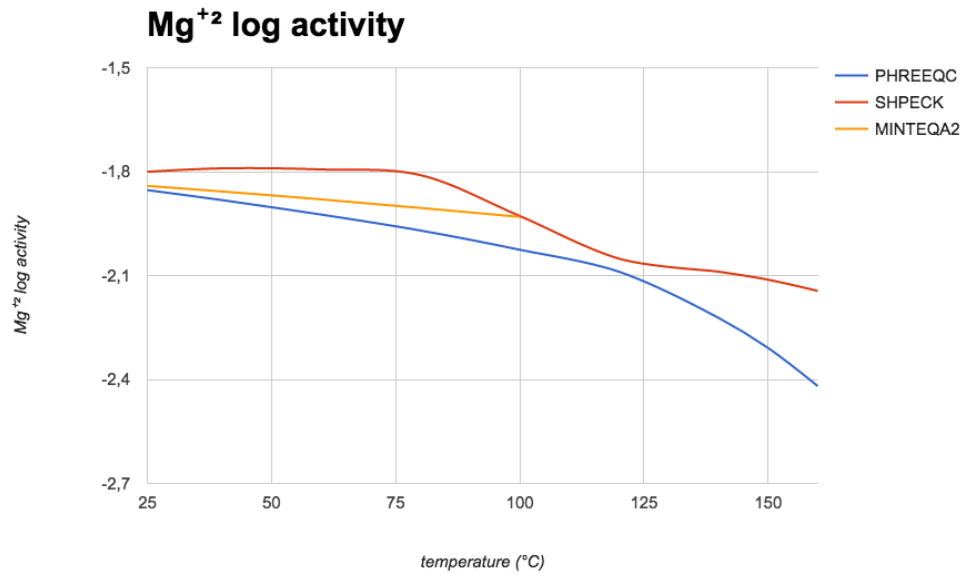
68

Figure 5.4 – Mg^{+2} log activity comparative study
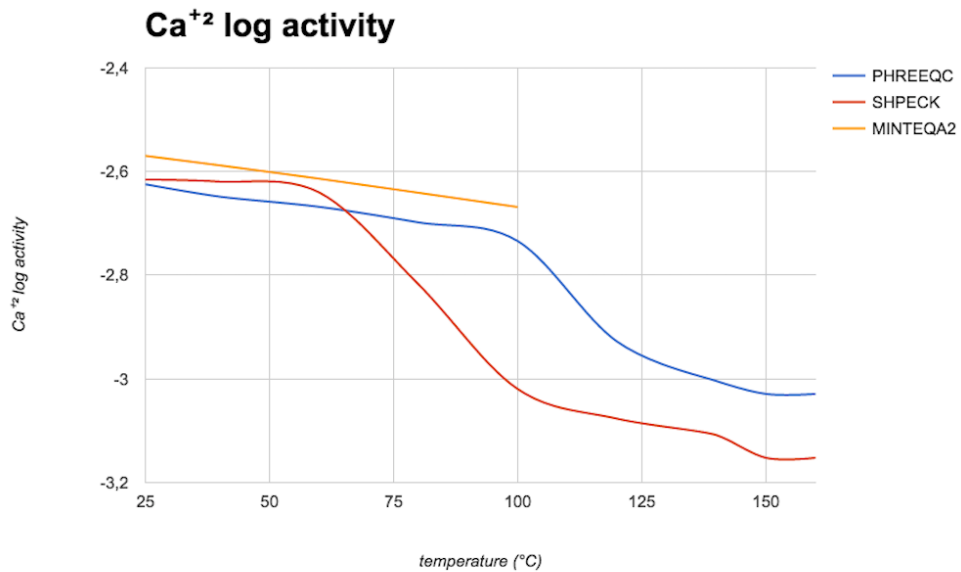
Figure 5.5 – Ca^{+2} log activity comparative study

and *SHPECK*'s databse.

Most of the information inside a geochemical database is related to each other (i.e. a mineral is described by a reaction (may be more than one), a reaction is composed by species, a specie is composed by elements). *SQLite* databases are naturally a structure where the data can be related to each other, this affects drastically the performance and robustness of the application. On *SQLite* databases, the data can be accessed using powerfull *SQL* queries that reduce the complexity and increase the speed on information retrieval.
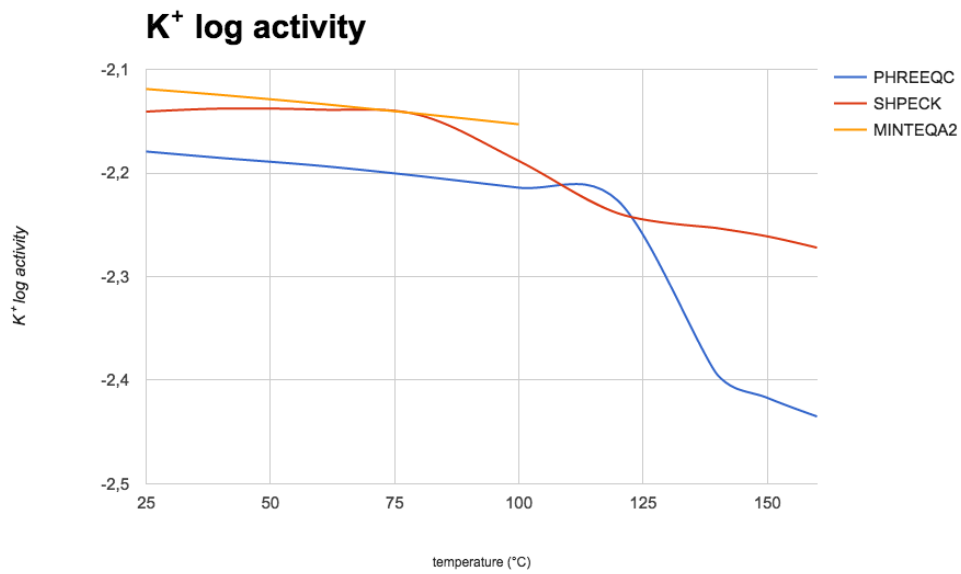
Figure 5.6 – K$^+$ log activity comparative study
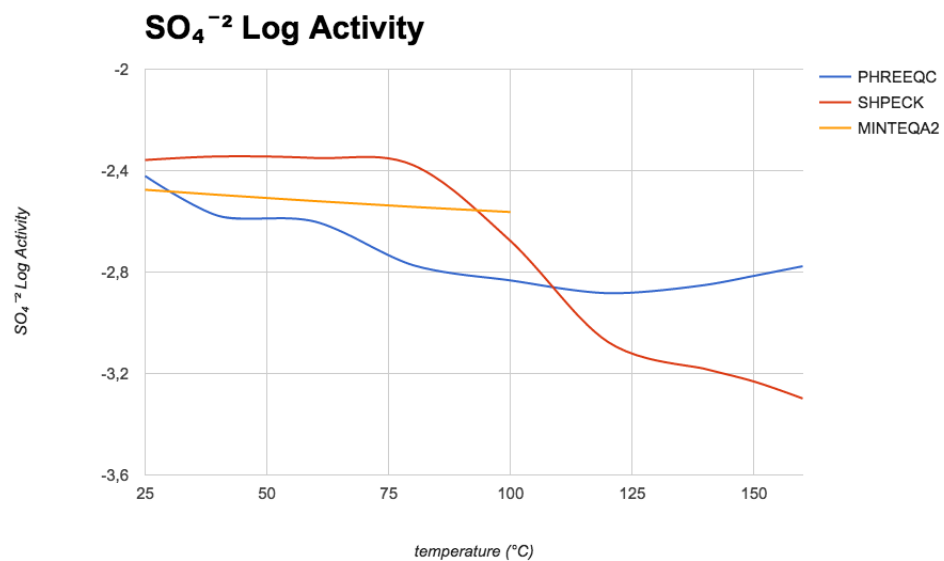
Figure 5.7 – SO$_4^{-2}$ log activity comparative study

With flat file database, we have regular continuous access to the information that it contains. In relation databases we use the *SQL* language. The studies in expressive power of query languages is one of the important fields in database studies - it studies the limitation and, on the other hand, the power of *SQL*. Due to this works scope, we will not address basic knowledges in *SQL* queries.

5.2.1 Time analysis

The response time is considered the sum of the processing time and the time waiting for the availabilit of the resource. It is fundamental for the performance of a geochemical modelling software to have fast access to the information since it is a bottleneck for the whole system. It is necessary to understand that until the software has received all the information requested from the database it will be doing nothing, completely stopped and waiting. This waste of CPU usage, if scaled to multiple simulations and long processing, is definitely something that can not be ignored. In order to analyze the response time with a geochemical analysis point of view, we discuss not only the access time but also what that information means.

When fetching any information from a thermodynamic dataset, it is important to take into consideration that no information actually matters if considered alone; there will always be supporting information to fetch aside - which means more database searches and consuming time. For example, when fetching a reaction (as expressed in equation 2.1), the basic informations are the compounds that take part on this reaction and the related stoichiometric values. Behind this action, the database must also provide informations about the compounds itself (i.e. charge, ion size, mole weigth, elements in that specie, formula, mole volume, thermodynamic equilibrium constant coefficients, etc).

Figure 5.8 indicate the time elapsed (in seconds) that it costs to retrieve all the information related to a chemical reaction from the database. In this example, we simulated from 20 to 580 randomly selected reactions in the database.
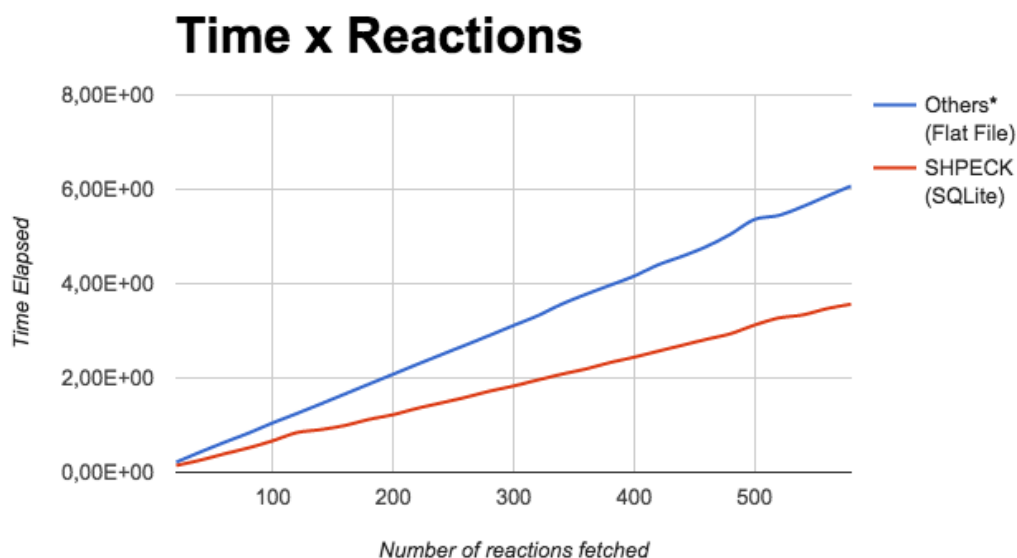

Figure 5.8 – Time elapsed in seconds X Reactions Accessed

It is possible to see clearly in 5.8 that the usage of *SQLite* improves considerably if compared to flat file databases. That is one of the reasons that flat file databases are considered deprecated nowadays.

5.3 Summary

- Diagenesis: This term refers to the physical-chemical reactions responsible for the compaction and precipitation of minerals during the generation, migration and storage of oil. Understand how these reactions actually occur is fundamental to be able to predict porosity, solubility and permeability of reservoirs and using geochemical modelling softwares is one effective and accessible way to learn and predict this complex environment.

- Study case: We reproduce the diagenetic reactions observed in Snorre Field reservoir sandstones, Norwegian North Sea. The environment modelled is described in (MORAD. S., 1990) and the chemical composition of the water in (NORDSTROM, 1979).

- Comparative study: We perform a comparative study with *SHPECK* side-by-side with others available geochemical speciation softwares. The results prove that *SHPECK* reach good results with high accuracy. The discrepancies are minimal and are expected since each software has its own mathematical and computational treatment to the set of equations, parameters, activity coefficient calculation methods and thermodynamic dataset interpretation.

- Database evaluation: The lack of a relational database in the geochemical modelling softwares makes clear that none of the existing options were really developed with a computer science emphasis. *SHPECK* uses a *SQLite* database specially developed to support a geochemical speciation modelling software. This software design option enables our software to achieve a higher performance, handle complex queries and fetch relevant information easier and faster. A comparative graph highlights the advantages in time elapsed by the number of reactions requested as an exercise to detect this.

6 CONCLUSION

This work provides details of *SHPECK* that has the purpose of increasing the quality of geochemical speciation modelling software. *SHPECK* is intended to raise the quality of geochemical speciation modelling software by including an interactive and intuitive user interface as well as a relational database as nowadays software require. It is also an important goal to implement a dynamic and useful tool which is ready to be introduced into any geochemical situation or operate as an academic didactic tool for students to help the understanding of geochemical speciation modelling software. Besides, it must be also mentioned here that *SHPECK* accepts any general combination of elements, species and reactions. *SHPECK* allows the user to create different environments, simulations, and, therefore, fully control any aspect and configuration of the model.

SHPECK has the purpose of raise and optimize in every possible way the concept of geochemical modelling, always taking into account the computer science's priorities, theories and power when applied to multidisciplinary domains. As would be expected, the first part of this work was to understand the area: geochemical modelling and geochemistry concepts. After that, we could study, use and analyze the available options to perform geochemical simulations. At this moment, the foundations for creating *SHPECK* were already solid. We could, therefore, clearly see that there was a colossal need for a solution that would bring either computer science and geochemical simulations side-by-side into the geochemical modelling *'world'*.

What has been noticed from all existing software studied is that they were created by geochemists to solve their daily challenge and questions. Our approach is different: computer science specialists do it with the support of geologists and geochemists.

The first challenge was to design the database that would fit the information of the geochemical speciation modelling software. This information was parsed from a well-stablished source and organised in a *SQLite* relational database that provides fast access to information and powerful queries. Once *SHPECK*'s database was ready to boost it with the data, the speciation technique itself was now on the light spot. Speciation is the calculation of the distribution of dissolved species between free ions and aqueous complexes and saturation indexes for different minerals. *SHPECK* is a geochemical speciation modelling software that calculates the speciation of a solution based on a set of mass-balance equations, which are solved iteratively using the Newton-Raphson's method. Once the implementation of both database and mathematical treatment was effectively finished, the new objective, was to implement a spontaneous, interactive and intuitive interface. *SHPECK*'s *GUI* allows interaction seamlessly and smoothly - this

way the user can focus on what matter: modelling the geochemical environment.

After the software is adequately concluded, the new target, as important as its implementation, is to test, analyze, compare and measure the quality and accuracy of the results provided. Possibly at this stage of the project any problem could appear that was undistinguished until this point of the work.

Fortunately, the results are positive when a series of tests hit both database, algorithm and mathematical accuracy. *SHPECK*'s database has improved approximately 40% in the average time elapsed to fetch the information if compared to the regular flat file database used by others. *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 if with the same inputs. Slight differences are found when simulating with temperatures higher than 100°C - which are acceptable and totally understandable taking into account 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).

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.

BURLEY S. D., K. J. D.; WAUGH, B. **Sedimentology: Recent developments and applied aspects**. [S.l.]: Geological Society Special Publication, 1985. 574 p.

DALE, N. B. **Programming and problem solving with C++**. [S.l.]: Jones Bartlett Publishers, 2004.

DEBRAAL, J. D.; KHARAKA, Y. K. Solinpu: A computer code to create and modify input files for the geochemical program solmineq.88. **U.S. Geological Survery Books and Open-File Reports**, 1989.

DEUTSCH, J. W. **Groundwater Geochemistry: fundamentals and applications to contamination**. [S.l.]: CRC Press, 1997.

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 ofdetrital 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 aquoues 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.

HELGESON, H. C.; KIRKHAM, D. H. Theoretical prediction of the thermodyanmic behavior of aqueous electrolytes at high pressures and temperatures: Ii. deby-hiickel parameters for activity coefficients and relative partial molal properties. **Am. Jour. Science**, 1974.

76

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 inifinite 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 Survery 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.

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 Survery Books and Open-File Reports**, 1990.

NORDSTROM, D. K. e. a. A comparison of computerized chemical models for equilibrium calculations in aqueous systems. **Americam 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 Modellig - 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.

TUCKER, M. E. **Sedimentary Petrology**. [S.l.: s.n.], 2001.

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.