NUMERICAL OPTIMISATION AND MACHINE LEARNING ADVANCEMENT IN SOLUTE GEOTHERMOMETRY

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN (Dr. rer. nat.)

von der KIT Fakultät für

Bauingenieur-, Geo- und Umweltwissenschaften

des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Lars Helge Yström

aus Aalen

Tag der mündlichen Prüfung:

13. November 2023

Referent: Prof. Dr. Thomas Kohl

Korreferent: Prof. Dr.-Ing. Michael Kersten

Karlsruhe 2023

WENN NICHT MEHR ZAHLEN UND FIGUREN

Wenn nicht mehr Zahlen und Figuren
Sind Schlüssel aller Kreaturen
Wenn die, so singen oder küssen,
Mehr als die Tiefgelehrten wissen,
Wenn sich die Welt ins freye Leben
Und in die Welt wird zurück begeben,
Wenn dann sich wieder Licht und Schatten
Zu ächter Klarheit werden gatten,
Und man in Mährchen und Gedichten
Erkennt die ewgen Weltgeschichten,
Dann fliegt vor Einem geheimen Wort
Das ganze verkehrte Wesen fort.

- Novalis, 1800

ABSTRACT

Geothermal energy is a sustainable and renewable energy source. It can be used to generate baseload electricity and to provide heating and cooling. The amount of energy available in the subsurface is related to the temperature of the reservoir, its permeability, and the presence of a heat transfer fluid. Reservoir temperature is therefore a critical parameter in the development of geothermal power plants. Solute geothermometry is a low-cost prospecting tool for estimating reservoir temperatures without the need for direct access to the source. To estimate the reservoir temperature of a hydrothermal fluid, its geochemical composition must be analysed. Due to temperature-dependent dissolution and precipitation reactions in the reservoir, the mineral assemblage of the reservoir rock and the aqueous solution are in equilibrium. This state of chemical equilibrium in the hydrothermal fluid contains information about the reservoir temperature. However, the chemical equilibrium cannot be preserved during the ascent to the surface. Therefore, the reconstruction of undisturbed chemical conditions is an essential part of solute geothermometry.
In this thesis, two approaches for solute geothermometers are developed, tested, improved, and validated on a global scale. In Chapter 3 - 5, MulT_predict is developed (Appendix B). It is an optimised solute multicomponent geothermometer based on the coupling of MATLAB with IPhreeqc. Through the integrated optimisation process, sensitive parameters can be back-calculated to reconstruct the undisturbed chemical conditions of the reservoir. This optimisation process is introduced in Chapter 3, further improved in Chapter 4, and tested on a global scale in Chapter 5. In addition, a universally valid mineral set and its individual refinement are introduced in Chapter 5 to further improve the applicability and accuracy of reservoir temperature predictions. Eventually, MulT_predict is a fully integrated, comprehensive multicomponent geothermometer.

The second solute geothermometer is based on a machine learning approach investigated in Chapter 6. Based on the results of a deep learning algorithm for reservoir temperature estimation, a solute artificial neural network geothermometer called AnnRG is developed (Chapter 7, Appendix C). A unique dataset is created, based on 208 data pairs of geochemical fluid parameters and in-situ temperature measurements. The processing of such heterogeneous data trains AnnRG complex thermodynamic water-rock interactions, resulting in accurate reservoir temperature estimates. Eventually, AnnRG represents a new generation of solute geothermometers, opening up a new vein of geochemistry.

ZUSAMMENFASSUNG

Geothermie ist eine nachhaltige und erneuerbare Energiequelle. Sie kann zur Erzeugung von grundlastfähigem Strom sowie zum Heizen und Kühlen genutzt werden. Die im Untergrund zur Verfügung stehende Energiemenge ist abhängig von der Reservoirtemperatur, der Durchlässigkeit des Gesteins und dem Vorhandensein eine Fluids. Die Reservoirtemperatur ist daher ein entscheidender Faktor bei der Entwicklung von Geothermiekraftwerken.
Die Löslichkeitsgeothermometrie ist eine kostengünstige Methode um Reservoirtemperaturen abzuschätzen zu können. Dabei ist bei dieser Prospektionmethode, kein direkter Zugang zum Reservoir erforderlich. Um die Reservoirtemperatur abschätzen zu können, muss die geochemische Zusammensetzung eines geothermalen Fluides analysiert werden. Aufgrund von temperaturabhängigen Lösungs- und Ausfällungsreaktionen in der Lagerstätte, befinden sich die Mineralien des Gesteins und die heiße, wässrige Lösung im chemischen Gleichgewicht. Somit enthält das chemischen Gleichgewicht des Fluides, Informationen über die Temperatur des Untergrunds. Während des Aufstiegs des Wassers an die Oberflächen, verändert sich das chemische Gleichgewicht des geothermalen Fluides. Daher spielt die Rekonstruktion des ungestörten Fluidchemismus eine wesentliche Rolle in der Geothermometrie.
In dieser Arbeit werden zwei Löslichkeitsgeothermometer entwickelt, getestet, verbessert und auf globaler Ebene validiert. In den Kapiteln 3 - 5 wird die Entwicklung von MulT_predict beschrieben (Anhang B). Dabei handelt es sich um ein optimiertes Multikomponentengeothermometer, welches auf der Verknüpfung von MATLAB und IPhreeqc basiert. Mithilfe eines integrierten Optimierungsprozess können sensitive Parameter zurückgerechnet werden, um die ungestörten chemischen Bedingungen innerhalb des Reservoirs zu rekonstruieren. Dieser Optimierungsprozess wird in Kapitel 3 vorgestellt, und im Laufe des 4. Kapitels verbessert sowie in Kapitel 5 weltweit getestet. Darüber hinaus wird in Kapitel 5 ein allgemeingültiges Mineralset und eine Methodik zu dessen individueller Anpassung an unbekannte Reservoire vorgestellt. Diese Entwicklungen verbessern die Anwendbarkeit und Genauigkeit der Reservoirtemperaturabschätzungen im weiteren Maße. Zusammenfassend handelt es sich bei MulT_predict um ein optimiertes Multikomponentengeothermometer.

Das zweite Löslichkeitsgeothermometer basiert auf Maschinellem Lernen, welches zur Reservoirtemperaturabschätzung genutzt wird. Basierend auf den Ergebnissen eines Deep-Learning-Algorithmus, welcher in Kapitel 6 untersucht wurde, wird eine neues Löslichkeitsgeothemometer namens AnnRG entwickelt (Kapitel 7, Anhang C). Hierzu wird ein einzigartiger Datensatz erstellt, der auf 208

Datenpaaren, bestehend aus geochemischen Fluidparametern und in-situ Temperaturmessungen, basiert. Durch das Trainieren an diesem heterogenen Datensatz, werden die komplexen thermodynamische Wasser-Gestein-Wechselwirkungen von AnnRG erlernt, die zur Reservoirtemperaturabschätzung benötigt werden. Zusammenfassend stellt AnnRG eine neue Generation von Löslichkeitsgeothermometern dar, die einen neues Anwendungsfeld in der Geochemie eröffnen.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ABBREVIATIONS

Adam adaptive moment estimation

ANN artificial neural network

AnnRG Artificial neural network Regression Geothermometer

COM component object model

CSV comma-separated values

DNN deep neural network

EGS enhanced geothermal system

GWO-MLP grey wolf optimiser multi-layer perceptron

IAP ion activity product

IQR interquartile range

IPhreeqc C++ class for PH REdox EQuilibrium

LLNL Lawrence Livermore National Laboratory

MATLAB MATrix LABoratory

MAPE mean absolute percentage error

MEAN arithmetic mean

ML machine learning

MLP multilayer perceptron

MSE mean square error

NaN not a number

R^2 coefficient of determination

ReLu rectified linear unit

RTEst Reservoir Temperature Estimator

RMED median

RMSE root-mean-square error

RMSprop root-mean-square propagation

SDEV standard deviation

SGD stochastic gradient descent

SHAP shapley additive explanation

SI	saturation index
TDS	total dissolved solids g L^{-1}
URG	Upper Rhine Graben

LATIN VARIABLES

a	stoichiometric coefficient of reagent
a_i	activity of a species i
$\mathring{a}_i$	effective hydrated ion size of the ith ion m
A	reagent of the reaction
A	Debye-Hückel parameter, considering temperature, density of the solvent, and the dielectric constant of the solvent (L mol^{-1})$^{0.5}$
A_{1-6}	analytical parameter to calculated $\log(K)$
A_{ij}	matix A consisting of saturation indices over a temperature range i of a corresponding mineral phase j
b	stoichiometric coefficient of reagent
B	reagent of the reaction
B	Debye-Hückel parameter, considering temperature, density of water, and the dielectric constant of water (L m^{-2} mol^{-1})$^{0.5}$
$\dot{B}$	empirical, ion-specific deviation function in the extended Debye-Hückel equation
B_{ij}	matix B consisting of evaluated saturation indices by signum function over a temperature range i of a corresponding mineral phase j
c	stoichiometric coefficient of product
C	product of the reaction
C	matix C consisting of the zero crossing of the saturation indices of corresponding mineral phases
C	number of components
$\Delta C^{\circ}_{P,r}$	standard heat capacity of the reaction J K^{-1}
d	stoichiometric coefficient of product
D	product of the reaction
D	array D consisting of the equilibrium temperatures
E	n-dimensional cell array E consisting of the equilibrium temperatures with corresponding mineral phases
E°	standard cell potential V
E_h	redox potential V

$E(r)$	electric field V m^{-1}
f_i	fugacity of pure gas or gas mixtures
f_i°	standard state fugacity of pure gas or gas mixtures
F	n-dimensional cell array F consisting of the equilibrium temperatures with corresponding mineral phases reduced by outliers
F	number of degrees of freedom
F	Faraday constant C mol^{-1}
G	Gibbs free energy J
G_i°	molal Gibbs free energy of a species i J
ΔG_r	Gibbs free energy of the reaction J
ΔG_r°	standard Gibbs free energy of the reaction J
H	enthalpy J
ΔH_r°	standard enthalpy of the reaction J
I	stoichometric ionic strength mol L^{-1}
k_B	Boltzmann constant J K^{-1}
K	equilibrium constant
m	parameter of regression
m_i	molality of the ith component
n	sample size
N_A	Avogadro constant mol^{-1}
p	pressure Pa
P	number of phases
pe	potential of free electrons in solution
q	point charge C
r	radius m
R	molar gas constant J K^{-1} mol^{-1}
s	sphere size
S	entropy J K^{-1}
ΔS_r°	standard entropy of the reaction J K^{-1}
T	thermodynamic temperature K
U	internal energy J
V	volume m^3
ΔV_r°	standard volume of the reaction m^3
x_i	molar fraction of the ith component

X	matrix consisting of geochemical input parameters for the ANN geothermometer
y	measured in-situ reservoir temperature °C
$\hat{y}$	predicted reservoir temperature °C
Y	vector consisting of in-situ reservoir temperature data for the ANN geothermometer
z_i	charge number of the ith species C
z_{e^-}	number of transferred electrons
Z	matrix consisting of aggregated geochemical input parameters and in-situ reservoir temperature data for the ANN geothermometer

GREEK VARIABLES

γ_i	activity coefficient of the ith solute species
ϵ	electric permittivity of the electrolyte F m^{-1}
ϵ_0	vacuum electric permittivity F m^{-1}
λ_D	Debye length m
λ_i	activity coefficient of the ith solide or liquid species
ν_i	stoichiometric coefficient of the ith component

1

INTRODUCTION

Geothermal energy is a sustainable and renewable source of energy. Depending on the temperature of the reservoir and the amount of thermal energy available, geothermal energy can be used to generate electricity and provide heating and cooling. Unlike most renewable energy sources, it can provide a reliable supply of energy independent of fuel supply, meteorological effects, and daily and seasonal cycles. In addition to its availability, geothermal energy is also adaptable to the demand for electricity or heat. With the development of geothermal energy, baseload electricity can be provided, reducing the volatility within the mains. A decentralised application has a low environmental impact and requires little surface space. However, there is a potential risk of induced seismicity, depending on the geology and stress state in the subsurface. In addition, high investment costs for exploration and construction as well as potential failure of the project limit the development of geothermal power plants.

1.1 SOLUTE GEOTHERMOMETRY

Flow rate and reservoir temperature are crucial parameters for the development of geothermal power plants. Solute geothermometry is a low-cost tool for estimating the temperature of a geothermal aquifer. By analysing the chemical composition of hydrothermal fluid, the temperature-dependent equilibrium reaction between the mineral assemblage of the reservoir and the aqueous solution gives the reservoir temperature. This relationship was first described by Morey et al. (1962) for quartz solubility over a temperature range of 25 °C to 300 °C. Ellis and Mahon (1964) carried out several hot water/rock interaction experiments and identified the measured silica concentration within the solubility range of amorphous silica and quartz. Finally, Fournier and Rowe (1966) first estimated subsurface temperatures from the silica content of water from hot springs and wet steam wells. Similarly, White (1965) discovered that the K/Na ratio of water served as a crude geothermometer for water temperature. Since then, new conventional solute geothermometers have been continuously developed. In 1983, Michard and Roekens (1983) modelled the chemical composition of the alkaline hot water at Thuès-les-Bains by plotting the saturation indices of several mineral phases against temperature. Reed and Spycher (1984) took up this idea and developed the first multicomponent geothermometer, calculating saturation indices of the reservoir mineral assemblage over a temperature range. Over the years, this method has been further developed and optimised (Palmer,

2014; Sonnenthal et al., 2013; Spycher and Finsterle, 2016). The latest development in solute geothermometry is introduced by Haklidir and Haklidir (2020). The authors used a machine learning approach to train an artificial neural network with element concentrations and system parameters to estimate reservoir temperatures.

1.2 MOTIVATION

This thesis pursues an advancement in solute geothermometry as a low-cost geothermal exploration tool. Conventional geothermometers are based on simple cation ratios (White, 1965) or element concentrations (Fournier and Rowe, 1966), which are used to estimate reservoir temperatures. Since the individual conventional geothermometers are established for specific geological and chemical conditions, the transfer to unknown sites leads to uncertainties and the need for adjustments (Fournier and Potter, 1979). Combining the predictions of several conventional geothermometers provides even higher uncertainties (Nitschke et al., 2018). Multicomponent geothermometry is therefore a statistically more robust application (Peiffer et al., 2014). These reservoir temperature estimations are based on multiple mineral phases representing the reservoir mineral assemblage (Reed and Spycher, 1984). However, the composition of the reservoir rock might be unknown, increasing the geochemical knowledge required a priori for the application of multicomponent geothermometers (Palmer, 2014).
Overall, the perturbation of the temperature-dependent equilibrium reaction between the reservoir mineral assemblage and the aqueous solution leads to high uncertainties in reservoir temperature estimation (Fournier et al., 1974). In particular, secondary processes such as phase precipitation, degassing, boiling, dilution or mixing with shallow and less mineralised waters, as well as reequilibration with surrounding rocks, perturb the chemical equilibrium as the fluid rises to the surface. To overcome this problem, extensive chemical analysis of element concentrations in the aqueous solution, as well as gas concentration and composition, and enthalpy measurement must be performed to reconstruct the geothermal fluid at reservoir conditions (Arnórsson et al., 1982). Such sophisticated sampling methods are costly and time-consuming, but can also introduce further uncertainties.
The focus of this thesis is to address these issues and to extend, facilitate and improve the applicability of solute geothermometry. To extend the applicability of an ideal solute geothermometer, the amount of input data needs to be reduced to a basic geochemical analysis consisting of only major cations, anions and system parameters. Any additional trace element and gas analyses should be avoided. In addition, the ideal geothermometer should be applicable to a wide range of geothermal settings. It should be easy to use to reduce the barriers of geochemical and geological preknowledge. In the case of multicomponent

geothermometry, a variety of aqueous species and mineral phases should be available to calculate estimations for a wide range of temperatures. Partially missing or perturbed data should also be suitable for temperature estimation. In addition, the uncertainties of the predictions should be reduced.

1.3 STRUCTURE OF THE THESIS

This thesis is structured in two main parts. First, the development of a solute multicomponent geothermometer MulT_predict (Chapters 3 - 5). This involves the development, automation, and optimisation of the multicomponent geothermometer (Chapter 3). In Chapter 4, the working principle of the optimisation processes is changed to one multidimensional numerical optimisation. Finally, the applicability of MulT_predict is tested on a global scale, while the integrated optimisation process is benchmarked internally (Chapter 5).
Second, the development of a solute artificial neural network geothermometer AnnRG (Chapters 6 & 7). In Chapter 6, the machine learning approach of deep learning is used to estimate reservoir temperatures, and the results are compared with the numerically optimised output of MulT_predict. Eventually, the solute artificial neural network geothermometer AnnRG is developed (Chapter 7).

1.3.1 *Development of a multicomponent geothermometer with integrated optimisation processes (Chapter 3)*

The first study shows the development of a multicomponent geothermometer based on the coupling of MATLAB (MATLAB, 2019) to IPhreeqc (Charlton and Parkhurst, 2011; Parkhurst, Appelo et al., 2013). An automated MATLAB script is implemented to statistically evaluate saturation curves of mineral phases calculated by IPhreeqc. Multiple sensitivity analyses on system parameters and element concentrations are performed to identify the most sensitive parameters. As a result, three individual optimisation processes for pH, aluminium concentration, and steam loss/dilution are implemented in the MATLAB code. To calculate temperature predictions above 300 °C, several thermodynamic databases are evaluated and the Lawrence Livermore National Laboratory database (llnl.dat) is extrapolated to temperatures up to 350 °C. The new multicomponent geothermometer, called MulT_predict, is tested on data from Krafla and Reykjanes to develop a site-specific basaltic mineral assemblage. The results of MulT_predict and conventional geothermometers based on the original and back-calculated fluid composition (via WATCH (Bjarnason, 2010)) are compared with each other and with the measured in-situ reservoir temperatures of the wells.

1.3.2 *Development of an automated interdependent optimisation process (Chapter 4)*

Ensuing from the first study (Chapter 3), the numerical optimisation process for the multicomponent geothermometer MulT_predict is improved. An automated interdependent optimisation process is implemented that combines the three individual optimisation processes for pH, aluminium concentration, and steam loss/dilution. The temperature dependence of pH is removed by adding organic acid (HNO_3) or base (NH_3) to fix the pH without disturbing the inorganic aqueous chemistry. In addition, a fourth optimisation process is added for an additional element concentration or system parameter. The MATLAB code is adapted to the new multidimensional IPhreeqc output for statistical evaluation of the data. In addition, the results are visualised by a four-panel output file showing the statistical evaluation of the reservoir temperature estimation. The interdependent optimisation process further improves the reservoir temperature predictions of MulT_predict.

1.3.3 *Application and internal benchmarking of MulT_predict (Chapter 5)*

This study develops a universally valid mineral assemblage for MulT_predict. This mineral assemblage applies to unknown reservoir compositions on a global scale. In addition, a three-step procedure for unknown temperature estimation is established. Outlier removal from the general mineral assemblage is introduced to refine reservoir temperature estimations. The new method is tested on data from eight different geothermal sites worldwide and validated using data of in-situ measured temperatures. Furthermore, an internal benchmark of the optimisation process is performed using a synthetic brine. A stepwise perturbation of the synthetic brine chemistry (pH, aluminium concentration, steam loss/dilution, and salinity) is calculated back to the initial reservoir conditions by applying the optimisation process. The performance of three individual optimisation processes (Chapter 3) and the interdependent optimisation process (Chapter 4) are calculated and compared.

1.3.4 *Numerical optimisation and machine learning in solute geothermometry (Chapter 6)*

The concept of numerical optimisation of multicomponent geothermometry (Chapter 5) is compared to a newly developed deep learning algorithm for reservoir temperature estimation. A Python-based (Van Rossum, Drake et al., 1995) deep learning algorithm is presented, which is trained on geochemical data and data of measured in-situ reservoir temperatures from Iceland. The results of the numerically optimised multicomponent geothermometer and the deep

learning algorithm are visualised and compared to each other and the measured in-situ reservoir temperatures of the wells.

1.3.5 *Development of an adequate solute artificial neural network geothermometer (Chapter 7)*

Based on the results of Chapter 6, the deep learning algorithm is redesigned to develop an adequate solute artificial neural network geothermometer, called AnnRG. A supervised feedforward multilayer perceptron is introduced. The development and testing of the network architecture and the hyperparameter optimisation of the artificial neural network are described step by step. To build AnnRG, a total of 208 data pairs of geochemical fluid parameters and in-situ temperature measurements from around the world are collected and compiled. Outliers are removed from the database to improve the quality of the input data for the artificial neural network. In addition, a sensitivity analysis of the required amount of data is performed to determine the minimum number of training data (>65 samples). 155 samples are used to build and train the solute artificial neural network geothermometer. To verify AnnRG, the geothermometer is transferred to 45 unknown samples to predict the reservoir temperature. AnnRG extends and refines machine learning applications in solute geothermometry and general geochemistry.

2

AQUEOUS GEOCHEMISTRY

2.1 THERMODYNAMICS

In solute geothermometry, the temperature of a geothermal reservoir is estimated by analysing the aqueous composition of its fluid. The fundamental assumptions of this method are described by Fournier et al. (1974):

1. "Temperature-dependent reactions occur at depth.
2. All constituents involved in a temperature-dependent reaction are sufficiently abundant (that is, supply is not a limiting factor).
3. Water-rock equilibration occurs at the reservoir temperature.
4. Little or no re-equilibration or change in composition occurs at lower temperatures as the water flows from the reservoir to the surface.
5. The hot water coming from deep in the system does not mix with cooler shallow ground water."

Thus, the geothermal fluid is governed by a thermal-chemical equilibrium reaction of the fluid with the mineral assemblage of the reservoir. In addition, the attained equilibrium state is mostly preserved while ascending to the surface. These mechanisms are based on the activity of the species (2.1.1), the concentrations of the dissolved elements at equilibrium (2.1.2), and the ionic strength of the solution (2.2.2). For constant reaction enthalpy and isobaric reaction heat capacity (fixed temperature and pressure at the reservoir), the Van 't Hoff-Equation (2.1.3) describes the temperature-dependency of the thermodynamic equilibrium constant. These thermodynamic parameters (2.3) can be used to calculate equilibrium state models for temperature estimation. For the geochemical modelling of equilibrium reactions the equilibrium state of system parameters (2.2.1) and the activity coefficients of aqueous species (2.2.2) have to be taken into account.

2.1.1 *Standard state*

The Gibbs free energy G is described by the enthalpy H, the thermodynamic temperature T, and the entropy S as the state of the system (Gibbs, 1873) (2.1).

$$G = H - T \cdot S = U + p \cdot V - T \cdot S \tag{2.1}$$

Since chemical reactions are generally governed by temperature and pressure, the Gibbs free energy can be also described by internal energy U, pressure p,

and volume V.
The variation of the state functions (G, H, and S) can be used to model changes in chemical reactions. Hence, the standard state of a thermodynamic potential is introduced, which is defined as a reference state at a given temperature, pressure, and concentration (298.15 K, 0.1 MPa, and 1 M). The standard state of the Gibbs free energy G_i° of one mole of a substance i is depending on its activity a_i as well as on the universal gas constant R, and the absolute temperature T (2.2).

$$G_i = G_i^\circ + RT \ln(a_i) \tag{2.2}$$

2.1.2 *Equilibrium constants*

The thermodynamics of a generalised geochemical reaction can be described by Equation 2.3.

$$\mathrm{a\ A} + \mathrm{b\ B} = \mathrm{c\ C} + \mathrm{d\ D} \tag{2.3}$$

In this formula A, B represent the reagents and C, D the products of the chemical reaction, where a, b, c, and d are their stoichiometric coefficients. The stoichiometric coefficients of reactants are negative, albeit the products are positive.
The equilibrium constant K is described by the activities of the reactants and the products, which is denoted by the parentheses (2.4).

$$K = \frac{(\mathrm{C})^{\mathrm{c}}(\mathrm{D})^{\mathrm{d}}}{(\mathrm{A})^{\mathrm{a}}(\mathrm{B})^{\mathrm{b}}} \tag{2.4}$$

At the most stable state, the equilibrium constant K is defined as equal to 1.
Regarding the activity of species, K is related to the Gibbs free energy in Equation 2.2. The energy differnce between the initial state and final state at any point in a given reaction is coupled to the equilibrium reaction. For the Gibbs free energy of the reaction ΔG_r, the equilibrium constant K can be inserted into the formula, linking it to the standard state Gibbs free energy of the reaction ΔG_r°, the absolute temperature T, and the ideal gas constant R (2.5).

$$\Delta G_r = \Delta G_r^\circ + RT \ln(K) \tag{2.5}$$

The Gibbs free energy of the reaction ΔG_r equals the sum of Gibbs free energy of the reactants and products multiplied by their stoichiometric coefficients (2.6).

$$\Delta G_r = \Delta G_{products} - \Delta G_{reactants} = \mathrm{c} \times \mathrm{G_C} + \mathrm{d} \times \mathrm{G_D} - \mathrm{a} \times \mathrm{G_A} - \mathrm{b} \times \mathrm{G_B} \tag{2.6}$$

In addition, the equation for the standard state molal Gibbs energy ΔG_r° is calculated similarly to Equation 2.6.
At equilibrium (K=1), the Gibbs free energy of the reaction ΔG_r is defined as equal to 0 (2.7).

$$\Delta G_r = 0 = \Delta G_r^\circ + RT \ln(K)_{eq} \tag{2.7}$$

Therefore, the Equation 2.7 can be reorganised to calculate the standard molal Gibbs energy ΔG_r° at equilibrium state (2.8).

$$\Delta G_r^\circ = -RT \ln(K)_{eq} = -RT \ln \left(\frac{(\mathrm{C})^{\mathrm{c}}(\mathrm{D})^{\mathrm{d}}}{(\mathrm{A})^{\mathrm{a}}(\mathrm{B})^{\mathrm{b}}} \right)_{eq} \tag{2.8}$$

The rearrangement of Equation 2.8 can be used to calculate the equilibrium constant at arbitrary temperatures (2.9).

$$\ln(K)_{eq} = -\frac{\Delta G_r^\circ}{RT} \qquad \text{thus,} \qquad \log(K)_{eq} = -\frac{\Delta G_r^\circ}{2.3RT} \tag{2.9}$$

2.1.3 *Effects of temperature and pressure*

Temperature changes ΔT and pressure changes ΔP affect the reaction between reactants and products. Similar to the activity coefficients, the variation in the standard state Gibbs free energy ΔG_r° depends on the species (solid, gaseous, and aqueous species). In addition, reference temperature T_{ref} and reference pressure P_{ref} are introduced as starting point of the reaction for the standard Gibbs free energy $\Delta G^\circ_{r_{ref},T_{ref},P_{ref}}$. For gas and solid species, the variation of the standard state Gibbs energy of the reaction is given by Equation 2.10.

$$\begin{aligned} \Delta G^\circ_{r,T,P} =& \Delta G^\circ_{r_{ref},T_{ref},P_{ref}} + \int_{T_{ref}}^{T} \frac{\partial \Delta G_r^\circ}{\partial T} dT + \int_{P_{ref}}^{P} \frac{\partial \Delta G_r^\circ}{\partial P} dP \\ =& \Delta G^\circ_{r_{ref},T_{ref},P_{ref}} - \int_{T_{ref}}^{T} \Delta S_r^\circ dT + \int_{P_{ref}}^{P} \Delta V_r^\circ dP \end{aligned} \tag{2.10}$$

The standard entropy of the chemical reaction ΔS_r° is temperature-driven, while the standard volume ΔV_r° is pressure-dependent. The calculation of both parameters ΔS_r°, ΔV_r° is similar to the calculation of ΔG_r described in Equation 2.6. For the aqueous species, the standard molal Gibbs free energy of a solute is given by Equation 2.11.

$$G^\circ_{P,T} = G^\circ_{P_{ref},T_{ref}} - S^\circ_{P_{ref},T_{ref}} \left(T - T_{ref}\right) + \int_{T_{ref}}^{T} C_P^\circ dT - \int_{T_{ref}}^{T} \frac{C_P^\circ}{T} dT + \int_{P_{ref}}^{P} V^\circ dP \tag{2.11}$$

For solids and gases, the standard entropy of the reaction ΔS_r° at constant pressure P is related to the standard heat capacity of the reaction $\Delta C^\circ_{P,r}$ for the temperature range T_{ref} to T (2.12).

$$\Delta S^\circ_{r,T} - \Delta S^\circ_{r,T_{ref}} = \int_{T_{ref}}^{T} \frac{\Delta C^\circ_{P,r}}{T} dT \tag{2.12}$$

The standard heat capacity of the reaction $\Delta C^\circ_{P,r}$ is also calculated like Equation 2.6 summing up the standard molal heat capacities of reactants and products

multiplied by their stoichiometric coefficients. Similar, the standard molal heat capacity of a solute is calculated for an aqueous species (2.13).

$$S^{\circ}_{P,T} = S^{\circ}_{P_{ref},T_{ref}} + \int_{T_{ref}}^{T} \frac{C^{\circ}_{P}}{T} dT - \int_{P_{ref}}^{P} \left[\left(\frac{\partial V^{\circ}}{\partial T} \right)_{P} \right]_{T} dP \tag{2.13}$$

The enthalpy changes ΔH°_{r} for a specific reaction at constant pressure are described by the standard molal heat capacity $\Delta C^{\circ}_{P,r}$ (2.14).

$$\Delta H^{\circ}_{r,T} - \Delta H^{\circ}_{r,T_{ref}} = \int_{T_{ref}}^{T} \Delta C^{\circ}_{P,r} dT \tag{2.14}$$

In addition, the standard molal enthalpy for aqueous species is described by Equation 2.15.

$$H^{\circ}_{P,T} = H^{\circ}_{P_{ref},T_{ref}} + \int_{T_{ref}}^{T} C^{\circ}_{P} dT - \int_{P_{ref}}^{P} \left[V^{\circ} - T \left(\frac{\partial V^{\circ}}{\partial T} \right)_{P} \right]_{T} dP \tag{2.15}$$

Van 't Hoff (1886) described the temperature dependence of a reaction coupled to enthalpy changes ΔH°_{r}. Thus, the reaction temperature is affecting the equilibrium constant K. This relation is expressed by Equation 2.16.

$$\ln \left(\frac{K_T}{K_{T_{ref}}} \right) = \int_{T_{ref}}^{T} \frac{\Delta H^{\circ}_{r}}{RT^2} dT = -\frac{\Delta H^{\circ}_{r}}{RT} + \frac{\Delta H^{\circ}_{r}}{RT_{ref}} \tag{2.16}$$

Rearranging Equation 2.16 comparable to 2.9, the variation of enthalpy can be used to calculate the changes in the equilibrium constant given by the temperature difference (2.17).

$$\log(K_T) = \log(K_{T_{ref}}) - \frac{\Delta H^{\circ}_{r}}{2.3R} \left(\frac{1}{T} - \frac{1}{T_{ref}} \right) \tag{2.17}$$

2.2 EQUILIBRIUM STATES

Multiple equilibrium states govern the geochemistry in a reservoir. Especially, properties of the system such as mineral assemblage of the reservoir rock and dissolved species, pH value (as the activity of hydrogen), reduction–oxidation potential, and amount of total dissolved solids (TDS), also known as salinity, characterise the chemical conditions of the reservoir. These properties as well as the activity coefficients given by thermodynamic parameters influence the equilibrium reactions in the reservoir fluid.

2.2.1 *System parameters*

The phase rule governs the degrees of freedom F, which are given by the number of components C and the number of phases P in the system. (2.18).

$$F = C - P + 2 \tag{2.18}$$

Both, the mineral assemblage of the reservoir rock and the dissolved mineral phases in the fluid are governed by the phase rule (2.18). Thus, the site-specific reservoir rock has a major impact on the chemical properties of the geothermal fluid. Consequently, temperature changes, water-rock interaction, and varying rock composition affect the number of components in the system.

The pH value is given by the negative decadic logarithm of the activity a_i of the hydrogen ion (2.19).

$$\mathrm{pH} = -\log(a_{\mathrm{H}^+}) \tag{2.19}$$

The amount of hydrogen ions within the aqueous solution is driven by the temperature-dependent dissociation of water (2.20).

$$\mathrm{H_2O} \rightleftharpoons \mathrm{OH^-_{(aq)}} + \mathrm{H^+_{(aq)}} \tag{2.20}$$

At neutral conditions, the pH value equals 7. For alkaline solutions, the pH increases above 7, whereas acidic fluids decreases below 7. The equilibrium state of acid–base reaction is given by the pH value, which is mainly controlling precipitation-dissolution reactions within the reservoir and while the ascendence of the fluid to the surface. Both, the geochemical equilibrium reaction of pH-dependant mineral phases and the solubility of $\mathrm{CO_{2(aq)}}$ and $\mathrm{H_2S_{(aq)}}$ are affected by the pH value in the reservoir. Therefore, the pH value can govern the presence of species in the fluid, likewise the concentration of sulphide (e.g. pyrite, galenite) and carbonate species (e.g. calcite, dolomite). In addition, the pH value is buffered by geochemical reactions in the solution. For example, the carbonate buffer system is able to regulate the pH value in the following manner (2.21).

$$\mathrm{CO_{2(aq)}} \rightleftharpoons \mathrm{HCO_3^-} + \mathrm{H^+} \rightleftharpoons \mathrm{CO_3^{2-}} + 2\,\mathrm{H^+} \tag{2.21}$$

Dissolved $\mathrm{CO_2}$ reacts with water to carbonic acid ($\mathrm{H_2CO_3}$), which is not stable in solution. The carbonic acid dissociates into bicarbonate ion ($\mathrm{HCO_3^-}$) and hydrogen ion ($\mathrm{H^+}$). Further deprotonation dissociates bicarbonate into carbonate ion ($\mathrm{CO_3^{2-}}$) and an additional hydrogen ion. These reversible acid-base reactions (2.21) govern the buffer capacity of the aqueous system by regulating the number of hydrogen ions and therefore, their activity (2.19).

The *pe* value is given by the negative decadic logarithm of the activity of the free electron concentration e^- in the solution (2.22). It regulates the reduction and oxidation potential of the system.

$$pe = -\log(a_{e^-}) \tag{2.22}$$

An increase in *pe* leads generally to a decrease in the pH. Similar to the pH buffer system (2.21), the reduction and oxidation reactions are reversible. In an aqueous solution, the redox potential E_h is given by the Nernst-Equation (2.23).

$$E_h = \frac{RT}{F} \log\, pe \tag{2.23}$$

The equation consists of the potential of free electrons in the solution *pe* as well as the thermal voltage $\left(\frac{RT}{F}\right)$ comprising the molar gas constant R, the absolute temperature T, and the Faraday constant F. The potential difference in the fluid is measured in volts. Positive voltage indicates oxidising conditions, whereas negative voltage indicates reducing conditions. In aqueous chemistry, the limits of a natural system are normally given by the activity of transferable oxygen (O_2) and hydrogen ions (H^+). Nevertheless in the subsurface, oxygen can be depleted fast resulting in reducing conditions in the system. The upper limit is governed by the oxidant, in this example oxygen (2.24).

$$E_h = E^\circ + \frac{RT}{z_{e^-} F} \log\left(\frac{[O_2][H^+]^4}{[H_2O]^2}\right) \tag{2.24}$$

In this case, E° is the standard half-cell potential of dissolved oxygen and the dissociation of water (cf. 2.20), where z_{e^-} indicates the number of transferred electrons. Similarly, the lower limit is given by the standard hydrogen electrode (2.25).

$$E_h = E^\circ + \frac{RT}{z_{e^-} F} \log\left(\frac{[H^+]^2}{[H_2]}\right) \tag{2.25}$$

Here, E_h indicates the potential difference between the aqueous system and the standard hydrogen electrode. Due to the hydrogen ion dependency, the redox potential is also a function of the pH value (2.19).

The salinity is also an influencing factor in the geochemistry of aqueous systems. It is specified as the amount of total dissolved solids (TDS) in a solution. The TDS is mostly governed by the alteration and dissolution of mienrals but also by runoff and soil solution. Thus, the origin of the water (e.g. meteoric water, seawater) determines the chemical properties of geothermal fluid. Increasing concentrations of salt, such as sodium chloride, function as electrolytes showing nonideal properties with increasing ionic strength. This nonideality is connected to the thermodynamic activity coefficients of aqueous species. Generally, these coefficients are calculated with models referring to infinitely diluted solutions.

2.2.2 *Activity coefficient*

For real substances, the difference between the standard state and the non-ideal state is given by their activity coefficients. The activity a_i depends on the species i. For solid and liquid phases, it is defined by the activity coefficient λ_i and the molar fraction of the component x_i (2.26).

$$a_i = \lambda_i \cdot x_i \tag{2.26}$$

For aqueous solutions, the activity is given by its activity coefficient γ_i and the molality of the components m_i (2.27).

$$a_i = \gamma_i \cdot m_i \tag{2.27}$$

In addition, for gaseous phases, the activity is related to the fugacities f_i of the gas mixture or single phase related to the standard state fugacity f_i° (2.28).

$$a_i = \frac{f_i}{f_i^\circ} \tag{2.28}$$

The activity coefficients are coupling the concentration to the thermodynamic activity function a_i. Thus, the activity in Equation 2.2 is linked to the aforementioned activity coefficients. Since geochemical reactions are based on their thermodynamic potential, only ratios and products of individual activity coefficients of species can be determined. For solute geothermometry, the equilibrium state of the aqueous phases is essential for temperature estimation. At equilibrium, the aqueous phases in the geothermal fluid and the solid phases of the mineral assemblage in the reservoir have the same chemical potential. The activity coefficient of an aqueous species γ_i is coupled to the amount of the components m_i and thus, the molal concentration of the ions in the solvent and their charge number z_i (2.29).

$$I = \frac{1}{2} \sum m_i z_i^2 \tag{2.29}$$

The ionic strength I describes the electrostatics of the dissolved ions in the electrolyte, where the electric field $E(r)$ of the ion is given by Coulomb's law (2.30).

$$E(r) = \frac{q}{4\pi\epsilon_0} \frac{1}{r^2} \tag{2.30}$$

Simplifying the ions as point charges q, the electric field is governed by the vacuum permittivity ϵ_0, and the distance from the source r. The electrostatic interaction between the ions shifts the solution from the ideal state. The influence of an ion on the electric field in an electrolyte is given by the Debye length λ_D (Debye and Hückel, 1923), where ϵ is the electrolyte permittivity and k_B is the Boltzmann constant (2.31).

$$\lambda_D = \sqrt{\frac{\epsilon\epsilon_0 k_B T}{2 z_i^2 I}} \tag{2.31}$$

Rearranging and substituting Equation 2.2 with Equation 2.31 and 2.30, we get the Debye–Hückel limiting law (Debye and Hückel, 1923) (2.32).

$$\ln(\gamma_i) = -\frac{z_i^2 q^2 \lambda_D^{-1}}{8\pi\epsilon\epsilon_0 k_B T} \tag{2.32}$$

The activity coefficient of an ion in a diluted electrolyte can be calculated via the charge number z_i of the ith ion species and the ionic strength I of the aqueous solution. To simplify Equation 2.32 all constants are unified in the Debye–Hückel constant A (2.33).

$$\log(\gamma_i) = -Az_i^2\sqrt{I} \tag{2.33}$$

Equation 2.33 is the foundation of the empirical calculation of activity coefficients for very dilute solutions with $I < 0.005$ mol L^{-1} (Sparks, 2018). For solutions with higher I up to 0.1 mol L^{-1} (Robinson et al., 1960), the extended Debye-Hückel-Equation (Hückel, 1925) can be used (2.34).

$$\log(\gamma_i) = -\frac{Az_i^2\sqrt{I}}{1 + \mathring{a}_i B\sqrt{I}} \tag{2.34}$$

The numerator quantifies the long-range Coulomb forces, whereas the denominator defines the short-range interactions between the ions and the solvent governed by the hydrated ion size $\mathring{a}_i$. In addition, the Debye–Hückel constant B is introduced (2.35), where the density-dependency is given by the Avogadro constant N_A.

$$B = \sqrt{\frac{2q^2 N_A}{\epsilon\epsilon_0 k_B T}} \tag{2.35}$$

To further expand the former validity limit in terms of the ionic strength (up to I = 1.0 mol L^{-1}), Helgeson (1969) modified the extended Debye-Hückel-Equation (2.34) by a linear term introducing the $\dot{B}$-Equation (2.36).

$$\log(\gamma_i) = -\frac{Az_i^2\sqrt{I}}{1 + \mathring{a}_i B\sqrt{I}} + \dot{B}I \tag{2.36}$$

The linear term $\dot{B}I$ includes the short-range ion-ion interaction related to the charge number z_i. For single ion-activity coefficients in Na-Cl solutions, the model fits data to at least I = 3.0 mol L^{-1} (Truesdell and Jones, 1974). In addition, Helgeson and Kirkham (1974) evaluated the effect of temperature and pressure on the constants A, B, and $\dot{B}$.

2.3 THERMODYNAMIC DATABASES

To determine aqueous geochemistry, dissolution reactions of mineral phases have to be calculated. Thus, thermodynamic databases were established, which are based on laboratory experiments, measurements, and extrapolations. In regard to hydrochemical equilibrium calculations, multiple thermodynamic parameters have to be available in such a database.

2.3.1 *Structure of a thermodynamic database*

In general, thermodynamic databases are constructed in an certain pattern. The elements, their chemical formulas, and their standard atomic weights are established. The ions of these elements are defined by their charge numbers at all possible oxidation states. Then, the thermodynamic data of these species are added. Redox couples of the ions are ascribed and predefined data, such as atomic weights and charge numbers, are shared and updated. Lastly, phases (e.g. minerals, gases) are assigned. Similarly, the chemical formula, the reaction species (ions), and charge numbers as well as thermodynamic data is implemented. Depending on the TDS of the solution, thermodynamic databases are coupled to different activity coefficients (2.2.2) to calculate the equilibrium of a reaction and its species. For example in the geochemical modelling program IPhreeqc (Charlton and Parkhurst, 2011; Parkhurst, Appelo et al., 2013), usually three major aqueous activity coefficients models are used.

1. The Davies-Equation (Davies, 1962), based on the Debye–Hückel-Equation (2.33), adding an empirical parameter $\frac{1}{1+\sqrt{I}} - 0.3I$ to the equation.
2. The $\dot{B}$-Equation (2.36) for salinity up to 1 mol L^{-1} (Helgeson, 1969).
3. The Pitzer-Equation (Pitzer, 1973) for salinity up to 6 mol L^{-1} (Pitzer and Mayorga, 1973).

Thus, the Debye-Hückel parameter respectively the Pitzer parameter are also included in the thermodynamic databases.
In case of thermodynamic data, multiple thermodynamic parameters can be used to calculate chemical reactions. In general, the $\log(K)$ is used to compute equilibrium reactions. Depending on the database, it is most commonly calculated using the standard molal Gibbs free energy of the reaction ΔG_r° (2.9) or the standard enthalpy of the reaction ΔH_r° given by the Van 't Hoff-Equation (2.17). Similarly, at constant pressure, the standard entropy of the reaction ΔS_r°, the standard heat capacity of the reaction $\Delta C_{p,r}^\circ$, and the standard volume of the reaction ΔV_r° can be used (2.11, 2.13). Moreover, so called 'analytic' polynomial expressions are used to determine the $\log(K)$ of the reaction (Parkhurst, Appelo et al., 2013) (2.37).

$$\log(K) = A_1 + A_2 T + \frac{A_3}{T} + A_4 \log(T) + \frac{A_5}{T^2} + A_6 T^2 \tag{2.37}$$

The analytical expression consists of up to six parameters A_{1-6}, which are used to calculate the $\log(K)$ for specific temperatures T. Equation 2.37 is derived from a formula representing high-temperature thermal heat capacity data on compounds for constant pressure proposed by Maier and Kelley (1932) (2.38).

$$\Delta C_{p,r}^\circ = a + bT + cT^{-2} \tag{2.38}$$

The Maier-Kelley-Equation can be applied to Equation 2.14 with a temperature range from T_{ref} to T if the pressure dependency is neglectable.

$$\Delta H^{\circ}_{r,T} = \Delta H^{\circ}_{r,T_{ref}} + \Delta a T + \frac{\Delta b}{2} T^2 + \Delta c T^{-1} \tag{2.39}$$

This formula is particularly useful to determine small temperature differences within the reaction, where $\partial \Delta H^{\circ}_{r}$ corresponds to $\Delta C^{\circ}_{P,r}$.

2.3.2 *Comparison of thermodynamic databases*

Since the herein introduced multicomponent geothermometer, called MulT_-predict, is coupled to IPhreeqc (Charlton and Parkhurst, 2011; Parkhurst, Appelo et al., 2013) only associated databases can be utilised. The application area of thermodynamic databases for solute geothermometry is given by a large temperature range of up to 300 °C, processing high ionic strength, and consisting of a variety of species and mineral phases. Five PhreeqC databases are coming within the ambit of usage: the llnl.dat (Daveler and Wolery, 1992) (converted by Greg Anderson and assisted by David Parkhurst (2017)), the Thermoddem.dat (Blanc et al., 2012), the core10.dat (Neveu et al., 2017), the carbfix.dat (Voigt et al., 2018), and the SupPHREEQC database (Zhang et al., 2020). The aqueous activity coefficient model of all databases is based on the $\dot{B}$-Equation (2.36) (Helgeson, 1969).

To assess the usability of the databases, the solubility of the main mineral phases corresponding to the main cations is calculated and compared. The monovalent cation, Na^+ is a major cation encountered in crustal rocks and geothermal water mostly controlled by the equilibrium state of albite (Ellis and Mahon, 1964) (Figure 2.1). In addition, K^+ is a major cation corresponding to microcline respectively to K-feldspar equilibrium state within the aqueous solution (Ellis and Mahon, 1964) (Figure 2.2). In hydrothermal systems, the Ca^{2+} concentration is mostly given by the temperature and salinity-dependent solubility of calcite and calcium-aluminium-silicates (Ellis, 1963; Giggenbach, 1981) (Figure 2.3). Similarly, the divalent Mg^{2+} is controlled by the solubility of K-Mg layer silicates, which is chlorite-dependent (Giggenbach, 1988). In addition, dolomite equilibrium is reached, when the origin of the geothermal fluid is seawater (Reed, 1982) (Figure 2.4). In geothermal fluids, the silica concentration is mostly driven by reservoir conditions regarding chalcedony or quartz equilibrium (Fournier and Rowe, 1966) (Figure 2.5). Following, the mineral phases low albite ($NaAlSi_3O_8$), microcline ($KAlSi_3O_8$), calcite ($CaCO_3$), dolomite ($CaMg(CO_3)_2$), and quartz (SiO_2) are compared. Therefore, the molality of the solute major element at equilibrium is plotted against the temperature (Figure 2.1 - 2.5).

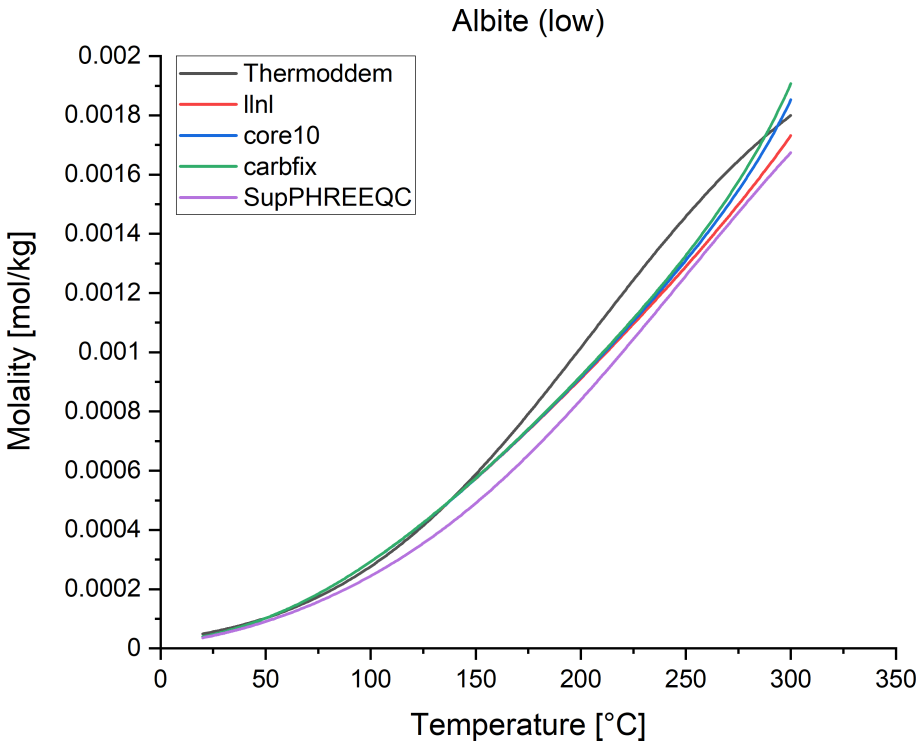


Figure 2.1: The molality of Na^+ of low albite against the temperature.

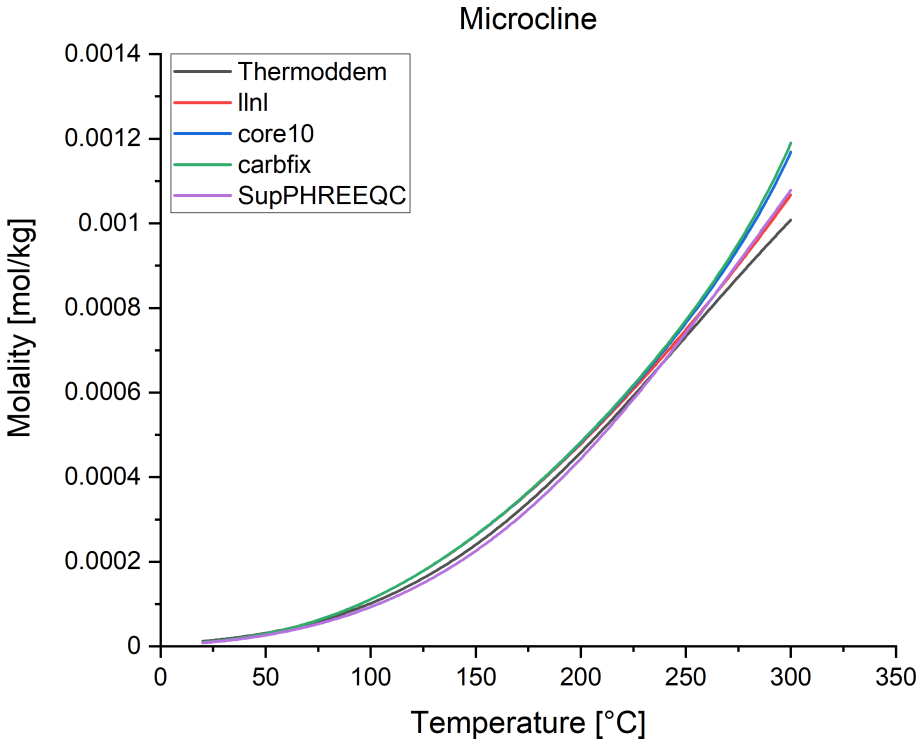


Figure 2.2: The molality of K^+ of microcline against the temperature.

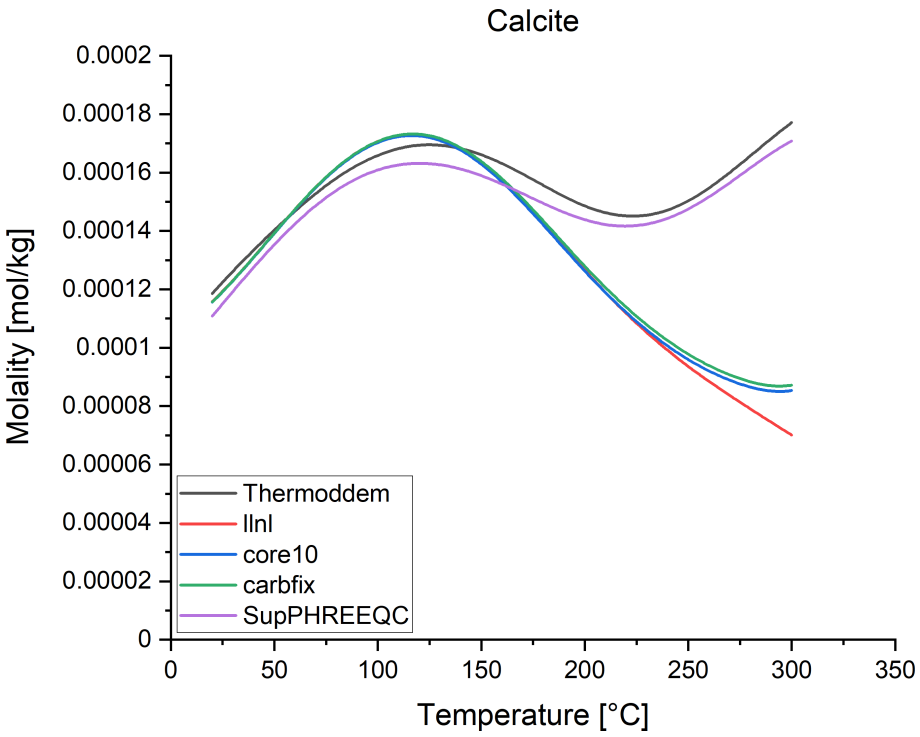


Figure 2.3: The molality of Ca^{2+} of calcite against the temperature.

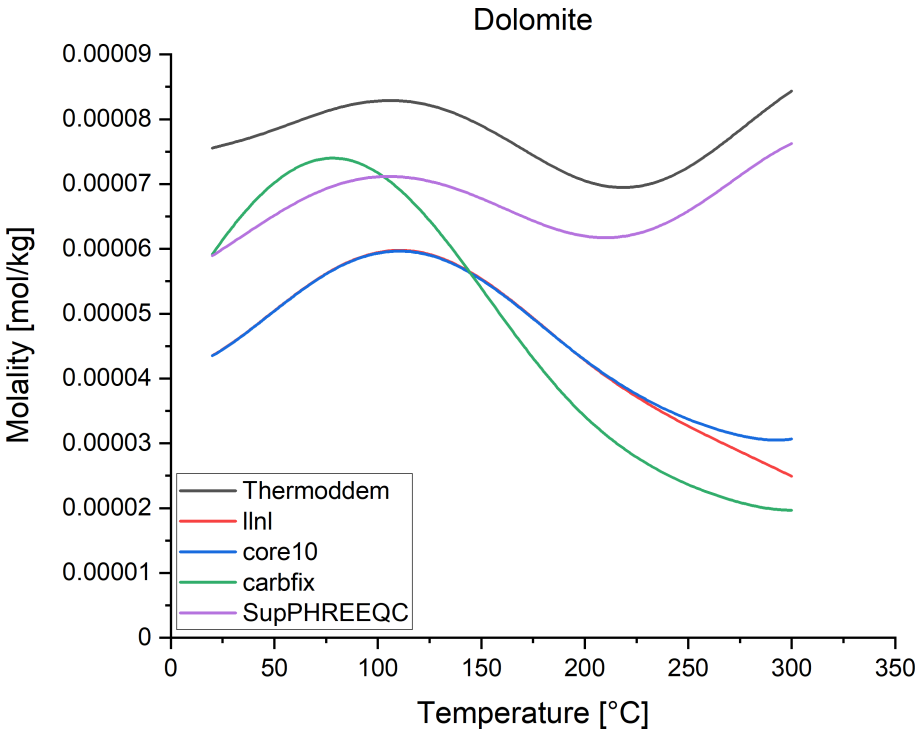


Figure 2.4: The molality of Mg^{2+} of dolomite against the temperature.

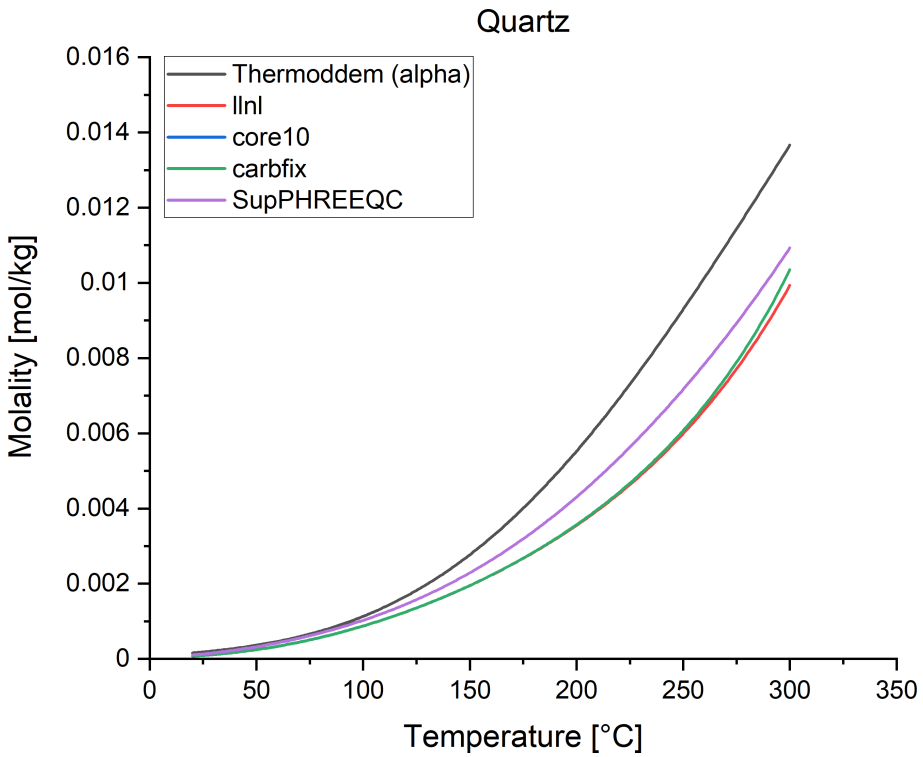


Figure 2.5: The molality of SiO_2 of quartz against the temperature.

Comparing the databases, the llnl.dat, core10.dat, and carbfix.dat show similar element solubilities for a temperature range from 25 °C up to 225 °C. Only the dolomite solubility shifts in its form. Since the activity model of the llnl.dat was adapted by the core10.dat, the $\log(K)$s were also transferred (Neveu et al., 2017). In addition, the carbfix.dat is based on the core10.dat, thus, the thermodynamic properties are mostly similar to the llnl.dat, only for carbonates, the carbfix.dat is modified to better fit these aqueous species (Voigt et al., 2018). In contrast, the SupPHRREQC.phr and the Thermoddem.dat have more variety in their curves' shapes. Especially for calcite, dolomite, and quartz, the solubility courses of the curves differ from the other databases. The SupPHREEQC.phr has implemented the thermodynamic data of Sverjensky et al. (1997), while the Thermoddem.dat is based on the thermodynamic properties implemented from Blanc (2008). However, the solubility of microcline and albite within the SupPHREEQC database follows the results of the llnl.dat due to the application of an equal framework (Zhang et al., 2020).

To test the databases for temperature prediction, the solubility curves of multiple mineral phases are plotted for a known geothermal sample and its reservoir temperature. Since the dominant mineral phases are Al-bearing silicate species, the influence of carbonates is mitigated. The Thermoddem.dat (2.6) and SupPHREEQC.phr show more deviation from the geochemical equilibrium temperature than the llnl.dat (2.7). Since llnl.dat, core10.dat, and carbfix.dat evolved from the 'thermo.com.V8.R6.230.dat' by Jim Johnson (2000), only the variety of

species and mineral phases determines, which database one would use. The llnl.dat comprises the most phases and hence, is used for the multicomponent geothermometer within this thesis.

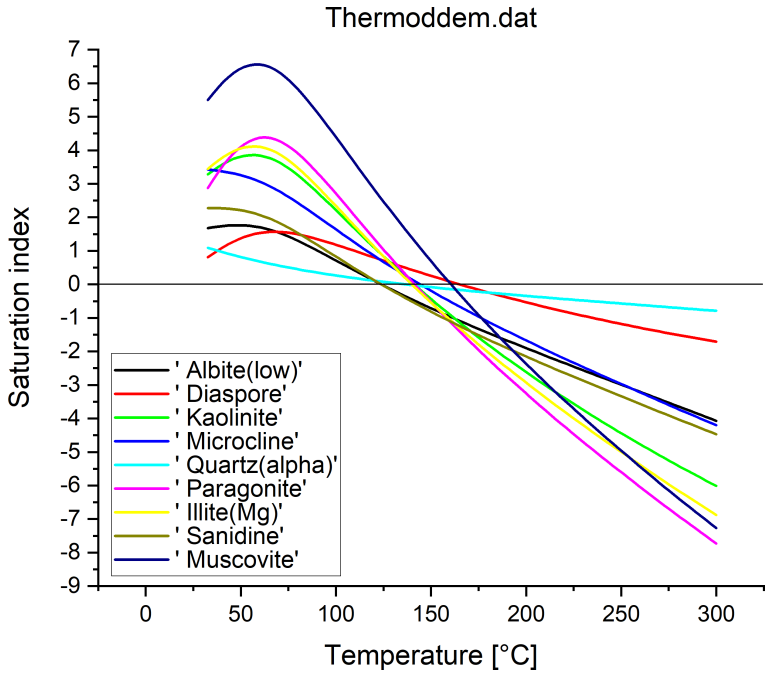


Figure 2.6: Thermoddem.dat: Solubility curves of multiple mineral phases (coloured lines) plotted against temperature.

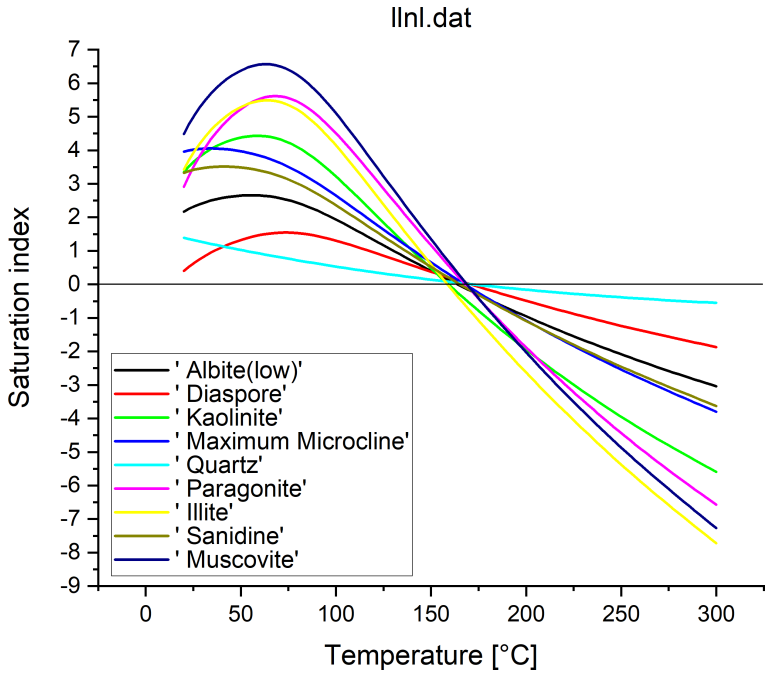


Figure 2.7: llnl.dat: Solubility curves of multiple mineral phases (coloured lines) plotted against temperature.

3

A MULTICOMPONENT GEOTHERMOMETER FOR HIGH-TEMPERATURE BASALT SETTINGS

This chapter was published in Geothermal Energy (2020) 8:2
DOI: 10.1186/s40517-020-0158-z

ABSTRACT

For successful geothermal reservoir exploration, accurate temperature estimation is essential. Since reservoir temperature estimation frequently involves high uncertainties when using conventional solute geothermometers, a new statistical approach is proposed. The focus of this study is on the development of a new multicomponent geothermometer tool, which requires a significantly reduced data set compared to existing approaches. The method is validated against reservoir temperature measurements in the Krafla and the Reykjanes geothermal systems. A site-specific basaltic mineral set was selected as the basis to compute the equilibrium temperatures. These high-enthalpy geothermal reservoirs are located in the neo-volcanic zone of Iceland, where the fluid temperatures are known to reach up to 350 °C at a depth of 2000 m. During ascent, the fluid composition is prone to changes as well as possible phase segregation due to depressurization and boiling. Therefore, to reduce the uncertainty of temperature estimations, reservoir temperature conditions are numerically reconstructed with sensitivity analyses considering pH, aluminium concentration, and steam loss. The evaluation of the geochemical data and the sensitivity analyses were calculated via a numerical in-house tool called MulT_predict. In all cases, the temperature estimations match with the in-situ temperatures measured at Krafla and Reykjanes. The development of this method tends to be a promising and precise tool for reservoir temperature estimation. The developed methodology is a fast and easy-to-handle exploration tool that can be applied to standard geochemical data without the need for a sophisticated gas analysis yet obtaining very accurate results.

3.1 INTRODUCTION

A reliable temperature estimation for a targeted geothermal reservoir, which lays the foundation for the prediction of producible energy, is essential for a successful exploration campaign. Conventional solute geothermometers are a commonly used tool for the deduction of reservoir temperature from geochemical compos-

ition of geothermal spring samples. These geothermometers were introduced in the 1960s and have been undergoing further development since then (Can, 2002; Ellis, 1970; Fouillac and Michard, 1981; Fournier, 1979; Fournier and Rowe, 1966; Fournier and Truesdell, 1973; Giggenbach, 1988; Sanjuan et al., 2014; Verma and Santoyo, 1997). These approaches use the temperature dependence of the saturation of mineral phases (e.g. silica) or certain cation ratios in the fluids. The measured concentrations of these fluid constituents are then directly linked to a reservoir temperature. The fundamental assumption of geothermometry is the overall chemical equilibrium of the fluid and the reservoir rock. Secondary processes may change the fluid composition and hence, the equilibrium while migrating to the earth surface. These variations can result in large uncertainties for the reservoir temperature determination using solute geothermometers (Nitschke et al., 2018). The more recently developed multicomponent geothermometry evaluates the equilibria of multiple mineral phases (Spycher et al., 2014). Numerical geochemical speciation codes facilitate this evaluation based on a large number of minerals, which leads to a statistically more robust method. Spycher et al. (2014) proposed a pre-selection of minerals representing the site-specific reservoir rocks to enhance accuracy. Corrections were established to overcome interferences from secondary processes such as dilution, boiling, and mixing of fluids affecting the temperature estimation (Cooper et al., 2013; Peiffer et al., 2014; Spycher et al., 2014). These methods need an additional gas analysis for precise temperature estimations. Thus, Nitschke et al. (2017) introduced a method to reconstruct in-situ conditions of the reservoir temperature by varying sensitive parameters, especially pH and aluminium concentration as well as steam loss, to further reduce the uncertainty of equilibrium temperatures.

The goals of this study are to refine and to validate the existing specific multicomponent approach according to Nitschke et al. (2017) and to expand it towards a highprecision exploration tool. This study devises a basalt-specific mineral set including secondary mineral phases for global application to basaltic stratigraphy. For the validation, geochemical data and in-situ temperature measurements from basalt-hosted geothermal systems, Krafla and Reykjanes, are used. These are high-enthalpy systems with near-boiling reservoir fluids and, thus, the effect of steam loss between the reservoir and the liquid sampled at the well-head has to be considered. Krafla hosts dilute meteoric fluids (Arnórsson, 1978), and Reykjanes a more saline reservoir fluid partially originating from seawater (Arnórsson et al., 1978). To validate the method, the temperature estimations are compared with direct in-situ temperature measurements of the wells published by Gudmundsson and Arnórsson (2002) for Krafla, and Óskarsson et al. (2015) for Reykjanes. The advantage of the validated method is the frugality in terms of input data. High-accuracy temperature estimations can be achieved based on standard fluid analyses and do not require comprehensive high-end fluid and gas analyses, which are required for other solute multicomponent geothermometer approaches.

3.2 METHOD AND DATA

The basis of the method is a standard fluid analysis comprising major cations and anions as well as aluminium concentration and pH. The water analysis is used to calculate equilibrium conditions between the dissolved constituents in the geothermal fluid and the reservoir rock minerals. For identification of the reservoir temperature conditions, sensitive parameters have to be evaluated statistically. The application relies on the following general assumptions: (i) the reservoir and the geothermal fluid are in equilibrium. Therefore, the ion activity product of a mineral phase equals its thermodynamic equilibrium constant. (ii) A temperature-dependent reaction between the host rock and the water leads to a specific amount of dissolved solids in the fluid phase.
The equilibrium reaction is based on the law of mass action. The state of the dynamic equilibrium between the reactants is expressed in terms of the saturation index SI (3.1)

$$SI(T) = log\left(\frac{IAP}{K(T)}\right) = log\left(\prod_i \gamma_i^{\nu_i} x_i^{\nu_i}\right) - log(K(T)) \tag{3.1}$$

with IAP being the ion activity product and K being the temperature-dependent thermodynamic equilibrium constant of one mineral phase.
IAP is the product of the activity coefficients γ_i and the mole fractions of the solute mineral phase x_i considering their stoichiometric coefficient ν_i. A positive saturation index indicates an oversaturation and a potential precipitation of the mineral. Though, if the ion activity product is smaller than the equilibrium constant, the saturation index will be negative. In this case, the solution is undersaturated with the potential to dissolve the mineral phase. Therefore, equilibrium is given at $SI = 0$.
Debye and Hückel (1923) established an equation for non-ideal electrolyte solutions taking into account the electrostatic interaction among the ions by using the activity coefficients γ_i (3.2). In order to fit the Debye–Hückel equation to experimental data, Robinson et al. (1960) extended the original equation by adding a linear concentration term $\dot{B}I$.

$$log\gamma_i = \frac{Az_i^2\sqrt{I}}{1 + \mathring{a}_i B\sqrt{I}} + \dot{B}I \tag{3.2}$$

where A and B are temperature-dependent constants, z_i is the charge number of the ion, I is the ionic strength, and $\mathring{a}_i$ is the hydrated ion size. The numerator quantifies the long-range Coulomb forces acting on the ion, whereas the

denominator defines the short-range interactions between the ions itself and with the solvent. The extended Debye–Hückel equation (3.2) expands the former validity limit in terms of the ionic strength to I = 1.0 mol/L for mixed electrolytes. Furthermore, there are application limits given by specific temperatures and pressures. The latter is negligible at least up to a temperature of 300 °C (Helgeson, 1969).

In this study, chemical speciation and saturation indices are computed with IPhreeqc 3.4.0-12927 (Parkhurst, Appelo et al., 2013). Ion activity coefficients are based on the extended Debye–Hückel equation. The required constants A, B, and $\dot{B}$ are obtained from the commonly used LLNL (Lawrence Livermore National Laboratory) database. Saturation indices are computed for all specified minerals for a given solution. Reed and Spycher (1984) plotted these saturation indices versus temperature to investigate the equilibrium temperature of the geothermal fluid and the reservoir mineral assemblage. For validation of the multicomponent geothermometer, the method is applied on wellstudied geothermal sites in Iceland and further developed to obtain an easy-to-handle and convenient high-precision exploration tool.

Krafla is a high-temperature geothermal field located in the NE of Iceland. The geothermal system is situated in the neo-volcanic zone Ármannsson et al., 1987. The in-situ temperature measurements (3.1) and the geochemical data of the wells (Appendix A.1) were published by Gudmundsson and Arnórsson (2002). The upper 1000 m of the stratigraphy are built up by alternating layers of basaltic lavas and hyaloclastite. The latter is subglacial erupted basaltic lava, which forms hydrated breccia once it is in contact with water. Below 500 m, the hyaloclastite layers form subhorizontal reservoirs. The following 1000 m are covering basaltic intrusives, where geothermal fluids of up to 310 °C are evident (Gudmundsson and Arnórsson, 2002). A more detailed stratigraphy of the field is given by Ármannsson et al. (1987). The mineralogical content of Icelandic geothermal systems and the geochemistry of the fluids are described by Arnórsson et al. (1983). Sampling methods and the geochemical analysis are given in Arnórsson et al. (2006).

Table 3.1: In-situ measurements of the temperature [°C] in the wells of Krafla (K) at specific depths [m] for permeable horizons presented in Gudmundsson and Arnórsson (2002).

K-11		K-24		K-28	
1330 m	>240 °C	580 m	190 °C	500 m	230 °C
1600 m	>240 °C	780 m	195 °C	800 m	240 °C
1700 m	>240 °C	920 m	210 °C		
2180 m	>300 °C	1150 m	225 °C		

3.3 RESULTS OF THE ANALYSIS

The application of this multicomponent geothermometer approach requires the evaluation of the equilibrium of each solute mineral phase and, thus, the calculation of saturation indices of the considered minerals versus temperature. The saturation indices are calculated from 20 to 300 °C. The calculations of the saturation indices are processed via MATLAB (MATLAB, 2019). Therefore, the MulT_predict tool was developed, which determines the intersection of the saturation index function for each mineral phase with the equilibrium line (3.1). Thus, the tool calculates all mineral-specific saturation indices functions throughout the temperature range by interacting with IPhreeqc. Only minerals having exactly one intersection with the equilibrium line are taken into account for the temperature determination procedure. The resulting intersection temperatures are combined in a box plot. This plot represents a first estimate of the reservoir temperature.

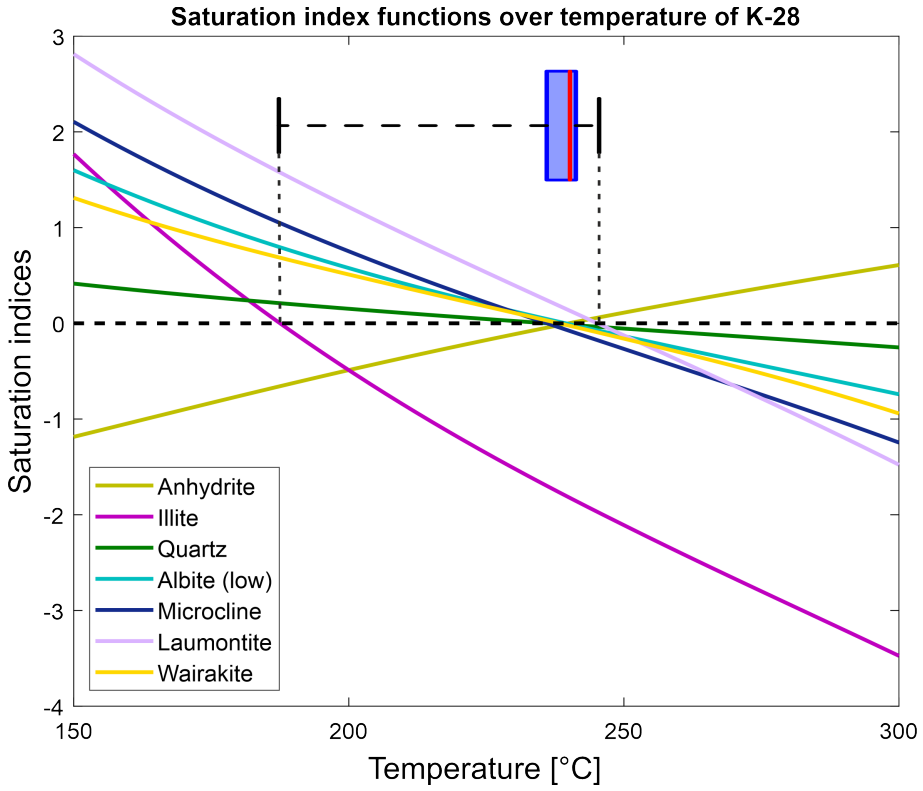


Figure 3.1: Example of the creation of an equilibrium temperature distribution box plot via the saturation indices over the temperature [°C] of the basalt-specific mineral phases for sample K-28. The box plot includes the temperature values of each mineral, where $SI = 0$ (intersection of the saturation index of a mineral phase and the equilibrium line)

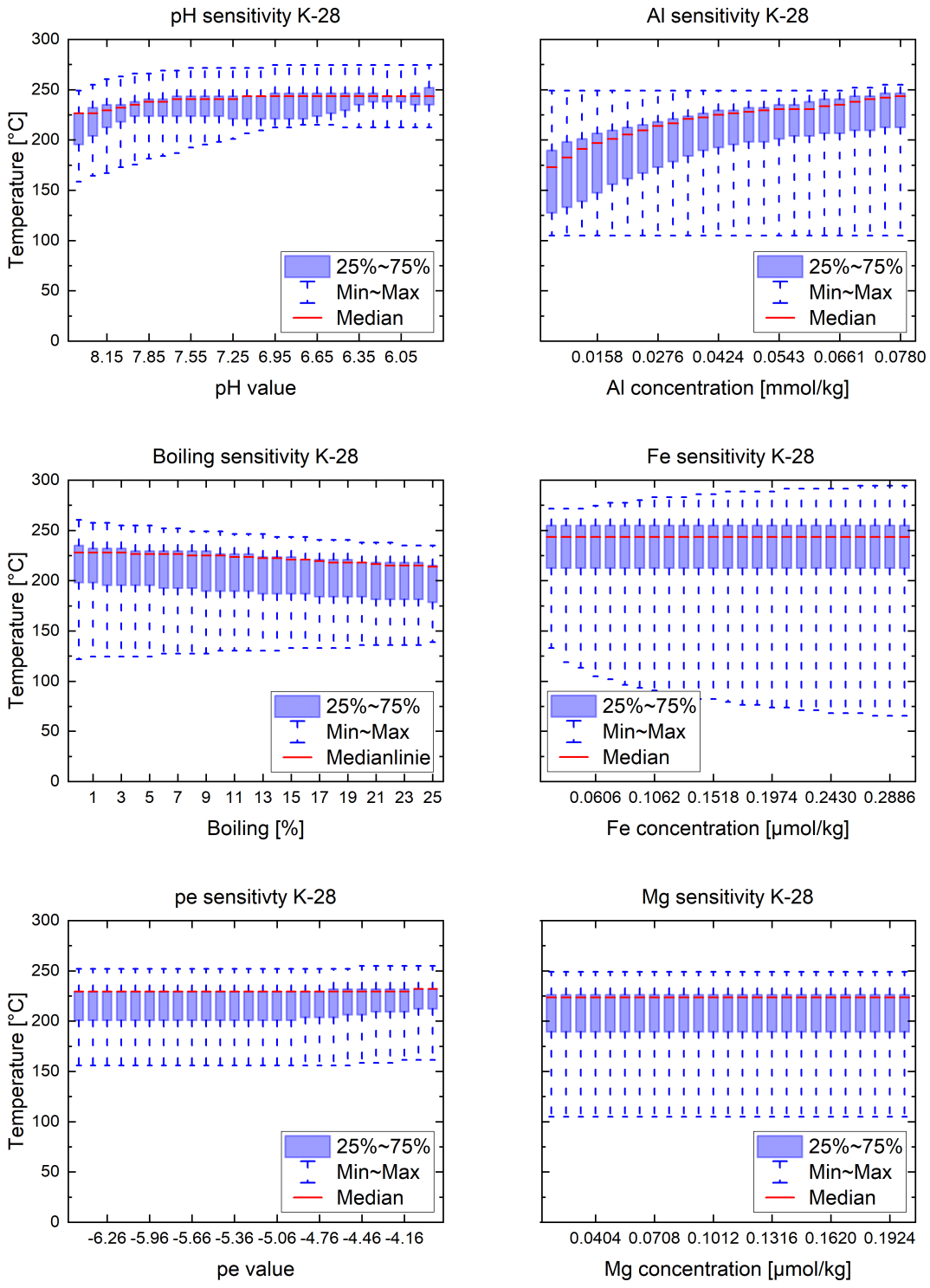


Figure 3.2: Various sensitivity analyses of several parameters (pH value, boiling, pe value, Al, Fe, and Mg concentration)

Secondary effects perturb the chemical equilibrium of a fluid sample while migrating to the earth surface. The chemistry may change due to boiling, degassing, precipitation of phases, dilution, or mixture with shallow and low-mineralized waters as well as reequilibration with the surrounding rocks (Cooper et al., 2013; Fournier, 1977; Fournier and Truesdell, 1974; Pang and Reed, 1998; Reed and Spycher, 1984).
To determine the most vulnerable sensitive parameters for later sequential sensitivity analysis, a series of variations on system parameters (pH, redox, and steam loss) have been computed. Similarly, the concentrations of aluminium, magnesium, and iron, being major components in the minerals but only trace elements in the fluid, have been examined. The equilibrium temperature distributions for K-28 were exemplarily plotted against these parameters (Figure 3.2). It is shown that the most important and vulnerable system parameter is the pH value, which is in good agreement with what is also assumed by Nitschke et al. (2017), as well as Reed and Spycher (1984). Changes have a significant impact on the solubility of mineral phases. The pH value is prone to phase segregation effects like degassing, boiling, and steam loss. Also, the pH is a temperature-dependent function, which decreases when temperature rises. In addition, regarding Figure 3.2, steam loss itself is another vulnerable system parameter. Thus, possible phase segregation due to boiling has to be taken into account. The loss of steam fraction corresponds to a loss of solvent and results in the concentration of the ascending fluid. The effect of steam loss needs to be compensated by adding back the lost water. Equally, the vulnerability of trace elements is shown in Figure 3.2. These constituents are particularly prone to interferences from secondary processes and measurement errors. Simultaneously, they have a high impact on the solubility product and hence, on the saturation index of the majority of reservoir minerals. Clearly, aluminium is the most vulnerable trace element. Its concentration is a crucial parameter when computing the saturation state of aluminosilicates, which represent the major mineral phases in most geothermal reservoirs (e.g. basalts, granitoids, sandstone, greywackes, etc.). Such aluminosilicate mineral assemblages contain phases like feldspars, zeolites, micas, and clay minerals. Due to the tendency of complex formation and precipitation processes (Brown, 2013), the determination of accurate aluminium concentrations is prone to large errors. Furthermore, the variations of the redox potential as well as magnesium and iron concentration were tested. As it is revealed that these parameters have only marginal effects on the temperature estimations, they are not further discussed in this study. In view of the above, the in-situ values of the most vulnerable sensitive parameters, pH, aluminium concentration, and steam loss, have to be reconstructed. For this optimization, a sequential sensitivity analysis for each parameter is used. This sensitivity analysis is executed by the tool. The statistically backed minimization of the temperature spread enables the back-calculation on the in-situ geochemical equilibrium between the geothermal fluid and the reservoir mineral assemblage, which is the basic assumption of the

method. The tool varies these parameters around the initial measured value such that a minimal temperature spread is found. In an ideal case, the equilibrium temperatures of each mineral phase of the reservoir assemblage converge to one discrete reservoir temperature. Also unknown parameters can be estimated in this manner. Thus, a geochemical foreknowledge of the geothermal system is needed to make an educated guess for the unknown parameter, which then can be estimated towards best-fit conditions. To statistically evaluate the resulting box plots, the mineral set has to remain unchanged throughout all variations. Minerals that do not equilibrate due to over- or undersaturation or have multiple intersections with the equilibrium line in any step of the sensitive analyses were discarded from further statistical processing. Changes within the set of consistent mineral phases during the sensitivity analysis would lead to false conclusions because the temperature estimations would then result from different basic conditions (i.e. different mineral sets). Concerning this, the spread of the overlaying boxes and the median differences of each neighbouring plot are matched to identify the most likely value for the sensitive parameter, indicated by the least equilibrium temperature spread. This procedure is done sequentially for the pH value, the aluminium concentration, and the percentage of steam loss. Afterwards, the best-fit values for all parameters are combined in a final temperature estimation. Nevertheless, the aim of this study is the reconstruction of reservoir temperatures, instead of the encompassing reconstruction of geochemical reservoir conditions.

As a generic example, the calculation and optimization will be shown in detail for well K-28 to give an understanding of the procedure. Therefore, the geochemical data (Appendix A.1) of the sample is used. The result of this first calculation is shown in Figure 3.6a. For the temperature estimation from the non-specific mineral set (Appendix A.2) without further optimization, a large temperature spread of 260 °C is obtained. Hence, in this study, a basalt-specific mineral set was devised to enhance the accuracy of this method. The basaltic minerals have been selected according to the mineralogical study of the Krafla reservoir rocks (Arnórsson et al. 1983). This set is extended for secondary mineral phases, occurring in geothermal reservoirs due to hydrothermal alteration processes. It is based on the stability of mineral phases at certain temperature and pressure levels, which were described by Giggenbach (1981). The resulting basalt-specific mineral set (Table 3.2) is used to evaluate the in-situ temperatures of the reservoir. After application of the multicomponent geothermometer based on the selected mineral set, the reservoir temperature estimation could be improved (Figure 3.6b), yet the temperature spread still exceeds 100 °C.

Table 3.2: Mineral phases contained in the basalt-specific mineral set devised and used in this study

mineral group	associated mineral phases
feldspar	albite (low), microcline, K-feldspar
SiO_2 phases	quartz, chalcedony
clays	smectite, clinochlore, illite
carbonate	calcite, aragonite
zeolite	analcime, laumontite, wairakite
sulphate	anhydrite, gypsum
halide	fluorite
soro-/ inosilicate	epidote, anthophyllite, tremolite, pargasite
Fe-phases	pyrite, marcasite, pyrrhotite, goethite

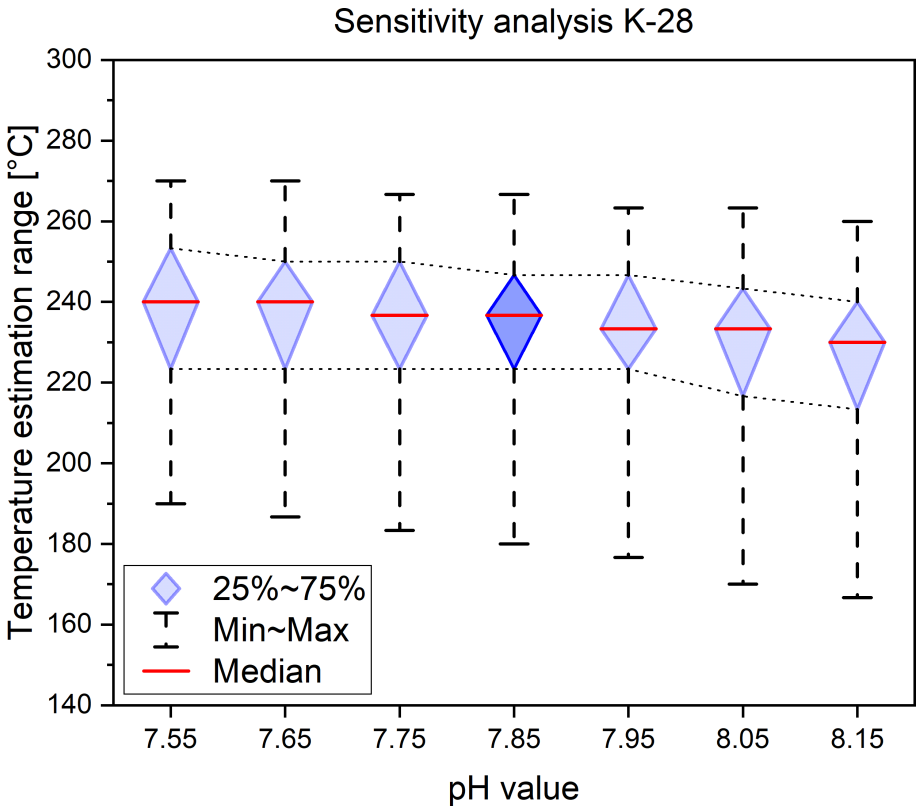


Figure 3.3: Sensitivity analysis of pH for sample K-28. The value was varied from 6.75 to 8.15 in increments of 0.1. The figure shows an extract of the data, where the pH value ranges from 7.55 to 8.15 in increments of 0.1. For pH 7.85, the statistical minimum of the box plot comparison is reached; it is highlighted in a darker blue colour

As a second step, the sensitivity analysis is conducted. Firstly, the pH value is optimized. The initial pH of 9.75 is varied in increments of 0.1 towards higher acidity and basicity. The result is shown in Figure 5.5, where the minimal spread of the box is reached at pH 7.85.

Separately, the aluminium concentration is evaluated. The average aluminium concentration in the Krafla geothermal fluids is about 1.2 ppm (0.04 mmol/L) (Gudmundsson and Arnórsson, 2002). Therefore, the aluminium concentration is varied in increments of 0.006 mmol/L. The initial aluminium concentration of the geothermal fluid composition for K-28 is 0.039 mmol/L. In Figure 5.6, an optimal concentration is also reached at 0.039 mmol/L.

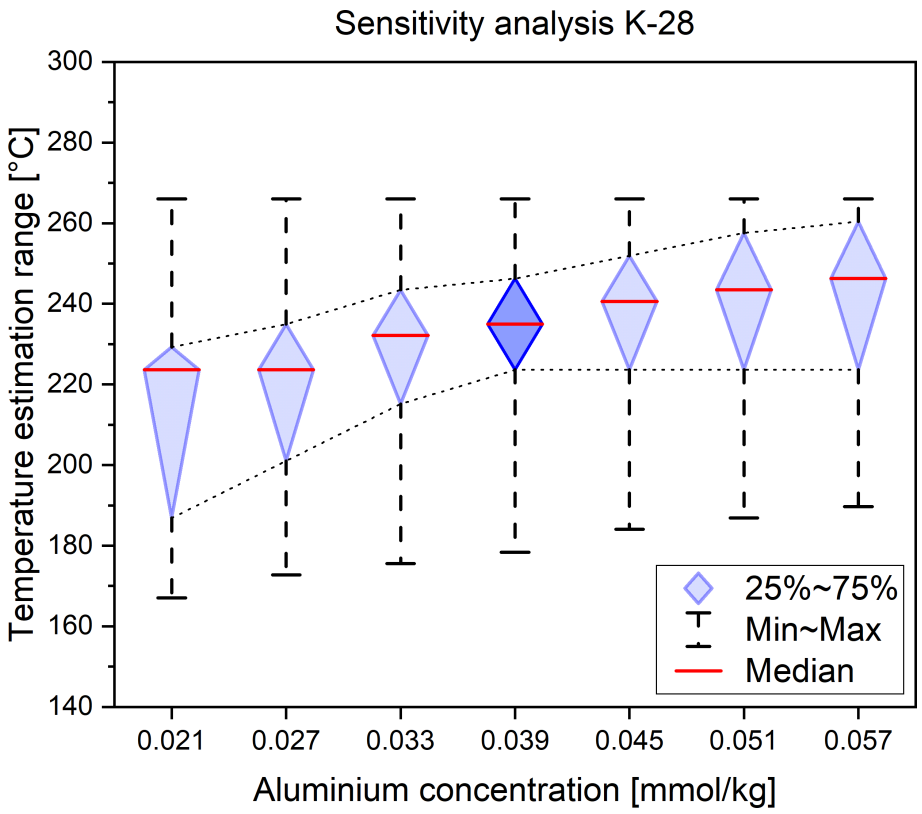


Figure 3.4: Sensitivity analysis of aluminium concentration for sample K-28. The value was varied from 0.009 to 0.117 mmol/L in increments of 0.006 mmol/L. The figure shows an extract of the data, where the aluminium concentration ranges from 0.021 to 0.057 mmol/L in increments of 0.006 mmol/L. The statistical minimum of the box plot comparison is reached for a concentration of 0.039 mmol/L highlighted in a darker blue colour

Lastly, the sensitivity of fluid composition to the magnitude of steam loss is considered. The amount of steam loss is unknown, but has to be back-calculated. Therefore, pure water is virtually added back in increments of 1%. In Figure 3.5, the optimum in steam loss is reached at 14%

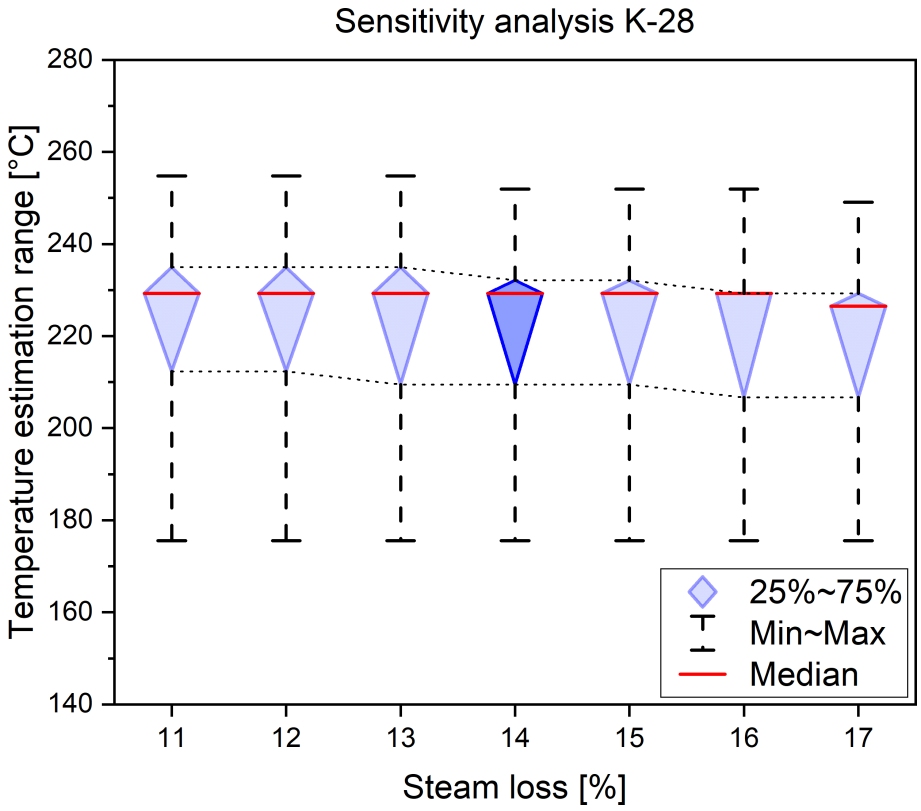


Figure 3.5: Sensitivity analysis of steam loss for sample K-28. The value was varied in increments of 1%. The figure shows an extract of the data, where the steam loss ranges from 11 to 17% in increments of 1%. For 14% steam loss, the statistical minimum of the box plot comparison is reached; it is highlighted in a darker blue colour

The final temperature estimation is then computed by combining the best-fit values for pH, aluminium concentration, and steam loss. Figure 3.6c displays the reduced spread of the calculated temperature after each optimization step. Simultaneously to the reduced uncertainty, the median of the temperature estimate has ascended.

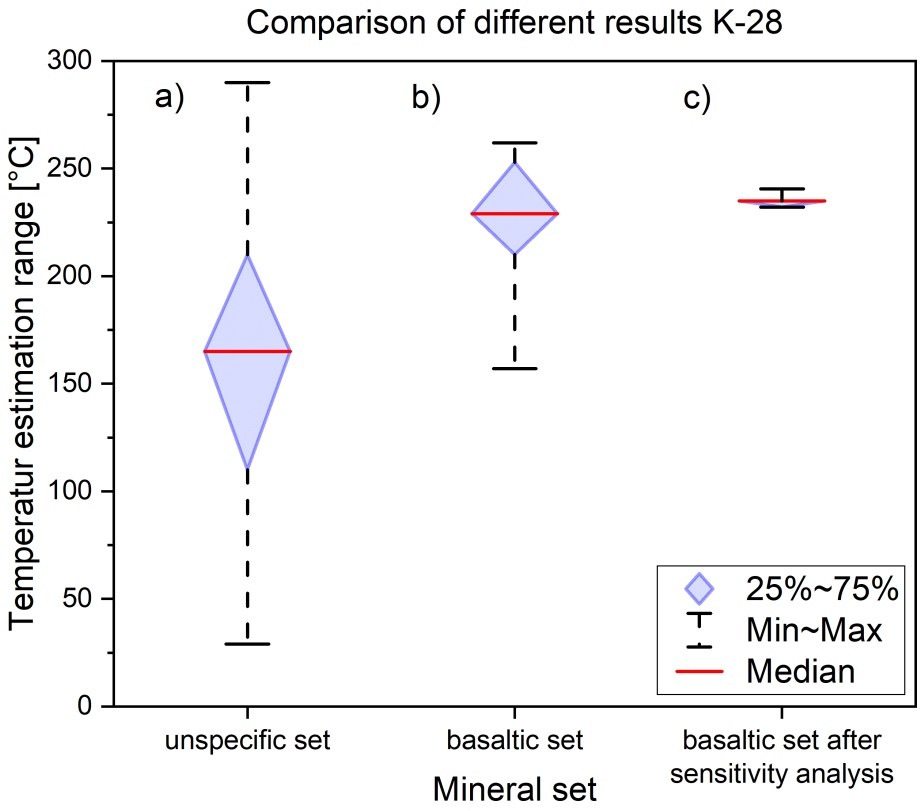
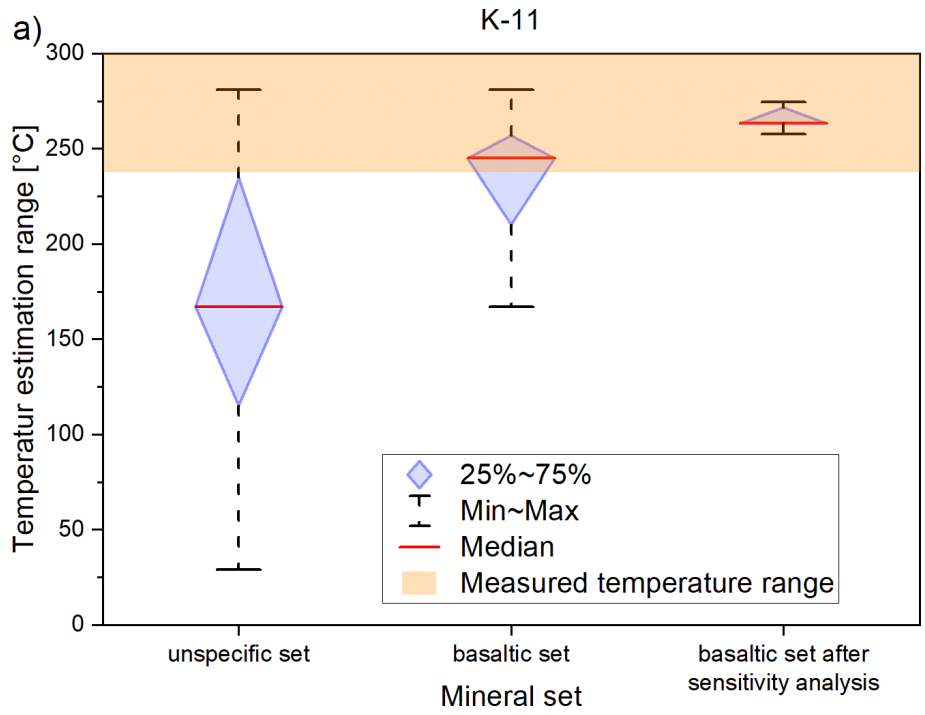
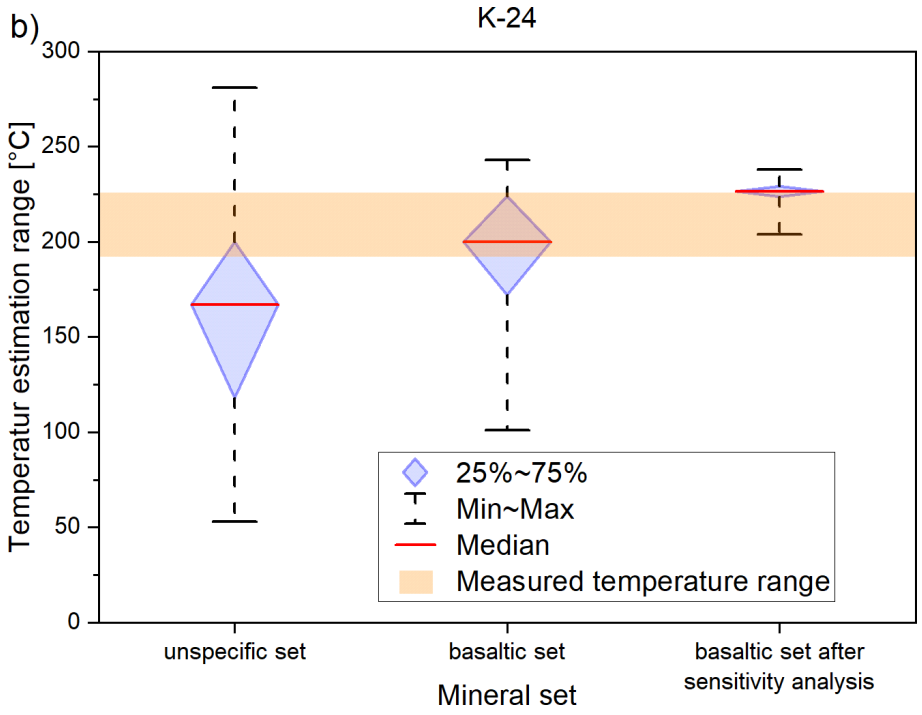
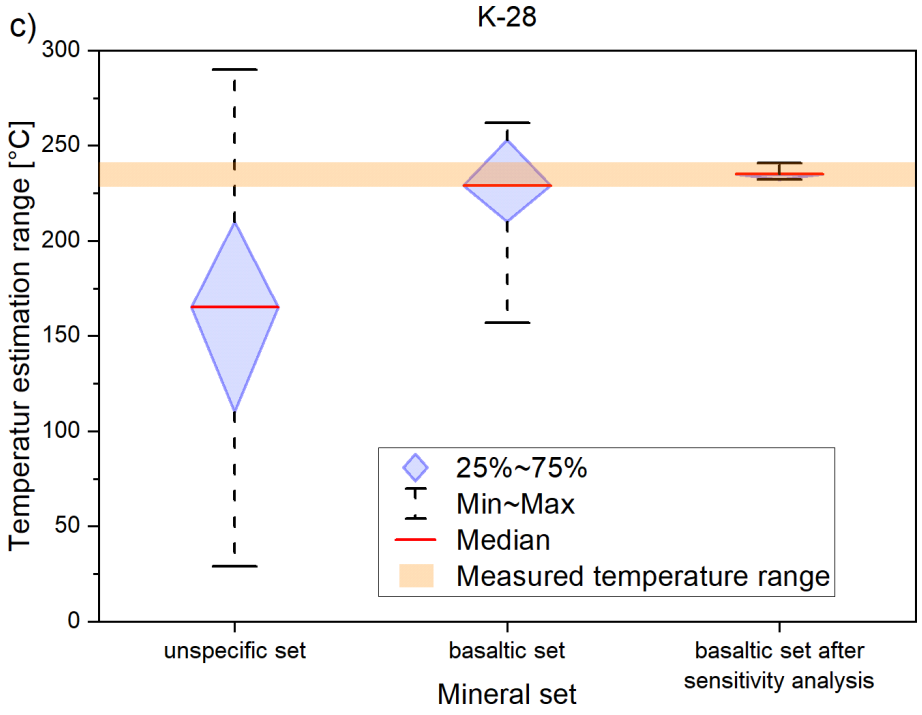


Figure 3.6: Comparison of an unspecific mineral set (a) with the developed basaltic set (b). The box plot in the third column (c) is the result of the combination of the pH, aluminium concentration, and steam loss sensitivity analysis. All analyses are done separately under static conditions for the remaining parameters, and all best-fit parameters were combined afterwards

For validation, the concluding reservoir temperature estimations are compared to downhole temperature measurements published by Gudmundsson and Arnórsson (2002) (Table 3.1). Figure 3.7 displays the temperature box plots for the wells K-11 a), K-24 b), and for K-28 of two consecutive years c) & d). The range of the measured in-situ temperatures (Table 3.1) is figured as an orange box. In each case, the estimated temperatures fit very well the measured borehole temperatures after the optimization procedure. Note that even very small temperature ranges are matched by the estimated temperatures (e.g. K-24 and K-28).




b)
K-24
Temperatur estimation range [°C]
300
250
200
150
100
50
0
25%~75%
Min~Max
Median
Measured temperature range
unspecific set
basaltic set
basaltic set after sensitivity analysis
Mineral set


c)
K-28
Temperatur estimation range [°C]
300
250
200
150
100
50
0
25%~75%
Min~Max
Median
Measured temperature range
unspecific set
basaltic set
basaltic set after sensitivity analysis
Mineral set

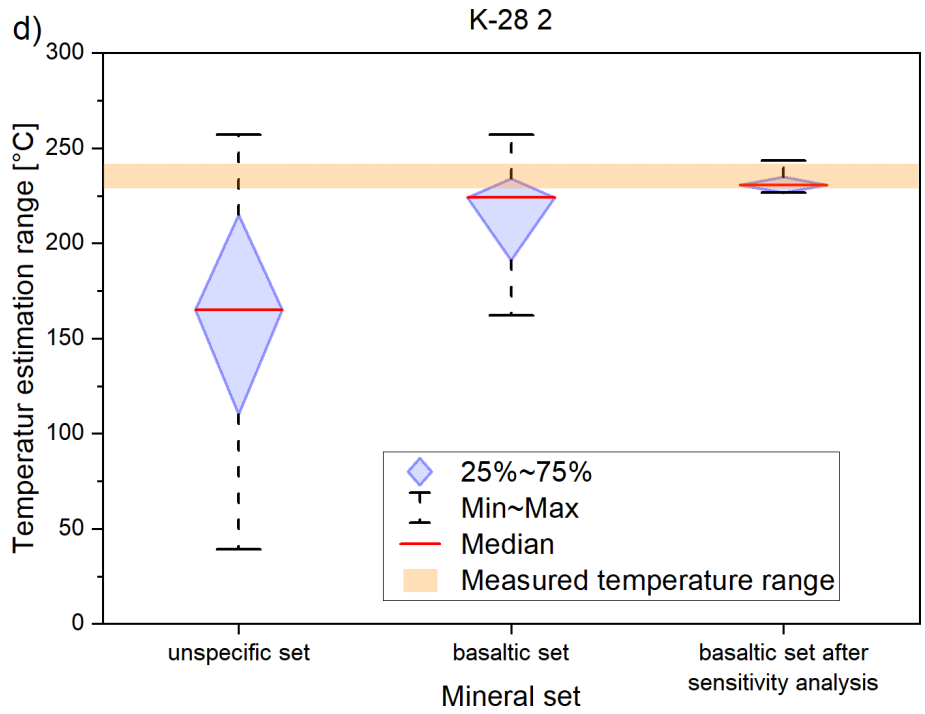


Figure 3.7: Results of wells K-11 a), K-24 b), and K-28 c) & d) for three stages of the analysis. The first column displays a temperature estimation calculated based on an unspecific mineral set. The box plot in the second column represents the specified basaltic mineral set. The last box plot shows the optimized temperature estimation via the pH, aluminium concentration, and steam loss sensitivity analysis. These box plots can be compared with the range of the measured temperatures in the boreholes, given by the orange box

3.4 DISCUSSION

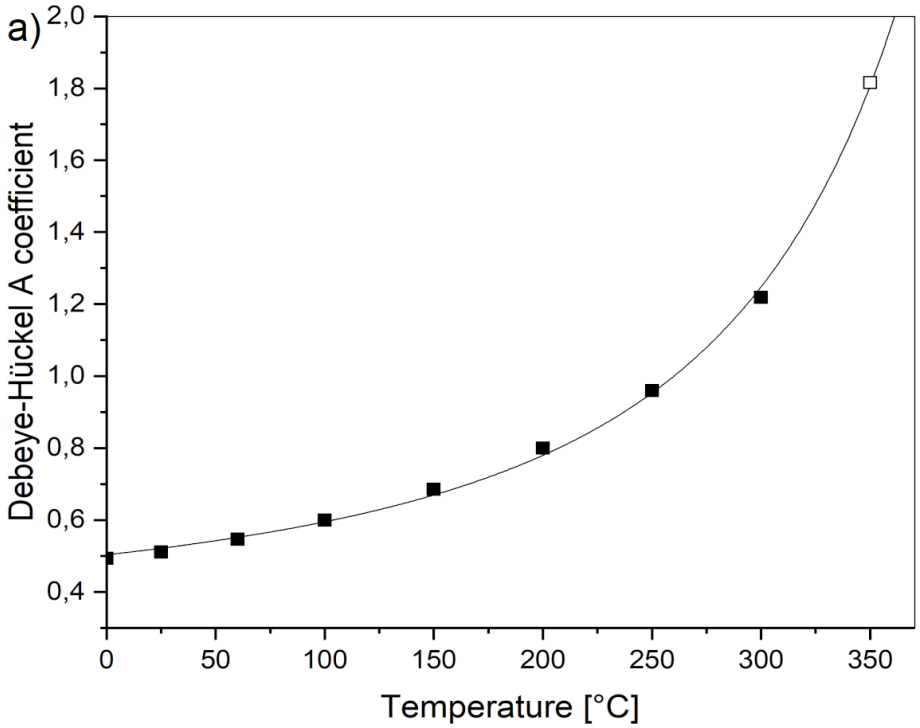
The comparison of the optimized temperature estimation and the measured downhole temperatures confirms the functionality of the application. Only for K-24, the median temperature shows a minor overestimation of 1 K above the highest measured inflow temperature, though, the estimations are also located in the measured temperature range. The overall spread of each final plot after the sensitivity analyses does not exceed 7% (K-24), but is on average 3.7% of the absolute median temperature. The uncertainty throughout the validation is at maximum 2.6% of the original absolute reservoir temperature. Thus, the validated developed tool shows a significant improvement compared to uncertainties of conventional solute approaches, following in the discussion.
The calculation of the saturation indices relies on the LLNL database, which is constrained to temperatures of 300 °C. Therefore, most of the geochemical modelling tools are also constrained for that p–T range. Icelandic geothermal systems have the potential to exceed these temperatures. To evaluate the validity limits, the investigations were extended to the Reykjanes geothermal system,

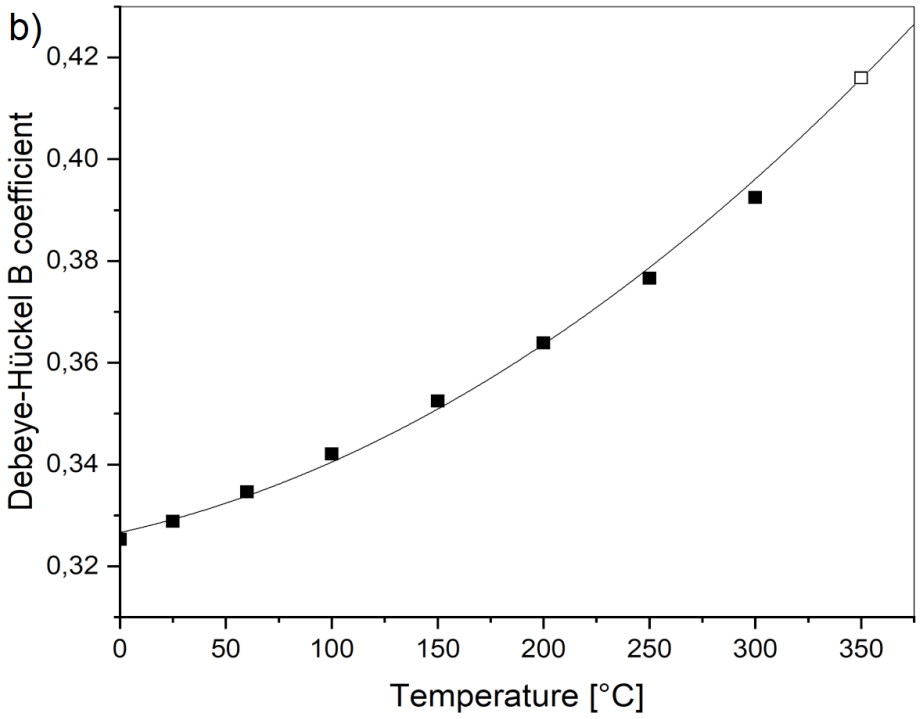
which is located in the SW of Iceland. Furthermore, as the system is recharged by seawater, also the effects of high salinities can be assessed. Reykjanes is also situated in the neo-volcanic zone and the stratigraphy equals the Krafla structure with alternating layers of basaltic lavas and hyaloclastite in the upper part, followed by basaltic intrusives at greater depth. However, since the geothermal system is located on a peninsula, seawater infiltrates the productive horizons of the reservoir. At Krafla, the dissolved solids content is generally up to 1500 ppm. Showing sea water concentrations, the salinities at Reykjanes are up to 58 times higher (RN-23: 87.160 ppm). Óskarsson et al. (2015) published the geochemical data (Appendix A.3) of the fluids from two production wells and the associated temperature logs (Table 3.3). In the following, the applicability of the numerical scheme will be tested at high-enthalpy geothermal fields with temperatures above 300 °C and elevated salinities.

Table 3.3: In-situ measurements of the temperture [°C] in the wells of Reykjanes (RN) at specific depths [m] presented by Óskarsson et al. (2015)

RN-12		RN-23	
1000 m	260 °C	900 m	255 °C
1200 m	270 °C	1200 m	280 °C
1300 m	290 °C	1700 m	300 °C
1700 m	310 °C		

Thus, the Debye–Hückel coefficients A, B, and $\dot{B}$ of Equation 3.2 have to be extrapolated towards higher temperatures and implemented into the database. The coefficients A and B were extrapolated by a quadratic fit, whereas $\dot{B}$ was extrapolated by a cubic fit. This extrapolation is similar to the scheme proposed by Helgeson (1969). Figure 3.8 shows the results of polynomial parameter estimation. The obtained values for the coefficients A and B herein are very close to the results computed by Helgeson et al. (1981). Estimations towards the critical temperature of water have to be used with care and, therefore, only exceed up to 350 °C.


a)
Debeye-Hückel A coefficient
2,0
1,8
1,6
1,4
1,2
1,0
0,8
0,6
0,4
0
50
100
150
200
250
300
350
Temperature [°C]


b)
Debeye-Hückel B coefficient
0,42
0,40
0,38
0,36
0,34
0,32
0
50
100
150
200
250
300
350
Temperature [°C]

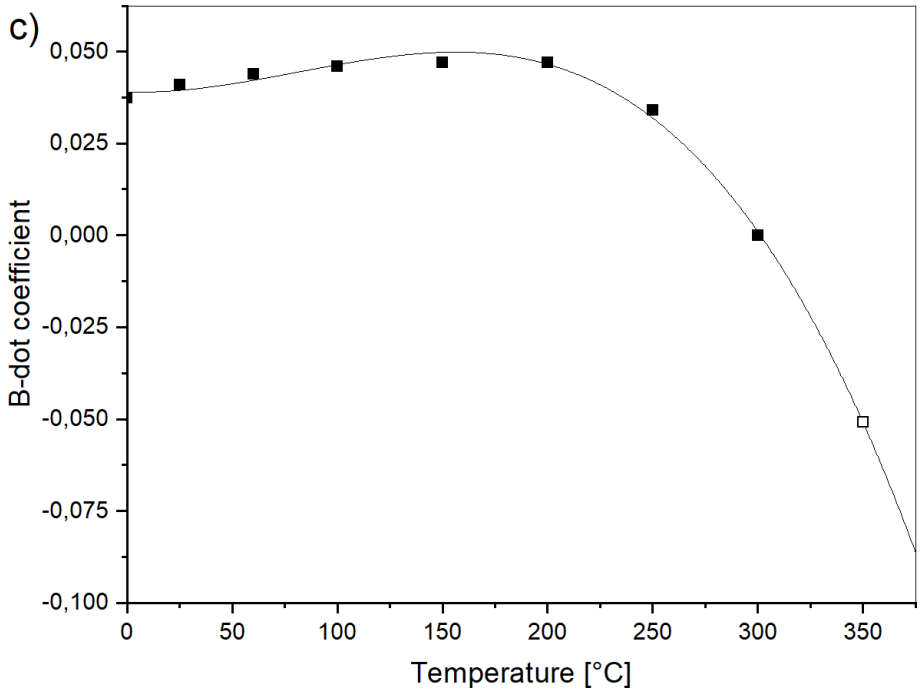


Figure 3.8: Extrapolation of the three parameters A a), B b), and $\dot{B}$ c) of the extended Debye–Hückel equation for temperatures beyond 300 °C. The values of the solid squares are acquired from the LLNL database in IPhreeqc. The hollowed squares are the results of a polynomial fit

These modifications of the Debye–Hückel parameters allow technically for the computation of saturation indices over an extended temperature range. To gain an overview, the saturation indices of the well RN-12 at Reykjanes were plotted to 350 °C (Fig. 3.9). The extrapolated coefficients follow the trend of the saturation curves. The results for the reservoir temperature estimation and the comparison against measured values are presented in Figure 3.10. Despite the high sodium chloride concentrations, the tool operates thoroughly. Analogous to the methodology presented for the Krafla site, the spread of the temperature box plots is minimized and eventually matches the measured temperatures (Table 3.3). The spread is about 4.7% of the measured absolute reservoir temperature, while the overall temperature accuracy is at 0.5%.

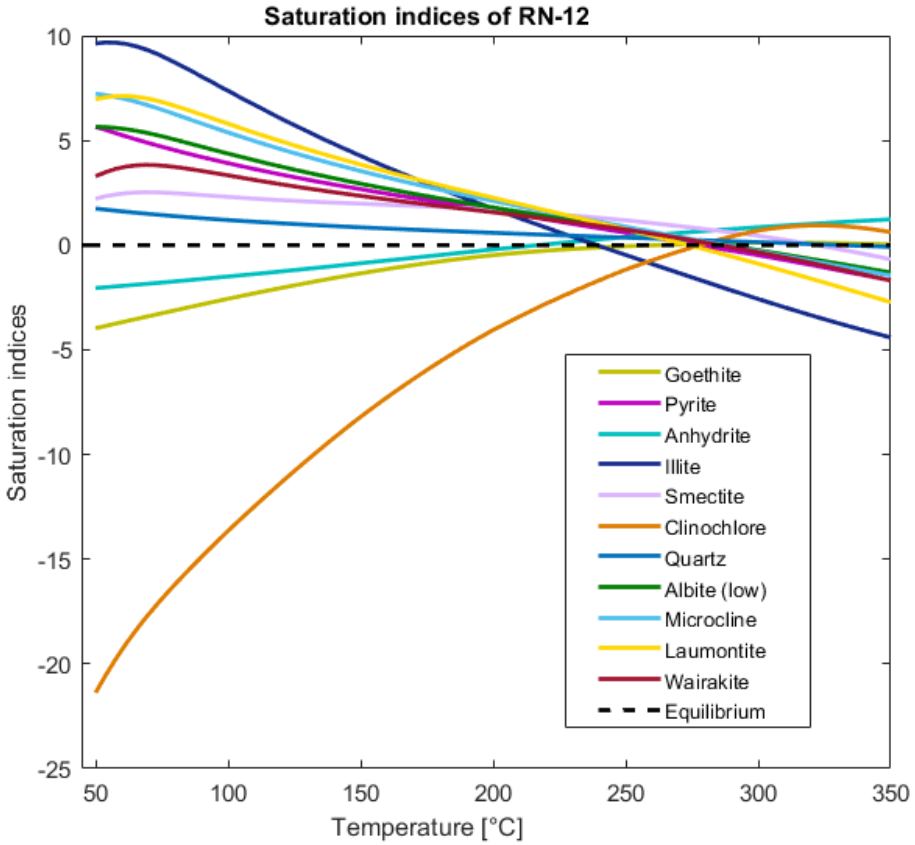


Figure 3.9: Initial saturation indices of the basalt-specific mineral set of RN-12 with a temperature range of up to 350 °C prior to the sensitivity analyses. Saturation curves approaching the critical point of water have to be used with care

To demonstrate the gain of accuracy, the MulT_predict temperature estimations for Krafla and Reykjanes are compared to conventional solute geothermometers. For the comparison, the original data of Krafla and Reykjanes (Appendices A.1, A.3) are corrected for steam loss via WATCH 2.4 (Bjarnason, 2010), requiring additional gas analysis data. Afterwards, the solute geothermometers are applied. The table in Appendix A.4 comprises quartz geothermometers according to Fournier, Potter et al. (1982), Arnórsson et al. (1983), and Verma (2000); Na/K geothermometers according to Truesdell (1976), Fournier (1979), Giggenbach (1988), Arnórsson (2000c), and Can (2002), as well as Na/K/Ca geothermometers according to Fournier and Truesdell (1973), Nieva and Nieva (1987), and Kharaka and Mariner (1989), and K^2/Mg geothermometer according to Giggenbach (1988). All geothermometers were checked for their applicability in these settings. The results of the conventional geothermometers are visualized in Appendix A.1 for Krafla and Appendix A.2 for Reykjanes together with the results of MulT_predict. In all cases, our application targeted the measured temperature more precisely with a lower overall spread of the temperature estimation without requiring additional gas analysis data. Compared to MulT_predict (Krafla: 3.7%; Reykjanes:

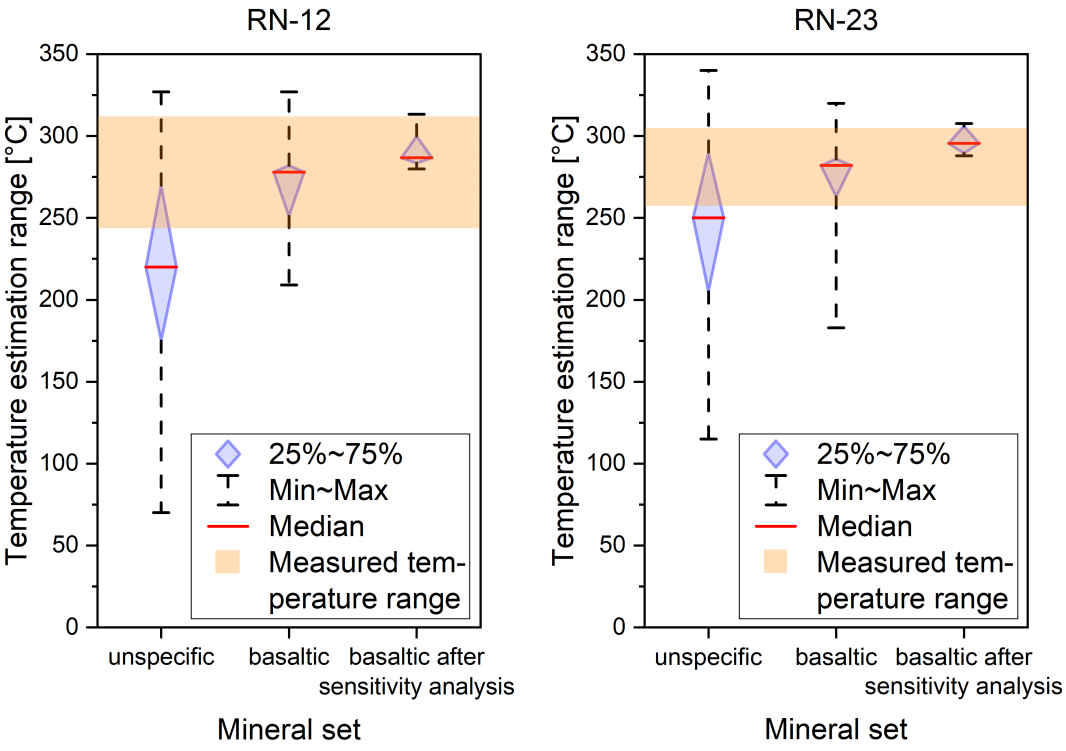


Figure 3.10: Results for wells RN-12 and RN-23 for the three stages of the analysis. The first box plot shows a temperature estimation calculated based on an unspecific mineral set. The second column displays the developed basaltic mineral set. The third box plot shows the minimized temperature estimation after the pH, aluminium concentration, and steam loss sensitivity analysis of the specified basaltic mineral set. These box plots can be compared with the range of the measured temperature in the borehole, given by the orange box

4.7%), the overall temperature spread of conventional solute geothermometers is 10.5% for Krafla and 12.3% for Reykjanes.

3.5 CONCLUSION AND OUTLOOK

This application of multicomponent geothermometry is a promising tool for reservoir temperature estimations. This specific approach comprises a devised basalt-specific mineral set and a subsequent sensitivity analysis based on a standard chemical analysis of the fluid composition without the need for a sophisticated gas analysis. The statistically robust temperature estimations of the reservoir are incorporated in a valuable tool for precise reservoir temperature estimation. Thus, the methodology enhances the usability, the applicability during geothermal exploration as an economically efficient tool for reservoir temperature determination.

The emphasis of this study was the validation of this optimized approach of multicomponent geothermometry. A basalt-specific mineral assemblage was devised to reduce temperature estimation uncertainties. These estimations were further improved by using a subsequent sensitivity analysis via the herein proposed MulT_predict tool. The optimization of pH, aluminium concentration, and steam loss reduces the uncertainty of the temperature estimations significantly. Hence, the back-calculations enable the reconstruction of the in-situ equilibrium temperature conditions between the geothermal fluid and the reservoir mineral assemblage. This equilibrium state corresponds with the underlying geochemical assumptions of geothermometry. The approach presented here would even allow for constraining unknown input parameters. An educated guess of the parameter can be varied to reach the best-fit value. For validation, the calculated reservoir temperatures are compared against measured in-situ reservoir temperatures and classic solute geothermometry. The maximum uncertainty of the temperature estimations is only 2.6% with respect to the in-situ reservoir temperature. The accuracy of the results shows the efficiency and credibility of the method. This multicomponent approach benefits from its statistical robustness due to the conjunction of the saturation indices of multiple mineral phases for temperature estimations. Therefore, it can be applied to diverse geothermal sites with different fluid origins. Furthermore, high-temperature systems can be investigated by extrapolation and modification of the Debye–Hückel parameters in the LLNL database, yet resulting in reservoir temperature estimations with small variances. The method is easy to apply because of the simplicity in terms of input data. A standard water analysis is sufficient for obtaining very accurate results, which facilitates the usability especially at less explored sites, where good-quality data is often missing.
Overall, the validation of the procedure was successful and improved temperature estimations via multicomponent geothermometry. In future, a broader application over different mineral sets is envisaged to expand the usability of the methodology towards other geological settings.

ACKNOWLEDGEMENTS

Thanks to the reviewers and the editor in chief for their constructive suggestions that helped us to improve the manuscript substantially. The study is part of the Helmholtz portfolio project “Geoenergy”. The support from the program “Renewable Energies”, under the topic “Geothermal Energy Systems”, is gratefully acknowledged. We also thank the EnBW Energie Baden-Württemberg AG for supporting geothermal research at KIT.

4

AN INTEGRATED SENSITIVITY ANALYSIS FOR THE BASALT SPECIFIC MULTICOMPONENT GEOTHERMOMETER FOR HIGH TEMPERATURE SETTINGS

This chapter was published in Proceedings World Geothermal Congress 2020+1

ABSTRACT

For a successful geothermal reservoir exploration, an in-situ temperature estimation is essential. Since geothermometric reservoir temperature estimations often entail high uncertainties, statistical approaches are used. The focus is on the application of sensitive analyses on a basalt specific mineral set as multicomponent geothermometer to estimate the reservoir temperatures in Krafla, high-temperature geothermal field, Iceland.

In quantitative geothermometry, the element ratios and mineral saturation of the geothermal fluid serve as single geothermometers. The geochemical equilibrium between mineral phases and the reservoir rock are used to obtain the reservoir temperature. The coupling of several minerals serving as a multicomponent geothermometer allows to get statistically robust temperature estimations. Herein, we set up a specific mineral set for basaltic reservoir rocks, which are calibrated by in-situ measurements of the reservoir temperature in Krafla. The developed method uses IPhreeqc to determine the geochemical equilibrium conditions, followed by the statistical evaluation conducted with a MATLAB-based in-house tool called MulT_predict. The results are presented via box plots. The evaluation of the dataset from Krafla allows the calibration of a basalt specific mineral set for the most accurate reservoir temperature estimation. As surface measurements of pH, aluminium concentration and steam loss do not reflect reservoir conditions, further sensitivity analyses are combined to back calculate these parameters in order to improve the temperature estimation. This statistical evaluation reflects the most plausible reservoir conditions. It is shown that, the variation of the redox potential, iron and magnesium concentration have only negligible effects and thus can be discarded, the correct determination of the in-situ pH, aluminium concentration and steam loss are essential for a robust temperature estimation. The calculated reservoir temperature matches the measured in-situ reservoir temperature with an overall spread of 1.7% of the total measured median temperature. In conclusion, the developed method is a promising tool for the estimation of reservoir temperatures. In addition, it is an

economical exploration tool that allows a high precision temperature estimation. Since the developed basalt specific multicomponent geothermometer also uses secondary mineralization it could be adapted to different geothermal settings, yet requiring further calibration and validation.

4.1 INTRODUCTION

Reservoir temperature estimation is a key technique in successful geothermal reservoir exploration. Quantitative solute geothermometry provides such temperature estimations. The coupling of multiple mineral phases as a multicomponent geothermometer was introduced by Reed and Spycher (1984). They plotted the saturation curves of multiple minerals against temperature. For this purpose, the saturation indices of multiple mineral phases have to be calculated over a certain temperature range (Equation 4.1).

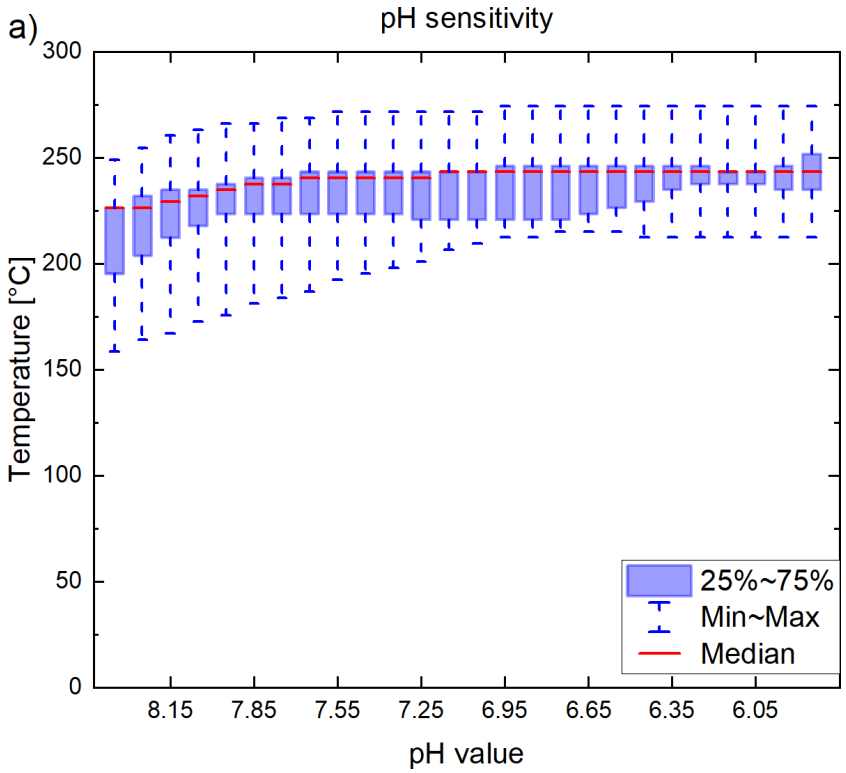
$$SI(T) = log\left(\frac{IAP}{K(T)}\right) \tag{4.1}$$

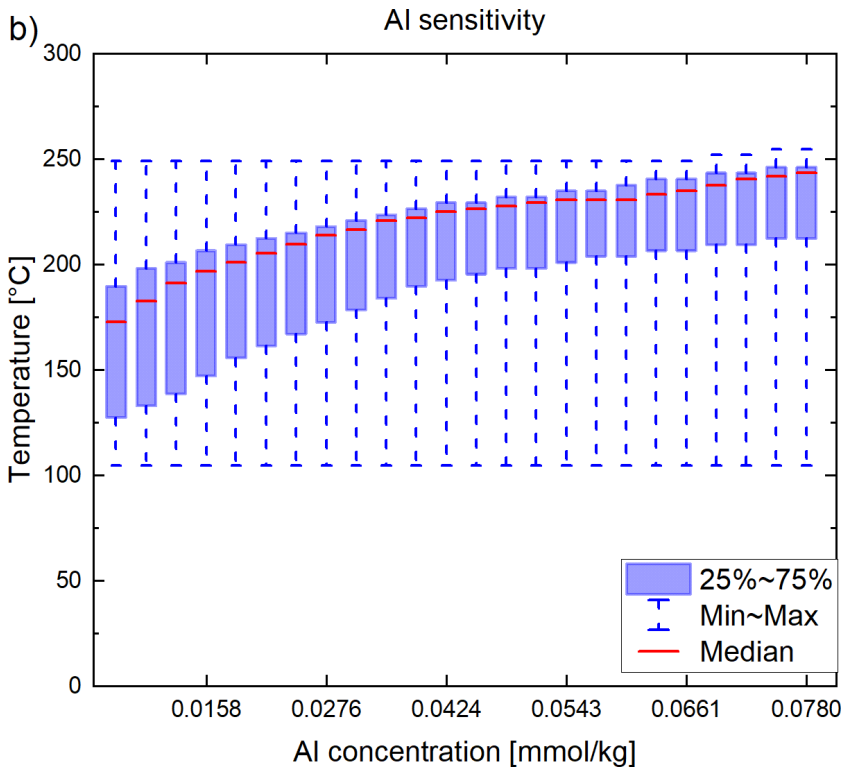
SI is the temperature-dependent saturation index of a mineral phase, IAP is the ion activity product and K the temperature-dependent thermodynamic equilibrium constant of one mineral phase. Only in the case, SI equals zero the equilibrium state between the reservoir fluid and the mineral phase is reached and can be evaluated. Thus, temperatures at which mineral phases reach equilibrium can be used for reservoir temperature estimation. Nevertheless, this result is prone to uncertainties. Changes in the chemistry of the fluid can lead to divergences of the initial equilibrium conditions in the reservoir. While the fluid ascends to the earth surface, it is vulnerable to secondary processes such as boiling, phase segregation, mixing or dilution and precipitation (Arnórsson et al., 1990; Cooper et al., 2013; Peiffer et al., 2014; Spycher et al., 2014). Therefore, additional sensitivity analyses can be performed to reconstruct reservoir conditions. The aim is to vary multiple sensitive parameters around their initial value to find a minimum in the calculated temperature estimation spread At this point, the equilibrium state $(SI = 0)$ of the mineral phases have the smallest distances between each other and thus, the minimum in the sensitivity field is reached. Phases, which do not reach the equilibrium are excluded from the calculations.

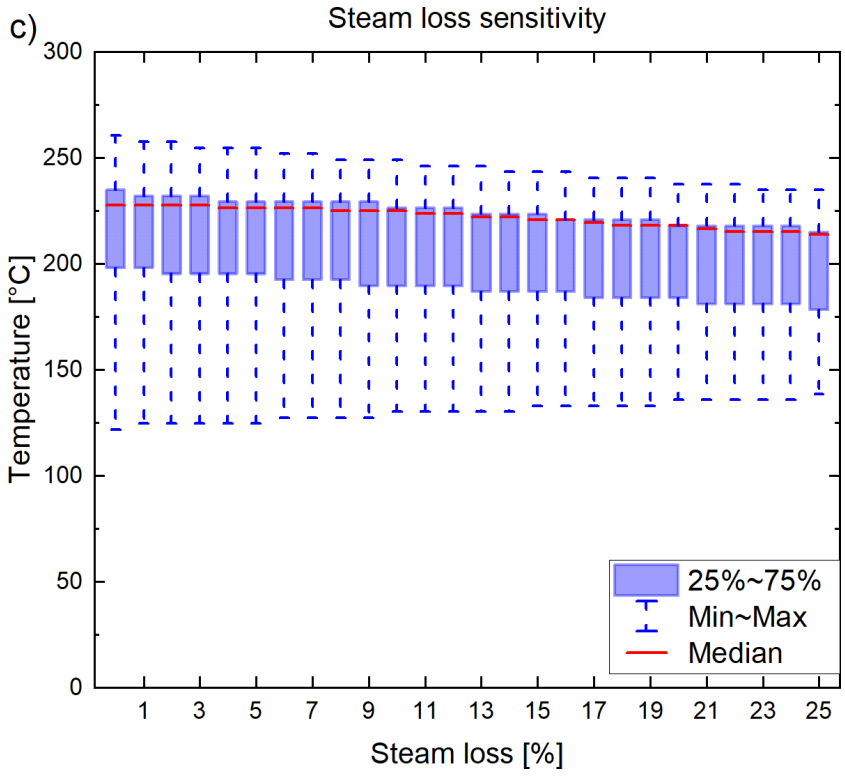
4.2 METHOD AND RESULTS

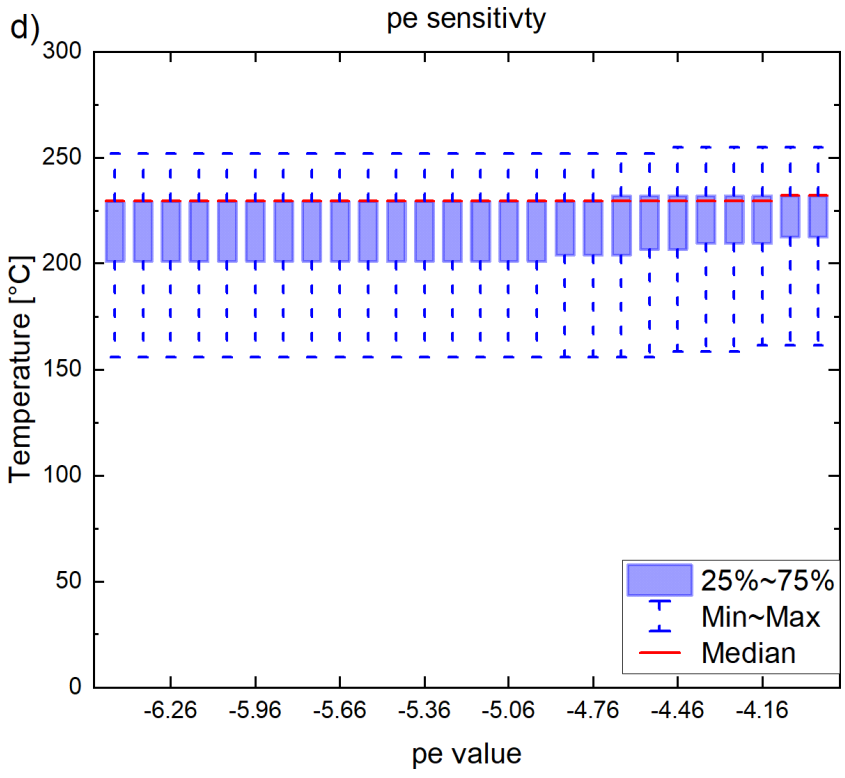
With the help of the developed basalt specific mineral set, it is possible to calculate robust temperature estimations. The set contains mineral phases found in the basaltic geology of Krafla, which is described by Arnórsson et al. (1983). These

mineral phases are extended by secondary minerals, which generally occur in geothermal systems (Giggenbach, 1981). The developed basalt specific mineral set is listed in Appendix A.5. For the sensitivity analyses, the used mineral phases have to be stable throughout the variation. Hence, the saturations indices of a mineral phase have to be calculates throughout a temperature range of 20 to 300 °C while each sensitive parameter is variated around its initial value. Static boundary conditions like the consistency of mineral phases are essential for the latter statistical evaluation. In the sensitivity analyses, the minimum in the temperature distribution of each parameter is determined. The variation of the sensitive parameter has to be chosen large enough to not only find a local minimum, but instead to find the global minimum. The integration of the sensitivity analyses considers the geochemical dependence of these sensitive parameters. Giggenbach (1981, 1988) proposed that changes in the partial pressure of CO_2 and H_2S results in pH changes and buffer reactions in the fluid. This degassing effect occurs due to temperature and pressure changes while ascent of the fluid and the sampling process. In addition, the pH value is temperature-dependent which also shifts the value before analysing in the laboratory. Furthermore, steam loss controls the pH value. This is taken into account by performing a sensitivity analysis for steam loss as well as pressure relief causes boiling. Furthermore, the pH value is vulnerable to measuring errors. Nitschke et al. (2017) and Reed and Spycher (1984) are aware of these problems and suggest to measure the pH value directly in the field and later on in the laboratory to reconstruct the pH value at reservoir conditions. Likewise, Giggenbach (1981) used the thermodynamic stability of aluminosilicates for reservoir temperature estimation. In conclusion, pH shifts can lead to changes in the aluminium concentration. Aluminosilicates have a strong tendency to precipitate in several phases. These buffer reactions in the fluid can form aluminium complexes, which are likely to precipitate. Therefore, complexes get lost while the ascent of the fluid. Also, the aluminium concentration is vulnerable to fluid sampling. While sampling, these complexes are filtered because their size exceeds the usually used 0.2 µm filter membrane. Likewise, the measured aluminium concentrations in the fluid are close to the detection limit, which leads to additional uncertainties. Hence, minimal changes in the concentration cause large effects of increasing or decreasing of the solubility product and thus, to huge uncertainties in temperature estimations. Furthermore, the redox potential is coupled to the pH value. The lower boundary of the redox potential in geothermal fluids is defined by the standard potential of hydrogen, consequently the pH value. The upper limit is given by the strongest naturally occurring oxidant which is oxygen. The iron-bearing minerals of the basalt specific mineral set (e.g. hematite, goethite, pyrite) are redox state-dependent. Also, the measured iron and magnesium concentrations are close to the detection limit. The variations of these sensitive parameters is visualised in Figure 4.1. The Figure 4.1 comprises the pH value a), aluminium concentration b), steam loss c), redox potential d) as well as iron e) and magnesium concentration f).


a)
pH sensitivity
Temperature [°C]
300
250
200
150
100
50
0
8.15
7.85
7.55
7.25
6.95
6.65
6.35
6.05
pH value
25%~75%
Min~Max
Median


b)
Al sensitivity
Temperature [°C]
300
250
200
150
100
50
0
0.0158
0.0276
0.0424
0.0543
0.0661
0.0780
Al concentration [mmol/kg]
25%~75%
Min~Max
Median


c)
Steam loss sensitivity
Temperature [°C]
300
250
200
150
100
50
0
1
3
5
7
9
11
13
15
17
19
21
23
25
Steam loss [%]
25%~75%
Min~Max
Median


d)
pe sensitivty
Temperature [°C]
300
250
200
150
100
50
0
-6.26
-5.96
-5.66
-5.36
-5.06
-4.76
-4.46
-4.16
pe value
25%~75%
Min~Max
Median

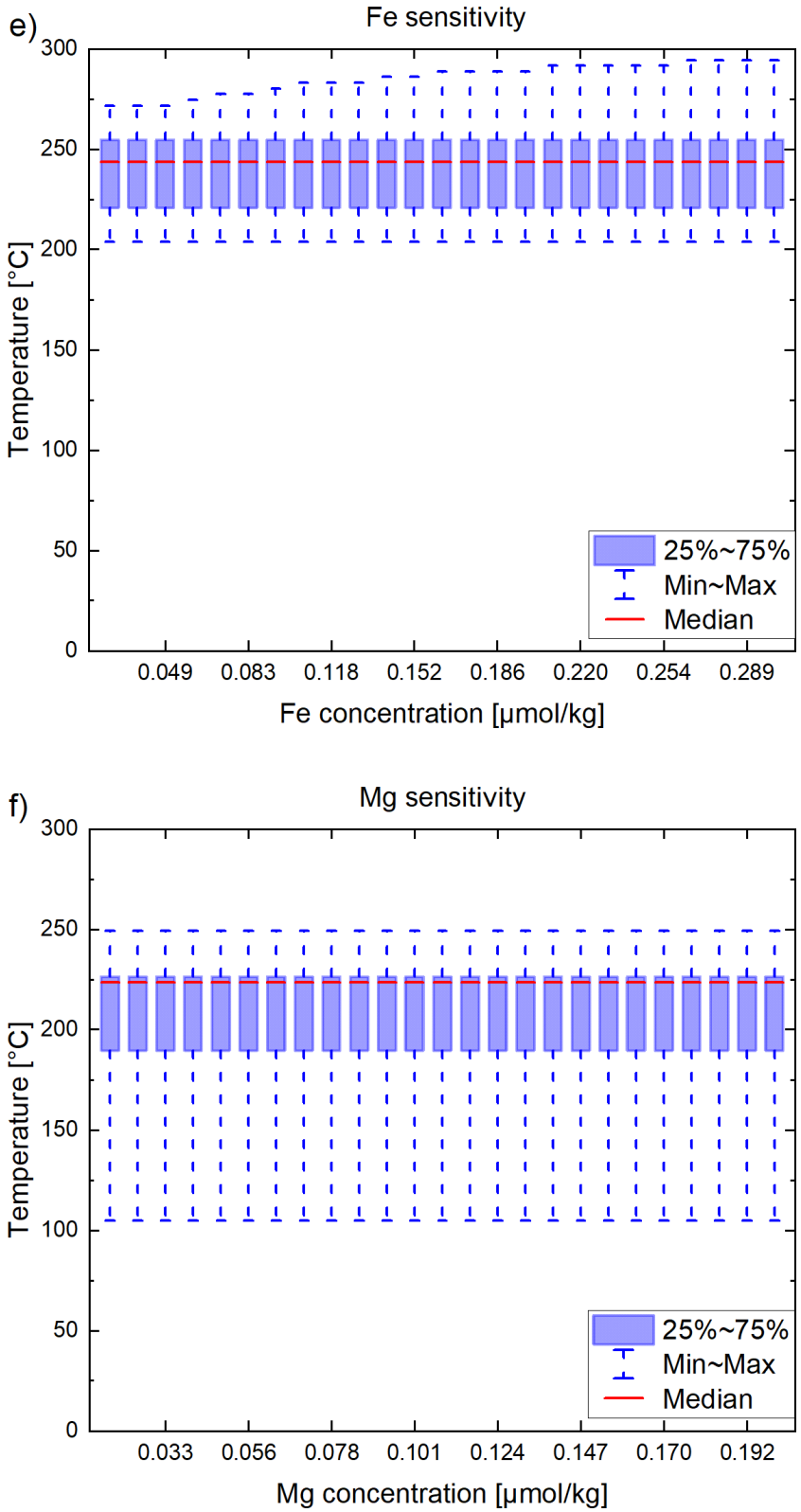


Figure 4.1: Visualisation of multiple sensitivity analyses for different sensitive parameters: pH a), aluminium concentration b), steam loss c), redox potential d), iron e) and magnesia concentration f) for the well K-28, Krafla (Iceland).

Especially the pH value, the aluminium concertation and the steam loss are vulnerable to changes. Whereas variations of the redox potential, the iron and magnesium concentration have negligible sensitivity on the equilibrium temperature distribution of the mineral set. Due to the marginal effects, the sensitivity analysis of the redox potential, as well as for the iron and magnesium concentration can be omitted for future temperature estimation. Therefore, pH, aluminium concentration and steam loss are the most sensitive parameters, which should be analysed to obtain the most realistic reservoir conditions. Due to the major amount of aluminosilicates as multicomponent geothermometer in the basalt specific mineral set (Appendix A.5), these three sensitivity analyses have to be combined and variated interdependently. The integrated sensitivity analysis of pH, aluminium concentration and steam loss is implemented in an in-house tool called MulT_predict (Ystroem et al., 2020). The MATLAB-based (MATLAB, 2019) software uses IPhreeqc (Parkhurst, Appelo et al., 2013) to calculate the saturation indices of the mineral phases of the basalt specific mineral set (Appendix A.5). Each sensitive parameter is variated over all variation steps of the other parameters. Hence, a three dimensional cell-array is calculated. An entry of the array contains the equilibrium temperatures of all consistent mineral phases. These temperature estimations were statistically evaluated to find the minimal distance within the temperature spread. Figure 4.2 shows the differences of this temperature spread for a static steam loss amount over the variation of aluminium concentration and pH. The morphology of the temperature distribution reflects one layer of the three dimensional cell-array. At the global minimum the temperature spread of the mineral phases is smallest. This point resembles the reservoir conditions for all sensitive parameters. Exemplarily, for the well K-28 (data by Gudmundsson and Arnórsson (2002)), the minimum is reached at an aluminium concentration of 0.079 mmol/L and a pH value of 7.95. The selected layer was calculated for a steam loss of 14%. Compared to the initial values of the geochemical analysis (Al concentration: 0.033 mmol/L, pH: 9.55) the parameters markedly shifted.

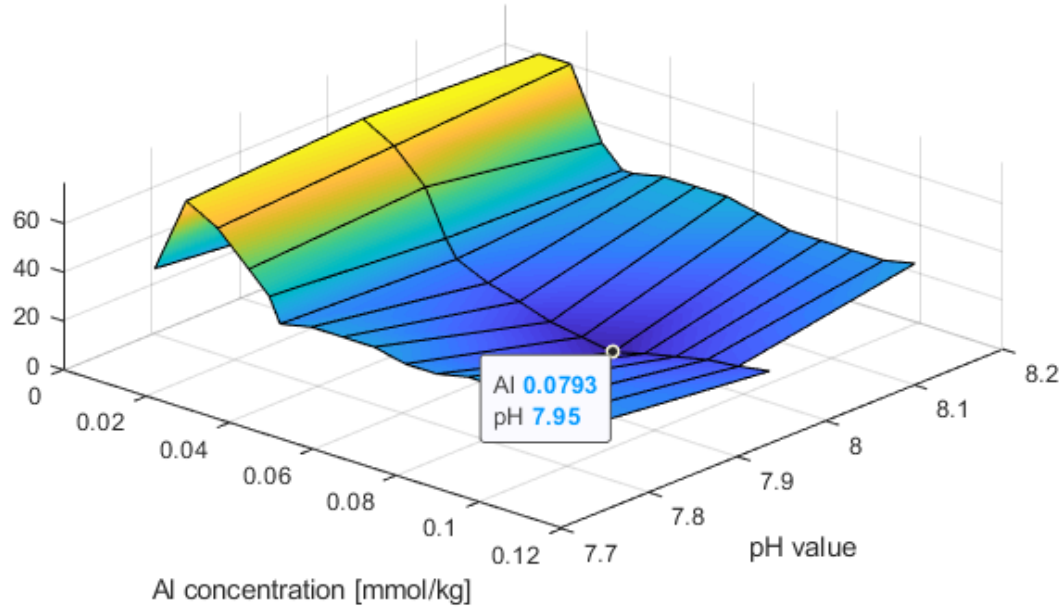


Figure 4.2: Sensitivity field for the well K-28, Krafla (Iceland). The field is spanned between the aluminium concentration and the pH value. The layer was calculated for a steam loss of 14%. The minimum resembles the reservoir conditions at pH 7.95 and an aluminuim concentration of 0.079 mmol/L.

4.3 DISCUSSION

Consequently, the calculated temperature estimation can be compared to the measured in-situ temperatures, exemplarily of the well K-28 in Krafla. Gudmundsson and Arnórsson (2002) identified two permeable horizons, which feed the borehole. These aquifers are located at a depth of 500 m and 800 m beneath ground level. Their measured inflow temperature is 230 °C respectively 240 °C. Figure 4.3 visualises the improvement of the reservoir temperature estimations by performing the integrated sensitivity analysis. The first box plot shows the temperature estimation by only using the basalt specific mineral set (Appendix A.5). After the application of the integrated sensitivity analysis the temperature estimations box plot gains a tremendous amount of precision. The overall spread of the box plot minimizes to 1.7% of the absolute median temperature estimation. The orange bar corresponds the measured in-situ temperature range. The optimized temperature estimation fits this range precisely.

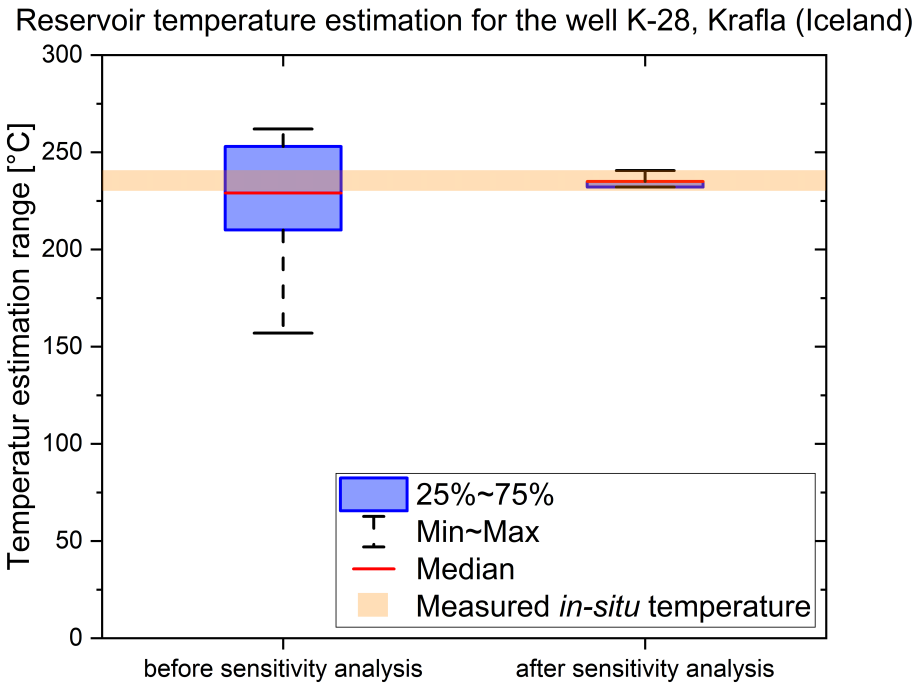


Figure 4.3: Reservoir temperature estimations for the well K-28 using the basalt specific mineral set (Appendix A.5). The first column shows the temperature estimation before applying the integrated sensitivity analysis (pH, aluminium concentration and steam loss). The second column visualises the temperature estimation after performing the integrated sensitivity analysis. The orange bar indicates the measured in-situ temperatures given by Gudmundsson and Arnórsson (2002).

In conclusion, the combination of multiple sensitivity analyses (pH, aluminium concentration and steam loss) results in a precise and robust application. By variating sensitive parameters around their initial value, reservoir conditions can be reconstructed. For this purpose, the in-house tool MulT_predict (Ystroem et al., 2020) statistically evaluates the calculated temperature estimations and finds the global minimum in temperature spread. Thus, the integrated sensitivity analysis markedly improves the temperature estimations of the basalt specific mineral set.

5

MULT_PREDICT - AN OPTIMISED COMPREHENSIVE MULTICOMPONENT GEOTHERMOMETER

This chapter was published in Geothermics (2022) Volume 105, 102548
DOI: 10.1016/j.geothermics.2022.102548

ABSTRACT

In this study, we introduce MulT_predict as a fully integrated solute multicomponent geothermometer, combining numerical optimisation processes for sensitive parameters to back-calculate to chemical reservoir conditions. This results in a state of the art geothermometer, providing an accurate reservoir temperature estimation validated by geothermal borehole measurements on a worldwide scale. In addition, a universally valid mineral assemblage for an unknown reservoir composition is developed, focusing on worldwide applicability. Using the evolved methodology, the limits of the optimisation processes are determined by using a synthetic brine (150 °C, pH 6, aluminium concentration 0.003 mmol/L) and successively perturbing its geochemical equilibrium state. Individual back-calculation of reservoir conditions lead to valid temperature estimations of 145 °C, 3.4% lower than the initial temperature while a simultaneous and interdependent optimisation reconstructs the sensitive parameters even more precisely with a deviation of 0.056 for the initial pH value, and 0.164 µmol/L for the aluminium concentration.

5.1 INTRODUCTION

Evaluating the quantity of energy in a reservoir is essential for an economical geothermal energy production. Therefore, the accurate determination of the reservoir temperature is one key factor in geothermal exploration besides flow rate. Chemical geothermometry is a common technique to predict the reservoir temperatures from the geochemical composition of geothermal fluids. Temperature-dependent cation ratios, as well as the saturation state of mineral phases, can be the basis for temperature estimation, assuming an in-situ chemical equilibrium of the reservoir rock and fluid (Ellis and Mahon, 1964). As a result, the first conventional solute geothermometers were introduced around 60 years ago by Fournier and Rowe (1966). In an ongoing development, new solute geothermometers were continuously developed but led to high uncertainties due to changes in fluid chemistry regarding different geothermal locations (Nitschke

et al., 2018). Reed and Spycher (1984) plotted saturation indices of mineral phases calculated from thermodynamic data against temperature. The resulting saturation curves illustrate the temperature-dependent solubility of the reservoir mineral assemblage in the geothermal fluid. Spycher et al. (2014) took up this methodology and presented a multicomponent geothermometer, which evaluates the geochemical equilibria of mineral phases and fluid for reservoir temperature estimation. Nitschke et al. (2017) analysed temperature predictions from conventional and multicomponent geothermometers, showing a large reduction in uncertainties for this modelling approach. Thus, multiple mineral phases lead to statistically more robust temperature estimations. Therefore, Spycher et al. (2011) introduced the multicomponent geothermometer GeoT, followed by RTEst by Palmer (2014). Also, software for the chemical reconstruction of downhole and reservoir conditions like WATCH (Bjarnason, 2010) improved the results of solute geothermometry. GeoT uses Newton-Raphson iteration to calculate the geochemical equilibria of mineral phases from external thermodynamic databases (Sonnenthal et al., 2013). To compute temperature estimations, a complete fluid analysis is necessary. The program allows reconstructing the deep fluid composition (pH value, gas loss, dilution, and mixing) needing additional gas chemistry as well as end-member solutions. Recently, a python script was developed to search for the most suitable mineral assemblage for reservoir temperature estimation (Olguín-Martínez et al., 2022). RTEst uses the weighted sum of squares of the saturation indices of mineral phases to calculate the reservoir temperature estimation optimising the calculation by CO_2 fugacity and the mass of water (Palmer, 2014). Developing MulT_predict, Ystroem et al. (2020) combined the multicomponent temperature estimation with a numerical reconstruction of the reservoir conditions. The scope of MulT_predict is to reconstruct the parameters, which control mineral solubility and are prone to secondary changes. Hence, sensitive parameters such as the aluminium concentration and pH value are back-calculated to reconstruct in-situ conditions (Ystroem et al., 2021). In addition, effects from secondary processes such as dilution and boiling are corrected. These parameters are prone to geochemical changes of the fluid during the ascent to the surface or while sampling (Giggenbach, 1981, 1988). The major advantage of MulT_predict, unlike other solute multicomponent geothermometers, is that a standard fluid analysis is sufficient and a sophisticated gas analysis is not required. In addition, all optimisation processes (pH value, aluminium concentration, boiling, and dilution) are calculated simultaneously and interdependently. Previously, MulT_predict was developed, applied, and validated for the basaltic setting of Krafla and Reykjanes fields in Iceland (Ystroem et al., 2020). The emphasis of this study is the development of a comprehensive multicomponent geothermometer for field exploration. At such an early stage, little reservoir knowledge and no borehole data is available. The reservoir rock mineralogy as the basis of multicomponent geothermometrical temperature estimation is typically lacking. To facilitate the application of MulT_predict, as

an early-stage exploration tool, for a broad range of geothermal sites, geological settings, and play types, the focus of the study is the development of a universally valid mineral set. Based on this, a general procedure of statistical outlier removal is introduced to acquire more precise temperature estimations and simplify the practicality of MulT_predict.

First, the applicability of the tool is transferred to a broad range of geothermal settings around the world including saline crystalline basements, marine, and continental basin facies as well as volcanic rock. In a first step, individual mineral sets are evolved for each geothermal setting. Afterwards, mineralogical coherences within the mineral sets are combined in a universally valid mineral set for general applicability. The subsequent outlier removal increases the precision of the temperature estimation. The results of the validation are shown for examples of boreholes. Furthermore, the characteristics of the site-specific brine are discussed to identify clustering and similarities concerning geochemical key parameters, providing insights into why a universally valid mineral set provides precise temperature predictions and where the limits of its application are. In the last part, MulT_predict's optimisation processes are benchmarked and discussed. Therefore, a predefined synthetic brine, initially in full equilibrium, is increasingly perturbed stepwise. This perturbation is then reconstructed using two different approaches. Firstly, each optimisation process is evaluated individually regarding the back-calculation of each sensitive parameter (boiling/dilution, pH value, and aluminium concentration). Secondly, the performance of the interdependent optimisation processes is evaluated by reconstructing all sensitive parameters together. Furthermore, the results of both reconstructions of reservoir conditions are discussed.

5.2 METHODS

In solute multicomponent geothermometry, the base assumption is the geochemical equilibrium between the geothermal fluid and the hosted reservoir rock (Fournier and Truesdell, 1974). Thus, in the temperature-dependent reaction mineral phases are dissolved until an equilibrium within the fluid is reached. The geochemical equilibrium of each mineral phase is attained when the determined ion activity product IAP of the fluid equals the temperature-dependent thermodynamic constant $K(T)$ (Equation 5.1). As a result, the temperature-dependent saturation index $SI(T)$ of a mineral phase is zero.

$$SI(T) = log\left(\frac{IAP}{K(T)}\right) \tag{5.1}$$

For reservoir temperature estimation, the saturation indices of several mineral phases are plotted against the temperature. The resulting equilibrium temperatures (at $SI = 0$) serve as a part of the geothermometer. Basing the temperature estimation on the SIs for several minerals has proven to be a robust geothermometric tool. As input parameters for MulT_predict, a standard geochemical fluid analysis covering the major cations and anions, the pH value, as well as silica and aluminium concentrations are needed to account for the chemical elements being the components of the reservoir rock mineral assemblage. Nevertheless, the geochemical equilibrium is still prone to uncertainties when the chemistry of the fluid changes. Divergences from the chemical equilibrium state are due to a) an immature fluid has not yet reached equilibrium or b) secondary processes while the fluid ascends to the surface or while sampling. Such processes refer to phase segregation, boiling, mixing, or dilution, as well as complex building and precipitation of mineral phases (Arnórsson et al., 1990; Cooper et al., 2013; Nitschke, 2018; Peiffer et al., 2014; Spycher et al., 2014). In a previous study, numerical optimisation processes are introduced to MulT_predict correcting sensitive parameters such as pH value, aluminium concentration, as well as changes in the fluid concentration from these effects (Ystroem et al., 2021). The goal is to reconstruct the in-situ chemical system as a basis to compute the equilibrium temperatures. Assuming that the temperature estimations for multiple mineral phases converge to an equal temperature. For these saturation index determinations, MulT_predict couples MATLAB (MATLAB, 2021) to an IPhreeqcCOM server introduced by Charlton and Parkhurst (2011). The current IPhreeqc version (3.7.0 – 15749) is used to calculate the equilibrium state of a chosen mineral set to serve as a multicomponent geothermometer. Thermodynamic data for mineral phases are taken from our updated Lawrence Livermore National Laboratory (LLNL.dat) database for high-temperature estimations (Ystroem et al., 2020). The major cation (Na^+, K^+, Ca^{2+}, and Mg^{2+}), and major anion (Cl^-, SO_4^{2-}, S^{2-}, and HCO_3^-) as well as SiO_2, $Fe^{2+/3+}$, and Al^{3+} concentrations are used as input data. In addition, to calculate the saturation indices of the mineral phases, the physicochemical parameters for the geothermal brine have to be defined. Therefore, the sampling pressure, temperature, as well as pH value are essential input data. MulT_predict distributes the input data to IPhreeqc to calculate the saturation indices $SI(T)$ over a predefined temperature range i for previously selected mineral phases j (e.g. a universally valid mineral set). To identify the equilibrium temperatures for the mineral phases ($SI = 0$), the $SI(T)$ returned from IPhreeqc are stored in a i x j matrix A_{ij} and evaluated by the signum function in matrix B_{ij} (Equation 5.2).

$$B_{ij} = \text{sign}(A_{ij}) := \lim_{k\to\infty} \frac{1 - 2^{-kA_{ij}}}{1 + 2^{-kA_{ij}}} = \begin{cases} - & 1 & \text{if } A_{ij} < 0 \\ & 0 & \text{if } A_{ij} = 0 \\ & 1 & \text{if } A_{ij} > 0 \end{cases} \tag{5.2}$$

Performing differences and approximate derivatives (Equation 5.3) upon the data returns the zero crossing of the saturation indices (Equation 5.1) of the chosen mineral set j over their defined temperature range i, which are stored in a new matrix C.

$$C = \text{diff } (B_{ij}) = \begin{bmatrix} B(1,0) - B(0,0) & \cdots & B(1,j) - B(0,j) \\ \vdots & \ddots & \vdots \\ B(i,0) - B(i-1,0) & \cdots & B(i,j) - B(i-1,j) \end{bmatrix} \tag{5.3}$$

To find the equilibrium temperature, the data of matrix C are indexed by Equation 5.4 and transferred into an array D.

$$D = \text{for diff } (C) \neq 0 \tag{5.4}$$

The resulting equilibrium temperatures D and their mineral phases are both stored as cell entries in E. When the optimisation process is enabled, the selected sensitive parameters (boiling/dilution, pH value, and aluminium concentration) are iterated interdependently in their predefined range. After each computation, the result is also stored in the cell array E. Depending on the number of selected parameters n, the cell array E is stocked up in n-dimensions. To get a consistent, statistically evaluable dataset, the mineral phases and their associated equilibrium temperatures are validated in each entry. Therefore, the cell entries of E have to be cleaned up. An n-dimensional loop compares all resulting mineral phases and their temperatures from beginning to end. Missing temperature data is filled up with NaN while the names of j are added as the associated phases. Simultaneously, redundant and excess information is removed. After the completion, the procedure is executed vice versa starting at the end to check for inconsistency within the updated array E. Then, the cell array E is analysed by comparing the 0.25 and 0.75 quantiles, the interquartile range IQR, as well as the 1.5*IQR outliers within each entry stored in an array F (Equation 5.5).

$$F_n = [(q_{0.75} + 1.5\ \mathrm{IQR}) - (q_{0.25} - 1.5\ \mathrm{IQR})]\,;\ \text{for all entries in } E \tag{5.5}$$

Afterwards, MulT_predict statistically evaluates all n-dimensional cell entries of Fn for the minimum element of the array excluding NaN values. The global minimum represents the maximal convergence of the equilibrium temperatures of the considered mineral phases (Equation 5.6).

$$\min(f(x_0)),\ F_n \subseteq \mathbb{R}^+,\ x_0 \in F_n,\ f : F_n \to \mathbb{R}^+,\ \text{if } (\forall x \in F_n)\ f(x_0) \leq f(x) \tag{5.6}$$

Determining the position of the global minima in the array F_n, the back-calculated in-situ values of the reservoir parameters are extracted. In the last step, MulT_predict re-evaluates the in-situ reservoir parameters and re-calculates the final temperature estimation. The final results are output graphically and tabular.

5.3 DATA AND RESULTS

To develop a comprehensive multicomponent geothermometer and a universally valid mineral assemblage MulT_predict has to be validated for a variety of geothermal sites. A thorough validation is only possible with datasets, where geochemical fluid analyses, as well as borehole temperatures, are available. Geothermal sites were intentionally selected in a way, that the data covers a broad range of different lithologies, temperatures, and geochemical characteristics to evolve a comprehensive tool. The geothermal fluid samples from different geothermal sites are mostly compiled from the literature. In sum, eight geothermal settings were evaluated (Table 5.1). To validate MulT_predict, geochemical data, as well as in-situ reservoir temperature measurements of the reservoir, must be available for each site. The sites are categorised by their geology and tectonic setting: crystalline basement of a rift basin, sedimentary back-arc basin dominated by continental facies, foreland basin dominated by marine facies, sedimentary basin dominated by marine facies, volcanic facies dominated by andesite to rhyolite at subduction zones, and basaltic facies on the mid-ocean ridge induced by a hotspot.

Table 5.1: Collocation of the lithology, geological setting, location, and references of the geothermal wells for the validation of MulT_predict.

Lithology	Setting	Location	Reference
Granite	Basement (rift basin)	Upper Rhine Graben	Sanjuan et al. (2001) Schindler et al. (2010) Dezayes et al. (2013) Sanjuan et al. (2016) Vidal and Genter (2018) Vidal et al. (2019)
Lacustrine facies	Back-arc basin	Pannonian Basin	Varsányi et al. (1997)
Marine facies, Malm	Foreland basin	German Molasse Basin	Internal communication
Marine facies, Dogger	Sedimentary basin	Paris Basin	Michard and Bastide (1988) Marty et al. (1988) Criaud et al. (1989)
Rhyolite, Andesite	Subduction zone	Waiotapu	Banwell (1959) Ellis and Mahon (1977) Giggenbach et al. (1994)
Andesite	Subduction zone	Miravalles	Dennis et al. (1989) Gherardi et al. (2002)
Dacite	Subduction zone	El Tatio	Ellis and Mahon (1977) Giggenbach (1978)
Basalt, Hyaloclastite	Hotspot	Iceland	Gudmundsson and Arnórsson (2002) Óskarsson et al. (2015)

The data representing the highly saline fluid of sedimentary origin hosted in the granitic crystalline basement in the Upper Rhine Graben (URG) is presented by Sanjuan et al. (2001) and Schindler et al. (2010) for GPK 1 and 2 in Soultz-sous-Forêts, Dezayes et al. (2013) and Vidal et al. (2019) for GRT-1 in Rittershoffen, and Sanjuan et al. (2016) as well as Vidal et al. (2019) for GTLA 1 in Landau and INSH in Insheim. The continental sedimentary facies is represented by lacustrine to fluvial sedimentary sequences in the Pannonian Basin published by Varsányi et al. (1997). Malm layers in the German Molasse Basin represent marine facies in a foreland basin (internal communication with power plant operators, Appendix A.6). Marine sedimentary facies from Dogger layers in the Paris Basin were published by Michard and Bastide (1988), Marty et al. (1988), and Criaud et al. (1989). The volcanic facies are summing up different types of reservoir rock around the Pacific Ring of Fire. For Waiotapu in New Zealand, rhyolite and andesite reservoir formations are hosting the geothermal wells published by Banwell (1959), Ellis and Mahon (1977), and Giggenbach et al. (1994). Further, data from the wells of the andesite reservoir formation in Miravalles, Costa Rica, are presented by Dennis et al. (1989), and Gherardi et al. (2002). Lastly, the geochemical data presented by Ellis and Mahon (1977), and Giggenbach (1978)

of a dacite reservoir in an ignimbrite formation at El Tatio, Chile, are used. The basalt and hyaloclastite facies are visualised by data from Krafla (Gudmundsson and Arnórsson, 2002), and Reykjanes (Óskarsson et al., 2015). For reference, the artificial standard of ocean water is illustrated (ASTM-Standard, 2013).

To establish the new procedure for unknown temperature estimations three steps had to be passed. First, setting-specific mineral sets were developed for known mineralogical and temperature data. Second, coinciding mineral phases of setting-specific mineral sets are combined into a universally valid set. Third, to refine the general temperature estimation, a procedure of outliner removal has to be implemented. The results of the three steps are visualised in Figure 5.1. For each geothermal setting, the mineral set for reservoir temperature estimation is compiled individually (Appendix A.7). Creating these unique sets is time-consuming. Therefore, if known, the observed mineral composition of the reservoir rock should be transferred into a mineral set. Hence, the main mineral phases of the reservoir rock, as well as secondary and accessory minerals, have to be considered. The resulting mineral assemblages have to be refined until the optimised saturation curves converge on each other, minimising the uncertainty of the temperature estimation (cf. Equation 5.5, 5.6). Therefore, mineralogical foreknowledge of the reservoir rock composition and its secondary mineralisation is preferable in general. Nevertheless, in some cases this information is not available or unknown, thus, a predefined universal mineral set is presented. This universally valid mineral set consists of the most common rock-forming minerals as well as secondary and accessory minerals. The focus of the universal set is its comprehensive applicability while providing still reasonable temperature estimations. Together with the upcoming procedure of outlier removal, the temperature estimations get more precise. Thus, all evolved setting-specific mineral sets were compared among themselves. As a result, in most of the sets, an overlap of mineral phases is repeatedly equilibrated. Combining them, a universally valid mineral set has been established, which can be applied on a worldwide scale independently of reservoir mineralogy and setting. This universally valid set consists of quartz, K-feldspar, microcline, albite, muscovite, illite, diaspore, analcime, scolecite, anhydrite, kaolinite, and pyrophyllite (Appendix A.7).

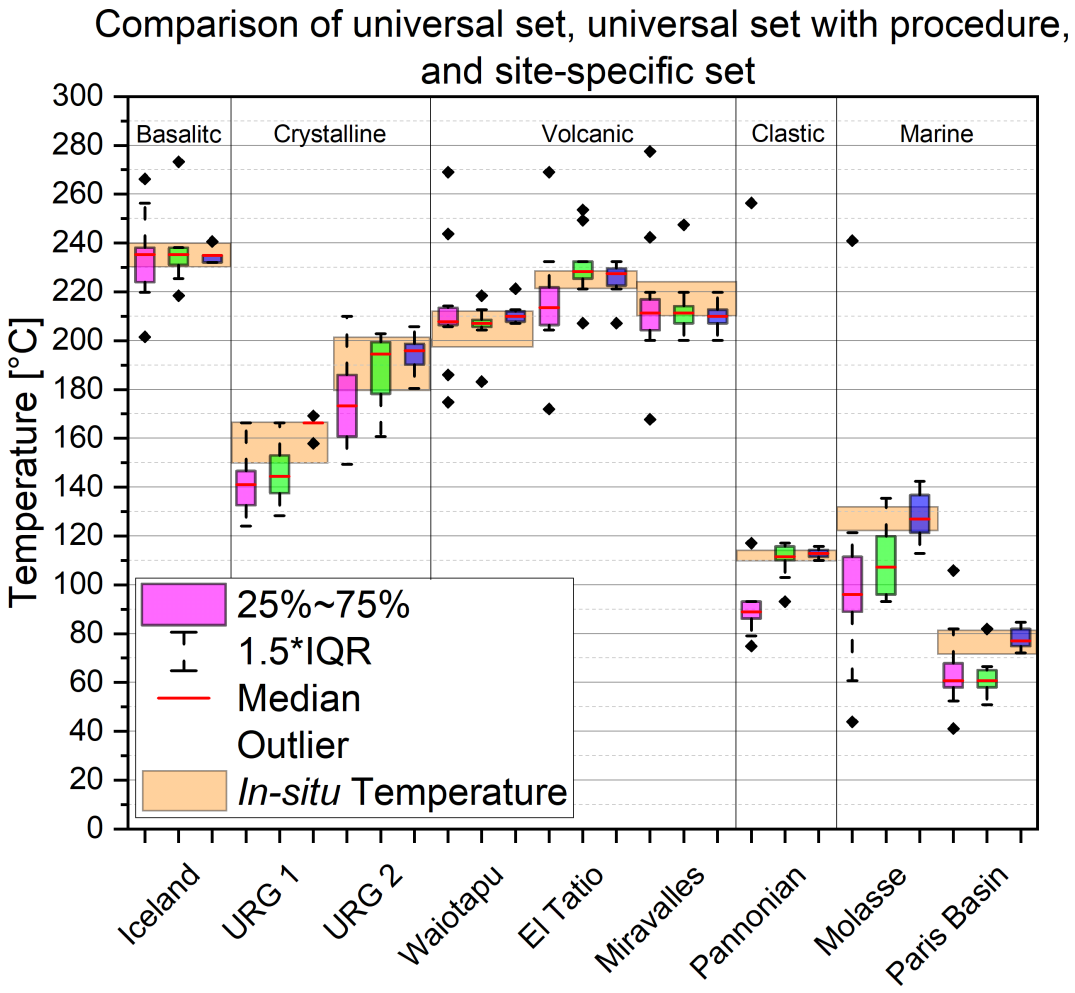


Figure 5.1: Comparison of three temperature estimations for an exemplary well of individual geothermal settings around the world (Krafla on Iceland, Soultz-sous-Forêts in the Upper Rhine Graben, Waiotapu in New Zealand, El Tatio in Chile, Miravalles in Costa Rica, Makó in the Pannonian Basin, Unterhaching in the German Molasse Basin, and Evry in the Paris Basin). Straight lines separate different reservoir rock compositions (basaltic, crystalline, volcanic, clastic, and marine facies). The orange box indicates the measured in-situ temperature in the open hole section of the well. The first box plot (magenta), visualises the temperature estimation with the original universally valid set. The secondary box plot is obtained by reducing the outermost mineral phases (black diamonds) within the universal set and recalculating the temperature estimation (outlier reduction in green). In comparison, the third box plot (blue) is the best-fitting mineral set individually developed for each setting (cf. Appendix A.7).

Figure 5.1 illustrates the outlier removal procedure to obtain the most reasonable temperature estimation. In the diagram, three temperature estimations per geothermal setting are shown. The orange box indicates the measured in-situ temperature corresponding to the open hole section within the example. The first box plot (magenta) illustrates the temperature estimation by just applying

the universally valid mineral set. In the next step, for the second box plot in green, two statistical outliers of the universal set are removed. This corresponds to mineral phases m, which are exceeding 1.5 times the interquartile range (IQR) the most (Equation 5.7).

$$M \subseteq \mathbb{R}^+, y \in M, \begin{cases} \text{if } y > y_{median} + 1.5 \cdot IQR \\ \text{if } y < y_{median} - 1.5 \cdot IQR \end{cases} \tag{5.7}$$

This outlier removal represents the procedure, which is used to evaluate new geochemical data with unknown mineralogy. The third box plot in blue visualises the best fitting result obtained by developing an individual mineral set for each setting, as mentioned in Appendix A.7. After the universally valid mineral set for reservoir temperature estimation has been introduced, it can be compared to the results of the outlier removal and each setting-specific mineral set (Appendix A.7). Comparing the three box plots in Figure 5.1, the shift of temperature estimation towards the measured in-situ temperature is evident. In each case, the overall spread of the plot is decreasing as well as a decreasing or steady IQR. Only for the carbonate facies of the Paris Basin and the Molasse Basin, calcite and dolomite have to be added to the universally valid mineral set to reconstruct reservoir conditions and temperature estimations more precisely. For Iceland, the outlier removal reduces the IQR by 50% to 7 K, while the median temperature remains at 235 °C, fitting the best-fit temperature and the temperature log. For URG, two samples of GPK2 in Soultz-sous-Forêts are presented. URG 1 was sampled in the year 1997 when the well reached 3876 m depth. URG 2 was sampled in the year 1999 after deepening the well down to 5093 m representing the actual temperature at the bottom of GPK2. For URG 1, the overall uncertainty decreases by 12% due to the outlier removal, but further refinement of the mineral set is necessary to reach best-fit conditions at 166 °C. For URG 2, the outlier removal decreases the spread of the box plot by 18 K. The best-fit temperature estimation matches the measured in-situ temperature of the open hole section. The two temperature estimations of GPK2 illustrate the importance of the equilibrium reaction between the hosted reservoir and the fluid, where the technical available temperatures are predicted. At Waiotapu, the necessity of outlier removal for the magenta box plot is clear. The removal reduces the spread by 59 K and emphasises the median temperature at 207 °C within the measured temperature. In addition, the IQR (4 K) of the adjusted universally valid mineral set equals the best-fitting result. For El Tatio, the overall spread is reduced by 50% due to the outliner removal and increases the median temperature to its final result at 228 °C, matching the measured in-situ temperature. The outlier removal for Miravalles diminishes the spread by 62 K and the IQR by 45%. Likewise, the adjusted universally valid mineral set

resembles the best-fitting plot. For the Pannonian Basin, the procedure shifts the median temperature by 22 K into the measured temperature range. Compared to the best-fitting plot, the mineral assemblage can be further refined to increase accuracy. For both marine facies, the outlier removal reduces the uncertainties but the temperature estimation does not fit the in-situ temperature. For the Molasse Basin, the spread diminishes by 79%, while for the Paris Basin its halves. Due to the lack of carbonate mineral phases within the universally valid mineral set, the median of the temperature estimation in the Molasse is 20 K beneath the best fit and in the case of Paris Basin 16 K.

5.4 DISCUSSION

For the illustration of coherence of the properties of brines from different geothermal settings, geochemical key parameters are analysed. These parameters are relevant to multicomponent geothermometry and can explain the plausibility of a generally valid mineral set. For this purpose, the major chemical components are used (Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, HCO_3^-, and additionally pH). To graphically cluster the brines according to the surrounding lithology, strongly temperature-sensitive components such as SiO_2 were omitted. Thus, the major chemical components of dissolved mineral phases used in geothermometry are evaluated. Thus, the graphical single-point method by Langelier and Ludwig (1942) is modified. In the original two-dimensional plot, the pH value of the fluid is added as the z-axis. Furthermore, the total dissolved solids TDS are projected as sphere size s (Equation 5.8).

$$s = \log\left(\mathrm{TDS} \cdot 10\right) \tag{5.8}$$

To display the differences in sphere sizes homogenously, the TDS is multiplied by ten and then the decadic logarithm is applied. Hence, standard fluid analysis of the geothermal brines are plotted in the modified Langelier-Ludwig diagram (Figure 5.2).

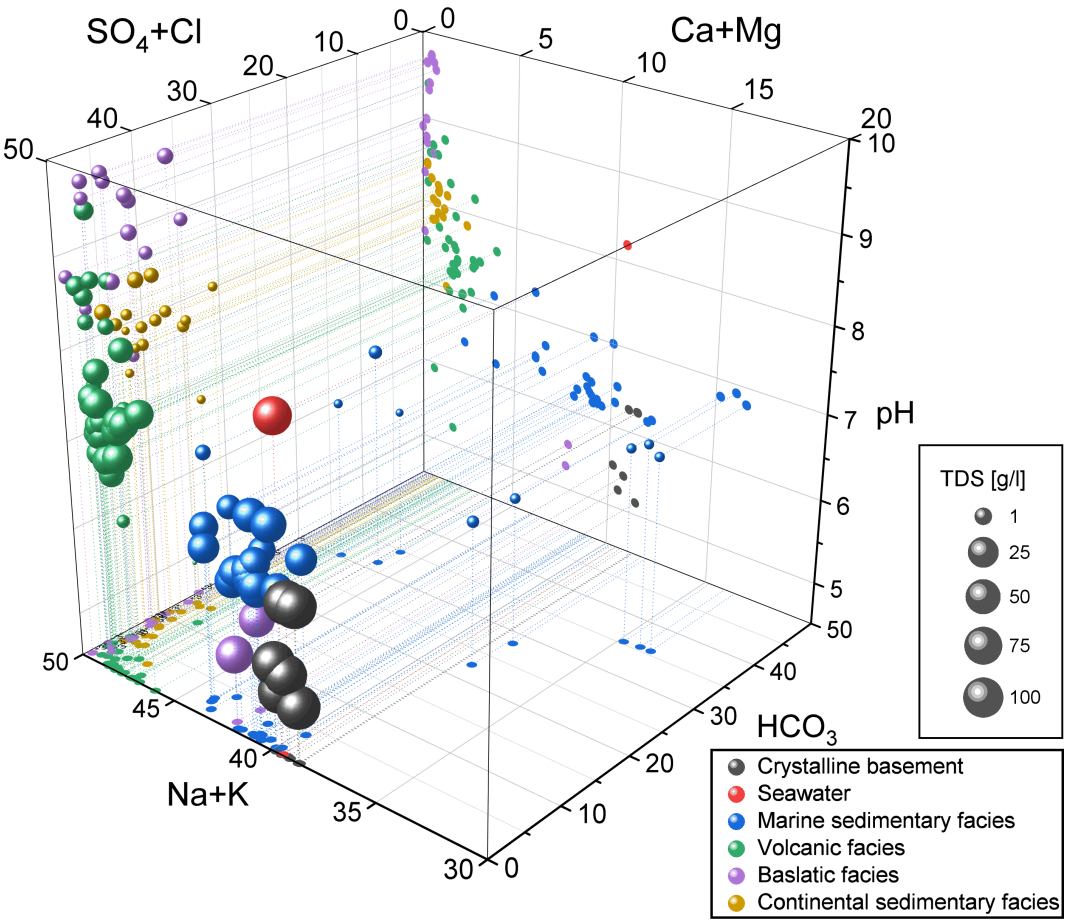


Figure 5.2: Modified Langelier-Ludwig diagram for classification of geothermal settings in different facies. The pH value was added as a third dimension (z-axis) as well as the total dissolved solids (TDS) visualised in sphere size (*s*).

By plotting the geochemistry of the different geothermal sites clustering can be observed. High saline brines from the Dogger formation in the Paris Basin, the saltwater intrusion in the wells of the Reykjanes Peninsula, as well as the samples from the Upper Rhine Graben, cluster below the salinity of seawater at Na^++K^+ 40% and SO_4^{2-}+Cl^- 50%. The shift in pH value is a function of temperature and salinity (Ellis, 1970). Increasing temperature and salinity cause the pH value to decrease. The marine facies samples cluster near-neutral pH values. This corresponds to a buffering reaction within the carbonates (Malm, Dogger) due to calcite solubility equilibrium and temperature-dependent auto-dissociation within the fluid (Ellis, 1963). Therefore, the brines in the URG (TDS ~100 g/L) and high-temperature settings in Reykjanes (exceeding 300 °C) have lower pH values. For the volcanic facies, Miravalles and El Tatio are forming the main cluster around Na^++K^+ 47% and SO_4^{2-}+Cl^- 50% with a pH of 7.5. The samples of Waiotapu have a shift to higher pH values. For Waiotapu, (Giggenbach et al., 1994) proposed a low content of CO_2 within the parent magma and removal of CO_2 as calcite through Ca-Al-silica interaction. In addition, there is degassing and dilution of the fluid before reaching the wells. The well waters in Krafla and

Námafjall cluster at the highest pH values. These waters are low mineralised and of meteoric origin. Due to degassing of CO_2 and H_2S the pH is increased as well as a bicarbonate concentration trend up to 15% of HCO_3^- corresponding to an increase in magmatic activity underneath well K-20 (Gudmundsson and Arnórsson, 2002). For samples from the Pannonian Basin, the TDS concentration is low. The fluid chemistry clusters next to Waiotapu samples. Similar to the bicarbonate trend in Icelandic samples, the Pannonian chemistry shows the HCO_3^- concentration increasing with temperature and depth corresponding to lacustrine carbonates within the formations (Varsányi et al., 1997).
The clustering of samples in the Langelier-Ludwig diagram (Figure 5.2) shows up similarities within the unique signature of each geothermal setting. In addition, the influence of evaporites and seawater can clearly be distinguished from samples of meteoric origin. The distribution of pH values represents a variety of geochemical processes coupled to the individual settings. The pH value sums up buffer reactions, degassing processes, and the impact of salinity and temperature. Therefore, the geochemistry of the geothermal fluid is an indicator of the mineral composition leached out of the reservoir rock. Thus, these similarities in the clustering are reflected in the universally valid mineral assemblage presented for reservoir temperature estimations.

Furthermore, the optimisation process within MulT_predict is evaluated. Therefore, a synthetic brine is set up and processed by MulT_predict to back-calculate the initial brine temperature. In the next step, all parameters of the optimisation processes are tested individually. In this case, the influence of salinity, steam loss, and dilution, the pH value as well as the Al concentration are examined. Therefore, the synthetic brine is increasingly perturbed stepwise for one of these parameters at once. This is resumed until the numerical limits of the optimisation of MulT_predict are reached. Thus, the individual optimisation processes can be evaluated. Lastly, the synthetic brine is perturbed for all parameters at once. The steam loss and dilution are increased stepwise, while random pH values and element concentrations in a predefined range are added. In this case, the MulT_predict's optimisation process is used interdependently to back-calculate reservoir conditions. The synthetic mineral assemblage is equilibrated at 150 °C and has a pH value of six. The mineral assemblage is composed of phases of the universally valid set: quartz, microcline, albite, calcite, anhydrite, muscovite, illite, and stilbite. The resulting equilibrated fluid lost some mass while reacting with stilbite (0.899 kg remaining) as well as the pH value increased to 6.743 due to changes in the activity of hydrogen ions. To revert to initial conditions, the fluid mass is set back to 1 kg. In the next step, each parameter is varied individually and stepwise around its initial value to perturb the equilibrium between mineral phases. Then, MulT_predict's optimisation process is used to reconstruct the equilibrium conditions of the fluid and determine the initial equilibrium temperature. First, the salinity is increased in 100 steps from zero

molar up to five molar by adding NaCl to the solution. In Figure 5.3, for each increment, the MulT_predict computes the temperature estimations which are plotted in the diagram. The computed temperature estimation is constantly at 148 °C close to the initial conditions of 150 °C, thus MulT_predict temperature estimations show virtually no sensitivity to increasing salinity.

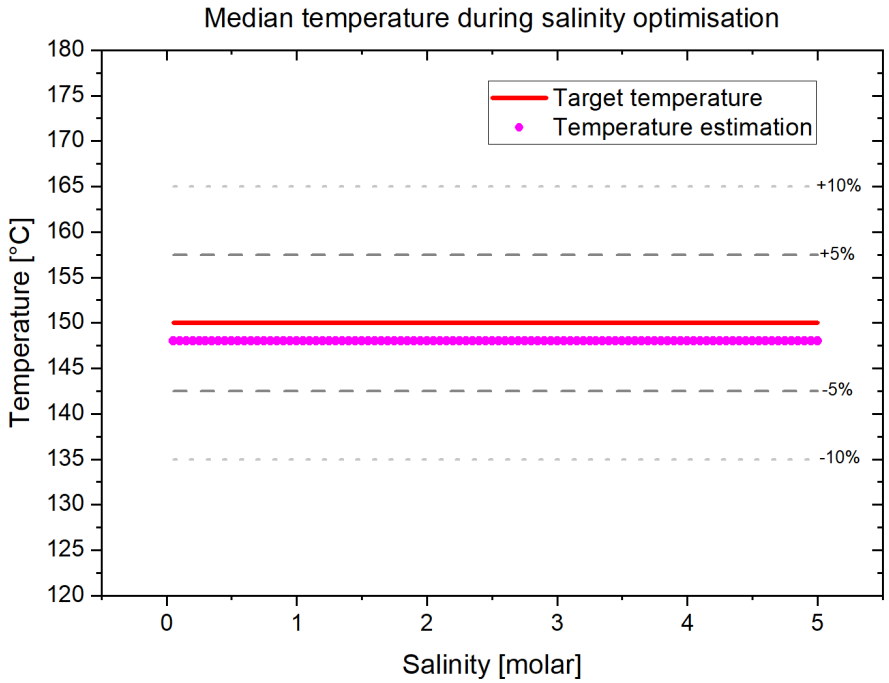


Figure 5.3: Temperature estimations corresponding to salinity changes from zero to five molar by adding NaCl to the solution. The red line is indicating the initial equilibrium temperature of the synthetic fluid. The red line indicates the equilibrium temperature of 150 °C.

The assessment of the sensitivity of changes by steam loss and dilution in the fluid is displayed in Figure 5.4. Therefore, the moles of the fluid are altered by 1% per step. For dilution, 0.555 moles of pure water are added per increment. In contrast, 0.555 moles of pure water are subtracted mimicking steam loss. In MulT_predict's optimisation process, these changes are back-calculated. In Figure 5.4a), MulT_predict's computed concentration error is plotted against the percentage of changes to the fluid concentration as well as the resulting temperature estimation in Figure 5.4b).

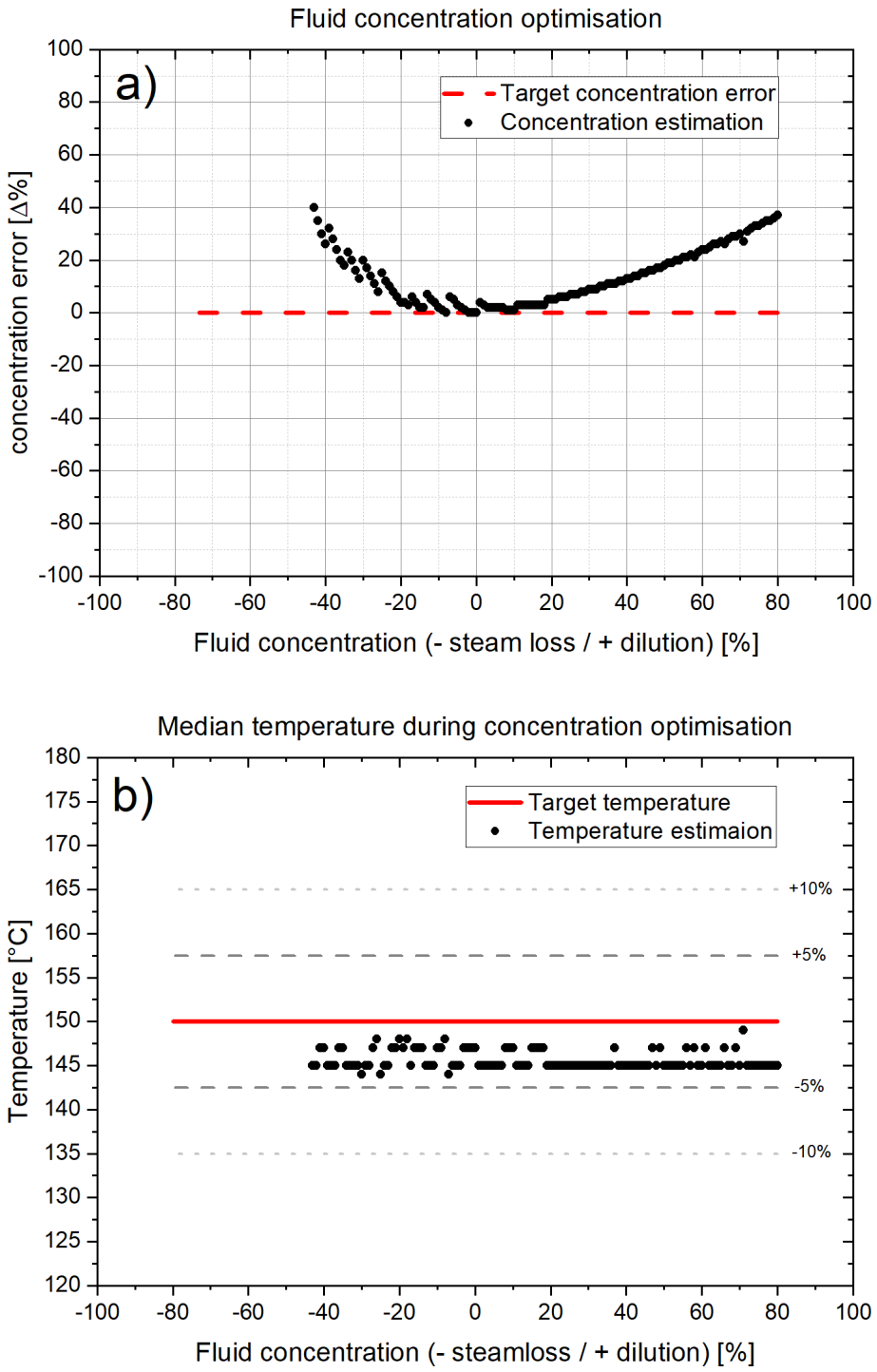
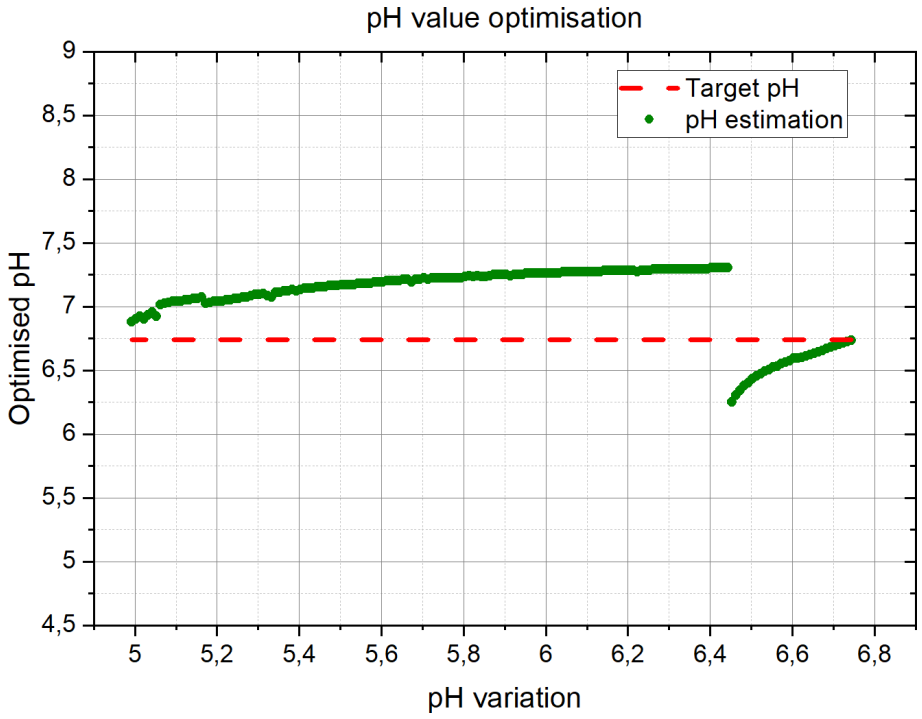


Figure 5.4: In graph a), the error of the reconstructed fluid concentration is plotted against the changes to the mass of water within the fluid. The dashed red line indicates the targeted error. In graph b), the corresponding temperature estimations are visualised, while the solid red line indicates the targeted temperature of 150 °C.

For perturbations within ±20% of dilution or steam loss, the optimisation process reconstructs the fluid with small errors of 3 percentage points on average (cf. dashed, red line in Figure 5.4). Regardless of whether steam loss or dilution, at higher changes MulT_predict starts to underestimate these perturbations. Thus, the difference between the synthetic changes in the fluid and the reconstruction is rising. This curve shape corresponds to the geochemistry in the fluid. Decreasing water by 50% doubles the element concentration. In contrast, dilution of 100% halves the concentration. Comparing the resulting element concentration leads to a continuous increase in error. Nevertheless, the temperature estimations in Figure 5.4a) have a maximum spread of 5 K (144 °C to 149 °C) and are on average 145 °C. The red line indicates the equilibrium temperature of the synthetic set. Furthermore, changes in the pH value are evaluated in Figure 5.5. The pH value varies between five and seven in 200 increments. The resulting pH value of the optimisation process is plotted in 5.5a) against the synthetic pH variation. In 5.5b), the resulting pH value is illustrated against the temperature estimation.



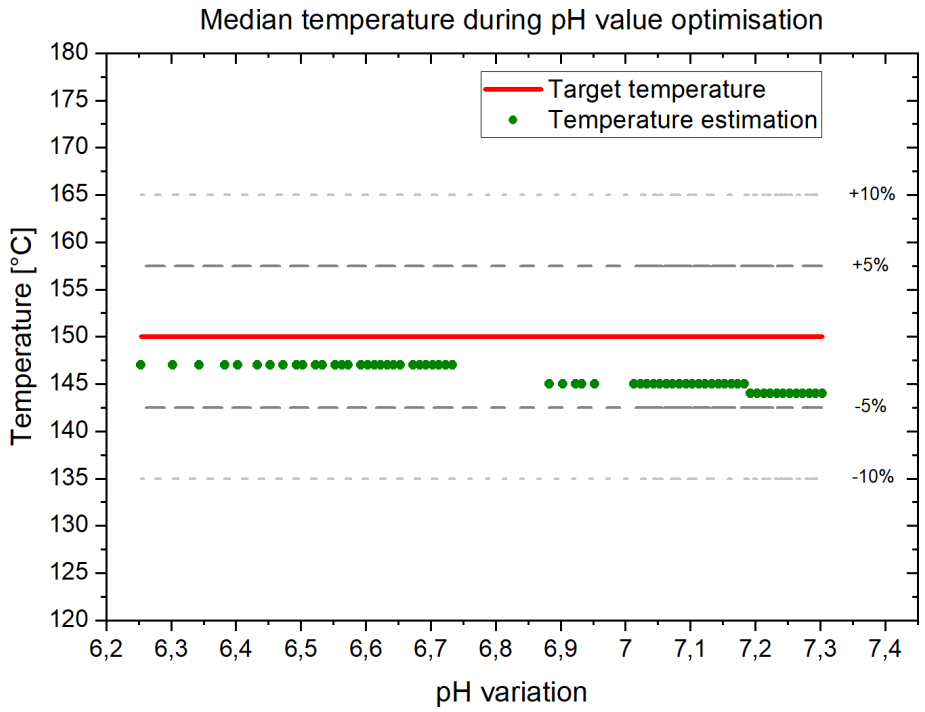


Figure 5.5: In graph a), the reconstructed pH value of the fluid is plotted against its variation between five and seven. The dashed red line indicates the initial equilibrium pH at 6.743. In graph b), the corresponding temperature estimations are visualised, while the solid red line indicates the targeted temperature of 150 °C.

Figure 5.5a) visualises the difference between the resulting optimised pH value and the initial equilibrate pH (6.743) indicated by the dashed, red line. pH values increasing 6.743 cannot be back-calculated by IPhreeqc because of a thermodynamic equilibrium gap for the aluminium concentration at higher pH. Therefore, MulT_predict is also capped at this level, not allowing interdependent optimisation yet. The estimations of the pH value show a pattern corresponding to geochemical adaptations of IPhreeqc iteration steps. Because of the loss in precision due to no redox buffering, IPhreeqc alters parameters automatically. Therefore, the redox potential is varied so the balance equations can be solved to obtain a chemical equilibrium (Parkhurst, Appelo et al., 2013). This leads to a recommencement at pH 6.443 and a steady decline in the optimised pH value towards the initial value while the redox potential is adjusted. In Figure 5.5b), optimised pH values in the range 7.303 to 7.193 reach a temperature estimation of 144 °C. For values from 7.183 to 6.883 the estimation is 145 °C and for a lower pH, a temperature of 147 °C is calculated.
Finally, variations in the aluminium concentration are examined. In 160 steps, 80% of the initial aluminium concentration is added and subtracted. Figure 5.6 illustrates the optimised aluminium concentration against the shift in concertation as well as the resulting temperature estimation.

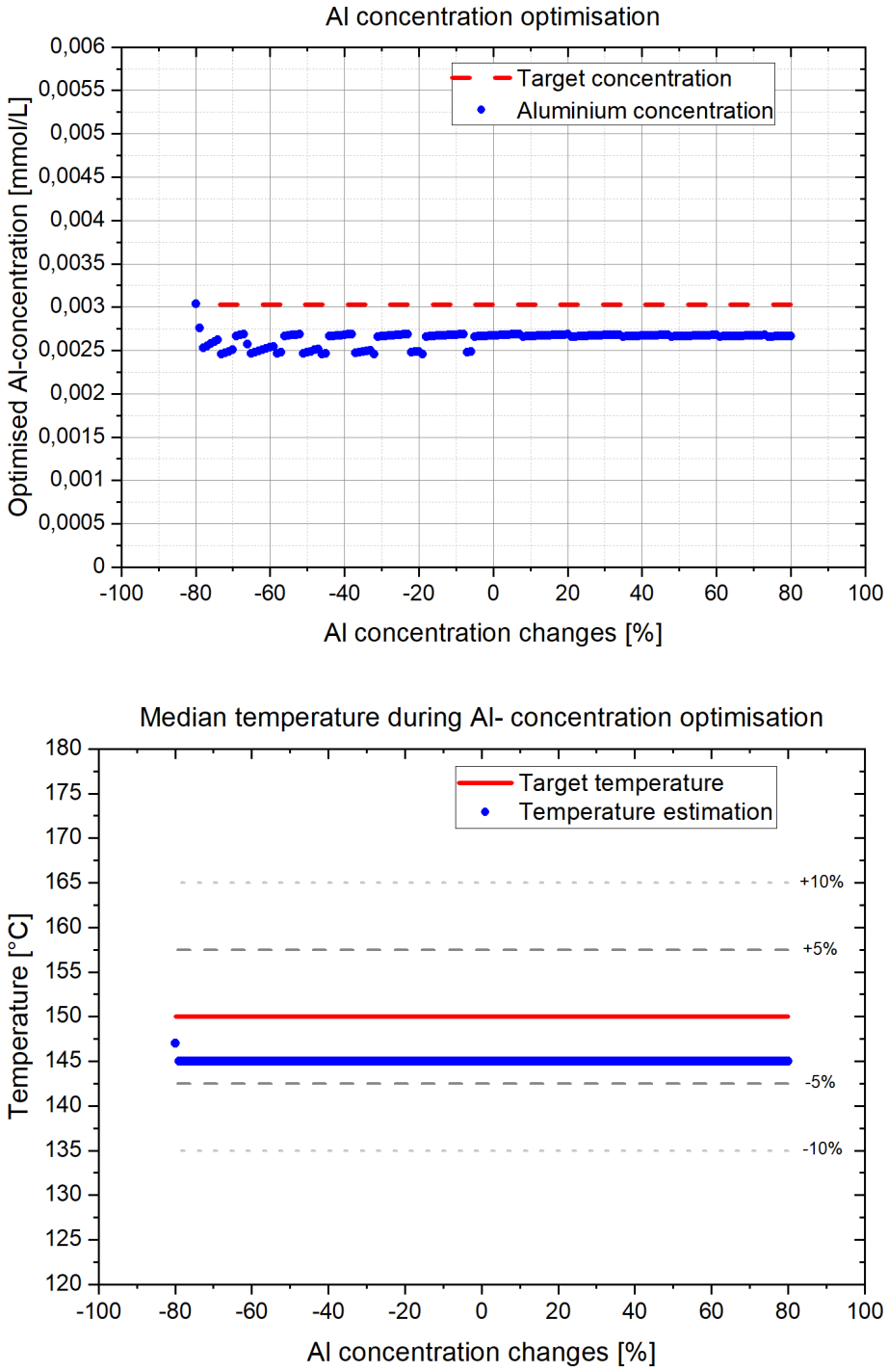
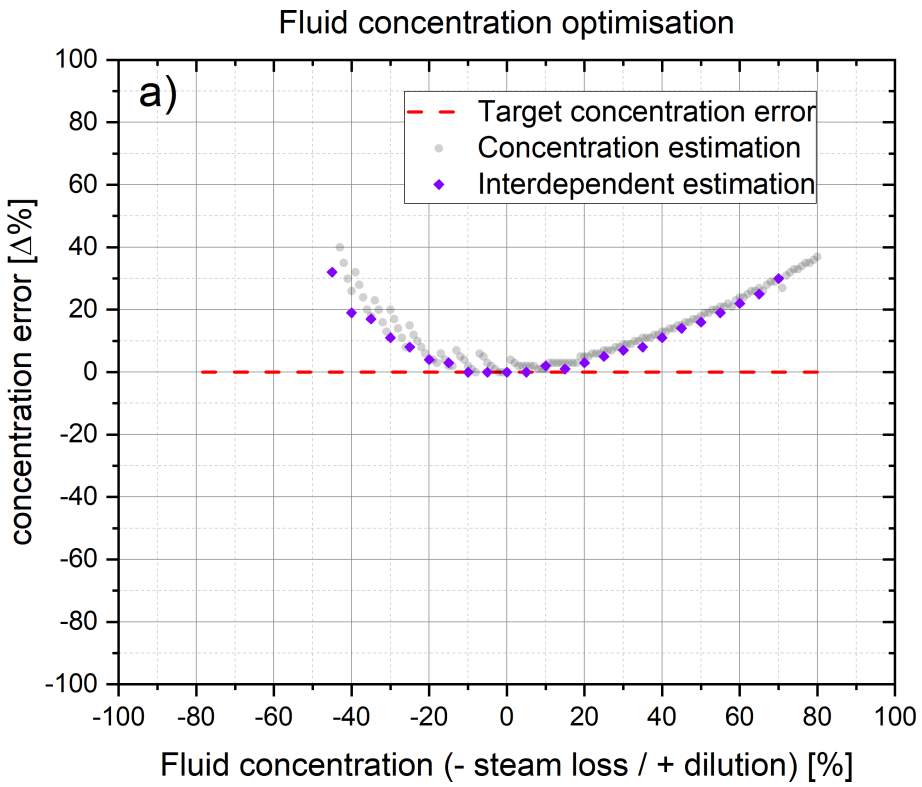
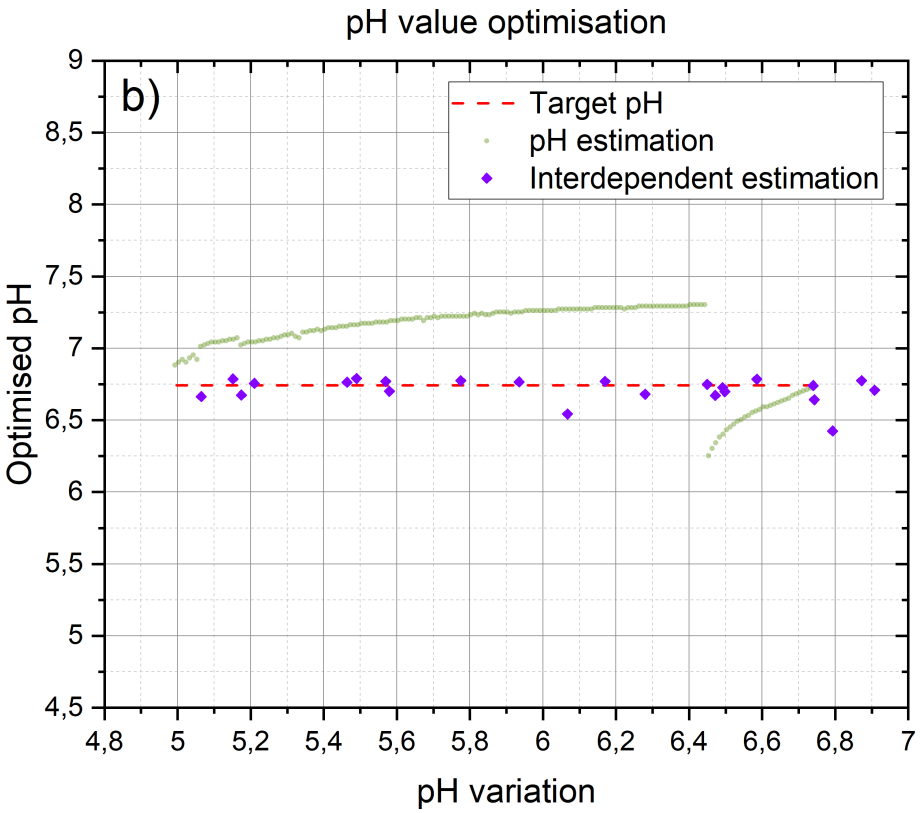
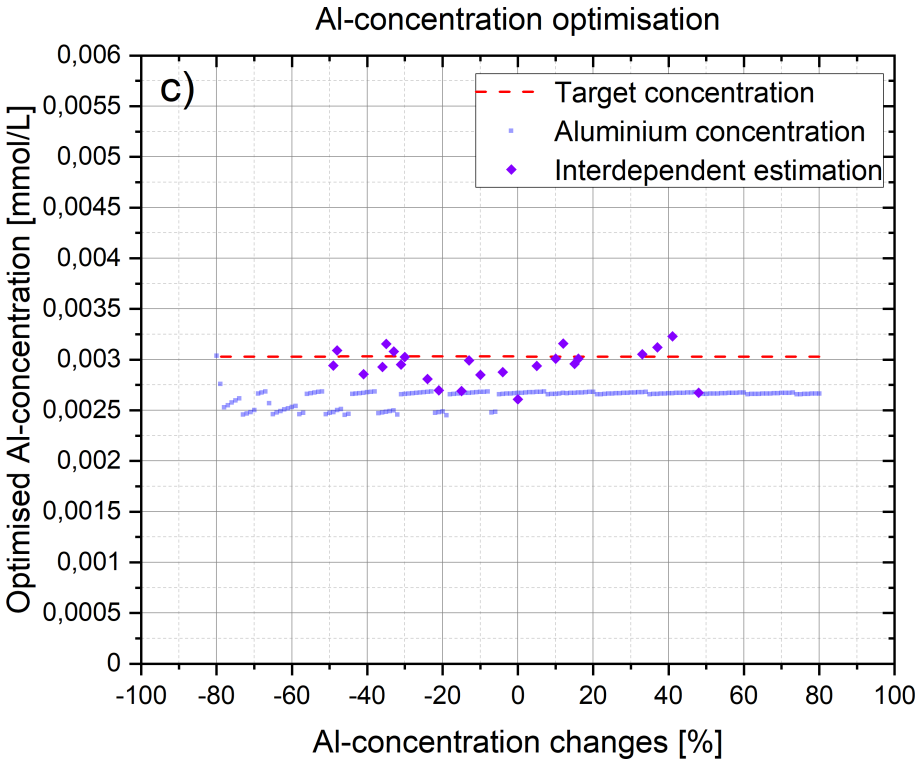


Figure 5.6: In graph a), the reconstructed aluminium concentration is plotted against its variation between -80% and +80%.The dashed red line indicates the initial aluminium concentration of 0.00303 mmol/L. In graph b), the corresponding temperature estimations are visualised, while the solid red line indicates the targeted temperature of 150 °C.

Figure 5.6a) shows a steady back-calculated aluminium concentration around 0.00267 mmol/L (11% less than the initial 0.00303 mmol/L) for raising concentrations. With decreasing aluminium concentration, IPhreeqc starts to iterate the equilibrium calculation to converge the chemistry. Due to the variation in the overall small aluminium concentration, errors accumulate while running the simulation. Thus, the program automatically attempts to solve the calculation by altering and combining tolerances and step sizes to fit the solution (Parkhurst, Appelo et al., 2013). Similarly, the redox potential is adjusted. Therefore, the aluminium concentration fluctuates from 0.00245 mmol/L to 0.00269 mmol/L and reaches the maximum of 0.00304 mmol/L at -80%. However, the temperature estimation is steady at 145 °C once rising to 147 °C at -80%. Comprising all cases, MulT_predict underestimates the temperature estimations, anyhow not exceeding 6 K (4% error). Only optimising individual parameters does not take chemical interactions into account. As mentioned in the methodology, these sensitive parameters are coupled due to secondary processes. Therefore, in a second step, the back-calculation process is tested by optimising the parameters interdependently. Therefore, in 5%-steps the fluid concentration was perturbed again between -45% to 70%, as well as the values for the pH and aluminium concentration have been randomised. For the pH, random values between pH 5 and 7 and for the aluminium concentration values between 0.001 mmol/L and 0.006 mmol/L were generated. Currently, MulT_predict is allowed to simultaneously optimise all sensitive parameters interdependently.




pH value optimisation
b)
Target pH
pH estimation
Interdependent estimation
Optimised pH
pH variation
9
8,5
8
7,5
7
6,5
6
5,5
5
4,5
4,8
5
5,2
5,4
5,6
5,8
6
6,2
6,4
6,6
6,8
7


Al-concentration optimisation
c)
Target concentration
Aluminium concentration
Interdependent estimation
Optimised Al-concentration [mmol/L]
Al-concentration changes [%]
0,006
0,0055
0,005
0,0045
0,004
0,0035
0,003
0,0025
0,002
0,0015
0,001
0,0005
0
-100
-80
-60
-40
-20
0
20
40
60
80
100

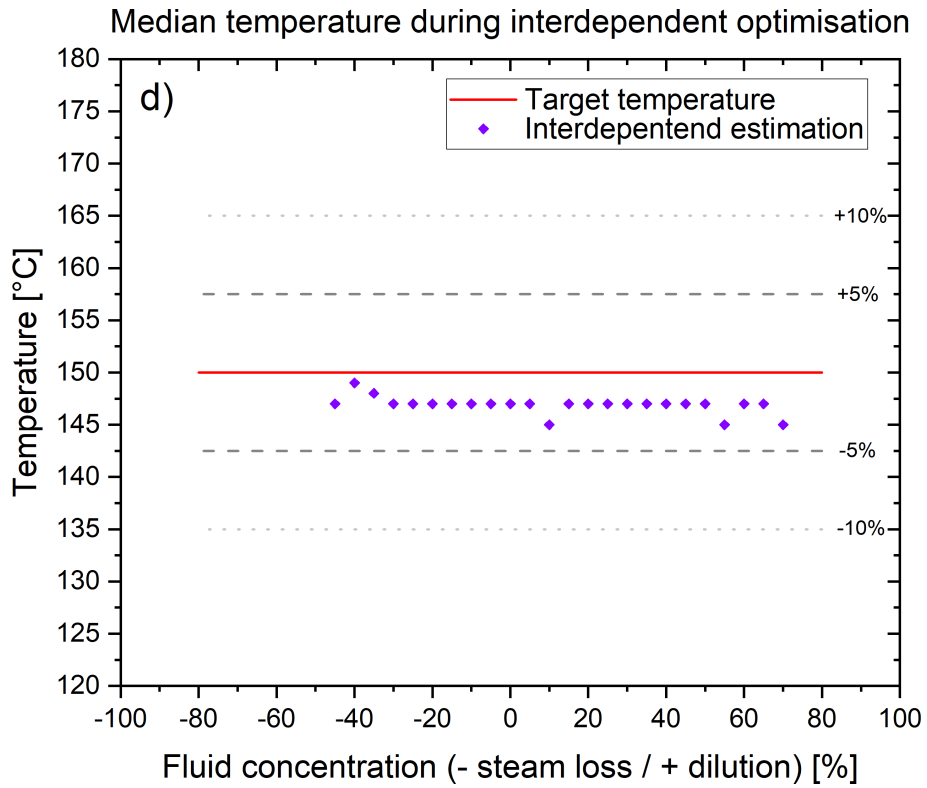


Figure 5.7: Simultaneous and interdependent optimisation (purple) of all sensitive parameters plotted into each optimisation process (pale colours). a) shows the reconstructed fluid concentration resulting from the variation of the mass of water within the fluid. b) visualises the back-calculated pH value against its change. In c), the interdependent reconstruction of the aluminium concentration is plotted against its variation. d) illustrates the temperature estimation of the overall optimisation process. The dashed red lines indicate the initial equilibrium conditions, as well as the solid red line which shows the equilibrium temperature of the synthetic fluid.

In Figure 5.7, the fluid concentration, pH value, and aluminium concentration are optimised simultaneously. The interdependent estimations from test step 1 were plotted into the results of the individual optimisation processes, which were paled out. In 5.7a), the average concentration error decreases by 2 percentage points compared to the result of the independent back-calculation. The pH values in 5.7b) and the aluminium concentrations in 5.7c) fit the target values more accurately than those resulting from the individual optimisation process. For the pH value, the maximum deviation is 0.3 less than the targeted pH. On average, the variance of the interdependent pH optimisation is 0.056, which is negligible compared to the mean deviation of 0.32 for the independent calculation. For the aluminium concentration, the average deviation is 0.164 µmol/L about 5% of the target concentration. Therefore, interdependent optimisation is circa 50% more precise than independent optimisation. In addition, the temperature estimations in 5.7d) improved to a mean value of 147 °C. Especially the reconstruction of

the pH value and the aluminium concentration show the importance of an interdependent optimisation process to reflect geochemical changes within the fluid.

5.5 CONCLUSIONS

MulT_predict is a robust multicomponent geothermometer with a built-in optimisation process to reconstruct the initial in-situ reservoir conditions and temperature from a chemically perturbed geothermal fluid sample. The temperature estimations are based on a standard chemical analysis of the geothermal fluid composition. No sophisticated sampling methods nor analyses are needed. A numerical reconstruction process of the in-situ chemical conditions is implemented to correct secondary perturbation of the fluid sample. The joint optimisation of coupled key parameters (boiling/dilution, pH value, and aluminium concentration) allows for the determination of precise reservoir temperatures. A universally valid mineral set is deduced, allowing the calculation of reservoir temperatures. Regardless of the setting and the reservoir mineralogy, the universally valid mineral set provides a great advantage when only little or no knowledge of the subsurface mineralogy is available. The implementation of a statistical outlier removal refines the mineral assemblage and improves the temperature estimation further. Therefore, MulT_predict can be used as an early phase greenfield exploration tool, which has been validated at multiple geothermal sites worldwide.

The emphasis of the study was the enhancement of the applicability of MulT_predict on a worldwide scale, as well as the focus on the performance of the optimisation processes and their validation. The newly developed universally valid mineral set allows a first temperature estimation for geothermal systems of an unknown subsurface. The set contains twelve mineral phases (quartz, K-feldspar, microcline, albite, muscovite, illite, diaspore, analcime, scolecite, anhydrite, kaolinite, and pyrophyllite). These mineral phases are selected according to their common existence in geothermal systems worldwide causing similar chemical signatures (in terms of key parameters for multicomponent geothermometry) to fluids of very different settings. They consist of key mineral phases of major rock types, accessory minerals, secondary mineral phases, as well as polymorphic mineral phases. The developed universally valid mineral set is tested and validated at eight different geothermal sites (Iceland, Upper Rhine Graben, German Molasse Basin, Paris Basin, Pannonian Basin, Waiotapu, Miravalles, and El Tatio). This mineral set combines mineral phases of the associated geology of the settings (basaltic, volcanic, marine, and continental facies as well as crystalline basement).

The functionality of the optimisation processes is tested by perturbing an equilibrated synthetic mineral assemblage. Thus, four critical parameters are invest-

igated: the salinity of the fluid, changes in the mass of water (steam loss and dilution), as well as variations in the pH value, and the aluminium concentration. First, each optimisation process is performed individually. Perturbation of the salinity does not affect the temperature estimation of MulT_predict. Perturbations in the mass of water, the pH value, and aluminium concentration lead to an average temperature estimation of 145 °C and therefore, a reasonable error of 3.4%. For the interdependent optimisation process, the back-calculations for pH value and aluminium concentration resemble the initial conditions. The average temperature estimation is 147 °C, 3 K less than the initial equilibrium temperature of 150 °C. Considering all cases, MulT_predict and its built-in optimisation processes provide precise temperature estimations. Thereby, the interdependent optimisation process can back-calculate reservoir conditions more accurately than an individual parameter optimisation. However, the calculation of the interdependent optimisation process is more computation-intensive. Therefore, the number of coupled sensitive parameters should be optimised jointly. Overall, the introduced universally valid mineral set expands the usability of MulT_predict and its applicability for the user. In addition, the effectiveness of interdependent optimisation processes is verified, resulting in improved temperature estimations. Eventually, MulT_predict is a fully integrated comprehensive multicomponent geothermometer.

ACKNOWLEDGEMENTS

This study is part of the subtopic “Geoenergy” in the program “MTET - Materials and Technologies for the Energy Transition” of the Helmholtz Association. The support from the program is gratefully acknowledged. We also thank multiple power plant operators in the German Molasse Basin for the internal communication and data. We also want to thank the Editor-in-Chief Dr. Chris Bromley and Associate Editor Dr. Halldór Ármannsson, as well as our two anonymous reviewers for their constructive remarks raising the quality of the manuscript.

6

DEEP LEARNING AND GEOCHEMICAL MODELLING AS TOOLS FOR SOLUTE GEOTHERMOMETRY

This chapter was published in Proceedings European Geothermal Congress 2022. The ISBN code is 978-2-9601946-2-3.

ABSTRACT

Geothermometry is constituted one of the most important geochemical tools for reservoir exploration and development. Solute geothermometers are used to estimate the temperature in the subsurface. Therefore, the chemical composition of a discharging geothermal fluid is used to infer the temperature of the reservoir. Changes in the chemical composition because of boiling, degassing, and dilution are disturbing the equilibrium state within the fluid leading to uncertainties in the temperature estimation. Especially, the pH value, the aluminium concentration, as well as boiling and dilution are parameters prone to changes. These parameters are elaborated in the geochemical modelling process to optimise these values to fit their in-situ reservoir conditions again. This geochemical modelling method can be used for multicomponent geothermometers leading to more robust and precise temperature estimations. However, this process is time-consuming, and geochemical as well as mineralogical knowledge is beneficial. Consequently, the field of artificial intelligence offers powerful methods to solve complex issues, even considering multiple unknowns. Therefore, a new solute geothermometer based on a deep learning algorithm is developed. This neural solute geothermometer is tested and compared to the optimised multicomponent geothermometer and in-situ temperature measurements concluding in a new generation of solute geothermometer as precise as an optimised multicomponent geothermometer but much easier and faster in its applicability.

6.1 INTRODUCTION

The temperature determination of the reservoir is a major factor in the assessment of geothermal reservoirs, especially during the exploration phase. Solute geothermometry can provide such temperature estimations assuming temperature-dependent equilibrium reactions between the geothermal fluid and the host rock minerals. Originally, conventional geothermometers were introduced using cation ratios or single mineral phases for temperature estimation (Arnórsson, 2000c; Fournier and Truesdell, 1973; Giggenbach, 1988).

Followed by a more robust method introduced by Reed and Spycher (1984) using temperature-dependent saturation indices of multiple mineral phases for reservoir temperature estimation. This led to the development of multicomponent geothermometers such as MulT_predict, GeoT, and RTEst (Palmer, 2014; Spycher et al., 2014; Ystroem et al., 2020). In an ongoing development, optimisation processes are implemented to improve the accuracy of these multicomponent geothermometers (Spycher et al., 2016; Ystroem et al., 2021). Especially, steam loss and dilution, as well as trace element concentrations (e.g. Al, Fe, or Mg), and pH value (e.g. degassing) are prone to perturbation of the in-situ equilibrium state, which therefore is back-calculated within the different multicomponent geothermometers. In some cases, these optimisation processes are computationally intensive, when applied interdependently while there are unknowns like the mineralogy of the reservoir (Ystroem et al., 2022). Thus, a new solute neural geothermometer is developed. Artificial neural networks (ANN) are designed to solve complex issues incorporating unknowns (Goodfellow et al., 2016). In addition, a trained network is able to handle a large amount of data efficiently conducting reservoir temperatures estimations. Further, both methodologies are compared to evaluate the temperature estimations. Therefore, a case study of temperature estimations is conducted based on high-quality data from Iceland. The dataset consists of fluid samples from geothermal wells and their in-situ temperature measurements are given by Arnórsson et al. (1983), Gudmundsson and Arnórsson (2002) and Óskarsson et al. (2015).

6.2 METHOD AND DATA

Solute geothermometry is based on the temperature-dependent solubility of mineral phases with the surrounding fluid. Under unperturbed conditions, an equilibrium state between the dissolved element concentration of the fluid and the reservoir rock is reached (Fournier and Truesdell, 1974). Therefore, element ratios, as well as individual solute mineral phases can be used to determine the temperature of the reservoir.

6.2.1 *Multicomponent geothermometry*

To increase the robustness and the precision of geothermometry the solubility of multiple mineral phases can be evaluated simultaneously. In this approach, the saturation indices SI of the mineral phases are evaluated over a predefined temperature range (Reed and Spycher, 1984). The geochemical equilibrium is reached when the measured ion activity product IAP is equal to the temperature-dependent thermodynamic constant $K(T)$ (Equation 6.1). In this case, the SI of

the mineral phase equals zero ($SI = 0$).

$$SI(T) = log\left(\frac{IAP}{K(T)}\right) \tag{6.1}$$

Immature fluids or secondary processes shift the fluid from its equilibrium state. Especially, secondary processes like phase segregation, boiling, mixing, dilution, as well as precipitation of mineral phases and complex building lead to perturbation and thus to uncertainties in the reservoir temperature prediction (Arnórsson et al., 1990; Cooper et al., 2013; Nitschke et al., 2017; Peiffer et al., 2014). Optimisation processes are able to reconstruct equilibrium state conditions assuming the individual mineral equilibrium temperatures converge to an equal overall reservoir temperature. This is achieved by varying sensitive parameters (pH value, aluminium concentration, and the fluid fraction) interdependently around the initial conditions until a global minimum between the equilibrium states of the mineral phases is reached (Ystroem et al., 2022). Figure 6.1 illustrates the output of MulT_predict. In a), the saturation indices of the reservoir mineralogy are plotted against temperature. The intersection with the dashed line represents the equilibrium state in the reservoir. Part b) shows the optimisation process; sensitive parameters (pH-value, aluminium concentration, dilution, steam loss) are simultaneously optimised and evaluated. The in-situ reservoir conditions are assumed to be the global minimum of temperature differences between the mineral phases. Plot c) shows the statistical evaluation of the optimisation. The root mean square (RMSE), standard deviations (SDEV), median (RMED), and the mean (MEAN) of the saturation indices are calculated and plotted against the temperature. In picture d), the result of the temperature estimation is shown. The box plot comprises the equilibrium temperatures of the best fitting reservoir conditions of the mineral set. Depending on the optimisation range, these optimisation processes can be computational time intensive. For each optimisation step, the calculations are computed interdependently increasing the time by the power of one for each sensitive parameter.

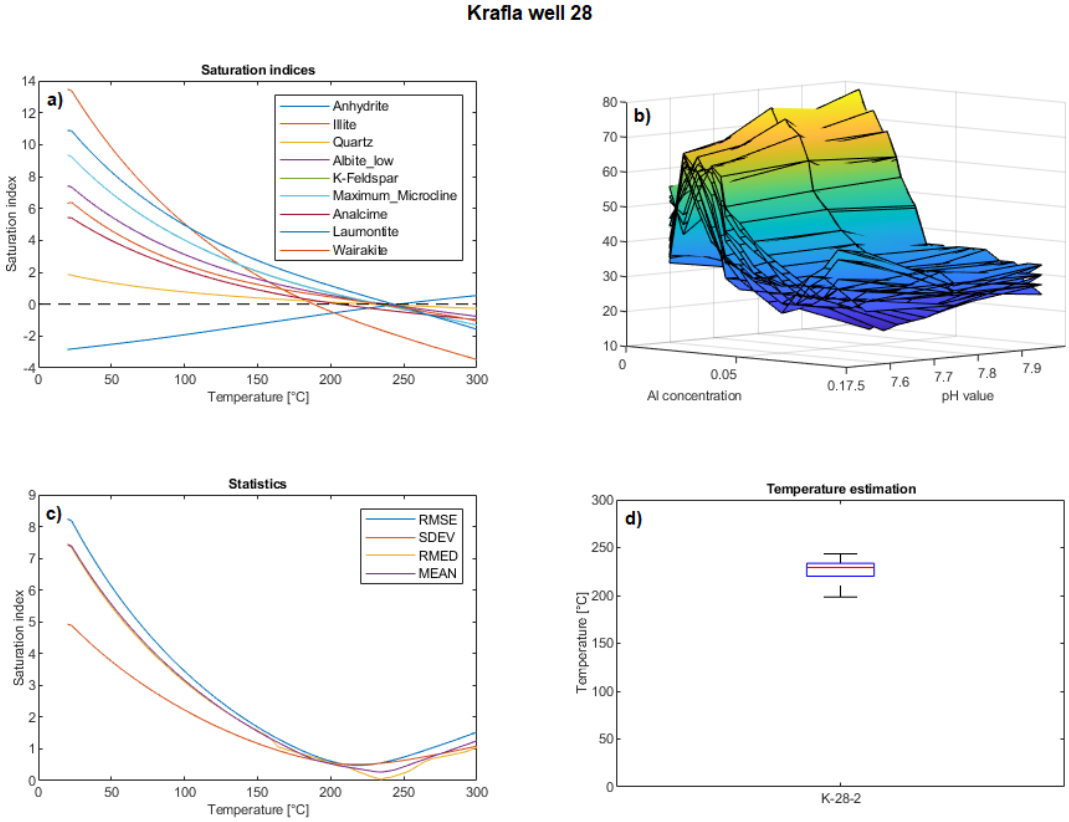


Figure 6.1: Example of the output of MulT_predict: a) Saturation indices of the reservoir mineralogy against temperature. The intersection with the dashed line represents the equilibrium state in the reservoir. b) Interdependent optimisation process of pH-value, aluminium concentration, dilution, and steam loss, the global minimum represents reservoir conditions of the sensitive parameters. c) Statistical evaluation of the optimisation (root mean square, standard deviations, median, and the mean of the saturation indices). d) Result of the best fitting temperature estimation as a box plot.

6.2.2 *Artificial neural network geothermometry*

Regarding the increasing computational time for interdependent optimisation processes of sensitive parameters in multicomponent geothermometry, artificial intelligence can perform calculations even for a large amount of data more efficiently (Goodfellow et al., 2016). Therefore, an ANN is trained with geochemical parameters of the fluid composition and in-situ temperature measurements of high-quality fluid data. A dataset of geothermal wells of Iceland given by Arnórsson et al. (1983), Gudmundsson and Arnórsson (2002) and Óskarsson et al. (2015) is compiled as input data for the network. After screening the input data, the selection of geochemical parameters as well as the network structure must be elaborated. Afterward, the network is trained with a majority (70%) of the data. The rest of the data is used for testing (20%) and validation (10%) of

the ANN. The goal is to train the ANN to estimate the reservoir temperature without overfitting the algorithm. The result of the training of the ANN, as well as the performance of the trained geothermometer, are illustrated in Figure 6.2.

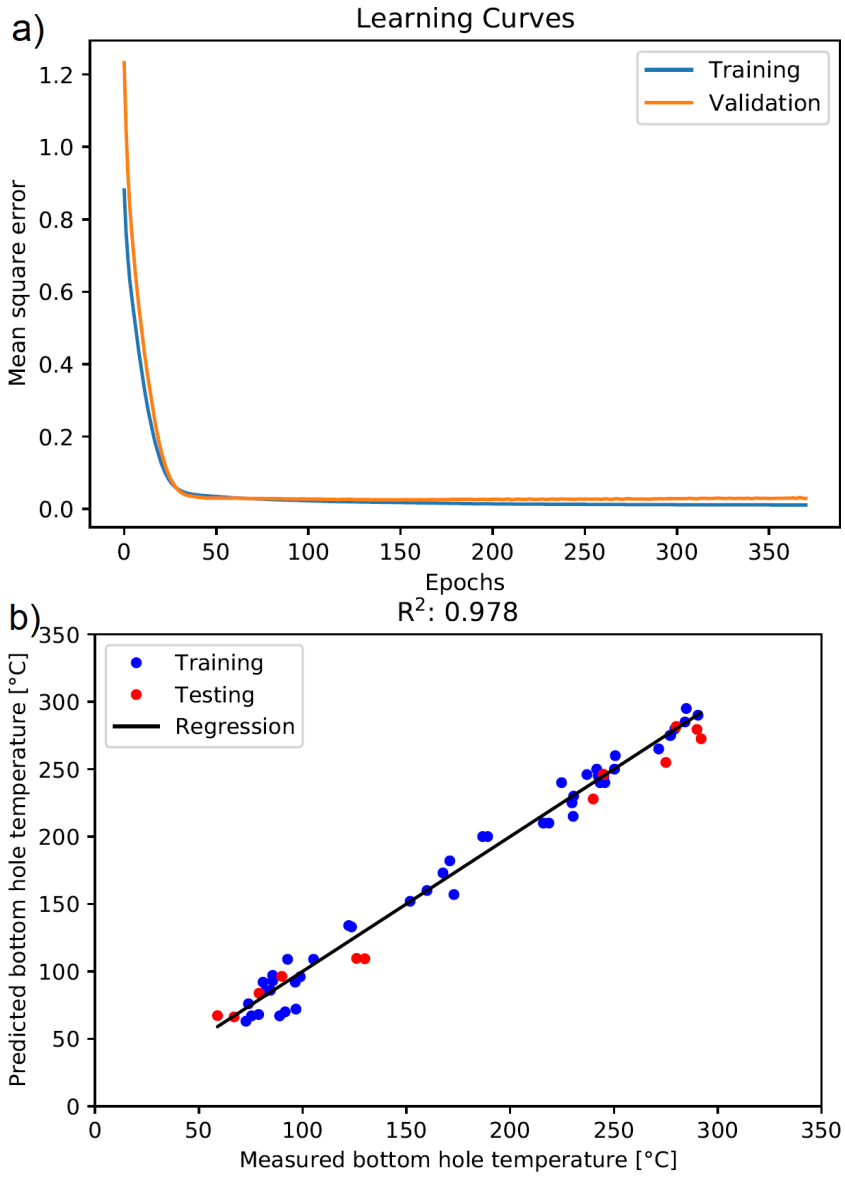


Figure 6.2: a) Mean square error against the epochs of the training. The early stopping function prevents the ANN from overfitting. b) Predicted versus measured bottom hole temperature. The testing data fits the ANN with R^2 = 0.978.

In Figure 6.2a), the mean square error is plotted over the epochs of the training phase. The even trend of the validation curve shows the adaption of the network, while not overfitting the ANN. In Figure 6.2b), the predicted bottom hole temperature is plotted over the measured bottom hole temperature. The blue dots represent the data used for the validation, fitting the trained geothermometer tool with a coefficient of determination R^2 of 0.978.

6.3 RESULTS

Both methods, the solute multicomponent geothermometer as well as the ANN geothermometer, are used to estimate the temperature of a known reservoir in Iceland. Therefore, four samples of Krafla and Reykjanes are computed. In Figure 6.3, the resulting temperature estimations are shown. The temperature estimation of the multicomponent geothermometer MulT_predict is visualised by blue box plots while the red line indicates the median temperature. The temperature estimation of the ANN is illustrated by a green circle with an inner black dot. The measured in-situ temperatures of the wells are indicated by an orange box given by the inflow temperatures of the geothermal fluid at permeable horizons in the open hole section.
In all cases, the median temperature of the multicomponent geothermometer is fitting the in-situ temperatures. For the ANN geothermometer, three of four temperature estimations are matching the measured temperature range. Only for well 28 at Krafla, the ANN is underestimating the temperature by a maximum of 15 K.

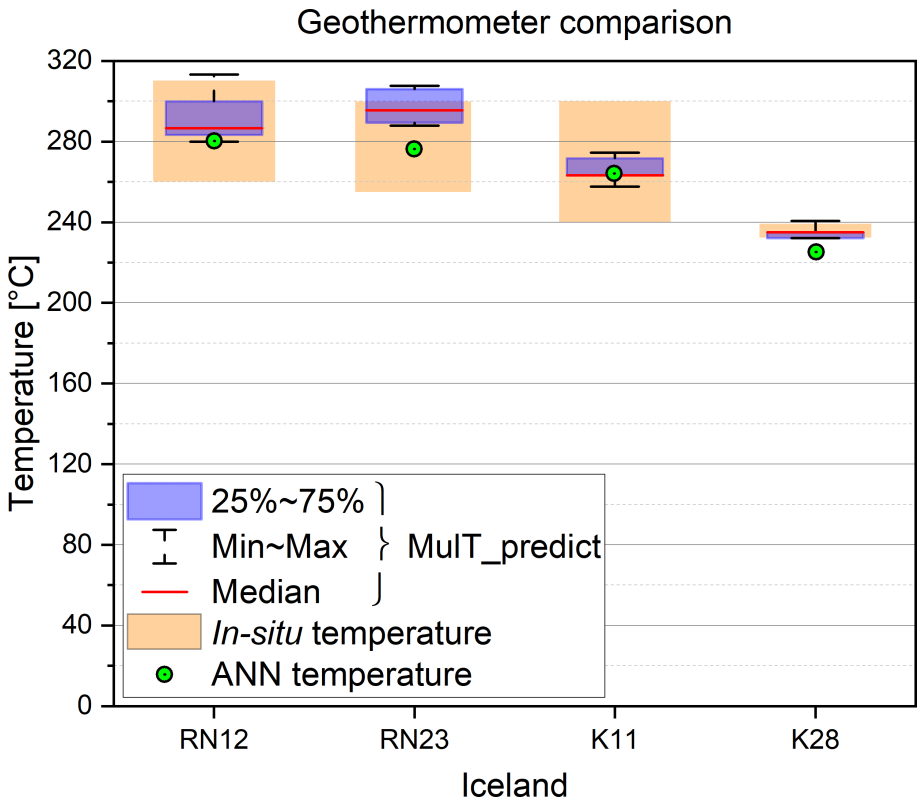


Figure 6.3: Comparison of multicomponent geothermometer MulT_predict (box plots) and the ANN geothermometer (green circles). The in-situ temperature measurement is indicated by the orange box for the four wells in Krafla and Reykjanes, Iceland.

6.4 DISCUSSION AND CONCLUSION

Comparing both geothermometer approaches in Figure 6.3, the solute multicomponent geothermometer is statistically more robust than the ANN. Nevertheless, MulT_predicts' temperature estimations have to be optimised to obtain precise results. Therefore, a mineralogical pre-knowledge of the hosting reservoir rock would be beneficial. In addition, multiple interdependent optimisation processes increase the computational time of the calculations. Regarding this, the newly developed ANN geothermometer can compute temperature estimations more efficiently while handling large amounts of data. In addition, no pre-knowledge nor optimisation is necessary. Nevertheless, the ANN has to be trained with high-quality data containing accurate in-situ temperature measurements.

While geochemical modelling of sensitive parameters in solute geothermometry is the key factor for accurate reservoir temperature estimation, a further improved and adequate ANN geothermometer is the next step in the evolution of solute geothermometry.

ACKNOWLEDGMENTS

This study is part of the subtopic "Geoenergy" in the program "MTET - Materials and Technologies for the Energy Transition" of the Helmholtz Association.

7

ANNRG - AN ARTIFICIAL NEURAL NETWORK SOLUTE GEOTHERMOMETER

This chapter was published in Applied Computing and Geosciences (2023) Volume 20, 100144; DOI: 10.1016/j.acags.2023.100144

ABSTRACT

Solute artificial neural network geothermometers offer the possibility to overcome the complexity given by the solute-mineral composition. Herein, we present a new concept, trained from high-quality hydrochemical data and verified by *in-situ* temperature measurements with a total of 208 data pairs of geochemical input parameters (Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, SiO_2, and pH) and reservoir temperature measurements being compiled. The data comprises nine geothermal sites with a broad variety of geochemical characteristics and enthalpies. Five sites with 163 samples (Upper Rhine Graben, Pannonian Basin, German Molasse Basin, Paris Basin, and Iceland) are used to develop the ANN geothermometer, while further four sites with 45 samples (Azores, El Tatio, Miavalles, and Rotorua) are used to encounter the established artificial neural network in practice to unknown data. The setup of the application, as well as the optimisation of the network architecture and its hyperparameters, are stepwise introduced. As a result, the solute ANN geothermometer, AnnRG (Artificial neural network Regression Geothermometer), provides precise reservoir temperature predictions (RMSE of 10.442 K) with a high prediction accuracy of R^2 = 0.978. In conclusion, the implementation and verification of the first adequate ANN geothermometer is an advancement in solute geothermometry. Our approach is also a basis for further broadening and refining applications in geochemistry.

7.1 INTRODUCTION

Geothermometry constitutes an important geochemical tool for reservoir temperature determination. The unperturbed reservoir temperature predictions are a key parameter for the exploration and development of geothermal resources in the subsurface (Arnórsson, 2000a). In solute geothermometry, the geochemical composition of the geothermal fluid reflects the temperature-dependent equilibrium state between reservoir and fluid (Ellis and Mahon, 1964). The saturation state of mineral phases as well as specific cation ratios are dependent on the

thermal conditions at depth and therefore can be used to infer the reservoir temperature (Fournier and Truesdell, 1974). In the 1960s the first conventional solute geothermometer was presented, which empirically linked the silica concentration in hot springs to the associated quartz equilibrium temperature at depth (Fournier and Rowe, 1966). Since then, continuously new geothermometers have been developed or improved. Most applications of solute geothermometers are based on SiO_2 concentration as well as the major cation ratios of Na/K, Na/K/Ca, and K^2/Mg (Arnórsson, 2000c; Fournier, Potter et al., 1982; Fournier and Potter, 1979; Fournier and Truesdell, 1973; Giggenbach, 1988). Na^+, K^+, Ca^{2+}, and Mg^{2+} are the major cations encountered in crustal rocks and geothermal water, whereas Cl^- and $SO4^{2-}$ are the major anions (Giggenbach, 1988). The monovalent cations, Na^+ and K^+, are mostly controlled by the ratio corresponding to the equilibrium state of albite and K-feldspar (Ellis and Mahon, 1964). In hydrothermal systems, the Ca^{2+} concentration is mostly given by the temperature and salinity-dependent solubility of calcite and calcium-aluminium-silicates (Ellis, 1963; Giggenbach, 1981). Likewise, Mg^{2+} is controlled by the solubility of K-Mg layer silicates, which is chlorite-dependent (Giggenbach, 1988). In geothermal fluids the silica concentration is driven by reservoir conditions regarding chalcedony or quartz equilibrium (Fournier and Rowe, 1966) with pH values in a broad range from acidic to alkaline conditions linked to the activity of hydrogen ions. Like the salinity, the pH value influences the ionic activity of the geothermal fluid (Davies, 1938; Debye and Hückel, 1923).

Due to the implementation of conventional geothermometers on regional geochemical data, the application is prone to variation in the chemical composition of geothermal fluid. This leads to high uncertainties regarding reservoir temperature predictions (Nitschke et al., 2017). Based on the evaluation of the saturation state of multiple aforementioned mineral phases, Reed and Spycher (1984) introduced an alternative approach by computing saturation curves based on thermodynamic solubility data of a set of reservoir minerals against temperature. In contrast to conventional geothermometers based on element concentrations and ratios, the saturation indices of multiple mineral phases need to be calculated to predict the reservoir temperature. Thereby, the aqueous ions concentration of the geothermal fluid is compared to the equilibrium concentration of the temperature-dependent solubility of the mineral phases. The clustering of the mineral's equilibrium temperatures in the chemical system of fluid and reservoir rock are indicating the reservoir temperature. Reed and Spycher (1984) already revealed that the results are statistically more robust than conventional geothermometers, which have often been proven to be afflicted with larger uncertainties (Nitschke et al., 2018, 2017; Pang and Reed, 1998; Pang, 1988).
Nevertheless, the multicomponent geothermometer is still prone to secondary perturbations of the equilibrium of the geothermal fluid, such as mixing, boiling and dilution, precipitation and dissolution, or analytical errors (Pang and Reed,

1998; Pang, 1988). The different tools have been developed with integrated optimisation processes correcting the error of fluid perturbation: 1) WATCH (Bjarnason, 2010), using analyses of the sampled water, gas, condensate, and excess enthalpy to compute the reservoir composition of the fluid (the aqueous speciation, the pH and redox potential, and the partial pressure of gas phases) (Arnórsson et al., 1982). 2) RTEst (Palmer, 2014) using temperature, CO_2 fugacity, and a mixing fraction by minimising the saturation indices of a suggested mineral set (Plamer et al., 2015). 3) iGeoT automatically estimates input parameters, such as the factor of concentration/dilution, the steam weight fraction, as well as input concentrations of aqueous and gas species, using the iTOUGH2 numerical optimisation engine (Spycher and Finsterle, 2016). With PyGeoT, a pre- and post-processing script is developed for automated mineral assemblage selection for GeoT/iGeoT (Olguín-Martínez et al., 2022). 4) MulT_predict being firstly introduced with distinct numerical optimisation for aluminium concentration, pH value, and steam loss/dilution, as well as an individual high temperature (up to 350°C) mineral assemblage for basalt settings (Ystroem et al., 2020). These optimisation processes were merged into an interdependent back-calculation of reservoir conditions (Ystroem et al., 2021). For worldwide applicability, a universally valid mineral assemblage for unknown reservoir composition, a procedure of outlier removal, and the limits of MulT_predict are developed (Ystroem et al., 2022), demonstrating the broad and accurate applicability of solute multicomponent geothermometers with integrated optimisation processes. However, the computational effort rises dramatically with the number of optimisation processes while the back-calculation of fluid perturbations requires necessary geochemical preknowledge.

This situation provides ideal starting conditions for the application of machine learning (ML) algorithms to support automated numerical optimisation programs. Compared to other natural sciences, ML is particularly rarely used in geochemical interpretation. This is largely due to the limited amount of data available, as geochemical analyses are usually difficult to obtain at great expense. At the same time, the heterogeneity of the subsurface and physical parameters limits the extrapolation of the collected data. Also, the application of ML techniques has been strongly increased in geosciences recently. This development is favoured by free software libraries such as *scikit-learn* (Pedregosa et al., 2011) or *TensorFlow* (Abadi et al., 2015) enabling easy access to ML in various programming environments. Likewise, literature and documentation such as Goodfellow et al. (2016) educate the structure and functionality of deep learning algorithms. ML is also favoured by the steady increase in geoscientific, complex data to be statistically evaluated (Dramsch, 2020) and the increase in computational power (Reichstein et al., 2019). The present paper supports this development with respect to geochemical analyses, which have been studied far less than other geoscience data because of their general complexity.

Geochemical fluid analyses are inflicted by a complexity of parameters that are

mostly coupled with each other and the underlying thermodynamics. In the early 2000s, the first attempts have been made to use artificial neural networks (ANNs) for geochemical data analysis. In geothermometry, Ferhat Bayram (2001), Can (2002) as well as Díaz-González et al. (2008) used ML for data regression improving conventional Na/K geothermometer methods. In addition, the first gaseous ANN geothermometer was introduced by Pérez-Zárate et al. (2019), using CO_2, H_2S, CH_4, and H_2 concentration as input parameters. Haklidir and Haklidir (2020) developed the first deep neural network (DNN) solute geothermometer approach based on chemical multi-parameter analysis (pH, electrical conductivity, K^+, Na^+, Li^+, B_{total}, SiO_2, and Cl^-). Nevertheless, this study is based on a small dataset of 83 samples. To increase the training data size to 66 samples, Haklidir and Haklidir (2020) predicted the reservoir temperature of 47 thermal springs by applying the conventional solute geothermometer of Fournier (1977) in advance. Therefore, this DNN approach is also prone to error propagation, since more than 70% of the training data are already reservoir temperature predictions comprising the uncertainties of conventional geothermometers. These predictions are then used to develop a DNN for reservoir temperature prediction for the regional case. Based on the data of Haklidir and Haklidir (2020), Ibrahim et al. (2023) tested five ML algorithms for their temperature prediction performance and the shapley additive explanation (SHAP) (Lundberg and Lee, 2017) to determine the contribution of each input parameter to the ML models. In addition, Altay et al. (2022) also used the data of Haklidir and Haklidir (2020) and others to test multiple machine learning methods to predict reservoir temperatures in Anatolia. A grey wolf optimiser multi-layer perceptron (GWO-MLP) showed good results, which then was further improved (Altay and Altay, 2023).

This manuscript demonstrates the development of an adequate solute ANN geothermometer, called AnnRG, which is only trained using geochemical data and *in-situ* temperature measurements of geothermal sites. No reservoir temperatures have to be predicted in advance to increase the data size, leading to less error propagation. This new ML study focuses on complex and heterogeneous geochemical data comprising thermodynamic coupling of system parameters and element concentrations. In regard to the numerically optimised multicomponent geothermometers, strong benefits in computation time are expected. The motivation is to establish an easy-to-apply ANN solute geothermometer for accurate reservoir temperature prediction without the need for sophisticated geochemical preknowledge such as geochemical equilibrium processes or reservoir mineral assemblages.

7.2 DATA AND METHOD

7.2.1 *Data acquisition*

A key aspect of the development of AnnRG (Artificial neural network Regression Geothermometer) is the acquisition of a high-quality dataset consisting of the chemical analysis of fluid samples and their*in-situ* measured reservoir temperature. This combination including high data quality and large datasets is unfortunately rare and such data are difficult to obtain. In sum, nine geothermal sites with a broad variety of geochemical characteristics and enthalpies are evaluated (cf. Table 7.1). This data can be distinguished into five sites according to their origin (Upper Rhine Graben, Pannonian Basin, German Molasse Basin, Paris Basin, and Iceland), which are used for the development of AnnRG, and further four sites (Azores, El Tatio, Miavalles, and Rotorua) to encounter the ANN in practice to unknown data. The location of the geothermal field is plotted in Figure 7.1.

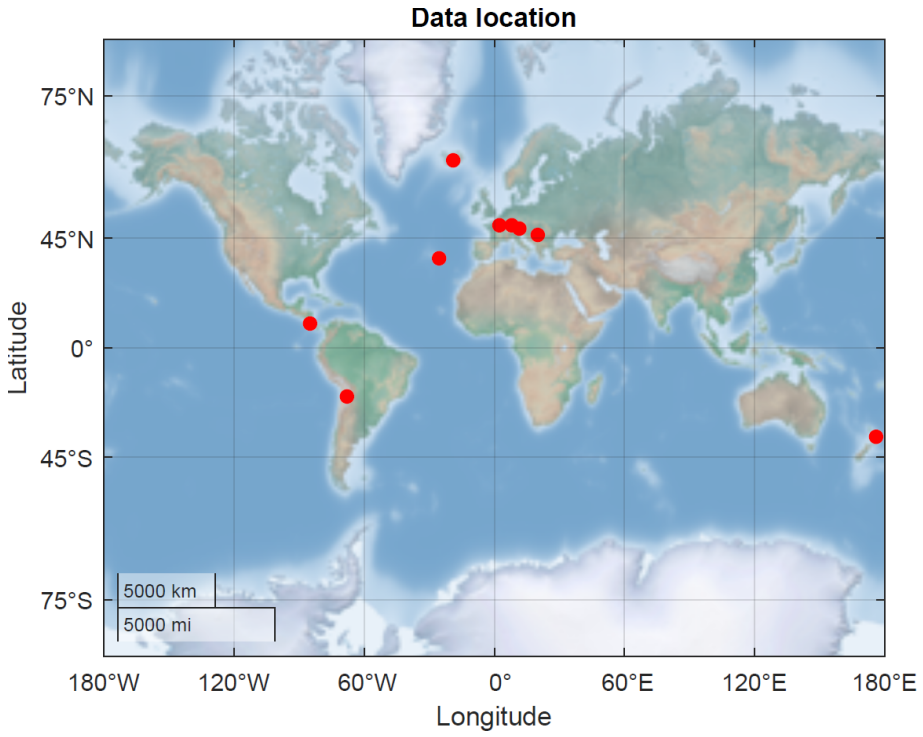


Figure 7.1: Location of the acquired data of the seven geothermal sites for training and verifying the ANN geothermometer.

Table 7.1: Collection of geochemical fluid analyses and measured *in-situ* temperatures of five geothermal sites to merge a dataset as the basis of the ANN geothermometer. In addition, four geothermal sites are used to verify the developed ANN geothermometer (separated by a line). In addition, the available sample size, the lithology, and the temperature range per site is given.

Data	Sample size	Lithology	Temperature range	Reference
Upper Rhine Graben	40	Granite, Germanic Trias facies, Permian facies	41 °C – 200 °C	Aquilina et al. (1997), Dezayes et al. (2013), Pauwels et al. (1993), Sanjuan et al. (2001), Sanjuan et al. (2004), Sanjuan et al. (2006), Sanjuan et al. (2010), Sanjuan et al. (2016), Schindler et al. (2010), Stober and Bucher (2015), Vaute (1998), Vidal and Genter (2018), Vidal et al. (2019)
Pannonian Basin	17	Fluvio-lacustrine facies	44 °C – 128 °C	Varsányi et al. (1997)
German Molasse Basin	9	Marine facies, Malm	36 °C – 153 °C	Ystroem et al. (2022)
Paris Basin	18	Marine facies, Dogger	47 °C – 76 °C	Criaud et al. (1989), Marty et al. (1988), Michard and Bastide (1988)
Iceland	79	Basalt, hyaloclastite	59 °C – 295 °C	Arnórsson (1977), Arnórsson (1978), Arnórsson et al. (1983), Arnórsson et al. (2007), Óskarsson et al. (2015), Gudmundsson and Arnórsson (2002)
Azores	19	Basalt	200 °C – 240 °C	Carvalho et al. (2006)
El Tatio	6	Dacite	180 °C – 253 °C	Ellis and Mahon (1977), Giggenbach (1978)
Miravalles	10	Andesite	215 °C – 255 °C	Dennis et al. (1989), Gherardi et al. (2002)
Rotorua	10	Rhyolite, andesite	121 °C – 211°C	Mroczek et al. (2003)

Well data consisting of the chemical fluid composition and the reservoir temperature is a key factor for avoiding the effects of data perturbation. To develop the geothermometer, five well-studied geothermal fields have been identified from which such data is available. In the Upper Rhine Graben, the geothermal fluids of the rift basin represent a highly saline Na^+-Cl^- type (up to 200 g/l) (Pauwels et al., 1993). They surpass Triassic to Permian sediments, though they are mostly produced from the deeper granitic crystalline basement (Sanjuan et al., 2016; Stober and Bucher, 2015). The Pannonian Basin is a back-arc basin filled with interbedded fluvial and lacustrine sediments. The total dissolved solids (TDS) of the samples are in the range of 1.2 to 4.4 g/l. The fluids are Na^+-HCO_3^- type (Varsányi et al., 1997). The German Molasse Basin is a foreland basin, where the hosting geothermal reservoirs are Jurassic marine facies, mostly Malm (Birner et al., 2011). The fluid chemistry is varying from a northern Na^+-Ca^{2+}-Mg^{2+}-HCO_3^- type to a more saline southern Na^+-(Ca^{2+})-HCO_3^--Cl^- type

(Birner et al., 2011). In the sedimentary Paris Basin, marine Dogger layers are the targeted facies for energy production (Criaud et al., 1989). The Na^+-Ca^{2+}-HCO_3^--Cl^- type fluids show high variability in their salinity (6.4 – 35 g/l) (Michard and Bastide, 1988). Iceland represents basaltic facies on the mid-ocean ridge induced by a hotspot. The high-enthalpy fields are either fed by meteoric or seawater ranging from a Na^+-HCO_3^--Cl^- to a Na^+-Cl^- type (Arnórsson et al., 1983). Based on these five geothermal fields, AnnRG is established.
To encounter AnnRG in practice, four additional sites of unknown data are aggregated and introduced to the established geothermometer. On the Azores, the geochemical data is from wells drilled in the Agua de Pau Massif in the vicinity of the Fogo Volcano. The eruption cycles show a sequence of effusive basaltic to Plinian activity (Carvalho et al., 2006). The TDS varies between 0.07 and 27.1 g/l corresponding to Na^+-HCO_3^- and Na^+-HCO_3^--Cl^- type waters (Cruz and França, 2006). The subduction zones of the Pacific Ring of Fire are dominated by a broad range of andesite to rhyolite. At El Tatio in Chile, the dacitic reservoir is implemented in an ignimbrite formation resulting in a Na^+-Cl^- type brine in the wells (Ellis and Mahon, 1977; Giggenbach, 1978). The geothermal site in Miravalles, Costa Rica, produces from an andesitic reservoir (Dennis et al., 1989; Gherardi et al., 2002). The downhole fluid samples are neutral Na^+-Cl^- type brines (Grigsby et al., 1989). At Rotorua Geothermal Field in New Zealand, the fluid is in contact with rhyolite and ignimbrite domes within the reservoir (Wood, 1992). The chemistry of the fluid is a Na^+-HCO_3^--Cl^- type (Mroczek et al., 2003).

7.2.2 *Implementation of the dataset and data editing*

To set up the dataset, the geochemical analyses of the references are digitised (Table 7.1). The element concentrations and system parameters of the fluid analyses are compiled into a CSV file. Afterwards, the geochemical dataset is sorted by its count of element concentrations and by the geothermal site. In the next step, the measured *in-situ* temperatures are ascertained from literature (Table 7.1) and matched to the associated wells. Then, the geochemical dataset and their reservoir temperatures are aggregated. This aggregated dataset is the foundation of the database, which is customised by selecting the required input parameters. Creating the database, intentionally only parameters such as element concentrations of major cations and anions as well as system parameters, like pH value, which are typically comprised in any standard geochemical fluid analysis, are used. This selection increases the data availability, whereas constituent trace elements such as aluminium or lithium, as well as elements with species of different oxidation states such as sulfur and carbon have been excluded. Such parameters are often not measured and therefore, would lead to an incomplete database and a decrease in sample number. As a compromise of sensitivity and

availability, the following input parameters are selected: Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, SiO_2, and pH. The sensitivity of temperature to the concentration of these parameters makes them also essential for conventional solute geothermometry (e.g. Arnórsson (2000b), Fournier and Truesdell (1973), Giggenbach (1988), Nieva and Nieva (1987) and Spycher et al. (2014)). Regarding the established parameter selection, the dataset is equalised. All unnecessary parameters are deleted. The distribution of each parameter comprising the data is visualised in Figure 7.2.

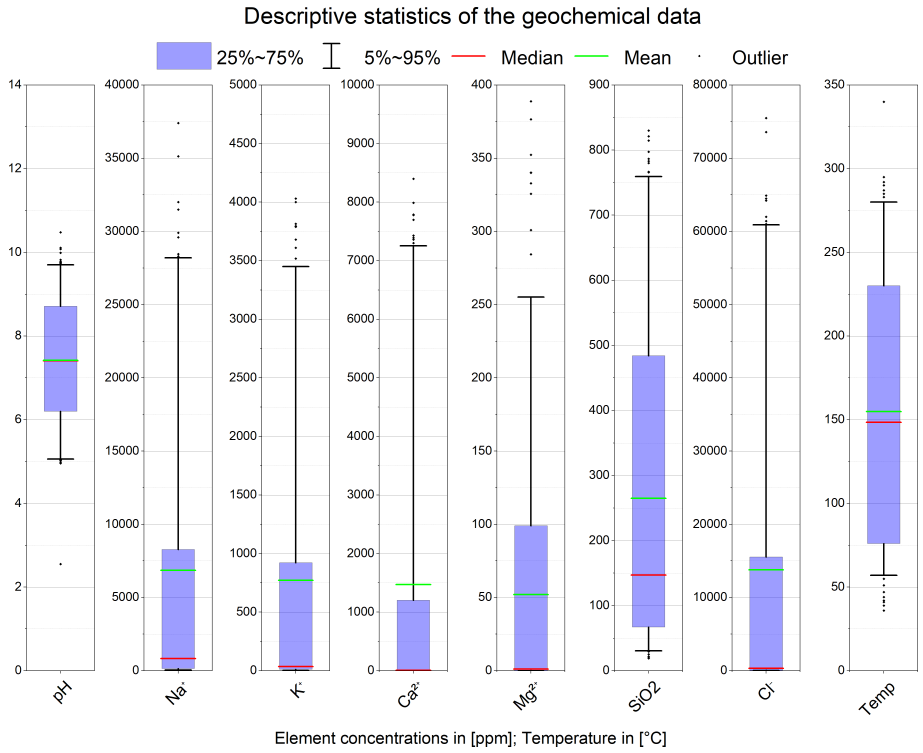


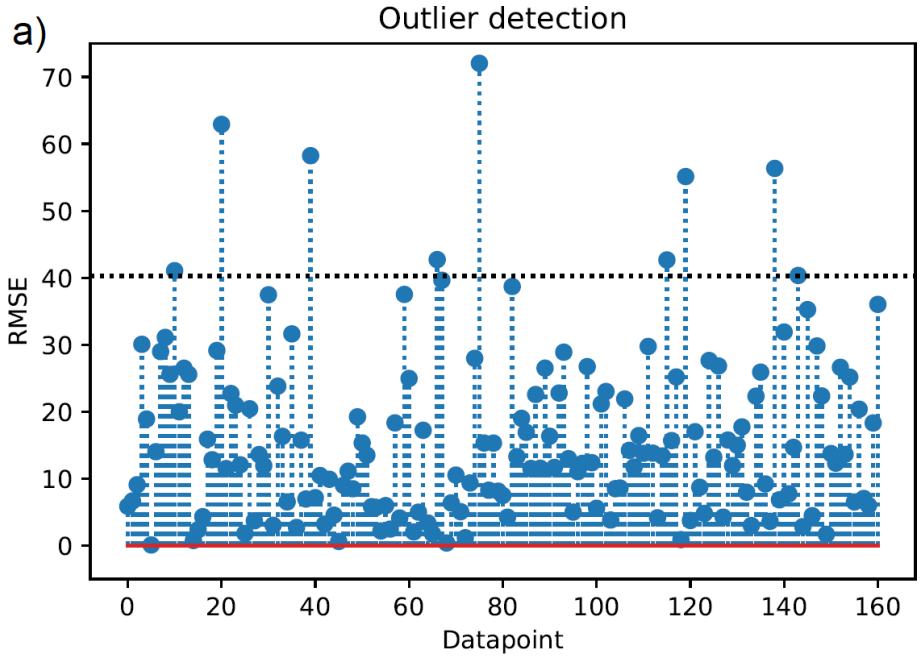
Figure 7.2: Parameter distribution of the data compiled from the literature (cf. Table 7.1) as boxplots, where the median (red line) and the mean (green line) of the dataset are visualised.

Statistical data editing is performed to evaluate the dataset. Throughout the dataset, 163 samples are identified representing geothermal fluid from boreholes with measured *in-situ* reservoir temperatures (c.f. Table 7.1). The data of these samples are stored in matrix $X \in \mathbb{R}$ 163x7 corresponding to the seven previously selected geochemical input parameters (Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, SiO_2, and pH). In addition, the matrix X is extended by the vector of the associated *in-situ* reservoir temperature $Y \in \mathbb{R}$ 163x1. After the merging, the matrix Z contains individual fluid samples in each row while the columns correspond to the geochemical features and the measured *in-situ* temperature. The matrix Z is the database for the training of the ANN. Validation is performed to ensure the quality of the database and remove outliers. Therefore, a basic neural network is trained and tested with standardised features. The range of the input parameters

is scaled and centred to not bias the net while training. In this regard, the mean of a parameter is subtracted from the value and divided by the standard deviation of the parameter. Afterwards, the temperature predictions are compared to the measured *in-situ* reservoir temperatures. To improve the quality of the database and thus, the accuracy of the ANN, wide reservoir temperature differences have to be removed. To identify these outliers, a threshold criterion is defined as the twofold initial root mean square error (RMSE) (Equation 7.1).

$$2\ \mathrm{RMSE} = 2\sqrt{\frac{\sum_0^{n-1}(y-\hat{y})^2}{n}} \tag{7.1}$$

In the outlier detection, the RMSE shows the average distance between the predicted reservoir temperatures $\hat{y}$ and the known *in-situ* reservoir temperatures y regarding the sample size n. The results of each sample are given in Figure 7.3.



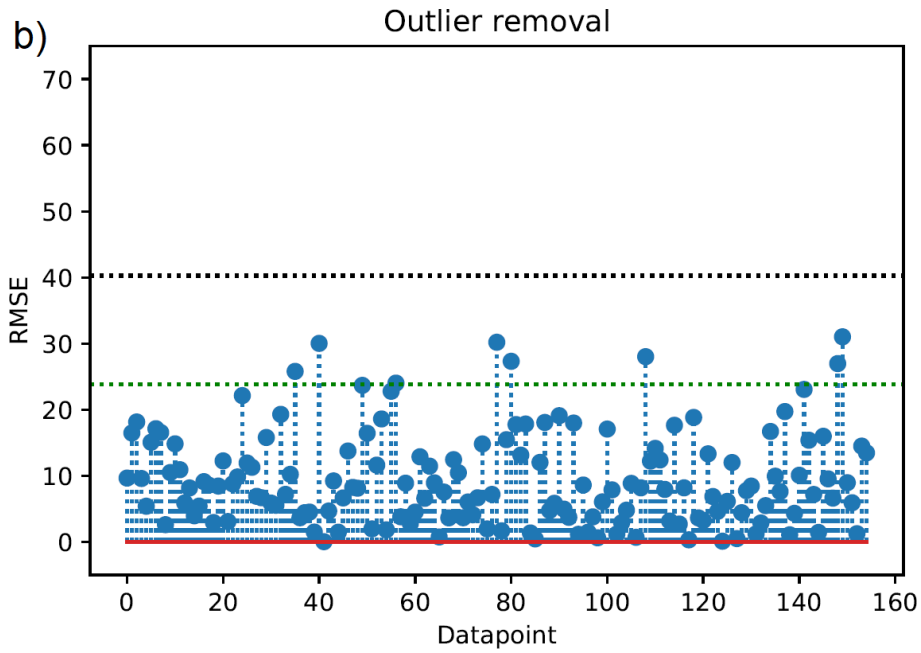


Figure 7.3: a) Outlier detection within the dataset. The threshold is defined as twofold root mean square error (RMSE [K], black dotted line) of the initial predicted temperature difference. b): Dataset after outlier removal with the new twofold RMSE [K] (green dotted line).

The result of the threshold criterion (Equation 7.1) is a temperature deviation of 40.2 K, which is indicated by the dotted black line in Figure 7.3. For each sample, the absolute deviation is plotted (blue data points) throughout the database. As a result, eight outliers exceed the threshold value (Figure 7.3a). After the outlier removal, the neural network is trained and tested again resulting in a more homogenous deviation distribution, where the new twofold RMSE is 23.8 K, which is visualised by the dotted green line (Figure 7.3b). The database (matrix $Z \in \mathbb{R}$ 155x8) is validated for the establishment of the baseline model of the neural network.

7.2.3 *Baseline model, network architecture, and hyperparameter optimisation*

AnnRG is processed in Python 3.8.5 (Van Rossum, Drake et al., 1995). Multiple libraries such as *NumPy* for scientific computation (Harris et al., 2020; Van der Walt et al., 2011), *pandas* for data analysis (Reback et al., 2020), and *Matplotlib* (Caswell et al., 2020; Hunter, 2007) as well as *seaborn* (Waskom, 2021) for visualisation are used. In addition, *scikit-learn* (Pedregosa et al., 2011), as well as *TensorFlow* (Abadi et al., 2015; TensorFlow, 2022) are used for the ML algorithm. For the latter, the high-level application programming interface *Keras* (Chollet et al., 2015) runs on top of it.

The data structure of chemical analyses and associated borehole temperatures qualifies the application of a supervised learning approach. A feedforward multilayer perceptron (MLP) is realised for the regression analysis of fluid chemistry and reservoir temperature. This regression problem is given by the function $f : \mathbb{R}^m \rightarrow \mathbb{R}$ fitting multiple parameter m to one scalar output. To train the MLP, the data is split into three groups: 70% training data, 20% validation data, and 10% testing data. This allocation is chosen to train the model properly, while the hyperparameters, which control the model capacity are not overfitted; as well as a separate test set to monitor the generalisation error. To run the code with the same random seed again, the global and the operation seed are set. This seed randomly assigns the data to each group. Hence, the developed networks are comparable and the progress of optimisations can be displayed. To not bias the network, the data is transformed in a centred and scaled manner like aforementioned. During the training, the MLP repeatedly adjusts weights within neurons while minimising the error between the predicted and measured reservoir temperature via gradient descent (Rumelhart et al., 1986). The gradient descent is calculated using back-propagation computing the sum of partial derivatives of the error, which are depending on the weights of the connected neurons throughout the neural net (Graves, 2012). This error surface is searched for its global minimum as the best-fitting result of the network (Rumelhart et al., 1986). As a result, the input parameters are processed within the hidden layer finding learning rules iteratively matching the output data.

The architecture of the neural network is evolved to establish a reasonable baseline model for the development of AnnRG. Afterwards, the baseline model is refined by hyperparameter optimisation minimising the error within the MLP. The data structure and the intrinsic properties of the temperature estimation problem constrain the network architecture. A simple supervised MLP with fully connected layers is chosen. The input layer is given by seven neurons, representing the predefined input parameters. For the output layer, a single neuron is implemented ensuing in the temperature prediction. The adjustment of the network design is coupled to the number of hidden layers as well as the number of neurons within these layers. Starting with this simple baseline model, the architecture design of the network is improved based on validation error minimisation. The number of hidden layers is tested stepwise via *scikit-learn's* GridSearchCV up to 20 layers. Simultaneously, the number of neurons is varied from 10 to 100 in five increments per layer. During the architecture optimisation, Early Stopping (Wahba, 1987; Yao et al., 2007) is used to avoid overfitting the neural net. The patience of GridSearchCV is set to 20 epochs while the MLP is able to perform up to 300 epochs. To solve the regression problem of the temperature estimation, different activation functions were tested: the rectified linear unit (ReLu) (Agarap, 2018; Hahnloser et al., 2000; Jarrett et al., 2009), sigmoid function, and softmax. As a result, the best-fitting architecture is achieved with

one hidden layer, 80 neurons, and ReLu as the activation function. The measured performance of the established network has a coefficient of determination (R^2) > 0.9, which is acceptable to continue the optimisation. The developed architecture design is used for further refinement of the hyperparameters.

In order to optimise the hyperparameter, multiple parameters were selected and stepwise analysed using *scikit-learn's* GridSearchCV. Overall, the network is trained while using Early Stopping to prevent overfitting. The patience is set to 20 epochs, before the GridSearchCV is active, while the MLP is able to perform up to 300 epochs. The optimisers, as well as their learning rate and batch size, were analysed interdependently.
In the case of *Keras'* optimisers, adaptive moment estimation (Adam) (Kingma and Ba, 2014), stochastic gradient descent (SGD) (Sutskever et al., 2013), as well as root mean square propagation (RMSprop) (Tieleman, Hinton et al., 2012) were compared. To fit the model to an optimal effective capacity, the learning rate within the optimisers is varied from 10^{-4} to 10^{-1} increasing by half an order of magnitude per step. The batch size is varied from 1 up to 32 in 2^n steps. All parameters are varied interdependently to lower the generalisation error while matching the training error. The training error, validation error, and testing error are monitored. The errors of the best-fitting MLP are presented in Table 7.2.

Table 7.2: Illustration of three error types of the best-fitting MLP. The mean absolute percentage error (MAPE), the root mean square error (RMSE), and R^2 are given for the training, validation, and testing of the neural net.

Error type	Training	Validation	Testing
MAPE	0.067	0.092	0.092
RMSE	9.965	11.269	10.091
R^2	0.983	0.972	0.978

The three error types are chosen due to their inherent information in the regression analysis. In Equation 7.2, the mean absolute percentage error (MAPE) indicates the precision of the regression function and its multiple input parameter m to its output $f : \mathbb{R}^m \rightarrow \mathbb{R}$, where $\hat{y}$ is the predicted reservoir temperature, y is the known *in-situ* reservoir temperature, and n is the sample size.

$$\text{MAPE} = \frac{1}{n} \sum_{0}^{n-1} \frac{y - \hat{y}}{|\hat{y}|} \tag{7.2}$$

The RMSE defines the average accuracy between the predicted and the *in-situ* temperature (Equation 7.3).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_0^{n-1}(y-\hat{y})^2}{n}} \tag{7.3}$$

The R^2 shows how accurately the model predicts the measured *in-situ* temperatures (Equation 7.4).

$$\mathrm{R^2} = 1 - \frac{\sum_1^n (y-\hat{y})^2}{\sum_1^n (y-(\frac{1}{n}\sum_1^n y))^2} \tag{7.4}$$

These errors are given for all three phases of training, validating, and testing the neural net. In addition, the shape of the loss function is visualised as learning curves in Figure 7.4.

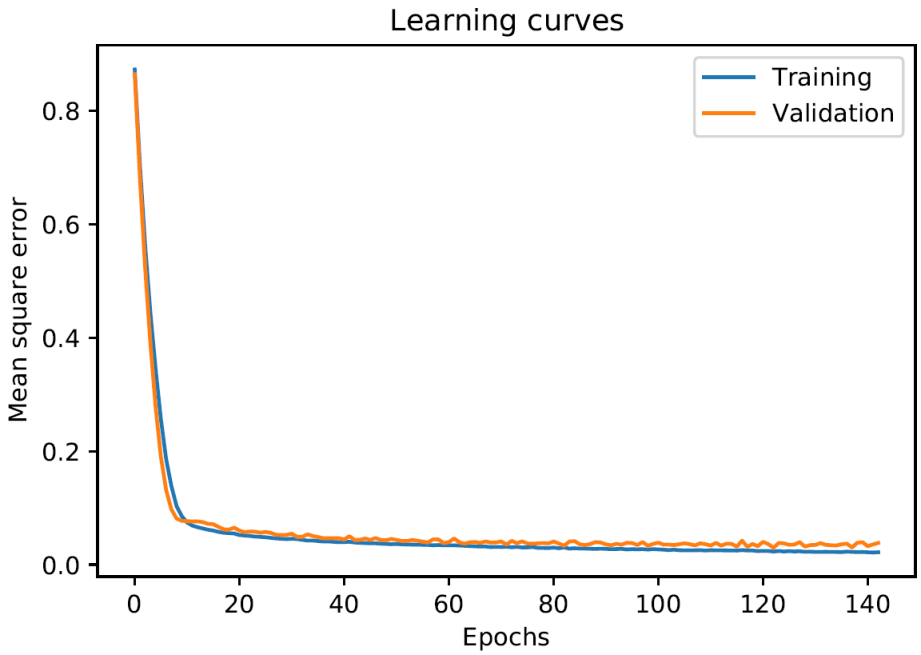


Figure 7.4: Learning curves of the training (blue line) and the validation (orange line) of the ANN. The Cross-entropy loss is visualised against the epochs until the Early Stopping function completes the learning phase.

The learning curves of the training phase (blue line) and the validation phase (orange line) are plotted over the epochs of the MLP. The error minimisation of the loss function is illustrated by the mean square error (MSE). Regarding

Figure 7.4, the Early Stopping function fixed the weights of the 142 epoch resulting in the best hyperparameter configuration of the predefined network architecture. As a result of the GridSearch CV, the activation function Adam (Kingma and Ba, 2014) with a learning rate of 10^{-3} and a batch size of 16 is fitting the regression problem of temperature estimation best. These hyperparameters are further used for AnnRG.

7.3 RESULTS AND DISCUSSION

After the establishment of the neural network architecture and the hyperparameter configuration, the final MLP geothermometer (Figure 7.5) is given by the following configuration (Table 7.3):

Table 7.3: Final network architecture and hyperparameter configuration of AnnRG.

Layers	Neurons	Activation	Optimiser	Learning rate	Batch size	Metrics	Losses
Input layer	7					RMSE	
Hidden layer	80	ReLu	Adam	10^{-3}	16	MAPE	MSE
Output layer	1					R^2	

One input layer with seven neurons, representing the input parameters (Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, SiO_2, and pH), one hidden layer with 80 neurons, and the output layer with one neuron representing the reservoir temperature prediction. The layers are fully connected with ReLu as the activation function and Adam as the optimiser. For the hyperparameter optimisation, a learning rate of 10^{-3} and a batch size of 16 is fitting the best, when the net is trained with Early Stopping (patience 20, and up to 300 epochs). Regarding the sufficient sample size of the dataset ($Z \in \mathbb{R}$ 155x8), this shallow network is less prone to overfitting compared to a deep neural network. This also corresponds to the errors while establishing the net (Table 7.2). The MAPE of the training (0.067), validation (0.092), and testing (0.092) show a close correlation between the input parameters and the regression problem. The slight increase in the MAPE from the training phase to the validation and testing phase is the result of the data splitting (70% training, 20% validation, and 10% testing) and therefore, the stepwise decrease in data. The low deviation of the RMSE for the three phases displays a good accuracy amongst predicted and measured temperatures on average. The variance of 1.304 K from training to validation, as well as the variance of -1.178 K from validation to testing, is in the order of randomisation of the data. In addition, the deviation of their R^2 values is also marginal representing a well-established model. Thus, the accuracy of the temperature prediction is comparable for each step. These results also correspond to the learning curves (Figure 7.4). The smooth shape of the validation curve implies an appropriate tuning of the hyperparameters

fitting the model. The close distance between the validation and training curve equals a small generalisation gap implying a well-balanced model capacity.

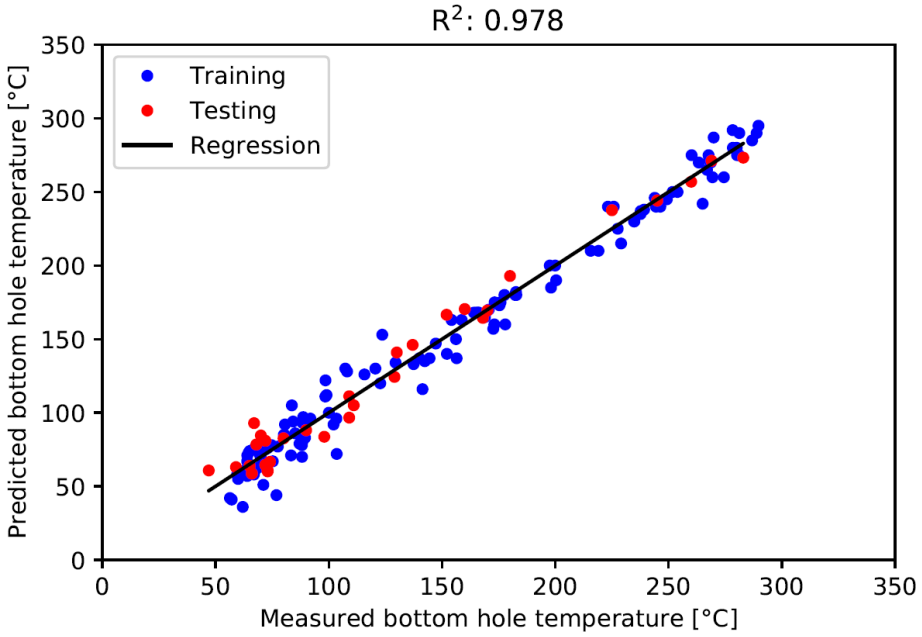


Figure 7.5: Result of the predicted temperature against the measured *in-situ* temperature with $R^2 = 0.978$. The training data (blue points) and testing data (red points) are visualised with the regression line (black line, $R^2 = 1$).

In Figure 7.5, the results of the regression problem of the temperature estimation are visualised for the best-fitting MLP geothermometer. The coloured dots illustrate the actual results of the training (blue), the testing (red), and their deviation of the regression line (cf. Figure 7.6). The homogeneous distribution of the dots, as well as an R^2 of 0.978, approve the developed MLP.

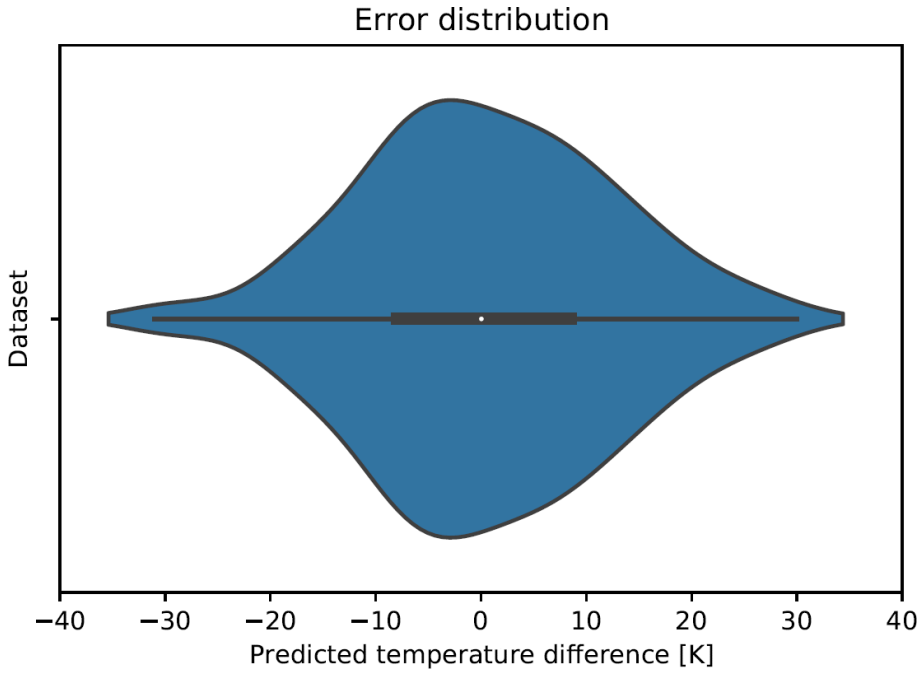


Figure 7.6: Error distribution of the temperature difference in kelvins between the predicted and measured temperature of the entire dataset.

In Figure 7.6, the overall error distribution of the dataset is displayed as a kernel density plot. It visualises the probability density of a random temperature prediction based on the weights. Hence, the curve comprises the positive and negative deviation of each temperature prediction in comparison to the measured reservoir temperature. The shape of the distribution has a slight positive skewness, implying a minor under-prediction of the reservoir temperature. In addition, a box plot gives the axis of symmetry. The maximum temperature differences are ±30 K in each case for two samples. The median is 1.9 K where the box has also a slight positive overhang implying a marginal over-prediction of the reservoir temperature. The interquartile range (IQR) of 17.3 K also shows the accuracy of AnnRG. To investigate the results of the testing phase, the testing error is decoupled and plotted as a histogram in Figure 7.7.

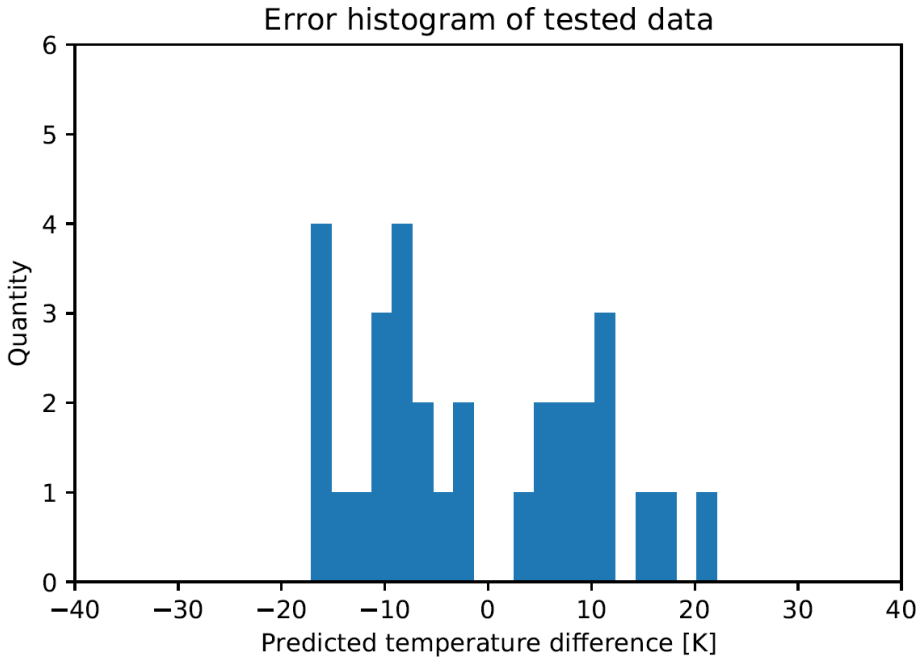
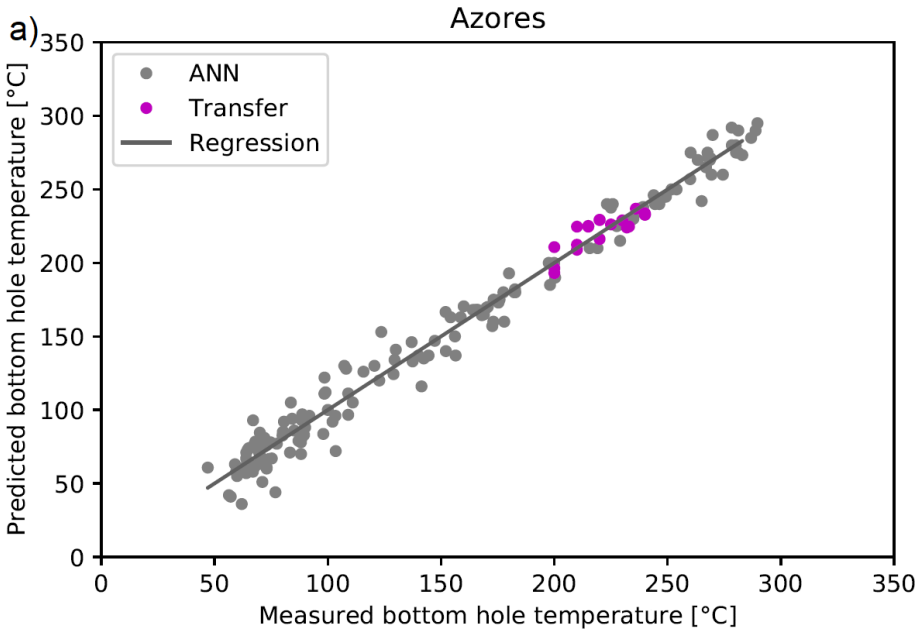
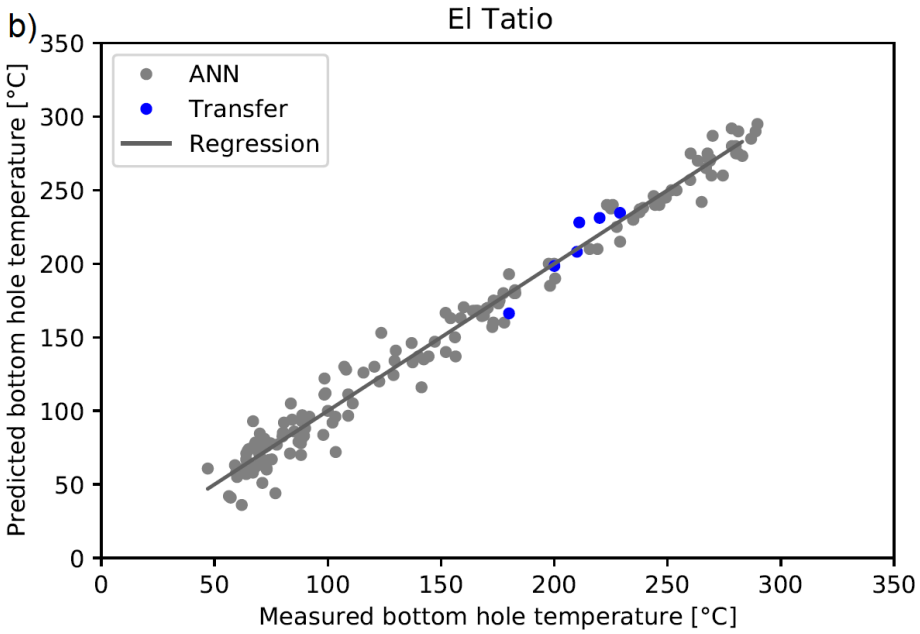


Figure 7.7: Error histogram of the temperature difference in kelvins between the predicted and measured temperature of the test dataset.

The error histogram of the test set (31 samples) is distributed between -18 and 22 K, which is more precise than the overall error distribution of the dataset (Figure 7.6). In line with the positive skewness of the error distribution (Figure 7.6), the temperature estimations of the test set are minorly underestimate the temperature.

To verify AnnRG and encounter the established geothermometer in practice, the geothermometer was applied to unknown data from other geothermal systems: The Azores, Portugal; El Tatio, Chile; Miravalles, Costa Rica; and Rotorua, New Zealand (Table 7.1). The geochemical data of these four geothermal sites are processed by the MLP and compared with the measured *in-situ* temperatures. In Figure 7.8, the results of each site are plotted on top of the developed geothermometer.


a)
Azores
ANN
Transfer
Regression
Predicted bottom hole temperature [°C]
Measured bottom hole temperature [°C]
0
50
100
150
200
250
300
350


b)
El Tatio
ANN
Transfer
Regression
Predicted bottom hole temperature [°C]
Measured bottom hole temperature [°C]
0
50
100
150
200
250
300
350

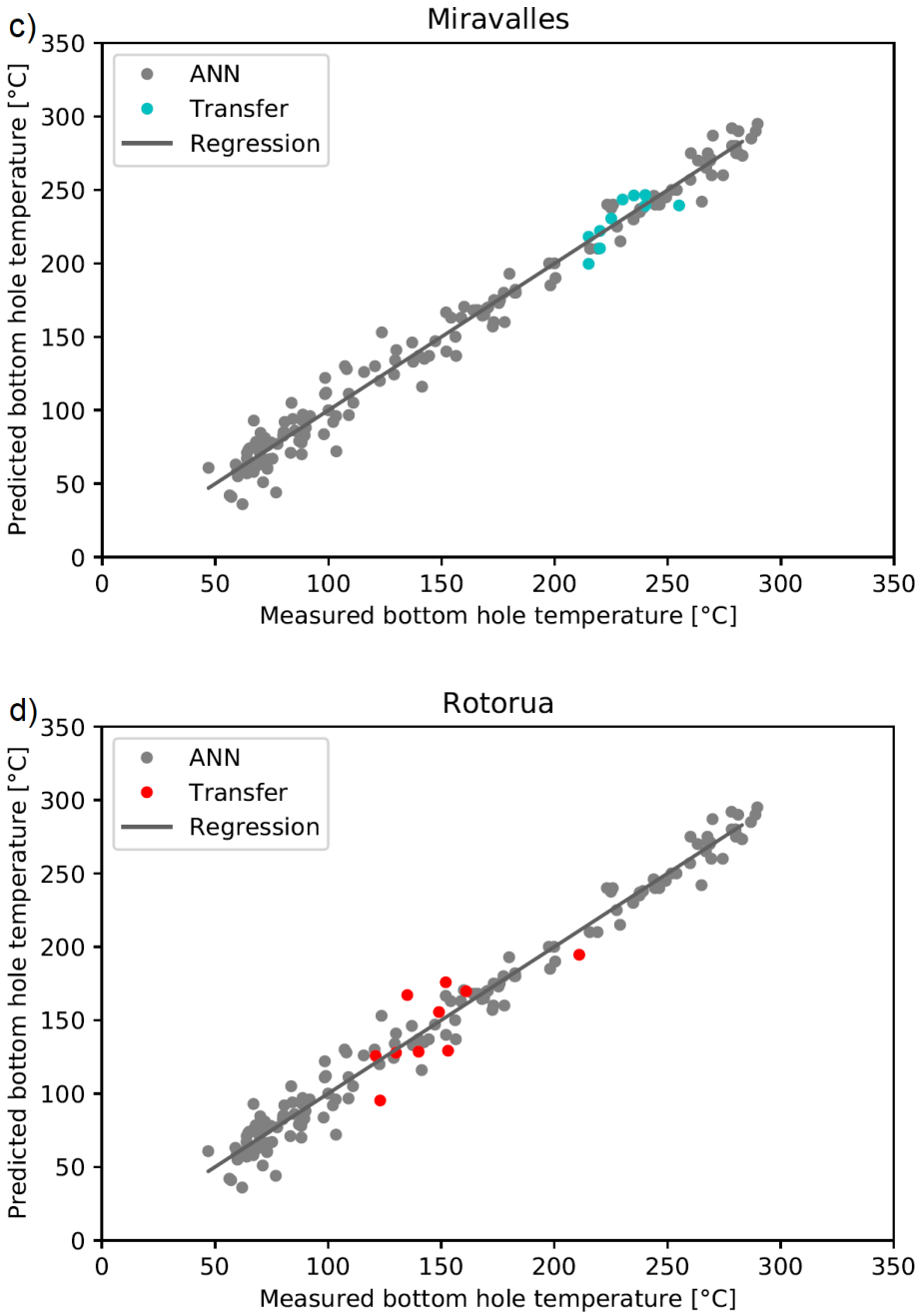


Figure 7.8: The introduction of AnnRG to unknown data: a) Azores, Portugal b) El Tatio, Chile c) Miravalles, Costa Rica d) Rotorua, New Zealand (c.f. Table 7.1). The data (grey points) are visualised with the regression line (grey line, $R^2 = 1$) and the results of the transferred data (coloured points).

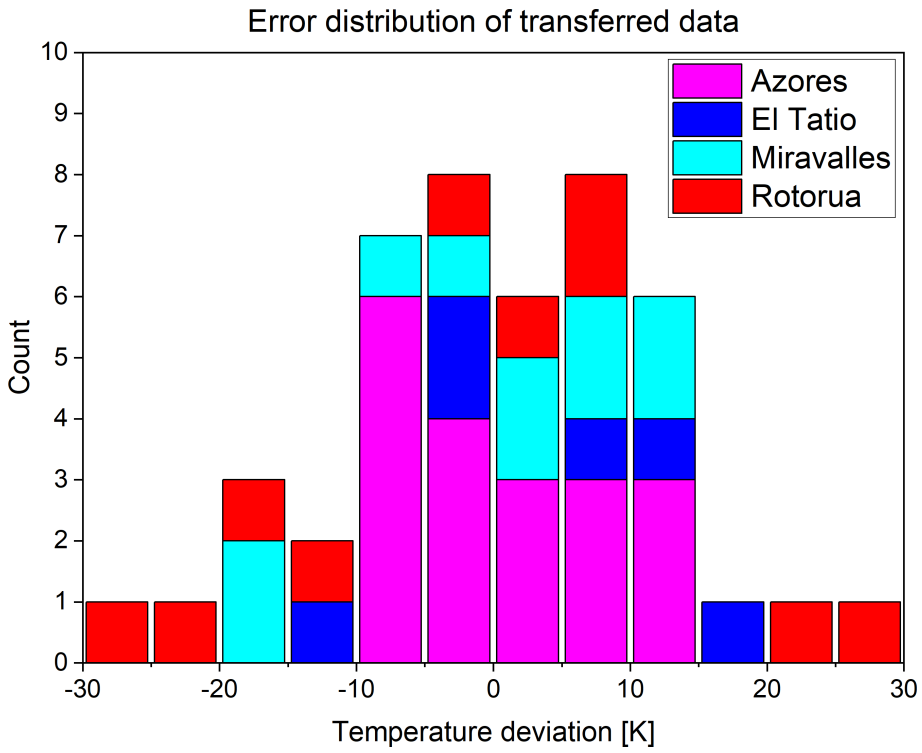


Figure 7.9: Error distribution of the introduced data with a batch size of 5 kelvins. Count of the samples against the temperature deviation of Azores, Portugal (magenta); El Tatio, Chile (blue); Miravalles, Costa Rica (cyan); Rotorua, New Zealand (red).

The predicted temperatures of all four sites (Figure 7.8) fit the measured temperatures in the same order of <±30 K as the MLP error distribution (Figure 7.6). The error distribution of each transferred sample is visualised in Figure 7.9. Especially for the Azores (±11 K), El Tatio (±13 K), and Miravalles (±15 K), the predictions are more precise than the error distribution of the test set (Figure 7.7). Nevertheless, the variation between measured and predicted temperatures at Rotorua (±27 K) is higher than the average test error but lower than the overall error distribution. In summary, the introduction of unknown data to the trained MLP verifies the applicability of AnnRG.

As mentioned in the introduction, the predictions of an ANN are sensitive to the size of the dataset. To evaluate the sufficiency of data pairs, a sensitivity analysis of the sample size of the database is conducted. The validated database is successively reduced in its sample size. In each iteration, the dataset is randomly reduced by one sample and the MLP is trained again with the same model parameters. The R^2 of each recalculation is plotted over the remaining sample size (Figure 7.10).

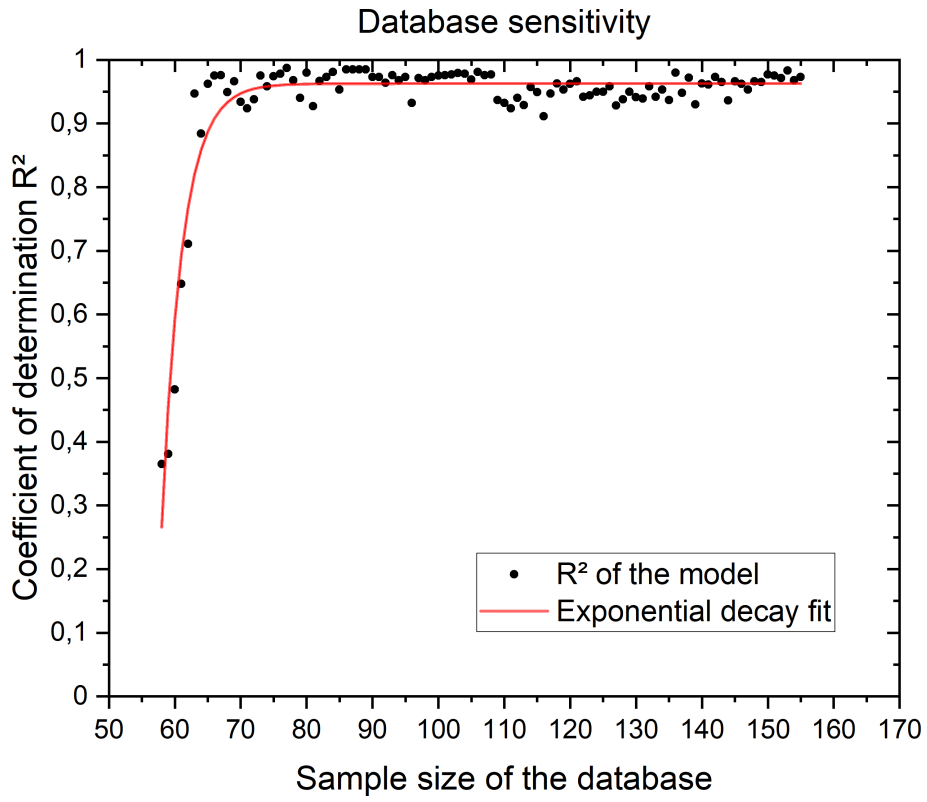


Figure 7.10: Sensitivity analysis of the sample size of the database. The R^2 is calculated stepwise reducing one random sample at a time. The distribution of the R^2 (black dots) is fitted by a curve of exponential decay (red curve).

The results of the R^2 score (black dots) are fitted by an exponential decay function (red line). For a sample size between 155 and 65, the R^2 is varying between 0.911 and 0.987. This spread is attributable to the random incremental reduction of the sample size, while the seed of the database is fixed. Therefore, the samples within the batches are reallocated, which means that the same amount of data is used to train the MLP in every iteration. The R^2 score declines exponentially with a sample size of less than 65 and reaches 0.365 at 58 samples. In conclusion, a sufficient sample size (>65) is required to obtain a suitable database to train, validate and test the MLP.

7.4 CONCLUSION

Up to now, due to typical limited data density and availability the application of ML approaches in geochemistry is very rare. Herein, for the first time, a unique dataset of 208 samples is compiled as the basis for the development of the first adequate solute ANN geothermometer, called AnnRG. This data comprises measured hydrochemical fluid data and corresponding measured *in-situ* temperature from a broad variety of geological settings with high complexity in fluid chemistry (Table 7.1). The MLP is suitable for processing such

heterogeneous data resulting from complex thermodynamic processes of subsurface water-rock interaction. AnnRG is built upon the evaluation of element concentrations and ratios similar to conventional geothermometry. Moreover, the tool performs a multi-parameter analysis training itself without explicitly programming the tool. In contrast to automated and numerical optimised multicomponent geothermometers, no implementation of sophisticated optimisation processes, or adaptation of the reservoir mineral assemblage is needed. In addition, AnnRG evaluates data more efficiently in a timely and computationally manner. Nevertheless, the development of AnnRG is sensitive to the size of the dataset. Regarding this issue, a data size sensitivity analysis was conducted. As a result, >65 data pairs are a sufficient sample size still yielding precise prediction accuracy. Fortunately, nowadays there is a continuous increase in geochemical data, which can be compiled by enhancing the database and improving AnnRG. For AnnRG a standard geochemical fluid analysis comprising system parameters and major element concentration (Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, SiO_2, and pH) is sufficient to predict reservoir temperatures without the need to sample and analyse trace elements, isotopes, or gas phases. After the removal of eight outliers, 155 data pairs are used to train the tool. Then, the applicability and accuracy of the geothermometer are tested on 61 samples comprising 45 samples of new unknown data from four different geothermal fields worldwide. The applicability of the trained MLP is successfully verified resulting in a similar average accuracy (RMSE 9.405 K) than the original dataset (10.442 K). AnnRG is applicable to regional-scale sies independent of geological settings. In addition, this application leads to less error propagation than approaches with predicted reservoir temperatures as input parameters. Eventually, the optimisation of the network architecture as well as the hyperparameters continuously improves the geothermometer throughout its development. AnnRG provides precise reservoir temperature predictions with an IQR of 17.3 K. In addition, the geothermometer has a high prediction accuracy of R^2 = 0.978. Overall, the implementation and verification of AnnRG is an advancement in solute geothermometry showing that ML can identify coherencies between hydrochemical data and temperature. Moreover, the potential for further development regarding new applications in geochemistry and increasing data size is vast. In the future, more advanced statistical evaluation regarding algorithms or optimisations may further enhance the accuracy of ML in geochemistry.

ACKNOWLEDGEMENTS

This research has been funded by the project "MALEG" of the German Federal Ministry for Economic Affairs and Climate Action (BMWK, No. 03EE4041B), implemented by the Project Management Jülich (PtJ). This study is part of the subtopic "Geoenergy" in the program "MTET – Materials and Technologies for the Energy Transition" of the Helmholtz Association. The support from the program is gratefully acknowledged.

CODE AVAILABILITY

Name of the code: AnnRG
Contact: lars.ystroem@kit.edu, Phone: +4972160845290
Program language: Python 3.8.5
Software required: Python libraries: *NumPy, pandas, Matplotlib, seaborn, scikit-learn, TensorFlow, Keras*
Program size: 8 KB
The source codes are available for download at the link: https://gitlab.kit.edu/lars.ystroem/annrg

8

CONCLUSION AND OUTLOOK

In the future, the expansion of geothermal energy will continue to increase due to the global demand for sustainable and renewable energy. Compared to other renewable energy sources, such as photovoltaic or wind power, geothermal energy has the ability to provide baseload electricity and heat independent of meteorological effects, as well as daily and seasonal cycles. The amount of heat energy available in the subsurface is related to the temperature of the reservoir, its permeability, and the presence of a heat transfer fluid. In particular, enhanced geothermal systems (EGS) offer the possibility of decentralised energy production on a global scale. However, the initial investment costs for exploration and construction are high, while the risk of project failure is present.

Geochemical prospection offers a risk minimisation approach to the development of geothermal energy. The chemical investigation of geothermal manifestations such as hot springs, mud pools, geysers, and fumaroles provides important information about the subsurface reservoir without the need for direct access to the source. Solute geothermometry can provide crucial information about the reservoir temperature. To estimate the reservoir temperature of a hydrothermal fluid, its geochemical composition must be analysed. Due to temperature-dependent dissolution and precipitation reactions in the reservoir, the mineral assemblage of the reservoir rock and the aqueous solution are in equilibrium. This chemical equilibrium state in the hydrothermal fluid contains information about the reservoir temperature.

Based on the fundamental assumptions of Fournier et al. (1974) for solute geothermometry, little or no reequilibration, nor mixing of the fluid occurs during migration to the surface. In reality, these ideal conditions could not be maintained. The chemical equilibrium can be changed by phase precipitation, degassing, boiling, dilution, or mixing with shallow and low-mineralised waters, as well as by reequilibration with surrounding rocks during ascent to the surface. These secondary effects can perturb the chemical equilibrium of the hydrothermal fluid. In addition, the sampling and analysis of geothermal fluids can introduce further uncertainties. These perturbations and uncertainties cause an error propagation that affects the reservoir temperature estimations of solute geothermometry. These issues are addressed in this thesis to extend, facilitate and improve the applicability of solute geothermometry.

8.1 MAJOR FINDINGS

In solute multicomponent geothermometry, the reconstruction of undisturbed chemical conditions is essential. Since the real reservoir conditions are only accessible in geothermal wells with wired-box sampling at depth, these reservoir conditions have to be numerically reconstructed. The first study (Chapter 3) addressed this issue and developed a multicomponent geothermometer based on the coupling of MATLAB (MATLAB, 2019) with IPhreeqc (Charlton and Parkhurst, 2011; Parkhurst, Appelo et al., 2013) with integrated optimisation processes for sensitive parameters, called MulT_predict (Appendix B). This geothermometer is based on a standard chemical analysis of the geothermal fluid composition, which includes the main concentrations of mineral-forming elements and system parameters. No sophisticated sampling methods (e.g. wired-box sampling, gas sampling) or analyses (e.g. gas composition, isotopes) are required. In the second study (Chapter 4) the optimisation process of MulT_predict has been further developed and in the third study (Chapter 5) the final solute multicomponent geothermometer is presented and tested on a global scale. MulT_predict improves the usability and applicability of geothermal exploration as an economically efficient tool for reservoir temperature determination.

A series of variations in system parameters and element concentrations were carried out (Chapter 3 & 4) to identify the most sensitive parameters for optimisation. For system parameters, pH, redox potential, dilution, and steam loss were investigated. For element concentrations, aluminium, magnesium, and iron are analysed as major components of the mineral phases, but only as trace elements in the solution. During these variations, all parameters other than the one being analysed are fixed. As a result, pH, aluminium concentration, steam loss, and dilution have the greatest effect on the chemical equilibrium. Therefore, MulT_predict has implemented individual optimisation processes for each sensitive parameter. These individual optimisation processes automatically perform a sensitivity analysis to determine the value that gives the most accurate temperature estimate. In Chapter 4, a more detailed evaluation of the individual optimisation processes revealed the interdependence of the sensitive parameters. Therefore, the working principle is improved to a single multidimensional numerical optimisation. In an internal benchmark (Chapter 5) using a synthetic brine, the individual optimisation processes, and the interdependent optimisation process are analysed and compared. As a result, the interdependent optimisation process is more accurate in back-calculating reservoir conditions, given more precise reservoir temperature predictions. The development and integration of the numerical optimisation process into MulT_predict has facilitated and improved the applicability of solute geothermometry.

To extend the applicability of MulT_predict (Appendix B), the underlying Lawrence Livermore National Laboratory thermodynamic database llnl.dat (Daveler and Wolery, 1992) is extrapolated to estimate reservoir temperatures up to 350 °C (Chapter 3). The database can normally calculate saturation indices from 20 °C to 300 °C. As the aqueous activity coefficient model of llnl.dat is based on the $\dot{B}$ Equation (Helgeson, 1969) (Chapter2), the Debye-Hückel constants A, B and $\dot{B}$ are extrapolated. The extrapolation scheme of Helgeson (1969) is used, resulting in values similar to those calculated and presented by Helgeson et al. (1981). The parameters in llnl.dat are thus implemented and adjusted to improve the ability of MulT_predict to estimate reservoir temperatures for high-enthalpy settings.

In addition, the third study (Chapter 5) also addressed the issue of applicability. Using the extended thermodynamic database (Chapter 3) in combination with the improved optimisation process (Chapter 4), several reservoir temperature estimations are performed on prominent geothermal settings. These settings cover a wide range of lithology, temperature, and salinity on a global scale. As a result, a universally valid mineral set is developed that is capable of estimating reservoir temperatures for unknown subsurface mineralogy. In addition, a statistical outlier removal is introduced to individually refine the universally valid mineral set, resulting in a more accurate multicomponent geothermometer. The combination of the developed mineral set and the outlier removal provides a great advantage where little or no knowledge of the subsurface mineralogy is available. These methods further reduce the barriers to the applicability of solute geothermometry.

In the fourth study (Chapter 6), the machine learning approach of deep learning is used to estimate reservoir temperatures. High-quality geochemical data from Iceland and associated in-situ temperature measurements are used to train and test the approach. As a result, the deep learning approach has estimated reservoir temperatures mostly in the range of the measured in-situ reservoir temperature comparable to the approach of multicomponent geothermometry. This method opens up a new vein in solute geothermometry.

The fifth study (Chapter 7) adopts the idea of the deep learning algorithm (Chapter 6) and develops a new adequate solute artificial neural network geothermometer. A supervised feedforward multilayer perceptron is implemented using the programming language Python (Van Rossum, Drake et al., 1995). This regression-based solute geothermometer is called AnnRG (Appendix C). Its development involves the establishment of a baseline model, the creation and adaptation of a network architecture, and the stepwise optimisation of hyperparameters. As a result, AnnRG represents the first adequate solute artificial neural network geothermometer trained only by in-situ temperature measure-

ments.

To develop AnnRG, a unique dataset is established in Chapter 7. A total of 208 data pairs of geochemical fluid parameters and in-situ temperature measurements are collected and compiled. The dataset covers a wide variety of geothermal settings and fluid chemistries from nine geothermal areas around the world (e.g. Upper Rhine Graben, Pannonian Basin). Such data density and quality are rare in this branch of geochemistry. The processing of such heterogeneous data trains the ANN complex thermodynamic water-rock interactions in the subsurface. Compared to multicomponent geothermometry, the input of AnnRG is reduced to only seven parameters: Na, K, Ca, Mg, Cl, SiO_2, and pH, avoiding the need to measure mineral-forming trace elements. This further enhances AnnRG's usability and applicability as a geothermal exploration tool.

8.2 OUTLOOK

MulT_predict is an optimised solute multicomponent geothermometer. Through the integrated optimisation process, the universally valid mineral set, and its individual refinement, the geothermometer can be used worldwide.
To further refine the reservoir temperature estimations of the tool, additional vulnerable parameters need to be investigated and implemented. In particular, the CO_2 fugacity is an important factor considering the degassing and hence multiphase flow during the ascent to the surface. Therefore, the thermodynamic database has to be revised as the llnl.dat has no correction for the pressure range and is based on the ideal gas law when dealing with fugacity coefficients. Therefore the new database had to be based on the Peng-Robinson equation of state.
To further improve MulT_predict, a real mixing model can be implemented. This mixing model is created by processing the end members of two aqueous solutions. In a stepwise optimisation, the two solutions are mixed in different fractions until the best-fitting ratio is back-calculated.
However, the implementation of additional optimisation processes increases the computational time and complexity of MulT_predict. Eventually, a balance has to be found between optimisation and accuracy.

AnnRG represents a new generation of solute geothermometers. As in other fields, artificial neural networks are becoming increasingly important in solute geothermometry and geochemistry in general. A ANN that estimates reservoir temperatures is the equivalent of an ideal solute geothermometer. The user of the geothermometer does not need any geochemical or geological knowledge. Only a basic, inexpensive geochemical analysis of the hydrothermal fluid is required to compute the reservoir temperature estimation. In addition, the ANN can analyse

large amounts of data within seconds. The potential for further improvements is huge, with increasing data size, new algorithms, and increasing computing power. At present, the accuracy of temperature predictions can be improved mainly by increasing the size of the database. However, this data processing is still done manually.

REFERENCES

ASTM-Standard (2013). 'D1141-98: Standard Practice for the Preparation of Substitute Ocean Water'. In: *ASTM International, West Conshohocken*.

Abadi, M. et al. (2015). *TensorFlow, Large-scale machine learning on heterogeneous systems*.

Agarap, A. F. (2018). 'Deep learning using rectified linear units (ReLu)'. In: *arXiv preprint arXiv:1803.08375*.

Altay, E. V., Gurgenc, E., Altay, O. and Dikici, A. (2022). 'Hybrid artificial neural network based on a metaheuristic optimization algorithm for the prediction of reservoir temperature using hydrogeochemical data of different geothermal areas in Anatolia (Turkey)'. In: *Geothermics* 104, p. 102476.

Altay, O. and Altay, E. V. (2023). 'A novel hybrid multilayer perceptron neural network with improved grey wolf optimizer'. In: *Neural Computing and Applications* 35.1, pp. 529–556.

Aquilina, L., Pauwels, H., Genter, A. and Fouillac, C. (1997). 'Water-rock interaction processes in the Triassic sandstone and the granitic basement of the Rhine Graben: Geochemical investigation of a geothermal reservoir'. In: *Geochimica et Cosmochimica Acta* 61.20, pp. 4281–4295. ISSN: 0016-7037.

Ármannsson, H., Gudmundsson, A. and Steingrímsson, B. (1987). 'Exploration and development of the Krafla geothermal area'. In: *Jökull* 37, pp. 13–30.

Arnórsson, S., Bjarnason, J. O., Giroud, N., Gunnarsson, I. and Stefánsson, A. (2006). 'Sampling and analysis of geothermal fluids'. In: *Geofluids* 6.3, pp. 203–216.

Arnórsson, S. (1977). 'Changes in the chemistry of water and steam discharged from wells in the Námafjall geothermal field, Iceland, during the period 1970–76'. In: *Jökull* 27, pp. 47–59.

Arnórsson, S. (1978). 'Major element chemistry of the geothermal sea-water at Reykjanes and Svartsengi, Iceland'. In: *Mineralogical Magazine* 42.322, 209–220.

Arnórsson, S. (2000a). 'Isotopic and chemical techniques in geothermal exploration, development and use'. In: *International Atomic Energy Agency*, pp. 109–111.

Arnórsson, S. (2000b). 'The quartz-and Na/K geothermometers. I. New thermodynamic calibration'. In: *Proceedings of the world geothermal congress*, pp. 929–934.

Arnórsson, S. (2000c). 'The quartz-and Na/K geothermometers: II. Results and application for monitoring studies'. In: *Proceedings of the world geothermal congress*, pp. 935–940.

Arnórsson, S., Grönvold, K. and Sigurdsson, S. (1978). 'Aquifer chemistry of four high-temperature geothermal systems in Iceland'. In: *Geochimica et Cosmochimica Acta* 42.5, pp. 523–536. ISSN: 0016-7037.

Arnórsson, S., Gunnlaugsson, E. and Svavarsson, H. (1983). 'The chemistry of geothermal waters in Iceland. II. Mineral equilibria and independent variables controlling water compositions'. In: *Geochimica et Cosmochimica Acta* 47.3, pp. 547–566. ISSN: 0016-7037.

Arnórsson, S., Sigurdsson, S. and Svavarsson, H. (1982). 'The chemistry of geothermal waters in Iceland. I. Calculation of aqueous speciation from 0 to 370 degree C'. In: *Geochimica et Cosmochimica Acta* 46.9, pp. 1513–1532. ISSN: 0016-7037.

Arnórsson, S., Stefánsson, A. and Bjarnason, J. O. (2007). 'Fluid-fluid interactions in geothermal systems'. In: *Reviews in Mineralogy and Geochemistry* 65.1, pp. 259–312.

Arnórsson, S., Björnsson, S., Muna, Z. W. and Bwire-Ojiambo, S. (1990). 'The use of gas chemistry to evaluate boiling processes and initial steam fractions in geothermal reservoirs with an example from the olkaria field, Kenya'. In: *Geothermics* 19.6, pp. 497–514. ISSN: 0375-6505.

Banwell, C. J. (1959). *Waiotapu Geothermal Field - Chapter VIII: Programming of Temperature Surveys*. Vol. 155. Bulletin.

Birner, J., Mayr, C., Thomas, L., Schneider, M., Baumann, T. and Winkler, A. (2011). 'Hydrochemie und Genese der tiefen Grundwässer des Malmaquifers im bayerischen Teil des süddeutschen Molassebeckens'. In: *Z Geol Wiss* 39, pp. 291–308.

Bjarnason, J. (2010). 'The chemical speciation program WATCH, version 2.4'. In: *ISOR–Iceland GeoSurvey, Reykjavik, Iceland*.

Blanc, P (2008). 'Thermoddem: Sélection de propriétés thermodynamiques pour les principales espèces aqueuses et minérales porteuses de fer'. In: *Rapport final. Rapport BRGM/RP-56587-FR*.

Blanc, P., Lassin, A., Piantone, P., Azaroual, M., Jacquemet, N., Fabbri, A. and Gaucher, E. (2012). 'Thermoddem: A geochemical database focused on low temperature water/rock interactions and waste materials'. In: *Applied Geochemistry* 27.10, pp. 2107–2116. ISSN: 0883-2927.

Brown, K. (2013). *Mineral scaling in geothermal power production*. United Nations University Reykjavik, Iceland.

Can, I. (2002). 'A new improved Na/K geothermometer by artificial neural networks'. In: *Geothermics* 31.6, pp. 751–760. ISSN: 0375-6505.

Carvalho, M., Forjaz, V. and Almeida, C. (2006). 'Chemical composition of deep hydrothermal fluids in the Ribeira Grande geothermal field (São Miguel, Azores)'. In: *Journal of Volcanology and Geothermal Research* 156.1. Volcanic geology of the Azores Islands, pp. 116–134. ISSN: 0377-0273.

Caswell, T. A., Droettboom, M., Lee, A., Hunter, J., Sales De Andrade, E., Firing, E., Hoffmann, T., Klymak, J., Stansby, D., Varoquaux, N., Hedegaard Nielsen, J., Root, B., May, R., Elson, P., Seppänen, J. K., Dale, D., Lee, J.-J., McDougall, D., Straw, A., Hobson, P., Gohlke, C., Yu, T. S., Ma, E., Vincent, A. F., Silvester, S., Moad, C., Kniazev, N., Hannah, Ernest, E. and Ivanov, P. (2020). *matplotlib/matplotlib: REL: v3.3.2*. Zenodo. Version v3.3.2.

Charlton, S. R. and Parkhurst, D. L. (2011). 'Modules based on the geochemical model PHREEQC for use in scripting and programming languages'. In: *Computers & Geosciences* 37.10, pp. 1653–1663. ISSN: 0098-3004.

Chollet, F. et al. (2015). 'Keras documentation'. In: *keras. io* 33.

Cooper, D. C., Palmer, C. D., Smith, R. W. and McLing, T. L. (2013). *Multicomponent equilibrium models for testing geothermometry approaches*. Tech. rep. Idaho National Lab.(INL), Idaho Falls, ID (United States).

Criaud, A., Fouillac, C. and Marty, B. (1989). 'Low enthalpy geothermal fluids from the paris basin. 2 — Oxidation-reduction state and consequences for the prediction of corrosion and sulfide scaling'. In: *Geothermics* 18.5, pp. 711–727. ISSN: 0375-6505.

Cruz, J. V. and França, Z. (2006). 'Hydrogeochemistry of thermal and mineral water springs of the Azores archipelago (Portugal)'. In: *Journal of Volcanology and Geothermal Research* 151.4, pp. 382–398. ISSN: 0377-0273.

Daveler, S. A. and Wolery, T. J. (1992). *EQPT, a data file preprocessor for the EQ3/6 software package: user's guide and related documentation (version 7.0)*. Tech. rep. Lawrence Livermore National Lab.

Davies, C. W. (1938). '397. The extent of dissociation of salts in water. Part VIII. An equation for the mean ionic activity coefficient of an electrolyte in water, and a revision of the dissociation constants of some sulphates'. In: *J. Chem. Soc.* (0), pp. 2093–2098.

Davies, C. W. (1962). *Ion association*. Butterworths scientific publications. London: Butterworths, p. 190.

Debye, P. and Hückel, E. (1923). 'Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen'. In: *Physikalische Zeitschrift* 24.185, p. 305.

Dennis, B., Lawton, R., Kolar, J. and Alvarado, A (1989). *Results of investigation at the Miravalles Geothermal Field, Costa Rica: Part 1, Well logging. Resultados de las investigaciones en el campo geotermico de Miravalles, Costa Rica: Parte 1, Registros de pozos*. Tech. rep. Los Alamos National Lab., NM (USA).

Dezayes, C., Sanjuan, B, Gal, F, Lerouge, C and Brach, M (2013). *Forage d'exploration géothermique GRT-1. Suivi géochimique des fluides et caractérisation des zones fracturées. Rapport final BRGM*. Tech. rep. RP-62546-FR.

Díaz-González, L., Santoyo, E. and Reyes-Reyes, J. (2008). 'Tres nuevos geotermómetros mejorados de Na/K usando herramientas computacionales y geoquimiométricas: aplicación a la predicción de temperaturas de sistemas geotérmicos'. In: *Revista mexicana de ciencias geológicas* 25.3, pp. 465–482.

Dramsch, J. S. (2020). '70 years of machine learning in geoscience in review'. In: *Advances in geophysics* 61, pp. 1–55.

Ellis, A. J. (1963). 'The solubility of calcite in sodium chloride solutions at high temperatures'. In: *American Journal of Science* 261.3, pp. 259–267. ISSN: 0002-9599. DOI: 10.2475/ajs.261.3.259.

Ellis, A. (1970). 'Quantitative interpretation of chemical characteristics of hydrothermal systems'. In: *Geothermics* 2, pp. 516–528. ISSN: 0375-6505.

Ellis, A. and Mahon, W. (1964). 'Natural hydrothermal systems and experimental hot-water/rock interactions'. In: *Geochimica et Cosmochimica Acta* 28.8, pp. 1323–1357. ISSN: 0016-7037.

Ellis, A. J. and Mahon, W. A. (1977). *Chemistry and geothermal systems*. Energy science and engineering: resources, technology, management. New York [u.a.]: Academic Press. ISBN: 0122374509.

Ferhat Bayram, A. (2001). 'Application of an artificial neural network model to a Na–K geothermometer'. In: *Journal of Volcanology and Geothermal Research* 112.1, pp. 75–81. ISSN: 0377-0273.

Fouillac, C. and Michard, G. (1981). 'Sodium/lithium ratio in water applied to geothermometry of geothermal reservoirs'. In: *Geothermics* 10.1, pp. 55–70.

Fournier, R. O. (1977). 'Chemical geothermometers and mixing models for geothermal systems'. In: *Geothermics* 5.1, pp. 41–50. ISSN: 0375-6505.

Fournier, R. O. (1979). 'Geochemical and hydrologic considerations and the use of enthalpy-chloride diagrams in the prediction of underground conditions in hot-spring systems'. In: *Journal of Volcanology and Geothermal Research* 5.1, pp. 1–16. ISSN: 0377-0273.

Fournier, R. O., Potter, I. et al. (1982). 'Revised and expanded silica (quartz) geothermometer'. In: *Bull., Geotherm. Resour. Counc.(Davis, Calif.);(United States)* 11.10.

Fournier, R. O. and Potter, R. (1979). 'Magnesium correction to the Na-K-Ca chemical geothermometer'. In: *Geochimica et Cosmochimica Acta* 43.9, pp. 1543–1550. ISSN: 0016-7037.

Fournier, R. O. and Rowe, J. J. (1966). 'Estimation of underground temperatures from the silica content of water from hot springs and wet-steam wells'. In: *American Journal of Science* 264.9, pp. 685–697. ISSN: 0002-9599.

Fournier, R. O. and Truesdell, A. (1973). 'An empirical Na-K-Ca geothermometer for natural waters'. In: *Geochimica et Cosmochimica Acta* 37.5, pp. 1255–1275. ISSN: 0016-7037.

Fournier, R. O. and Truesdell, A. (1974). 'Geochemical indicators of subsurface temperature: Part 2, estimation of temperature and fraction of hot water mixed with cold water'. In: *Journal of Research of the US geological Survey* 2.3, pp. 263–270.

Fournier, R. O., White, D. and Truesdell, A. (1974). 'Geochemical indicators of subsurface temperature: Part 1, basic assumptions'. In: *Journal of Research of the US Geological Survey* 2.3, pp. 259–262.

Gherardi, F., Panichi, C., Yock, A. and Gerardo-Abaya, J. (2002). 'Geochemistry of the surface and deep fluids of the Miravalles volcano geothermal system (Costa Rica)'. In: *Geothermics* 31.1, pp. 91–128. ISSN: 0375-6505.

Gibbs, J. W. (1873). 'Graphical Methods in the Thermodynamics of Fluids'. In: *Transactions of the Connecticut Academy* 2, pp. 309–342.

Giggenbach, W. F. (1978). 'The isotopic composition of waters from the El Tatio geothermal field, Northern Chile'. In: *Geochimica et Cosmochimica Acta* 42.7, pp. 979–988. ISSN: 0016-7037.

Giggenbach, W. F. (1981). 'Geothermal mineral equilibria'. In: *Geochimica et Cosmochimica Acta* 45.3, pp. 393–410. ISSN: 0016-7037.

Giggenbach, W. F. (1988). 'Geothermal solute equilibria. Derivation of Na-K-Mg-Ca geoindicators'. In: *Geochimica et Cosmochimica Acta* 52.12, pp. 2749–2765. ISSN: 0016-7037.

Giggenbach, W., Sheppard, D., Robinson, B., Stewart, M. and Lyon, G. (1994). 'Geochemical structure and position of the Waiotapu geothermal field, New Zealand'. In: *Geothermics* 23.5, pp. 599–644. ISSN: 0375-6505.

Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep learning*. MIT press.

Graves, A. (2012). *Supervised sequence labelling*. Springer. ISBN: 978-3-642-24797-2.

Grigsby, C., Goff, F, Trujillo Jr, P., Counce, D., Dennis, B, Kolar, J and Corrales, R (1989). *Results of investigation at the Miravalles geothermal field, Costa Rica. Resultados de las investigaciones en el campo geotermico de Miravalles, Costa Rica; Parte 2, Muestreo de fluidos pozo abajo*. Tech. rep. Los Alamos National Lab., NM (USA).

Gudmundsson, B. T. and Arnórsson, S. (2002). 'Geochemical monitoring of the Krafla and Námafjall geothermal areas, N-Iceland'. In: *Geothermics* 31.2, pp. 195–243. ISSN: 0375-6505.

Hahnloser, R. H., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J. and Seung, H. S. (2000). 'Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit'. In: *nature* 405.6789, pp. 947–951.

Haklidir, F. S. T. and Haklidir, M. (2020). 'Prediction of reservoir temperatures using hydrogeochemical data, Western Anatolia geothermal systems (Turkey): a machine learning approach'. In: *Natural Resources Research* 29.4, pp. 2333–2346.

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J. et al. (2020). 'Array programming with NumPy'. In: *Nature* 585.7825, pp. 357–362.

Helgeson, H. C. (1969). 'Thermodynamics of hydrothermal systems at elevated temperatures and pressures'. In: *American Journal of Science* 267.7, pp. 729–804. ISSN: 0002-9599.

Helgeson, H. C. and Kirkham, D. H. (1974). 'Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures; II, Debye-Huckel parameters for activity coefficients and relative partial molal properties'. In: *American Journal of Science* 274.10, pp. 1199–1261. ISSN: 0002-9599.

Helgeson, H. C., Kirkham, D. H. and Flowers, G. C. (1981). 'Theoretical prediction of the thermodynamic behavior of aqueous electrolytes by high pressures and temperatures; IV, Calculation of activity coefficients, osmotic coefficients, and apparent molal and standard and relative partial molal properties to 600 °C and 5 kb'. In: *American Journal of Science* 281.10, pp. 1249–1516. ISSN: 0002-9599.

Hückel, E (1925). 'Zur Theorie konzentrierterer wässeriger Lösungen starker Elektrolyte'. In: *Phys. Z* 26, pp. 93–147.

Hunter, J. D. (2007). 'Matplotlib: A 2D Graphics Environment'. In: *Computing in Science & Engineering* 9.3, pp. 90–95.

Ibrahim, B., Konduah, J. O. and Ahenkorah, I. (2023). 'Predicting reservoir temperature of geothermal systems in Western Anatolia, Turkey: A focus on predictive performance and explainability of machine learning models'. In: *Geothermics* 112, p. 102727.

Jarrett, K., Kavukcuoglu, K., Ranzato, M. and LeCun, Y. (2009). 'What is the best multi-stage architecture for object recognition?' In: *2009 IEEE 12th international conference on computer vision*. IEEE, pp. 2146–2153.

Kharaka, Y. K. and Mariner, R. H. (1989). 'Chemical geothermometers and their application to formation waters from sedimentary basins'. In: *Thermal history of sedimentary basins: methods and case histories*. Springer, pp. 99–117.

Kingma, D. P. and Ba, J. (2014). 'Adam: A method for stochastic optimization'. In: *arXiv preprint arXiv:1412.6980*.

Langelier, W. F. and Ludwig, H. F. (1942). 'Graphical Methods for Indicating the Mineral Character of Natural Waters'. In: *Journal AWWA* 34.3, pp. 335–352.

Lundberg, S. M. and Lee, S.-I. (2017). 'A unified approach to interpreting model predictions'. In: *Advances in neural information processing systems* 30.

MATLAB (2019). *Version 9.7.0.1190202 (R2019b)*. Natick, Massachusetts: The MathWorks Inc.

MATLAB (2021). *Version 9.10.0.1613233 (R2021a)*. Natick, Massachusetts: The Mathworks, Inc.

Maier, C. G. and Kelley, K. K. (1932). 'An equation for the representation of high-temperature heat content data'. In: *Journal of the American Chemical Society* 54.8, pp. 3243–3246.

Marty, B., Criaud, A. and Fouillac, C. (1988). 'Low enthalpy geothermal fluids from the Paris sedimentary basin — 1. Characteristics and origin of gases'. In: *Geothermics* 17.4, pp. 619–633. ISSN: 0375-6505.

Michard, G and Roekens, E (1983). 'Modelling of the chemical composition of alkaline hot waters'. In: *Geothermics* 12.2-3, pp. 161–169.

Michard, G. and Bastide, J.-P. (1988). 'Etude geochimique de la nappe du dogger du bassin parisien'. In: *Journal of Volcanology and Geothermal Research* 35.1, pp. 151–163. ISSN: 0377-0273.

Morey, G., Fournier, R. and Rowe, J. (1962). 'The solubility of quartz in water in the temperature interval from 25 to 300 degree C'. In: *Geochimica et Cosmochimica Acta* 26.10, pp. 1029–1043. ISSN: 0016-7037.

Mroczek, E., Stewart, M. and Scott, B. (2003). 'Chemistry of the Rotorua Geothermal Field Part 2: Discharging Wells - Update of chemical and isotopic compositions and comparison with historical data'. In: *GNS Confidential Client Report* 94.

Neveu, M., Desch, S. J. and Castillo-Rogez, J. C. (2017). 'Aqueous geochemistry in icy world interiors: Equilibrium fluid, rock, and gas compositions, and fate of antifreezes and radionuclides'. In: *Geochimica et Cosmochimica Acta* 212, pp. 324–371. ISSN: 0016-7037.

Nieva, D. and Nieva, R. (1987). 'Developments in geothermal energy in Mexico — Part twelve. A cationic geothermometer for prospecting of geothermal resources'. In: *Heat Recovery Systems and CHP* 7.3, pp. 243–258. ISSN: 0890-4332.

Nitschke, F. (2018). 'Numerical and Experimental Characterization of Dissolution and Precipitation Processes in Deep Geothermal Reservoirs'. 35.14.01; LK 01. PhD thesis. Karlsruher Institut für Technologie (KIT). 153 pp.

Nitschke, F., Held, S., Neumann, T. and Kohl, T. (2018). 'Geochemical characterization of the Villarrica geothermal system, Southern Chile, part II: Site-specific re-evaluation of SiO2 and Na-K solute geothermometers'. In: *Geothermics* 74, pp. 217–225. ISSN: 0375-6505.

Nitschke, F., Held, S., Villalon, I., Neumann, T. and Kohl, T. (2017). 'Assessment of performance and parameter sensitivity of multicomponent geothermometry applied to a medium enthalpy geothermal system'. In: *Geothermal Energy* 5, pp. 1–20.

Olguín-Martínez, M. G., Peiffer, L., Dobson, P. F., Spycher, N., Inguaggiato, C., Wanner, C., Hoyos, A., Wurl, J., Makovsky, K. and Ruiz-Aguilar, D. (2022). 'PyGeoT: A tool to automate mineral selection for multicomponent geothermometry'. In: *Geothermics* 104, p. 102467. ISSN: 0375-6505.

Óskarsson, F., Fridriksson, T. and Thorbjörnsson, D. (2015). 'Geochemical monitoring of the Reykjanes geothermal reservoir 2003 to 2013'. In: *World geothermal congress*. Vol. 2015.

Palmer, C. D. (2014). *Reservoir Temperature Estimator, Version 00*.

Pang, Z.-H. and Reed, M. (1998). 'Theoretical Chemical Thermometry on Geothermal Waters: Problems and Methods'. In: *Geochimica et Cosmochimica Acta* 62.6, pp. 1083–1091. ISSN: 0016-7037.

Pang, Z.-H. (1988). *Multiple fluid-mineral equilibrium calculations and their applications to geothermometry and hydrochemical processes in geothermal systems*. 5. United Nations University.

Parkhurst, D. L., Appelo, C. et al. (2013). 'Description of input and examples for PHREEQC version 3 — A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations'. In: *US geological survey techniques and methods* 6.A43, p. 497.

Pauwels, H., Fouillac, C. and Fouillac, A.-M. (1993). 'Chemistry and isotopes of deep geothermal saline fluids in the Upper Rhine Graben: Origin of compounds and water-rock interactions'. In: *Geochimica et Cosmochimica Acta* 57.12, pp. 2737–2749. ISSN: 0016-7037.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al. (2011). 'Scikit-learn: Machine learning in Python'. In: *the Journal of machine Learning research* 12, pp. 2825–2830.

Peiffer, L., Wanner, C., Spycher, N., Sonnenthal, E., Kennedy, B. and Iovenitti, J. (2014). 'Optimized multicomponent vs. classical geothermometry: Insights from modeling studies at the Dixie Valley geothermal area'. In: *Geothermics* 51, pp. 154–169. ISSN: 0375-6505.

Pérez-Zárate, D., Santoyo, E., Acevedo-Anicasio, A., Díaz-González, L. and García-López, C. (2019). 'Evaluation of artificial neural networks for the prediction of deep reservoir temperatures using the gas-phase composition of geothermal fluids'. In: *Computers & Geosciences* 129, pp. 49–68. ISSN: 0098-3004.

Pitzer, K. S. (1973). 'Thermodynamics of electrolytes. I. Theoretical basis and general equations'. In: *The Journal of Physical Chemistry* 77.2, pp. 268–277.

Pitzer, K. S. and Mayorga, G. (1973). 'Thermodynamics of electrolytes. II. Activity and osmotic coefficients for strong electrolytes with one or both ions univalent'. In: *The Journal of Physical Chemistry* 77.19, pp. 2300–2308.

Plamer, C., Ohly, S., Smith, R., Neupane, G, McLing, T and Mattson, E (2015). 'Mineral selection for multicomponent equilibrium geothermometry'. In: *Transactions-Geothermal Resources Council* 38.INL/JOU-15-34930.

Reback, J., McKinney, W., Jbrockmendel, Van Den Bossche, J., Augspurger, T., Cloud, P., Gfyoung, Sinhrks, Hawkins, S., Klein, A., Roeschke, M., Tratner, J., Petersen, T., She, C., Ayd, W., MomIsBestFriend, Garcia, M., Schendel, J., Hayden, A., Saxton, D., Jancauskas, V., McMaster, A., Battiston, P., Seabold, S., Chris-B1, H-Vetinari, Dong, K., Hoyer, S., Overmeire, W. and Winkel, M. (2020). *pandas-dev/pandas: Pandas 1.1.3*. Version v1.1.3.

Reed, M. H. (1982). 'Calculation of multicomponent chemical equilibria and reaction processes in systems involving minerals, gases and an aqueous phase'. In: *Geochimica et Cosmochimica Acta* 46.4, pp. 513–528.

Reed, M. and Spycher, N. (1984). 'Calculation of pH and mineral equilibria in hydrothermal waters with application to geothermometry and studies of boiling and dilution'. In: *Geochimica et Cosmochimica Acta* 48.7, pp. 1479–1492. ISSN: 0016-7037.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J. and Carvalhais, N. (2019). 'Deep learning and process understanding for data-driven Earth system science'. In: *Nature* 566.7743, pp. 195–204.

Robinson, R. A., Stokes, R. H. and Bates, R. G. (1960). 'Electrolyte solutions: the measurement and interpretation of conductance, chemical potential and diffusion in solutions of simple electrolytes'. In: *Journal of The Electrochemical Society* 107.8, p. 205.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986). 'Learning representations by back-propagating errors'. In: *nature* 323.6088, pp. 533–536.

Sanjuan, B, Jacquot, E, Guigues, N, Foucher, J., Brach, M and Braibant, G (2001). 'Field geochemistry contribution, Annex 2 (BRGM contribution) of the final EEC report (European hot dry rock geothermal research programme April 1998 - June 2001)'. In: *Contract no JOR3-CT98-0313*.

Sanjuan, B., Millot, R., Ásmundsson, R., Brach, M. and Giroud, N. (2014). 'Use of two new Na/Li geothermometric relationships for geothermal fluids in volcanic environments'. In: *Chemical Geology* 389, pp. 60–81. ISSN: 0009-2541.

Sanjuan, B., Millot, R., Innocent, C., Dezayes, C., Scheiber, J. and Brach, M. (2016). 'Major geochemical characteristics of geothermal brines from the Upper Rhine Graben granitic basement with constraints on temperature and circulation'. In: *Chemical Geology* 428, pp. 27–47. ISSN: 0009-2541.

Sanjuan, B., Millot, R., Dezayes, C. and Brach, M. (2010). 'Main characteristics of the deep geothermal brine (5 km) at Soultz-sous-Forêts (France) determined using geochemical and tracer test data'. In: *Comptes Rendus Geoscience* 342.7-8, pp. 546–559.

Sanjuan, B., Pinault, J.-L., Rose, P., Gerard, A., Brach, M., Braibant, G., Crouzet, C., Foucher, J.-C., Gautier, A. and Touzelet, S. (2006). 'Geochemical fluid characteristics and main achievements about tracer tests at Soultz-sous-Forêts (France)'. In: 13.

Sanjuan, B., Rose, P., Foucher, J.-C., Brach, M. and Braibant, G. (2004). 'Tracer testing at Soultz-sous-Forêts (France) using Na-benzoate, 1, 5 and 2, 7-naphthalene disulfonate'. In: 8–pages.

Schindler, M., Baumgärtner, J., Gandy, T., Hauffe, P., Hettkamp, T., Menzel, H., Penzkofer, P., Teza, D., Tischner, T. and Wahl, G (2010). 'Successful hydraulic stimulation techniques for electric power production in the Upper Rhine Graben, Central Europe'. In: *Proceedings World geothermal congress*. Vol. 2010.

Sonnenthal, N. S. et al. (2013). 'GeoT User's Guide: A Computer Program for Multicomponent Geothermometry and Geochemical Speciation, Version 1.4'. In.

Sparks, D. L. (2018). 'Kinetics and mechanisms of chemical reactions at the soil mineral/water interface'. In: *Soil physical chemistry*. CRC press, pp. 135–192. ISBN: 1351415786.

Spycher, N., Peiffer, L., Sonnenthal, E., Saldi, G., Reed, M. and Kennedy, B. (2014). 'Integrated multicomponent solute geothermometry'. In: *Geothermics* 51, pp. 113–123. ISSN: 0375-6505.

Spycher, N., Sonnenthal, E. and Kennedy, B. (2011). 'Integrating multicomponent chemical geothermometry with parameter estimation computations for geothermal exploration'. In: vol. 35 1, 663 – 666.

Spycher, N. and Finsterle, S. (2016). *iGeoT v1. o: Automatic Parameter Estimation for Multicomponent Geothermometry, User's Guide*. Tech. rep. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

Spycher, N., Finsterle, S. and Dobson, P. (2016). 'New developments in multicomponent geothermometry'. In: *Proc. 41nd Workshop on Geothermal Reservoir Engineering Stanford University, Stanford, California.*

Stober, I. and Bucher, K. (2015). 'Hydraulic and hydrochemical properties of deep sedimentary reservoirs of the Upper Rhine Graben, Europe'. In: *Geofluids* 15.3, pp. 464–482.

Sutskever, I., Martens, J., Dahl, G. and Hinton, G. (2013). 'On the importance of initialization and momentum in deep learning'. In: *International conference on machine learning*. PMLR, pp. 1139–1147.

Sverjensky, D., Shock, E. and Helgeson, H. (1997). 'Prediction of the thermodynamic properties of aqueous metal complexes to 1000 deegre C and 5 kb'. In: *Geochimica et Cosmochimica Acta* 61.7, pp. 1359–1412. ISSN: 0016-7037.

TensorFlow, D. (2022). 'TensorFlow'. In: *Zenodo*.

Tieleman, T., Hinton, G. et al. (2012). 'Lecture 6.5 - rmsprop: Divide the gradient by a running average of its recent magnitude'. In: *COURSERA: Neural networks for machine learning* 4.2, pp. 26–31.

Truesdell, A. (1976). 'Summary of section III-geochemical techniques in exploration'. In: *Proc. 2nd UN Symp. on the Development and Use of Geothermal Resources*. Vol. 1, pp. 1iii–1xxix.

Truesdell, A. H. and Jones, B. F. (1974). 'WATEQ, a computer program for calculating chemical equilibria of natural waters'. In: *J. Res. US Geol. Surv* 2.2, pp. 233–248.

Van Rossum, G., Drake, F. L. et al. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Van der Walt, S., Colbert, S. C. and Varoquaux, G. (2011). 'The NumPy Array: A Structure for Efficient Numerical Computation'. In: *Computing in Science & Engineering* 13.2, pp. 22–30.

Van 't Hoff, J. H. (1886). 'Une propriété générale de la matière diluée'. In: *Kungliga Svenska Vetenskaps-Akademiens Handlingar* 21.17, pp. 42–49.

Varsányi, I., Matray, J.-M. and Ó.Kovács, L. (1997). 'Geochemistry of formation waters in the Pannonian Basin (southeast Hungary)'. In: *Chemical Geology* 140.1, pp. 89–106. ISSN: 0009-2541.

Vaute, L (1998). 'Tests de traçage réalisés sur le site géothermique de Soultz-sous-Forêts (juillet-novembre 1997)'. In: *Final BRGM report n* 40230, p. 39.

Verma, M. P. (2000). 'Revised quartz solubility temperature dependence equation along the water-vapor saturation curve'. In: *World Geothermal Congress, Kyushu-Tohoku, Japan*, pp. 1927–1932.

Verma, S. P. and Santoyo, E. (1997). 'New improved equations for NaK, NaLi and SiO2 geothermometers by outlier detection and rejection'. In: *Journal of Volcanology and Geothermal Research* 79.1, pp. 9–23. ISSN: 0377-0273.

Vidal, J. and Genter, A. (2018). 'Overview of naturally permeable fractured reservoirs in the central and southern Upper Rhine Graben: Insights from geothermal wells'. In: *Geothermics* 74, pp. 57–73. ISSN: 0375-6505.

Vidal, J., Hehn, R., Glaas, C. and Genter, A. (2019). 'How can temperature logs help identify permeable fractures and define a conceptual model of fluid circulation? An example from deep geothermal wells in the Upper Rhine Graben'. In: *Geofluids* 2019.

Voigt, M., Marieni, C., Clark, D. E., Gíslason, S. R. and Oelkers, E. H. (2018). 'Evaluation and refinement of thermodynamic databases for mineral carbonation'. In: *Energy Procedia* 146. Carbon in natural and engineered processes: Selected contributions from the 2018 International Carbon Conference, pp. 81–91. ISSN: 1876-6102.

Wahba, G. (1987). 'Three Topics in Ill-posed Problems'. In: *Inverse and Ill-Posed Problems*. Ed. by H. W. Engl and C. Groetsch. Academic Press, pp. 37–51. ISBN: 978-0-12-239040-1.

Waskom, M. L. (2021). 'seaborn: statistical data visualization'. In: *Journal of Open Source Software* 6.60, p. 3021.

White, D. E. (1965). 'Saline waters of sedimentary rocks'. In: *AAPG Special Volumes*, pp. 342–366.

Wood, C. (1992). 'Geology of the Rotorua geothermal system'. In: *Geothermics* 21.1, pp. 25–41. ISSN: 0375-6505.

Yao, Y., Rosasco, L. and Caponnetto, A. (2007). 'On Early Stopping in Gradient Descent Learning'. In: *Constructive Approximation* 26.2, pp. 289–315. ISSN: 0176-4276.

Ystroem, L. H., Nitschke, F., Held, S. and Kohl, T. (2020). 'A multicomponent geothermometer for high-temperature basalt settings'. In: *Geothermal Energy* 8, pp. 1–21.

Ystroem, L. H., Nitschke, F., Held, S. and Kohl, T. (2021). 'An Integrated Sensitivity Analysis for the Basalt Specific Multicomponent Geothermometer for High Temperature Settings'. In: *Proceedings World Geothermal Congress*, p. 4.

Ystroem, L. H., Nitschke, F. and Kohl, T. (2022). 'MulT_predict - An optimised comprehensive multicomponent geothermometer'. In: *Geothermics* 105, p. 102548. ISSN: 0375-6505.

Zhang, G., Lu, P., Zhang, Y., Tu, K. and Zhu, C. (2020). 'SupPhreeqc: A program for generating customized Phreeqc thermodynamic datasets from Supcrtbl and

extending calculations to elevated pressures and temperatures'. In: *Computers & Geosciences* 143, p. 104560. ISSN: 0098-3004.

DECLARATION OF AUTHORSHIP

CHAPTER 3: A MULTICOMPONENT GEOTHERMOMETER FOR HIGH TEMPERATURE BASALT SETTINGS

Ystroem, L. H., Nitschke, F., Held, S. and Kohl, T. (2020). 'A multicomponent geothermometer for high-temperature basalt settings'. In: *Geothermal Energy 8:2*, pp. 1–21. DOI: 10.1186/s40517-020-0158-z.

This study was conducted within the Helmholtz Association's portfolio project 'Geoenergy' supported by the program 'Renewable Energies', under the topic 'Geothermal Energy Systems' and was supported by EnBW, Energie Baden-Württemberg AG, Germany.
In this study, I set up the conception of a multicomponent geothermometer based on the geochemical modelling tool IPhreeqc. I wrote the MATLAB code interacting with IPhreeqc. I implemented an automated script, statistically evaluating the IPhreeqc output within MATLAB. I performed sensitivity analyses on vulnerable system parameters and implemented three separate optimisation processes for pH, aluminium concentration, and steam loss/dilution in the MATLAB code. For the calculation of temperature predictions exceeding 300 degrees Celsius, I evaluated multiple thermodynamic databases and extrapolated the Lawrence Livermore National Laboratory database. I tested the multicomponent geothermometer on data from Krafla and Reykjanes developing a site-specific basalt mineral assemblage. I compared the results to conventional geothermometers based on the original and back-calculated fluid composition. I visualised and interpreted the results. I wrote the manuscript of the paper.

CHAPTER 4: AN INTEGRATED SENSITIVITY ANALYSIS FOR THE BASALT SPECIFIC MULTICOMPONENT GEOTHERMOMETER FOR HIGH TEMPERATURE SETTINGS

Ystroem, L. H., Nitschke, F., Held, S. and Kohl, T. (2021). 'An Integrated Sensitivity Analysis for the Basalt Specific Multicomponent Geothermometer for High Temperature Settings'. In: *Proceedings World Geothermal Congress 2020+1*, p. 4.

This study was conducted within the Helmholtz Association's portfolio project 'Geoenergy' supported by the program 'Renewable Energies', under the topic 'Geothermal Energy Systems' and was supported by EnBW, Energie Baden-Württemberg AG, Germany.

In this study, I set up the conception of an interdependent sensitivity analysis for the multicomponent geothermometer. I restructured the MATLAB code and combine the three separate optimisation processes for pH, aluminium concentration, and steam loss/dilution. I added a fourth optimisation process for an additional element concentration or system parameter. I created a multidimensional IPhreeqc output of the interdependent optimisation processes and adapted the statistical evaluation of this data. I developed a visualisation of the statistical evaluation of the reservoir temperature estimation as a four-panel output file. I exemplarily recalculated the Krafla data, illustrated it, and interpreted it. I wrote the manuscript of the proceeding.

CHAPTER 5: MULT_PREDICT - AN OPTIMISED COMPREHENSIVE MULTICOMPONENT GEOTHERMOMETER

Ystroem, L. H., Nitschke, F. and Kohl, T. (2022). 'MulT_predict - An optimised comprehensive multicomponent geothermometer'. In: *Geothermics 105*, p. 102548. ISSN: 0375-6505. DOI: 10.1016/j.geothermics.2022.102548

This study was conducted as part of the subtopic "Geoenergy" in the program "MTET - Materials and Technologies for the Energy Transition" of the Helmholtz Association and was supported by multiple power plant operators in the German Molasse Basin with their geochemical fluid analyses.
In this study, I set up the conception of a universally valid mineral assemblage for the multicomponent geothermometer. This mineral assemblage is applicable for unknown reservoir composition. I established a three-step procedure for unknown temperature estimations using the universally valid mineral assemblage and perform an outlier removal for temperature estimation refinement. I tested the new method on data from eight geothermal sites worldwide and validated using in-situ measured temperatures. I performed an internal benchmark of the optimisation process with a synthetic brine. I stepwise perturbed the chemistry of the synthetic brine (pH, aluminium concentration, steam loss/dilution, and salinity) and back-calculated the initial conditions via the optimisation process. I compared the performance of three separate optimisation processes (pH, aluminium concentration, and steam loss/dilution) as well as the interdependent optimisation process. I visualised and interpreted the results. I wrote the manuscript of the paper.

CHAPTER 6: DEEP LEARNING AND GEOCHEMICAL MODELLING AS TOOLS FOR SOLUTE GEOTHERMOMETRY

Ystroem, L. H., Vollmer, M., Nitschke, F. and Kohl, T. (2022). 'Deep learning and geochemical modelling as tools for solute geothermometry'. In: *Proceedings*

European Geothermal Congress 2022, p. 4. ISBN: 978-2-9601946-2-3.

This study was conducted as part of the subtopic "Geoenergy" in the program "MTET - Materials and Technologies for the Energy Transition" of the Helmholtz Association.
In this study, I set up the conception of comparing the reservoir temperature estimation of the numerically optimised multicomponent geothermometer and a newly developed deep learning algorithm. I improved the Python script of the deep learning algorithm, which was developed by Mark Vollmer in his master thesis. I acquired more data to improve the deep learning algorithm. I performed the calculation of Icelandic data via the numerically optimised multicomponent geothermometer and the deep learning algorithm. I visulised, compared, and interpreted the results. I wrote the manuscript of the proceeding.

CHAPTER 7: ANNRG - AN ARTIFICIAL NEURAL NETWORK SOLUTE GEOTHERMOMETER

Ystroem, L. H., Vollmer, M., Kohl, T. and Nitschke, F. (2023). 'AnnRG - An artificial neural network solute geothermometer'. In: *Applied Computing and Geosciences 20:100144*, DOI: 10.1016/j.acags.2023.100144.

This study was conducted as part of the project "MALEG" of the German Federal Ministry for Economic Affairs and Climate Action (BMWK, No. 03EE4041B), implemented by the Project Management Jülich (PtJ). This study is part of the subtopic "Geoenergy" in the program "MTET - Materials and Technologies for the Energy Transition" of the Helmholtz Association.
In this study, I developed an adequate solute artificial neural network geothermometer, called AnnRG. I remodelled the deep learning algorithm of Mark Vollmer to a feedforward multilayer perceptron. I developed and tested the new architecture of the artificial neural network. I performed a hyperparameter optimisation on the artificial neural network. I acquired 208 data pairs of geochemical fluid parameters and in-situ temperature measurements worldwide to establish the geothermometer. I conducted an outlier removal of the database to improve the input for the artificial neural network. I performed a sensitivity analysis of the required amount of data necessary to train the geothermometer (>65 samples). I used 155 samples to set up the geothermometer. To verify the geothermometer, I transferred 45 unknown samples to the trained artificial neural network and calculated the reservoir temperature. I visulised and interpreted all results. I wrote the manuscript of the paper.

PUBLICATIONS

PEER-REVIEWED PUBLICATIONS

Ystroem, L. H., Vollmer, M., Kohl, T. and Nitschke, F. (2023). 'AnnRG - An artificial neural network solute geothermometer'. In: *Applied Computing and Geosciences 20*, 100144. p.10.

Ystroem, L. H., Nitschke, F. and Kohl, T. (2022). 'MulT_predict - An optimised comprehensive multicomponent geothermometer'. In: *Geothermics 105*, pp. 102548. ISSN: 0375-6505.

Ystroem, L. H., Nitschke, F., Held, S. and Kohl, T. (2021). 'An Integrated Sensitivity Analysis for the Basalt Specific Multicomponent Geothermometer for High Temperature Settings'. In: *Proceedings World Geothermal Congress 2020+1*, p. 4.

Ystroem, L. H., Nitschke, F., Held, S. and Kohl, T. (2020). 'A multicomponent geothermometer for high-temperature basalt settings'. In: *Geothermal Energy 8:2*, p. 21.

FURTHER PUBLICATIONS

Nitschke, F., Ystroem, L. H., Bauer, F. and Kohl, T. (2023). 'Geochemical Constraints on the Operations of High Temperature Aquifer Energy Storage (HT-ATES) in Abandoned Oil Reservoirs'. In *Proceedings 47th Workshop on Geothermal Reservoir Engineering. Stanford University*, p. 10.

Ystroem, L. H., Vollmer, M., Nitschke, F. and Kohl, T. (2022). 'Deep learning and geochemical modelling as tools for solute geothermometry'. In *Proceedings European Geothermal Congress 2022*, p. 4. ISBN: 978-2-9601946-2-3.

Ystroem, L. H., Nitschke, F., Held, S. and Kohl, T. (2020). 'MulT_predict: Ein numerisches Multikomponentengeothermometer zur präzisen Bestimmung der Reservoirtemperatur'. In: *Geothermische Energie 95*, p. 4.

PRESENTATIONS WITH ABSTRACT

Ystroem, L. H., Trumpp, M., Goldberg, V., Eichinger, F., Amtmann, J., Winter, D., Koschikowski, J., Kohl, T. and Nitschke, F. (2023). 'MALEG - Machine Learning for Enhancing Geothermal Energy'. *10th European Geothermal Workshop 2023*. Utrecht, Netherlands.

Ystroem, L. H., Trumpp, M., Goldberg, V., Eichinger, F., Amtmann, J., Winter, D., Koschikowski, J., Kohl, T. and Nitschke, F. (2023). 'MALEG - Maschinelles Lernen zur Verbesserung der Geothermischen Energienutzung'. *Der Geothermiekongress 2023*. Essen, Germany.

Ystroem, L. H., Nitschke, F. and Kohl, T. (2023). 'Geothermometer temperature predictions for DeepStor'. *KIT Energy Center 10th annual meeting*. Karlsruhe, Germany.

Ystroem, L. H., Vollmer, M., Nitschke, F. and Kohl, T. (2023). 'Solute Geothermometry - Applications and Development'. *Focus on Geothermal (Webinar) by Enerchange and Think Geoenergy*. Online.

Ystroem, L. H., Vollmer, M., Nitschke, F. and Kohl, T. (2022). 'Deep learning and geochemical modelling as tools for solute geothermometry'. *European Geothermal Congress 2022*. Berlin, Germany.

Ystroem, L. H., Vollmer, M., Nitschke, F. and Kohl, T. (2021). 'Reservoirtemperaturbestimmung mittels künstlicher neuronaler Netze (KNN)'. *Der Digitale Geothermiekongress 2021*. Online.

Ystroem, L. H., Vollmer, M., Nitschke, F. and Kohl, T. (2021). 'An artificial neural network (ANN) as solute geothermometer'. *9th European Geothermal Workshop 2021*. Karlsruhe, Germany.

Ystroem, L. H., Nitschke, F., Held, S. and Kohl, T. (2021). 'An integrated sensitivity analysis for multicomponent geothermometer for high temperature settings'. *World Geothermal Congress 2020+1*. Online.

Ystroem, L. H., Nitschke, F. and Kohl, T. (2021). 'Optimised multicomponent geothermometer – MulT_predict'. *12th European Geothermal PhD Days*. Online.

Ystroem, L. H., Nitschke, F., Held, S. and Kohl, T. (2020). 'Zeitliche und tiefenabhängige Effekte auf die Fluidchemie - Untersuchung granitischer Reservoire im Oberrheingraben mittels Multikomponentengeothermometrie'. *Der Digitale Geothermiekongress 2020*. Online.

Ystroem, L. H., Nitschke, F., Held, S. and Kohl, T. (2020). 'An optimised multicomponent geothermometer - Application of MulT_predict for the Upper Rhine Graben'. *Student Technical Congress 2020*. Karlsruhe, Germany.

Ystroem, L. H., Nitschke, F., Held, S. and Kohl, T. (2019). 'Entwicklung & Validierung eines Tools zur Reservoirtemperaturabschätzung basierend auf Multikomponentengeothermometrie'. *Der Geothermiekongress 2019*. Munich, Germany.

Ystroem, L. H., Nitschke, F., Held, S. and Kohl, T. (2019). 'MulT_predict – A multicomponent geothermometer optimized by sensitivity analyses'. *7th European Geothermal Workshop 2019*. Karlsruhe, Germany.

Ystroem, L. H., Nitschke, F., Held, S. and Kohl, T. (2019). 'MulT_predict - A multicomponent geothermometer for precise reservoir temperature estimation'. *KIT Energy Center 8th annual meeting*. Karlsruhe, Germany.

ACKNOWLEDGMENTS

During my PhD I have met many people who have influenced and inspired me and my work. Thanks to all of them.

First of all, I would like to thank Thomas Kohl for giving me the opportunity to do my PhD. I am grateful for his support over all these years.

Futhermore, I like to thank Michael Kersten for being my co-supervisor during this thesis.

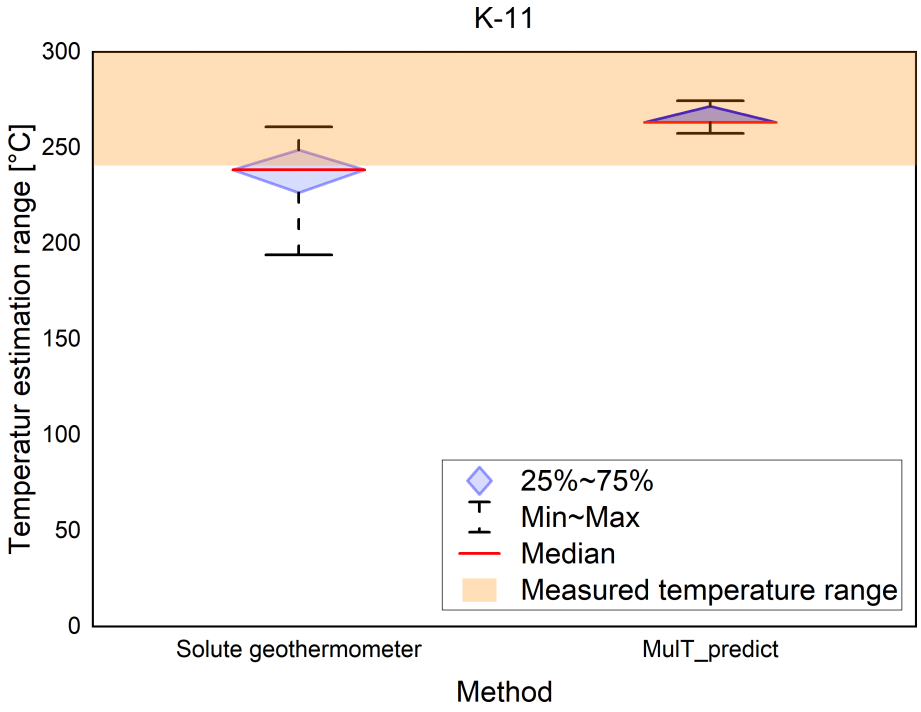
I would also like to personally thank Fabian and Sebastian, who already supervised my master thesis and supported me during my PhD.

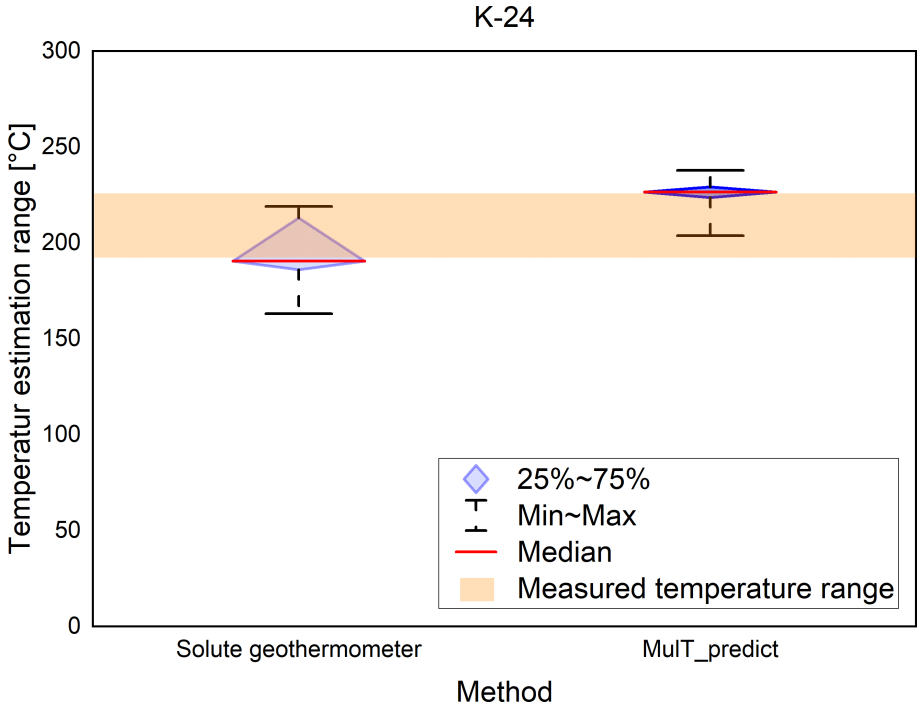
And of course a big thank you to the whole Geothermal Energy Group. I enjoyed the good conversations during coffee breaks and on Thursday afternoons. Therefore, a special thank you to the ASWLS group. We had inspiring discussions that solved almost every challenging problem at some point.

Finally, I would like to thank Silke for all her support with the administration of KIT.

A

APPENDIX


K-11
Temperatur estimation range [°C]
300
250
200
150
100
50
0
25%~75%
Min~Max
Median
Measured temperature range
Solute geothermometer
MulT_predict
Method


K-24
Temperatur estimation range [°C]
300
250
200
150
100
50
0
25%~75%
Min~Max
Median
Measured temperature range
Solute geothermometer
MulT_predict
Method

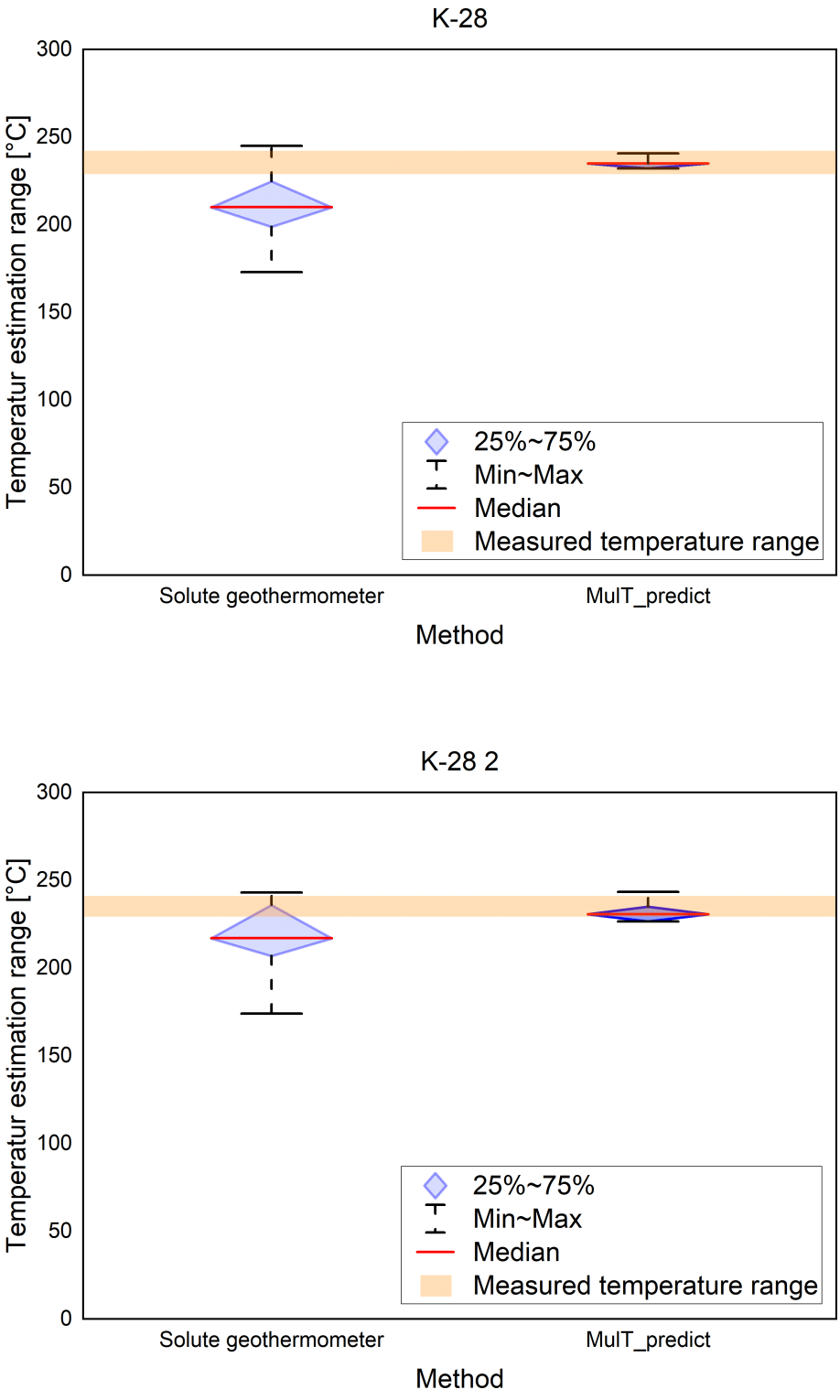


Figure A.1: Comparison of conventional qualitative solute geothermometers based on back-calculated element concentrations using WATCH 2.4 (Bjarnason, 2010) (cf. Appendix A.4) with the end results of MulT_predict (Fig. 3.7) for the wells at Krafla

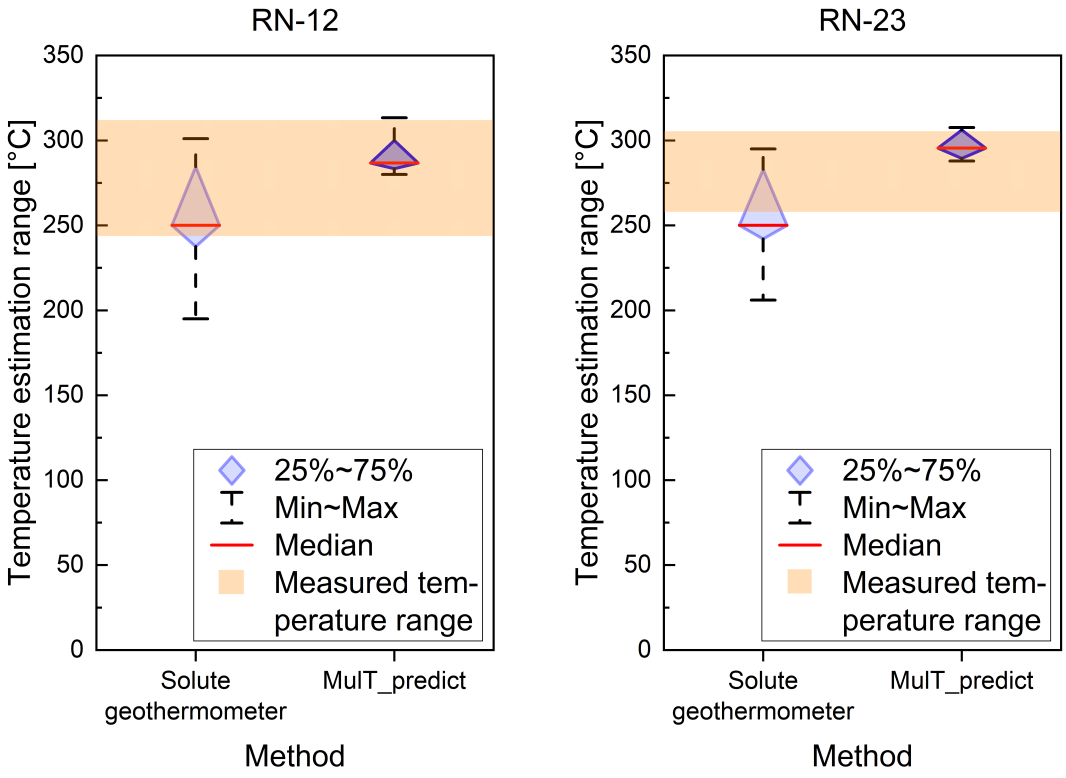


Figure A.2: Comparison of conventional qualitative solute geothermometers based on back-calculated element concentrations using WATCH 2.4 (Bjarnason, 2010) (cf. Appendix A.4) with the end results of MulT_predict (Fig. 3.10) for the wells at Reykjanes

Table A.1: Chemical analysis of water samples by Gudmundsson and Arnórsson (2002) from wells at Krafla collected in 1997 and 1998. Concentrations are in mmol/kg.

Sample No.	Well	Sampling date	Pressure	pH value	Temperature of pH measurement	Element concentration of the sample [mmol/kg]											
			[bar]		[°C]	SiO_2	Na	K	Ca	Mg	Cl	F	SO_4	H_2S	Fe	Al	CO_2
97-3103	K-28	26.10.1997	6.5	9.75	20.9	7.98	9.66	0.66	0.10	0.0001	0.49	0.048	2.99	1.08	0.00015	0.039	0.70
98-3206	K-11	24.06.1998	1.9	9.65	23.2	10.81	9.70	0.80	0.05	0.0001	1.19	0.105	2.34	1.03	0.00009	0.038	2.15
98-3208	K-24	25.06.1998	2.3	9.73	24.2	6.04	9.48	0.44	0.09	0.0000	0.86	0.040	3.29	0.85	0.00006	0.031	0.41
98-3210	K-28	25.06.1998	5.2	9.55	24.8	7.89	9.86	0.65	0.09	0.0001	0.72	0.042	3.58	1.29	0.00013	0.033	1.23

Table A.2: The unspecific mineral set, ordered in groups of minerals and their associated phases

mineral group	associated mineral phases
carbonates	calcite, aragonite, dolomite
oxides and hydroxides	spinel, hematite, goethite, diaspore, gibbsite
sulphides and sulphates	pyrite, marcasite, pyrrhotite, anhydrite, gypsum
nesosilicates	forsterite, grossular, andradite, andalusite, sillimanite, kyanite
sorosilicates	gehlenite, lawsonite, epidote, zoisite
inosilicates	wollastonite, diopside, hedenbergite, ferrosilite, enstatite, anthophyllite, tremolite, pargasite
phyllosilicates	talc, muscovite, paragonite, phlogopite, illite, smectite, clinochlore, kaolinite
tectosilicates	quartz, chalcedony, anorthite, albite, sanidine, K-feldspar, microcline, analcime, laumontite, wairakite

Table A.3: Chemical analysis of water samples by Óskarsson et al. (2015) from the Reykjanes field

Well	Sampling date	Pressure	pH value	Temperature of pH measurement	Element concentration of the sample [mmol/kg]											
		[bar]		[°C]	SiO_2	Na	K	Ca	Mg	Cl	F	SO_4	H_2S	Fe	Al	CO_2
RN-12	30.10.2012	29.7	5.57	22.3	13.81	465.65	39.74	47.25	0.0382	602.25	0.012	0.18	0.14	0.00553	0.004	1.45
RN-23	21.05.2012	26.3	5.29	21.9	13.03	561.30	49.74	56.25	0.0913	711.54	0.012	0.29	0.02	0.01208	0.006	0.52

Table A.4: Temperature estimations via conventional solute geothermometers (quartz, Na/K, Na/K/Ca, K^2/Mg) for the wells at Krafla (K-28, K-11, K-24 and K-28-2) as well as Reykjanes (RN-12 and RN-23) based on back-calculated element concentrations using WATCH 2.4 (Bjarnason, 2010)

		Solute geothermometer [°C]											
Sample No.	Well	quartz			Na/K					Na/K/Ca			K^2/Mg
		Fournier & Potter	Arnósson	Verma	Truesdell	Fournier	Giggenbach	Arnósson	Can	Fournier & Truesdall	Nieva-Nieva	Kharaka & Mariner	Giggenbach
		(1982)	(1983)	(2000)	(1976)	(1979)	(1988)	(2000)	(2002)	(1973)	(1987)	(1989)	(1988)
97-3103	K-28	197	181	201	204	230	245	211	218	209	173	220	240
98-3206	K-11	250	221	259	227	248	261	232	237	226	194	248	240
98-3208	K-24	213	194	219	163	198	214	169	187	186		187	
98-3210	K-28-2	235	210	243	200	227	242	207	215	207	174	219	237
	RN-12	301	249	300	231	251	264	236	240	239	195	269	301
	RN-23	293	245	295	236	255	267	240	244	244	206	278	288

Table A.5: Mineral phases of the basalt specific mineral set, grouped according to the structure.

mineral group	associated mineral phases
carbonates	calcite, aragonite
oxides and hydroxides	hematite, goethite
sulfides and sulfates	pyrite, marcasite, pyrrhotite, anhydrite, gypsum
sorosilicates	epidote
inosilicates	anthophyllite, tremolite, pargasite
phyllosilicates	illite, smectite, clinochlore
tectosilicates	quartz, chalcedony, albite, K-feldspar, microcline, analcime, laumontite, wairakite

Table A.6: Anonymised data from power plant operators from the German Molasse Basin. Element concentrations are presented in ppm.

pH	Sampling temperature	Element concentration of the samples [ppm]										
	[°C]	SiO_2	Na	K	Ca	Mg	Cl	SO_4	H_2S	HCO_3	Fe	Al
7.1	46.5	54	284	15.5	14.0	4.4	172.0	0.5	1.0	509.0	0.02	0.0870
7.3	34.9	20	353	13.2	30.7	7.3	400.9	0.0	0.2	390.4	0.16	0.0072
6.9	49.2	40	104	12.3	28.4	10.1	70.0	0.5	4.5	314.0	0.11	-
6.8	38.1	44	105	12.2	27.5	9.1	68.9	1.4	5.6	308.5	0.03	0.0060
6.2	41.3	103	162	30.5	38.9	5.7	169.0	35.0	11.3	289.2	0.09	-
6.4	24.8	130	111	16.3	3.2	1.1	68.3	16.3	6.5	192.8	0.03	0.0150

Table A.7: Individual setting-specific minerals sets for reservoir temperature estimation. (Montmor. = Montmorillonite)

URG	Iceland	El Tatio	Waiotapu	Miravalles	Pannonia	Paris	Molasse
Quartz	Quartz	Chalcedony	Quartz	Chalcedony	Quartz	Quartz	Quartz
K-feldspar	K-feldspar	K-feldspar	K-feldspar	K-feldspar	K-feldspar	K-feldspar	K-feldspar
Microcline	Microcline	Microcline	Microcline	Microcline	Microcline	Microcline	Microcline
Albite	Albite	Albite	Albite	Albite	Albite	Albite	Albite
Muscovite	Muscovite	Muscovite	Muscovite	Muscovite	Muscovite	Muscovite	Muscovite
Illite	Illite	Illite		Illite	Illite	Illite	Illite
Diaspore	Diaspore	Diaspore	Diaspore	Diaspore	Diaspore	Diaspore	Diaspore
Analcime	Analcime	Analcime	Analcime	Analcime	Analcime	Analcime	
Scolecite		Scolecite	Scolecite	Scolecite	Scolecite	Scolecite	Scolecite
Anhydrite	Anhydrite	Anhydrite	Anhydrite	Anhydrite		Anhydrite	Anhydrite
Kaolinite		Kaolinite		Kaolinite	Kaolinite	Kaolinite	Kaolinite
Pyrophyllite		Pyrophyllite		Pyrophyllite	Pyrophyllite	Pyrophyllite	Pyrophyllite
Montmor.	Montmor.	Montmor.	Montmor.	Montmor.	Montmor.	Montmor.	Montmor.
Beidellite		Beidellite	Beidellite	Beidellite	Beidellite	Beidellite	Beidellite
	Calcite			Calcite	Calcite	Calcite	Calcite
	Sanidine		Sanidine	Sanidine	Sanidine	Sanidine	
Smectite	Smectite		Smectite		Smectite		Smectite
		Mesolite	Mesolite	Mesolite			Mesolite
		Stiblite	Stiblite	Stiblite	Stiblite		Stiblite
			Saponite	Saponite		Saponite	Saponite
Paragonite		Paragonite			Paragonite	Paragonite	
Tremolite	Tremolite						
	Clinochlore				Clinochlore		Clinochlore
Laumontite	Laumontite			Laumontite			
Gibbsite					Gibbsite	Gibbsite	Gibbsite
					Dolomite	Dolomite	Dolomite
			Clinoptilolite		Clinoptilolite		
	Enstatite				Enstatite		
	Wollastonite			Wollastonite			
	Wairakite						

B

MULT_PREDICT

```
%=========================================================================
% MulT_predict.m
%
% FUNCTION:
% - Version 0.22: L. H. Ystroem (KIT) - March 2023
% MulT_predict is a MATLAB tool to facilitate the application of multicom-
% ponent geothermometry on geochemical data. Saturation indices of
% mineral phases were automatically calculated via IPhreeqC and evaluated
% by sensitivity analyses. The result is a best-fit temperature estimation.
%
% INPUT:
% - xlsx-file:[Sample No., Name, Pressure, pH value, Temperature, Element
%             concentration]
% An xlsx-file, in a given template, can be read automatically and the geo-
% chemical data is passed to IPhreeqC.
%
% OUTPUT:
% - 'Name'.fig: [Saturation indices vs. temperature plot, statistics,
%               temperature estimation]
% A saved MATLAB figure, which contains a diagram of the saturation indi-
% ces of each utilized mineral phase plotted against the temperature, as
% well as the statistics (Root mean square error, standard deviation, etc)
% and the temperature estimation box plot of the best-fit values.
%
% - Text message: [steamloss/dilution [%], pH, and Al-concentration]
% A text message in MATLABs Command Window with the details about the best
% fit parameters (-steamloss/+dilution 'w'%, pH value is 'x', and Aluminium
% concentration 'y', and 'individual element' concentration is 'z'[unit]).
%
% WARNING:
% - Only one xlsx-file can be read and calculated at the time
% - The xlsx-file must be written in the given template
% - Mineral phases which do not intersect or multiply intersect with the
%   equilibrium line is not taken into account.
%
% PLEASE NOTE:
% - MulT_predict was programmed in MATLAB R2020b;
% - To run MulT_predict, IPhreeqcCOM server (3.7.3) has to be installed
% - The instruction of MulT_predict is described in the following
%   '*'-framed section
%=========================================================================
```

```
%************************************************************************
%* INSTRUCTION                                                          *
%* - To use MulT_predict, MATLAB and IPhreeqC must be installed         *
%* - Start MulT_predict via the Run-button in the Editor Window in MATLAB *
%* - Select an xlsx-file, in the recommended template, as an input file *
%* - Select the samples which should be calculated                      *
%* - Select the temperature-, pH-, Al concentration- and dilution range *
%* - Enter, if needed pH-, Al-, -steamloss/+dilution- fixation or leave *
%*   it empty                                                           *
%* - Select mineral phases, which are taken into account for the tempera- *
%*   ture estimation                                                    *
%* - Runtime, status updates, and the best-fit parameters are printed in *
%*   the Command Window in MATLAB                                       *
%* - The final results are given in a four-parted figure which is saved *
%*   as a fig-file named like the selected well in the list             *
%*                                                                      *
%* - Changes in the used mineral phases or the size and numbers of      *
%*   sensitivity analysis steps must NOT be varied in the MATLAB Code!  *
%************************************************************************

% Reading the xlsx-inputfile and listing all selectable samples
[filename,pathname] = uigetfile('*.xlsx', 'Select an Inputfile');
    if filename == 0
        fprintf('No Inputfile found \n')
        return
    else
        [numdata,header,raw] = xlsread([pathname,filename]);
        [rows,columns] = size(numdata);
        fprintf('Inputfile read \n')
        [cName,dName] = find(contains(header,'List')==1);
        b = size(header)- size(numdata);
        list1 = header((cName+b(1)):end,dName);
        listin = listdlg('ListString',list1);
    end

% Reading the values of the chosen sample (e.g. element concentration)
a = listin(1:end);
clear('SenMa','Sen1','Sen1m','Sen2','Sen2m')
ccon = find(contains(header(1,:),'['));
bcon = strfind(header(1,ccon),'[');
econ = strfind(header(1,ccon),']');
con = header{1,ccon}(1,cell2mat(bcon)+1:cell2mat(econ)-1);
Clist = {'Pressure','Temperature','pH','Al','Ba','CO2','Ca','Cl','Fe',...
    'H2S','K','Mg','Na','SiO2','SO4'};
for bc = 1:numel(Clist)
[~,db] = find(contains(header(1:2,:),Clist{1,bc})==1);
Clist{2,bc} = numdata(a,(db-b(2)));
```

```
clear('db')
end
struct = cell2struct(Clist(2,:),Clist(1,:),2);

% Inpute values for temperature, pH etc.
prompt = {'Enter minimum temperature [C]:',...
'Enter maximum temperature [C]:','Enter temperature steps:',...
'pH change per step +/-:','pH steps combined:',...
'Aluminium steps:','Boil/Dilu [%] per step +/-:',...
'Boil/Dilu steps combined:','pH fixation','Aluminium fixation',...
'-Boil/+Dilu fixation [%]','Threshold mineral phases'}; %,'Element steps:'
dlgtitle = 'Input';
definput = {'20','300','200','0.1','10','20','0','0','','','','6'};
dims = [1 32];
opts = 'off';
answer = inputdlg(prompt,dlgtitle,dims,definput,opts);
values = str2double(answer); % to activate change values(x)+1
aaa = values(5)+1; % number of pH sensitivity steps
bbb = values(6)+1; % number of Al-con. sensitivity steps
ccc = values(8)+1; % number of steamloss/dilution sensitivity steps
ddd = 1;           % number of element-concentration sensitivity steps
if values(1) >= 1 && values(1) <= 250
tmin = values(1);
else; tmin = 20; % temperature lower limit
end
if values(2) >= 1 && values(2) <=351 % temperature upper limit
tmax = values(2);
else; tmax = 300;
end
tstp = values(3);
pHc = values(4);
cstp = values(7)/100;
if isempty(answer{11,1}) == 0 % fixed concentration
cfix = abs((values(11)/100)-1);
else; cfix = 1;
end
if isempty(answer{10,1}) == 0 % fixed Al-con.
struct.Al = values(10);
else; struct.Al = struct.Al;
end
if isempty(answer{9,1}) == 0 % fixed pH value
pHn = values(9);
else; pHn = struct.pH;
end
threshold = values(12); % threshold minerals

% Selection of possible mineral phases
%============================== llnl.dat ===================================
```

```
list2 = {'Calcite ','Araggonite ','Dolomite ','Spinel ','Hematite ',...
'Goethite ','Pyrite ','Pyrrhotite ','Diaspore ','Gibbsite ',...
'Anhydrite ','Gypsum ','Forsterite ','Grossular ','Andradite ',...
'Andalusite ','Sillimanite ','Kyanite ','Gehlenite ','Lawsonite ',...
'Epidote ','Zoisite ','Wollastonite ','Diopside ','Hedenbergite ',...
'Ferrosilite ','Enstatite ','Anthophyllite ','Tremolite ',...
'Pargasite ','Talc ','Muscovite ','Paragonite ','Phlogopite ',...
'Illite ','Smectite-high-Fe-Mg ','Smectite-low-Fe-Mg ',...
'Clinochlore-14A ','Clinochlore-7A ','Kaolinite ','Quartz ',...
'Chalcedony ','Anorthite ','Albite_high ','Albite_low ',...
'Sanidine_high ','K-Feldspar ','Maximum_Microcline ','Analcime ',...
'Laumontite ','Wairakite ','Barite ','Pyrophyllite ','Dolomite-dis ',...
'Dolomite-ord ','Beidellite-Ca ','Beidellite-K ','Beidellite-Mg ',...
'Beidellite-Na ','Montmor-Ca ','Montmor-K ','Montmor-Mg ','Montmor-Na ',...
'Annite ','Nepheline ','Nontronite-Ca ','Nontronite-H ',...
'Nontronite-K ','Nontronite-Mg ','Nontronite-Na ','Saponite-Ca ',...
'Saponite-H ','Saponite-K ','Saponite-Mg ','Saponite-Na ','Scolecite ',...
'Sepiolite ','Clinoptilolite-Ca ','Clinoptilolite-K ',...
'Clinoptilolite-Na ','Stilbite ','Natrolite ','Mordenite ','Mesolite ',...
'Chamosite-7A ','Ripidolite-14A ','Ripidolite-7A ','Margarite ',...
'SiO2(am) '};
[indx,tf] = listdlg('ListString',list2,'InitialValue',...
[1 11 32 35 40 41 45 47 48 49 76 81]);
%================================================================================
minerals = list2(indx);
minC = strcat('si_',minerals(:));
minC = regexprep(minC, '\s', '');

% Defining the size of later on used matrices and vectors
Sensitivity = cell(aaa*2,bbb,ccc,ddd);
ValuepH = zeros(aaa,1);
ValueAl = zeros(bbb,1);
ValueC = zeros(ccc,1);
ValueK = zeros(ddd,1);
starttime = cputime;
fprintf('Sensitvity analysis started \n')

% For-loops of the several sensitivity analyses calculated via IPhreeqC
for bb = 1:bbb   % numbre of aluminium concentration sensitivity steps
for cc = 1:ccc   % number of steamloss/dilution sensitivity steps
for aa = 1:aaa   % number of pH sensitivity steps
for dd = 1:ddd   % number of individual concentration sensitivity steps
    iphreeqc = actxserver('IPhreeqcCOM.Object');
    iphreeqc.LoadDatabase(['C:\Program Files\USGS\IPhreeqcCOM' ...
        ' 3.7.3-15968\database\llnl.dat']); % pathname to IPhreeqcCOM
    iphreeqc.ClearAccumulatedLines;
    iphreeqc.AccumulateLine ('SOLUTION 1');
    iphreeqc.AccumulateLine (['-units ' con]);
```

```
iphreeqc.AccumulateLine (['-pressure ' (num2str(struct.Pressure))]);
iphreeqc.AccumulateLine (['-temperature ' (num2str(...
    struct.Temperature))]);
iphreeqc.AccumulateLine (['-pH ' (num2str(struct.pH))]);

if bb == 1  % size of aluminium concentration sensitivity steps
    iphreeqc.AccumulateLine (['Al ' (num2str(struct.Al))]);
    ValueAl(bb,:) = (struct.Al);
elseif (bb > 1) && (bb <= bbb)
    iphreeqc.AccumulateLine (['Al ' (num2str((struct.Al)...
        *(0.1+1.9/(bbb-1)*(bb-1))))]);
    ValueAl(bb,:) = (struct.Al*(0.1+1.9/(bbb-1)*(bb-1)));
end

if cc == 1 % size of steamloss/dilution sensitivity steps or fix
    iphreeqc.AccumulateLine (['C(4) ' (num2str(struct.CO2/cfix))]);
    iphreeqc.AccumulateLine (['Si ' (num2str(struct.SiO2/cfix))]);
    iphreeqc.AccumulateLine (['Cl ' (num2str(struct.Cl/cfix))]);
    iphreeqc.AccumulateLine (['Ca ' (num2str(struct.Ca/cfix))]);
    iphreeqc.AccumulateLine (['S(-2) ' (num2str(struct.H2S/cfix))]);
    iphreeqc.AccumulateLine (['K ' (num2str(struct.K/cfix))]);
    iphreeqc.AccumulateLine (['Fe ' (num2str(struct.Fe/cfix))]);
    iphreeqc.AccumulateLine (['Na ' (num2str(struct.Na/cfix))]);
    iphreeqc.AccumulateLine (['Mg ' (num2str(struct.Mg/cfix))]);
    iphreeqc.AccumulateLine (['S(6) ' (num2str(struct.SO4/cfix))]);
    iphreeqc.AccumulateLine (['Ba ' (num2str(struct.Ba/cfix))]);
    ValueC(cc,:) = cfix;
elseif (cc > 1) && (cc <= ((ccc-1)/2)+1)
    iphreeqc.AccumulateLine (['C(4) ' (num2str(struct.CO2/abs(cfix...
        +cstp*(cc-1))))]);
    iphreeqc.AccumulateLine (['Si ' (num2str(struct.SiO2/abs(cfix...
        +cstp*(cc-1))))]);
    iphreeqc.AccumulateLine (['Cl ' (num2str(struct.Cl/abs(cfix...
        +cstp*(cc-1))))]);
    iphreeqc.AccumulateLine (['Ca ' (num2str(struct.Ca/abs(cfix...
        +cstp*(cc-1))))]);
    iphreeqc.AccumulateLine (['S(-2) ' (num2str(struct.H2S/abs(cfix...
        +cstp*(cc-1))))]);
    iphreeqc.AccumulateLine (['K ' (num2str(struct.K/abs(cfix...
        +cstp*(cc-1))))]);
    iphreeqc.AccumulateLine (['Fe ' (num2str(struct.Fe/abs(cfix...
        +cstp*(cc-1))))]);
    iphreeqc.AccumulateLine (['Na ' (num2str(struct.Na/abs(cfix...
        +cstp*(cc-1))))]);
    iphreeqc.AccumulateLine (['Mg ' (num2str(struct.Mg/abs(cfix...
        +cstp*(cc-1))))]);
    iphreeqc.AccumulateLine (['S(6) ' (num2str(struct.SO4/abs(cfix...
        +cstp*(cc-1))))]);
```

```
        iphreeqc.AccumulateLine (['Ba ' (num2str(struct.Ba/abs(cfix...
            +cstp*(cc-1))))]);
        ValueC(cc,:) = (cfix+cstp*(cc-1));
    elseif (cc > ((ccc-1)/2)+1) && (cc <= ccc)
        iphreeqc.AccumulateLine (['C(4) ' (num2str(struct.CO2/abs(cfix...
            -cstp*(cc-((ccc-1)/2+1))))]);
        iphreeqc.AccumulateLine (['Ca ' (num2str(struct.Ca/abs(cfix...
            -cstp*(cc-((ccc-1)/2+1))))]);
        iphreeqc.AccumulateLine (['Cl ' (num2str(struct.Cl/abs(cfix...
            -cstp*(cc-((ccc-1)/2+1))))]);
        iphreeqc.AccumulateLine (['Fe ' (num2str(struct.Fe/abs(cfix...
            -cstp*(cc-((ccc-1)/2+1))))]);
        iphreeqc.AccumulateLine (['S(-2) ' (num2str(struct.H2S/abs(cfix...
            -cstp*(cc-((ccc-1)/2+1))))]);
        iphreeqc.AccumulateLine (['K ' (num2str(struct.K/abs(cfix...
            -cstp*(cc-((ccc-1)/2+1))))]);
        iphreeqc.AccumulateLine (['Mg ' (num2str(struct.Mg/abs(cfix...
            -cstp*(cc-((ccc-1)/2+1))))]);
        iphreeqc.AccumulateLine (['Na ' (num2str(struct.Na/abs(cfix...
            -cstp*(cc-((ccc-1)/2+1))))]);
        iphreeqc.AccumulateLine (['Si ' (num2str(struct.SiO2/abs(cfix...
            -cstp*(cc-((ccc-1)/2+1))))]);
        iphreeqc.AccumulateLine (['S(6) ' (num2str(struct.SO4/abs(cfix...
            -cstp*(cc-((ccc-1)/2+1))))]);
        iphreeqc.AccumulateLine (['Ba ' (num2str(struct.Ba/abs(cfix...
            -cstp*(cc-((ccc-1)/2+1))))]);
        ValueC(cc,:) = (cfix-cstp*(cc-((ccc-1)/2+1)));
    end

    if dd == 1  % size of individual concentration sensitivity steps
        iphreeqc.AccumulateLine (['Si ' (num2str(struct.SiO2))]);
        ValueK(dd,:) = (struct.SiO2);
    elseif (dd > 1) && (dd <= ddd)
        iphreeqc.AccumulateLine (['Si ' (num2str((struct.SiO2)*(0.8+0.4/...
            (ddd-1)*(dd-1))))]);
        ValueK(dd,:) = (struct.SiO2*(0.8+0.4/(ddd-1)*(dd-1)));
    end

    if aa == 1  % size of pH sensitivity steps
        if struct.pH <= pHn
        iphreeqc.AccumulateLine ('EQUILIBRIUM_PHASES 1');
        iphreeqc.AccumulateLine (['Fix_pH -' (num2str(pHn)) ' NH3']);
        else
        iphreeqc.AccumulateLine ('EQUILIBRIUM_PHASES 1');
        iphreeqc.AccumulateLine (['Fix_pH -' (num2str(pHn)) ' HNO3']);
        end
        ValuepH(aa,:) = (pHn);
    elseif (aa > 1) && (aa <= ((aaa-1)/2)+1)
```

```
        if struct.pH < pHn-pHc*(aa-1)
        iphreeqc.AccumulateLine ('EQUILIBRIUM_PHASES 1');
        iphreeqc.AccumulateLine (['Fix_pH -' (num2str(pHn-pHc...
            *(aa-1))) ' NH3']);
        else
        iphreeqc.AccumulateLine ('EQUILIBRIUM_PHASES 1');
        iphreeqc.AccumulateLine (['Fix_pH -' (num2str(pHn-pHc...
            *(aa-1))) ' HNO3']);
        end
        ValuepH(aa,:) = (pHn-pHc*(aa-1));
    elseif (aa > ((aaa-1)/2)+1) && (aa <= aaa)
        if struct.pH <= pHn+pHc*(aa-(((aaa-1)/2)+1))
        iphreeqc.AccumulateLine ('EQUILIBRIUM_PHASES 1');
        iphreeqc.AccumulateLine (['Fix_pH -' (num2str(pHn+pHc...
            *(aa-(((aaa-1)/2)+1)))) ' NH3']);
        else
        iphreeqc.AccumulateLine ('EQUILIBRIUM_PHASES 1');
        iphreeqc.AccumulateLine (['Fix_pH -' (num2str(pHn+pHc...
            *(aa-(((aaa-1)/2)+1)))) ' HNO3']);
        end
        ValuepH(aa,:) = (pHn+pHc*(aa-(((aaa-1)/2)+1)));
    end

% pH fixation in IPhreeqC
    iphreeqc.AccumulateLine ('PHASES');
    iphreeqc.AccumulateLine ('Fix_pH');
    iphreeqc.AccumulateLine ('H+ = H+');
    iphreeqc.AccumulateLine ('-log_k 0');
    iphreeqc.AccumulateLine ('REACTION_TEMPERATURE 1');
    iphreeqc.AccumulateLine ([(num2str(tmin)) ' ' (num2str(tmax))...
        ' in ' (num2str(tstp)) ' steps']);
    iphreeqc.AccumulateLine ('SELECTED_OUTPUT 1');
    iphreeqc.AccumulateLine (['-si ' minerals{:}]);
    iphreeqc.AccumulateLine ('-temperature');
    iphreeqc.AccumulateLine ('-pH');
    iphreeqc.AccumulateLine ('-totals Al');
    try
    iphreeqc.RunAccumulated;

% Extraction of temperature data and saturation indices of mineral phases
        out_PHREEQC = iphreeqc.GetSelectedOutputArray;
        names = out_PHREEQC(1,:);
        Te = find(strncmp('temp',names,4));
        te = 2:size(out_PHREEQC);
        Temperature = cell2mat(out_PHREEQC(te,Te));
        Si = find(strncmp('si_',names,3));
        Matrix = cell2mat(out_PHREEQC(te,Si));
        Matrix(Matrix == 0) = 0.001;
```

```
% Clean up of non- or multiple-intersecting mineral phases
        Data = diff(sign(Matrix));
        d = any(Data(2:end,:));
        e = sum(Data(2:end,(d == 1)));
        f = find(e == 0);
        Matrix(:,(d == 0)) = [];
        Matrix(:,f) = [];
        NameLegend = names(Si(1):end);
        NameLegend(d == 0) = [];
        NameLegend(f) = [];
        Data = diff(sign(Matrix));
        g = sum(abs(Data(2:end,:)));
        Matrix(:,(g > 2)) = [];
        NameLegend(g > 2) = [];

% Calculation of equilibrium temperature of each mineral phase
            Temp = zeros(1,numel(NameLegend));
            h = 1;
            for j = 1:numel(NameLegend)
                Temp(h) = Temperature(find(diff(sign(Matrix(2:end,j)))));
                h = h+1;
            end

% Saving equilibrium data (temperatures of mineral phases) in a cell-array
        Sensitivity{aa*2-1,bb,cc,dd} = NameLegend;
        Sensitivity{aa*2,bb,cc,dd} = num2cell(Temp);
        if numel(cell2mat(Sensitivity{aa*2,bb,cc,dd})) <= threshold
            Sensitivity{aa*2-1,bb,cc,dd} = minC(:).';
            Sensitivity{aa*2,bb,cc,dd} = {nan(1)};
        end
        catch

    end
end
end
end
Time = cputime - starttime
end
fprintf('Sensitvity analysis completed \n')

% Clean up mineral phases for a consistent, statistically evaluable dataset
for rr = 1:size(Sensitivity,2)
   for tt = 1:2:size(Sensitivity,1)
      for oo = 1:size(Sensitivity,3)
         for uu = 1:size(Sensitivity,4)
            try
               x = setdiff(Sensitivity{1,1,1,1},Sensitivity{tt,rr,oo,uu});
```

```
            z = setdiff(Sensitivity{tt,rr,oo,uu},Sensitivity{1,1,1,1});
                 if  isempty(x) == 0
                   for ee = 1:size(Sensitivity,2)
                      for ff = 1:2:size(Sensitivity,1)
                         for pp = 1:size(Sensitivity,3)
                            for vv = 1:size(Sensitivity,4)
                               y = find(contains(Sensitivity...
                                   {ff,ee,pp,vv},x));
                               for gg = length(y):-1:1
                                 Sensitivity{ff,ee,pp,vv}...
                                     (:,y(1,gg)) = [];
                                 Sensitivity{ff+1,ee,pp,vv}...
                                     (:,y(1,gg)) = [];
                               end
                            end
                         end
                      end
                   end
                 elseif isempty(z) == 0
                   for ee = 1:size(Sensitivity,2)
                      for ff = 1:2:size(Sensitivity,1)
                         for pp = 1:size(Sensitivity,3)
                            for vv = 1:size(Sensitivity,4)
                               y = find(contains(Sensitivity...
                                   {ff,ee,pp,vv},z));
                               for gg = length(y):-1:1
                                 Sensitivity{ff,ee,pp,vv}...
                                     (:,y(1,gg)) = [];
                                 Sensitivity{ff+1,ee,pp,vv}...
                                     (:,y(1,gg)) = [];
                               end
                            end
                         end
                      end
                   end
                 end
          catch
          end
        end
      end
   end
end
fprintf('Mineral comparison completed \n')

% Statistical evaluation of the matched data and extraction in a matrix
SenMa = zeros(size(Sensitivity,1)/2,size(Sensitivity,2),...
    size(Sensitivity,3),size(Sensitivity,4));
for ll = 1:size(Sensitivity,2)
```

```
    for ii = 2:2:size(Sensitivity,1)
        for qq = 1:size(Sensitivity,3)
            for ww = 1:size(Sensitivity,4)
                if isempty(Sensitivity{ii,ll,qq,ww}) == 1
                    Sensitivity{ii,ll,qq,ww} = {nan(1)};
                end
                SenMa(ii/2,ll,qq,ww) = range(cell2mat(Sensitivity...
                    {ii,ll,qq,ww}));
            end
        end
    end
end
[Mmin,Imin] = min(SenMa(:),[],'omitnan');
[Ib_row, Ib_col, Ib_pag, Ib_cub] = ind2sub(size(SenMa),Imin);
NameLegende = Sensitivity{2*Ib_row-1,Ib_col,Ib_pag,Ib_cub};
NameLegende = regexprep(NameLegende, 'si_', ' ');
fprintf('Statistics completed \n')

% Rerun IPhreeqC with the best fit values obtained by sensitivity analyses
% Enter ValueK(Ib_cub) for individual element concentration
iphreeqc = actxserver('IPhreeqcCOM.Object');
iphreeqc.LoadDatabase(['C:\Program Files\USGS\' ...
    'IPhreeqcCOM 3.7.3-15968\database\llnl.dat']);% pathname to IPhreeqcCOM
iphreeqc.ClearAccumulatedLines;
iphreeqc.AccumulateLine ('SOLUTION 1');
iphreeqc.AccumulateLine (['-units ' con]);
iphreeqc.AccumulateLine (['-pressure ' (num2str(struct.Pressure))]);
iphreeqc.AccumulateLine (['-temperature ' (num2str(struct.Temperature))]);
iphreeqc.AccumulateLine (['Al ' (num2str(ValueAl(Ib_col)))]);
iphreeqc.AccumulateLine (['C(4) ' (num2str(struct.CO2/ValueC(Ib_pag)))]);
iphreeqc.AccumulateLine (['Ca ' (num2str(struct.Ca/ValueC(Ib_pag)))]);
iphreeqc.AccumulateLine (['Cl ' (num2str(struct.Cl/ValueC(Ib_pag)))]);
iphreeqc.AccumulateLine (['Fe ' (num2str(struct.Fe/ValueC(Ib_pag)))]);
iphreeqc.AccumulateLine (['S(-2) ' (num2str(struct.H2S/ValueC(Ib_pag)))]);
iphreeqc.AccumulateLine (['K ' (num2str(struct.K/ValueC(Ib_pag)))]);
iphreeqc.AccumulateLine (['Mg ' (num2str(struct.Mg/ValueC(Ib_pag)))]);
iphreeqc.AccumulateLine (['Na ' (num2str(struct.Na/ValueC(Ib_pag)))]);
iphreeqc.AccumulateLine (['Si ' (num2str(struct.SiO2/ValueC(Ib_pag)))]);
iphreeqc.AccumulateLine (['Ba ' (num2str(struct.Ba/ValueC(Ib_pag)))]);
iphreeqc.AccumulateLine ('EQUILIBRIUM_PHASES 1');
    if struct.pH <= ValuepH(Ib_row)
    iphreeqc.AccumulateLine (['Fix_pH -' (num2str(ValuepH(Ib_row)))...
        ' NH3']);
    else
    iphreeqc.AccumulateLine (['Fix_pH -' (num2str(ValuepH(Ib_row)))...
        ' HNO3']);
    end
iphreeqc.AccumulateLine ('PHASES');
```

```
iphreeqc.AccumulateLine ('Fix_pH');
iphreeqc.AccumulateLine ('H+ = H+');
iphreeqc.AccumulateLine ('-log_k 0');
iphreeqc.AccumulateLine ('REACTION_TEMPERATURE 1');
iphreeqc.AccumulateLine ([(num2str(tmin)) ' ' (num2str(tmax))...
    ' in ' (num2str(tstp)) ' steps']);
iphreeqc.AccumulateLine ('SELECTED_OUTPUT 1');
iphreeqc.AccumulateLine (['-si' NameLegende{:}]);
iphreeqc.AccumulateLine ('-temperature');
iphreeqc.AccumulateLine ('-pH');
iphreeqc.RunAccumulated;

% Extraction of temperature data and saturation indices of mineral phases
out_PHREEQCB = iphreeqc.GetSelectedOutputArray;
namesB = out_PHREEQCB(1,:);
TeB = find(strncmp('temp',namesB,4));
teB = 2:size(out_PHREEQCB);
Temperature = cell2mat(out_PHREEQCB(teB,TeB));
SiB = find(strncmp('si_',namesB,3));
MatrixB = cell2mat(out_PHREEQCB(teB,SiB));
MatrixB(MatrixB == 0) = 0.001;

% Clean up of non- or multiple-intersecting mineral phases
DataB = diff(sign(MatrixB));
e = any(DataB(2:end,:));
f = sum(DataB(2:end,(e == 1)));
g = find(f == 0);
MatrixB(:,(e == 0)) = [];
MatrixB(:,g) = [];
NameLegendB = namesB(SiB(1):end);
NameLegendB(e == 0) = [];
NameLegendB(g) = [];
DataB = diff(sign(MatrixB));
i = sum(abs(DataB(2:end,:)));
MatrixB(:,(i > 2)) =[];
NameLegendB(i > 2) = [];

% Calculation of equilibrium temperature of each mineral phase
        Best = zeros(1,numel(NameLegendB));
        jj = 1;
        for kk = 1:numel(NameLegendB)
            Best(jj) = Temperature(find(diff(sign(MatrixB(2:end,kk)))));
            jj = jj+1;
        end

% Visulisation of saturation curves
o = 2:size(Temperature);
figure
```

```
u = header(a+b(1),dName);
sgtitle(u);
plotbrowser('on')
plot(Temperature(o),MatrixB(o,:));
xlabel('Temperature [C]')
ylabel('Saturation index')
line([0 max(Temperature)], [0 0],'Linestyle','--','Color','k')
h = legend(NameLegendB,'Location','best');
h.Interpreter = 'none';
keyboard % possibility to delete false phases or just click on "RUN" again
fc = get(gca,'Children');
hLegend = findobj(gcf,'Type','Legend');
LegendB = hLegend.String;
MatrixB = get(fc,'YData');
MatrixB = (flip(cell2mat(MatrixB(2:end,:))))';
Best = zeros(1,numel(LegendB));
ss = 1;
    for tt = 1:numel(LegendB)
        Best(ss) = Temperature(find(diff(sign(MatrixB(2:end,tt)))));
        ss = ss+1;
    end

% Statistics and output of the best fit parameters
RMSE = rms(MatrixB');
SDEV = std(MatrixB');
RMED = median(abs(MatrixB'));
MEAN = mean(abs(MatrixB'));
fprintf('Best fit completed \n')
Results = [(1-ValueC(Ib_pag))*100, ValuepH(Ib_row),ValueAl(Ib_col),...
    ValueK(Ib_cub)];
formatSpec = ['-Steamloss/+Dilution %2.0f%%, pH is %4.3f, Aluminium' ...
    ' concentration %5.6f, and SiO2 concentration %.3f %s \n'];
fprintf(formatSpec,Results,con)

% Ploting and saving of the best fit data
LegendB = regexprep(LegendB, 'si_', '');
o = 2:size(MatrixB);
figure
u = header(a+b(1),dName);
sgtitle(u);
subplot(2,2,1) %[1,2]
plot(Temperature(o),MatrixB(o,:));
title('Saturation indices')
xlabel('Temperature [C]')
ylabel('Saturation index')
line([0 max(Temperature)], [0 0],'Linestyle','--','Color','k')
h = legend(LegendB,'Location','best');
h.Interpreter = 'none';
```

```
legend('boxoff');

subplot(2,2,3)
t = plot(Temperature(o),RMSE(o),Temperature(o),SDEV(o),Temperature(o),...
         RMED(o),Temperature(o),MEAN(o));
title('Statistics')
xlabel('Temperature [C]')
ylabel('Saturation index')
legend('RMES','SDEV','RMED','MEAN');

subplot(2,2,4)
boxplot(Best,'whisker',8);
ylim([0 max(Temperature)])
title('Temperature estimation')
xticklabels(strjoin(u))
ylabel('Temperature [C]')
drawnow;

subplot(2,2,2)
for xx = 1:numel(ValueC)
surf(sort(ValueAl),ValuepH,SenMa(:,:,xx),'FaceColor','interp'); hold on
end
xlabel('Al concentration')
ylabel('pH value')
zlabel('\Delta T')
saveas(gcf, strjoin(u),'fig');
fprintf('Sample finished \n')
```

C

ANNRG

```
"""
AnnRG - An artificial neural network solute geothermometer

FUNCTION:
    - Version 0.03: M. Vollmer (KIT), L. H. Ystroem (KIT) - June 2023
An artificial neural network solute geothermometer trained by data from
the measured reservoir temperatures worldwide. Using a feedforward
multilayer perceptron to solve the regression analysis of fluid chemistry
and reservoir temperature.

INPUT:
    - cvs-file:['pH','Na','K','Ca','Mg','SiO2','Cl','Temperature']

OUTPUT:
    - graphical output of predicted vs. measured data plus error diagrams
    - array of errors and predictions

PLEASE NOTE:
    - The solute ANN geothermometer was programmed in Python 3.8 with
      associated libraries: pandas, matplotlib, numpy, seaborn, tensorflow,
      keras

INSTRUCTION:
- To use the geothermometer, Python (3.8) and associated libraries
  must be installed
- Training data and new unknown data must be in the recommended csv-file
  template: ['pH','Na','K','Ca','Mg','SiO2','Cl','T']
- csv-input-files must be renamed within the code (lines 59 & 151)
- Start the ANN via the Run-button
- Results are visualised in "plots" and output text on the console
- Further results can be picked from the variables
"""
# Libraries
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from numpy.random import seed

# Preprocessing
from sklearn.model_selection import train_test_split
```

```
from sklearn.preprocessing import StandardScaler

# Metrics
from sklearn import metrics
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_percentage_error

#Tensorflow & Keras
from tensorflow.keras import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.callbacks import EarlyStopping
from tensorflow.keras.callbacks import ModelCheckpoint
import tensorflow as tf

# Reading the csv-inputfile and delet all nan/0 from data
data= pd.read_csv("Training_correct.csv", delimiter=",")
data= data.dropna()

# Fixing up global and local seed
seed(0)
tf.random.set_seed(0)

# Splitting the input data
temp, test = train_test_split(data, test_size=0.2)
train, val = train_test_split(temp, test_size=0.1)

# Define input variables and output variable
X_train = train[['pH','Na','K','Ca','Mg','SiO2','Cl']]
y_train = train[['T']]
X_val = val[['pH','Na','K','Ca','Mg','SiO2','Cl']]
y_val = val[['T']]
X_test = test[['pH','Na','K','Ca','Mg','SiO2','Cl']]
y_test = test[['T']]

# Scale and centre data
scaler_input, scaler_target = StandardScaler(), StandardScaler()
scaler_input.fit(X_train)
scaler_target.fit(y_train)
X_train = scaler_input.transform(X_train)
y_train = scaler_target.transform(y_train)
X_test = scaler_input.transform(X_test)
y_test = scaler_target.transform(y_test)
X_val = scaler_input.transform(X_val)
y_val = scaler_target.transform(y_val)

# Determine the input features
n_features = X_train.shape[1]
```

```
# Set initializer with optimiser
kernel_initializer = 'normal'
opt = tf.keras.optimizers.Adam(learning_rate=0.001)

# Implementing Early Stopping
es = EarlyStopping(monitor='val_loss', mode='auto', verbose=1, patience=20,
     restore_best_weights=True)

# Save the trained model
checkpoint_filepath = './checkpoint.hdf5'
checkpoint = ModelCheckpoint(filepath = checkpoint_filepath, verbose = 1,
    save_best_only = True,
                                    monitor ='val_loss', save_weights_only =
                                         True, mode = "auto")

# Define model architecture
model = Sequential()
model.add(Dense(80, activation='relu', kernel_initializer=
    kernel_initializer, input_shape=(n_features,)))
model.add(Dense(1))

# Compile the model
model.compile(optimizer=opt, loss='mean_squared_error')

# Hyperparameter optimisation
history = model.fit(X_train, y_train, epochs=300, batch_size=16, verbose=2,
     validation_data=(X_val,y_val),
                    callbacks=[es])

# Prediction of the test set
yhat = model.predict(X_test)
X_train_p = model.predict(X_train)
X_val = model.predict(X_val)

# Plot learning curves
plt.xlabel('Epochs')
plt.ylabel('Mean square error')
plt.plot(history.history['loss'], label='Training')
plt.plot(history.history['val_loss'], label='Validation')
plt.legend()
plt.title('Learning curves')
plt.savefig('loss.png')
plt.show()

# Inverse transform scaled and centred data
y_test = scaler_target.inverse_transform(y_test)
ypred = scaler_target.inverse_transform(yhat)
```

```
X_train_p = scaler_target.inverse_transform(X_train_p)
y_train = scaler_target.inverse_transform(y_train)
X_val = scaler_target.inverse_transform(X_val)
y_val = scaler_target.inverse_transform(y_val)
X_train = scaler_input.inverse_transform(X_train)

# Metric scores (change for other sets)
mse = mean_squared_error(y_test, ypred)
rmse = np.sqrt(metrics.mean_squared_error(y_test,ypred))
mape =mean_absolute_percentage_error(y_test, ypred)
r2 = r2_score(y_test, ypred)
print('MAPE: %.3f' % mape)
print('MSE: %.3f' % mse)
print('RMSE: %.3f' % np.sqrt(mse))
print('R_squared: %.3f' % r2)

# Reading in csv-inputfile of new transferable data
extra = pd.read_csv("Verification.csv", delimiter=",")
x_a = extra[['pH','Na','K','Ca','Mg','SiO2','Cl']]
y_a = extra[['T']]

# Scale, centre, and predict new transferable centre data
scaler_x = StandardScaler()
scaler_y = StandardScaler()
scaler_x.fit(x_a)
scaler_y.fit(y_a)
Xnew = scaler_x.transform(x_a)
y_pred_a = model.predict(Xnew)
y_pred_a = scaler_y.inverse_transform(y_pred_a)
y_true = y_a

# Polt of predicted temperature vs measured temperature plus transferred
    data
plt.figure()
plt.plot(X_train_p, y_train,'.b', label='Training',markersize=8)
#color='#808080', marker='.', markersize=8, label='ANN', linewidth=0
plt.plot(X_val, y_val,'.b',markersize=8)
plt.plot(y_test, ypred,'.r', label='Testing',markersize=8)
plt.plot(y_true, y_pred_a,'.g', label='Transfer',markersize=8)
plt.plot(y_test,y_test,'k', label ='Regression') #color='#606060'
plt.xlabel('Measured bottom hole temperature [C]')
plt.ylabel('Predicted bottom hole temperature [C]')
plt.xlim(0,350)
plt.ylim(0,350)
plt.legend(loc='upper left')
plt.title('R$^2$: %.3f' % r2)
plt.savefig('regression.png')
plt.show()
```

```
# Error histogram of test set
error = ypred - y_test
plt.hist(error, bins=20)
plt.xlabel('Predicted temperature difference [K]')
plt.ylabel('Quantity')
plt.xlim(-40,40)
plt.ylim(0,6)
plt.title('Error histogram of tested data')
plt.savefig('histogram.pdf')
plt.show()

error_train = y_train - X_train_p
error_val = y_val - X_val
error_main = np.concatenate([error,error_val,error_train])

# Violinplot of the error distribution
sns.violinplot(error_main, cut=1)
plt.title('Error distribution')
plt.ylabel('Dataset')
plt.xlabel('Predicted temperature difference [K]')
plt.xlim(-40,40)
plt.savefig('disribution.pdf')
plt.show()

# Plot of the outlier removal
abs_error_main = np.absolute(error_main)
measured = np.concatenate([y_test , X_train_p, X_val])
predict = np.concatenate([ypred , y_train, y_val])
rmse_error = np.sqrt(metrics.mean_squared_error(measured, predict))
plt.stem(abs_error_main, linefmt=':')
plt.axhline(y=2*rmse_error, c='black', ls=':')
plt.title('Outlier detection')
plt.ylabel('RMSE')
plt.xlabel('Datapoint')
plt.ylim(-5,75)
plt.savefig('outlier removal.pdf')
plt.show()
```