AnnRG - An artificial neural network solute geothermometer

Lars H. Ystroem *, Mark Vollmer, Thomas Kohl, Fabian Nitschke

Karlsruhe Institute of Technology (KIT), Institute of Applied Geoscience (AGW), Division of Geothermal Energy and Reservoir Technology, Campus South, Adenauerring 20b, 76131, Karlsruhe, Germany

ARTICLE INFO

Keywords:
AnnRG
Machine learning
Artificial neural network
Solute geothermometry
Geochemical exploration
Reservoir temperature prediction

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⁺, Ca²⁺, Mg²⁺, Cl⁻, SiO₂, 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 R² = 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.

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 (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 and therefore can be used to infer the reservoir temperature (Fournier and Truesdell, 1974). Since then, continuously new geothermometers have been developed or improved. Most applications of solute geothermometers are based on SiO₂ concentration as well as the major cation ratios of Na/K, Na/K/Ca, and K²/Mg (Fournier and Truesdell, 1973; Fournier and Potter, 1979, 1982; Giggenbach, 1988; Arnórsson, 2000b). Na⁺, K⁺, Ca²⁺, and Mg²⁺ are the major cations encountered in crustal rocks and geothermal water, whereas Cl⁻ and SO₄²⁻ 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 Mabon, 1964). In hydrothermal systems, the Ca²⁺ concentration is mostly given by the temperature and salinity-dependent solubility of calcite and calcium-aluminium-silicates (Ellis, 1963; Giggenbach, 1981). Likewise, Mg²⁺ is controlled by the solubility of K-Mg layer silicates, which is chloride-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 (Debye and Hückel, 1923; Davies, 1938).

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 (Pang, 1988; Pang and Reed, 1998; Nitschke et al., 2017, 2018).

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, 1988; Pang and Reed, 1998). 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, CO₂ fugacity, and a mixing fraction by minimising the saturation indices of a suggested mineral set (Palmer et al., 2014). 3) GeoT 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/GeoT (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 processes. 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 machine learning 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 (Dransch, 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 afflicted 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-Zarate et al. (2019), using CO₂, H₂S, CH₄, and H₂ 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, SiO₂, 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 geo-thermometers, 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.

2. Data & methods

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 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 Fig. 1.

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., 2020).
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 are given.

<table>
<thead>
<tr>
<th>Data</th>
<th>Sample size</th>
<th>Lithology</th>
<th>Temperature range</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper Rhine Graben</td>
<td>40</td>
<td>Granite, Germanic Triassic facies, Permian facies</td>
<td>41 °C–260 °C</td>
<td>Panuwels et al. (1993); Aquilina et al. (1997); Vaute et al. (1998); Sanjuan et al. (2001); Sanjuan et al. (2004); Sanjuan et al. (2006); Sanjuan et al. (2010); Schindler et al. (2010); Dezayes et al. (2015); Stober and Bucher (2015); Sanjuan et al. (2016); Vidal and Genter (2018); Vidal et al. (2019)</td>
</tr>
<tr>
<td>German Molasse Basin</td>
<td>9</td>
<td>Marine facies, Dogger</td>
<td>36 °C–153 °C</td>
<td>Ystroem et al. (2022)</td>
</tr>
<tr>
<td>Paris Basin</td>
<td>18</td>
<td>Marine facies, Andesite</td>
<td>47 °C–76 °C</td>
<td>Marty et al. (1988); Richard and Bastide (1988); Criaud et al. (1989); Aronsson (1977); Aronsson et al. (1983); G语mundsson and Aronsson (2002); Aronsson et al. (2007); G语erson et al. (2015)</td>
</tr>
<tr>
<td>Iceland</td>
<td>79</td>
<td>Basalt, hyaloclastite</td>
<td>59 °C–295 °C</td>
<td>Carvalho et al. (2006)</td>
</tr>
<tr>
<td>Azores</td>
<td>19</td>
<td>Basalt</td>
<td>200 °C–240 °C</td>
<td>Carvalho et al. (2006)</td>
</tr>
<tr>
<td>El Tatio</td>
<td>6</td>
<td>Dacite</td>
<td>180 °C–253 °C</td>
<td>Ellis and Mahon (1977); Gigenbach (1978)</td>
</tr>
<tr>
<td>Miravalle</td>
<td>10</td>
<td>Andesite</td>
<td>215 °C–255 °C</td>
<td>Dennis et al. (1989); Gherardi et al. (2002)</td>
</tr>
<tr>
<td>Rotorua</td>
<td>10</td>
<td>Rhyolite, andesite</td>
<td>121 °C–211 °C</td>
<td>Mroczek et al. (2003)</td>
</tr>
</tbody>
</table>

Table 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 are given.

1993). They surpass Triassic to Permian sediments, though they are mostly produced from the deeper granitic crystalline basement (Stober and Bucher, 2015; Sanjuan et al., 2016). 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 1.2–4.4 g/l. The fluids are Na⁺-HCO₃ 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⁺-Ca²⁺-Mg²⁺-HCO₃ type to a more saline southern Na⁺-(Ca²⁺)-HCO₃-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⁺-Ca²⁺-HCO₃-Cl type fluids show high variability in their salinity (6.4–35 g/l) (Richard 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⁺-HCO₃-Cl to a Na⁺-Cl type (Aronsson 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 Água 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⁺-HCO₃ and Na⁺-HCO₃-Cl type waters (Cruz and France, 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; Gigen!ach, 1978). The geothermal site in Miravalle, 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⁺-HCO₃-Cl type (Mroczek et al., 2003).

2.2. Implementation of the dataset and data editing

To set up the dataset, the geochemical analyses of the references are digitised (Table 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 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⁺, Ca²⁺, Mg²⁺, Cl⁻, SiO₂, 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; 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 Fig. 2.

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 1). The data of these samples are stored in matrix $x \in \mathbb{R}^{163 \times 7}$ 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 $y \in \mathbb{R}^{163 \times 1}$. 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 (1)).

$$2RMSE = 2 \sqrt{\frac{\sum_{n=1}^{n} (\hat{y} - y)^2}{n}}$$

(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 Fig. 3.

The result of the threshold criterion (Equation (1)) is a temperature deviation of 40.2 K, which is indicated by the dotted black line in Fig. 3. For each sample, the absolute deviation is plotted (blue data points) throughout the database. As a result, eight outliers exceed the threshold value (Fig. 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, which is visualised by the dotted green line (Fig. 3b). The database (matrix $z \in \mathbb{R}^{155 \times 8}$) is validated for the establishment of the baseline model of the neural network.

2.3. Baseline model, network architecture, and hyperparameter optimisation

AnnRG is processed in Python 3.8.5 (van Rossum, 1995). Multiple libraries such as NumPy for scientific computation (van der Walt et al., 2011; Harris et al., 2020), pandas for data analysis (Reback et al., 2020), and Matplotlib (Hunter, 2007; Caswell et al., 2020) 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 Developers, 2022) are used for the ML algorithm. For the latter, the high-level application programming interface Keras (Chollet, 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 optimisation 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 defined by hyperparameter optimisation minimising the error within the MLP. The data structure and the intrinsic properties of the temperature prediction 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 prediction, different activation functions were tested: the rectified linear unit (ReLU) (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 and Hinton, 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 $2^s$ 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 2.

The three error types are chosen due to their inherent information in the regression analysis. In Equation (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.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (2)$$

The RMSE defines the average accuracy between the predicted and the in-situ temperature (Equation (3)).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (4)$$

<table>
<thead>
<tr>
<th>Error type</th>
<th>Training</th>
<th>Validation</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE</td>
<td>0.067</td>
<td>0.092</td>
<td>0.092</td>
</tr>
<tr>
<td>RMSE</td>
<td>9.965</td>
<td>11.269</td>
<td>10.091</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.983</td>
<td>0.972</td>
<td>0.978</td>
</tr>
</tbody>
</table>
\[ \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \]  

(3)

The \( R^2 \) shows how accurately the model predicts the measured in-situ temperatures (Equation (4)).

\[
R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}
\]

(4)

These errors are given for all three phases of training, validating, and testing the neural network. In addition, the shape of the loss function is visualised as learning curves in Fig. 4.

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. Regarding Fig. 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.

3. Results & discussion

After the establishment of the neural network architecture and the hyperparameter configuration, the final MLP (Fig. 5) is given by the following configuration (Table 3):

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 (\( \varepsilon \in \mathbb{R} 155 \times 8 \)), 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 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 from training to validation, as well as the variance of -1.178 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 (Fig. 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.

In Fig. 5, the results of the regression problem of the temperature prediction are visualised for the best-fitting MLP. The coloured dots illustrate the actual results of the training (blue), the testing (red), and their deviation of the regression line (cf. Fig. 6). The homogeneous distribution of the dots, as well as an \( R^2 \) of 0.978, approve the developed MLP.

In Fig. 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 \( \pm 30 \) 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 Fig. 7.

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 (Fig. 6). In line with the positive skewness of the error distribution (Fig. 6), the temperature predictions 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 1). The geochemical data of these four geothermal sites are processed by the MLP and compared with the measured in-situ temperatures. In Fig. 8, the results of each site are plotted on top of the developed geothermometer.

The predicted temperatures of all four sites (Fig. 8) fit the measured
temperatures in the same order of $< \pm 30$ K as the MLP error distribution (Fig. 6). The error distribution of each sample is visualised in Fig. 9. Especially for the Azores ($\pm 11$ K), El Tatio ($\pm 13$ K), and Miravalles ($\pm 15$ K), the predictions are more precise than the error distribution of the test set (Fig. 7). Nevertheless, the variation between measured and predicted temperature at Rotorua ($\pm 27$ K) is higher than the average test error ($\pm 11$ K), El Tatio ($\pm 27$ K) is higher than the average test error ($\pm 27$ K) is higher than the average test error distribution of the test set ($\pm 27$ K) is higher than the average test error distribution (Fig. 10).

### 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 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 ($\text{Na}^+$, $\text{K}^+$, $\text{Ca}^{2+}$, $\text{Mg}^{2+}$, $\text{Cl}^-$, $\text{SO}_4^2-$, and $\text{pH}$) 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 ($\text{RMSE} 9.405$) than the original dataset ($10.442$). AnnRG is applicable to regional-scale sites 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.

Fig. 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 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).

Fig. 9. Error distribution of the introduced data with a batch size of 5 K. Count of the samples against the temperature deviation of Azores, Portugal (magenta); El Tatio, Chile (blue); Miravalles, Costa Rica (cyan); Rotorua, New Zealand (red).

Fig. 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).

Authorship statement

Lars Ystroem: Conceptualisation, Data curation, Investigation, Methodology, Software, Writing - Original Draft. Mark Vollmer: Data curation, Methodology, Software. Thomas Kohl: Writing - Review & Editing, Supervision. Fabian Nitschke: Conceptualisation, Writing -
Decloration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data is cited within the manuscript.

Acknowledgement

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.

References


