Computers & Geosciences 143 (2020) 104560

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

SupPhreeqc: A program for generating customized Phreeqc thermodynamic datasets from Supcrtbl and extending calculations to elevated pressures and temperatures

Guanru Zhang [a,b], Peng Lu [c,*], Yilun Zhang [b], Kevin Tu [d], Chen Zhu [b,**]

[a] College of Ecology and Environment, Chengdu University of Technology, Chengdu, 610059, China
[b] Department of Earth and Atmospheric Sciences, Indiana University, Bloomington, IN, 47405, USA
[c] EXPEC Advanced Research Center, Saudi Aramco, Dhahran, 31311, Saudi Arabia
[d] School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, 47405, USA

ARTICLE INFO

Keywords:
Supcrt
Geochemical modeling
Gas-water-mineral interaction
Phreeqc
Carbon sequestration

ABSTRACT

Supcrt and Phreeqc are two software packages widely used for thermodynamic calculations and geochemical modeling of gas-water-mineral reactions, respectively. The versatility and easy-to-use functions of Phreeqc have earned its popularity, but the software is limited by the default thermodynamic datasets it is distributed with, which has limited Phreeqc applications to mostly at ambient temperature. To expand the applicability of Phreeqc to wider temperature and pressure conditions, this study developed the software utility SupPhreeqc that extracts thermodynamic data from Supcrtbl and translates them into Phreeqc-readable formats. By linking Supcrtbl and Phreeqc, this study helps harness high-quality thermodynamic data for minerals and a large number of inorganic and organic aqueous species with the Helgeson-Kirkham-Flowers equation of state parameters for speciation and solubility modeling at elevated *T-P*.

SupPhreeqc can generate both formats of Phreeqc datasets, namely, *phreeqc.dat* and *llnl.dat*. These two formats differ in activity coefficient and gas-mixing models, as well as pressure corrections for log *K*. As examples, three datasets were made from SupPhreeqc: *diagenesis.dat* (0.01–200 °C and up to 1000 bar), *geothermal.dat* (0.01–300 °C at P_{SAT}), and *bl.dat* at 0.5, 1, 2, and 5 kb 1000 °C. The *diagenesis.dat* and *geochermal.dat* datasets have also incorporated the library of callable BASIC scripts for rate equations and kinetic parameters for 53 minerals. These databases can be downloaded or run together with Phreeqc online at https://models.earth.indiana.edu.

1. Introduction

Zhu and Anderson (2002) state that any geochemical modeling exercise requires the "parts of three":

(1) A computer code that is an assembly of numerical techniques, bookkeeping, and control language that represents the model from acceptance of input data and instruction to delivery of output;
(2) an input file that defines the conceptual model and provides the chemical and physical measurements of the system of concern; and
(3) a dataset of thermodynamic, kinetic, and surface properties that supplies the computer code the necessary equilibrium constants and kinetic parameters for the calculations.

While many thermodynamic datasets are distributed with geochemical modeling software, it is generally the modeler's responsibility to ensure the quality of the thermodynamic data and the dataset's applicability with respect to the ranges of temperature (*T*), pressure (*P*), chemical components (*X*), salinity, and other constraints for a particular study.

Phreeqc is a widely used geochemical modeling computer program developed by the United States Geological Survey and freely distributed to the public (Parkhurst and Appelo, 1999, 2013). The program can be used for calculations of speciation, mineral or gas solubility, reaction

* Corresponding author.
** Corresponding author.
E-mail addresses: peng.lu@aramco.com (P. Lu), chenzhu@indiana.edu (C. Zhu).

https://doi.org/10.1016/j.cageo.2020.104560
Received 30 December 2019; Received in revised form 28 May 2020; Accepted 7 July 2020
Available online 2 August 2020

path, and one-dimensional coupled reactive mass transport. However, the temperature and pressure range of the software's applicability is partly limited by the *T-P* coverage of the accompanying thermodynamic and kinetics datasets.

Table 1 compares the *T-P* range, coverage of chemical systems, pressure correction, activity coefficient models, and fugacity coefficient models in the most widely used PHREEQC datasets (*phreeqc.dat* and *llnl.dat*). The earlier versions of *phreeqc.dat* are suitable for geochemical modeling at ambient conditions only (Parkhurst and Appelo, 1999), but recent improvement allows for calculations up to ~200 °C and ~1000 bars (Appelo et al., 2014). However, the parameters necessary for *T-P* extrapolation are available for only a small set of reactions; many species in *phreeqc.dat* lack the necessary parameters and cannot be used for calculations above 25 °C and 1 bar (Appelo et al., 2014). The dataset *llnl.dat* partially fulfills the need for using PHREEQC to model at elevated temperatures by covering a *T-P* range of 0.01–100 °C at 1 bar and 100–300 °C at P_{SAT}. However, not all chemical species listed in *llnl.dat* have the parameters necessary for calculations above 25 °C. *llnl.dat* also does not consider the effects of pressure changes on log *K* at pressures higher than P_{SAT}, for example, 2 kb. Also, the fugacity estimation for gases at higher temperatures and pressures is not accurate because it uses the ideal gas law.

It is desirable to have the capability to generate thermodynamic datasets for PHREEQC that cover a wide selection of chemical species and are reliably applicable to elevated *T-P* of interest. The sources of log *K*s in the dataset must be well documented and traceable. The dataset must also be updatable when new thermodynamic properties for minerals, aqueous species, and gases become available. This will greatly expand the utility of the PHREEQC program to a broader geological and environmental science community. Linking PHREEQC with SUPCRT is one way to achieve this goal.

For many decades, the SUPCRT program has been widely used in the geochemistry community for computing standard state thermodynamic properties for minerals, aqueous species, and gases at elevated temperatures and pressures up to 1000 °C and 5000 bars, respectively. The original version, SUPCRT92 (Johnson et al., 1992), was recently updated to SUPCRTBL (Zimmer et al., 2016), which includes a comprehensive set of thermodynamic data. The standard state thermodynamic properties for most mineral endmembers are from Holland and Powell (2011), Al-bearing aqueous species from Tagirov and Schott (2001), SiO_2^0(aq) and$HSiO_3^-$ from Rimstidt (1997) and Stefánsson (2001). For environmental applications, SUPCRTBL also includes As-bearing species from Nordstrom and Archer (2003), metal-arsenate and metal-arsenate aqueous complexes from Marini and Accornero (2010), As-bearing minerals including arsenopyrite (Ball and Nordstrom, 1991), scorodite and ferric-As(am) (Langmuir et al., 2006), barium-As and barium-H-As (Zhu et al., 2005), and other As-bearing solids (Nordstrom and Archer, 2003). The thermodynamic properties of boehmite, gibbsite, and dawsonite in the literature were also re-calculated to be consistent with the minerals and aqueous species mentioned above (Zimmer et al., 2016). There are about 1630 (~1060 inorganic substances) species in the *spronsbl.dat* dataset that accompanies SUPCRTBL.

Although SUPCRTBL is capable of computing the thermodynamic properties of a single species (mineral, aqueous, or gas) and reactions at *T-P* of interest, it cannot generate a formatted dataset that can be directly read by geochemical modeling software. In addition, it cannot stand alone for directly modeling fluid-related geochemical processes. Therefore, it is necessary to develop software to extract thermodynamic data calculated from SUPCRTBL and translate it into a PHREEQC-readable format.

The SUPPHREEQC program presented in this study fulfills this much-needed function. Both PHREEQC and SUPCRTBL have specific requirements in their data input and output files. Below, we will first describe these specific requirements and how SUPPHREEQC handles these requirements. Then, we will present three example datasets and demonstrate the types of datasets that SUPPHREEQC can generate. Finally, we will demonstrate how these datasets can be applied to geological and environmental processes.

2. The design of SUPPHREEQC

2.1. PHREEQC dataset structures and formats

Two types of dataset structures for PHREEQC are defined in Parkhurst and Appelo (2013) - the *phreeqc.dat* and *llnl.dat* frameworks. Each type requires a specific format and structure for master species, secondary species, activity coefficient parameters, parameters for pressure corrections of log *K*, and Equation of State (EOS) of gas for geochemical calculations. Table 2 lists four data blocks of interest in the framework of *phreeqc.dat*: SOLUTION_MASTER_SPECIES, SOLUTION_SPECIES, PHASES, and RATES. SOLUTION_MASTER_SPECIES defines element names, aqueous primary and secondary master species, and related alkalinity contribution of the master species. The SUPCRTBL reaction files discussed below must be consistent with this list.

The SOLUTION_SPECIES block defines aqueous species-related chemical reactions. The master species used in the solution species must be consistent with what is listed in the SOLUTION_MASTER_SPECIES. The "-analytic" line supplies parameters for calculating log *K* at elevated temperatures. The lack of these parameters for most entries in *phreeqc.dat* is our primary motivation for establishing the link to SUPCRTBL. The PHASES data blocks define the names of mineral and gas phases, chemical reactions, log *K*, parameters for log *K* pressure corrections, and Peng-Robinson EOS for gases. The mineral phase names in SUPCRTBL must be compatible with those in *phreeqc.dat*-type datasets.

Table 1
Comparison of thermodynamic datasets discussed in this study.

Dataset	*T-P* range	Pressure correction	*P* range	Aqueous activity coefficient model	Fugacity coefficients	Number of species	Sources
llnl.dat	0.01–100 °C at 1 bar, 100–300 °C along P_{SAT}	SUPCRT92	P_{SAT}	B-dot	Ideal gas law	~2590	Greg M. Anderson; EQ3/6[a]
phreeqc.dat	<200 °C <1 kb	Appelo et al. (2014)	up to ~1000 bar	Davies	Peng-Robinson	~310[b]	Appelo et al. (2014)
diagenesis.dat	0.01–100 °C at 1 bar, 100–200 °C along P_{SAT}	Appelo et al. (2014) & log *K* in SUPCRTBL	up to ~1000 bar	Davies	Peng-Robinson	~1040	SUPPHREEQC (this study)
geothermal.dat	0.01–100 °C at 1 bar, 100–300 °C along P_{SAT}	SUPCRTBL	P_{SAT}	B-dot	Ideal gas law	~1040	SUPPHREEQC (this study)
bl.dat	Up to 1000 °C and 5000 bar (variable *T* isobaric)	SUPCRTBL	up to 5000 bars	B-dot	Ideal gas law	~1040	SUPPHREEQC (this study)

[a] Fom 'thermo.com.V8.R6.230' prepared by Jim Johnson at Lawrence Livermore National Laboratory in Geochemist's Workbench format. Converted to PHREEQC format by Greg Anderson with assistance from David Parkhurst. The actual number of species with parameters for calculations at elevated *T-P* is much smaller than the total number of species in the dataset.

[b] The number of species with parameters for calculations at elevated *T-P* is much smaller. See Appelo et al. (2014). Shaded entries are example datasets generated in this study by SUPPHREEQC.

Table 2
The framework of the *phreeqc.dat* dataset.

Data block	Parameters
SOLUTION_MASTER_SPECIES	element name, master species, alkalinity, formula, gram formula weight of element
SOLUTION_SPECIES	association reaction - analytic: coefficients for an analytical expression of log *K* - log_k: log *K* at 25 °C - delta_h: enthalpy of reaction at 25 °C. - Vm: parameters used to calculate the specific volume (cm^3/mol) of aqueous species
PHASES	phase name dissolution reaction - analytic: coefficients for an analytical expression of log *K* - Vm: molar volume of the solid phase - Omega: acentric factor of the gas - P_c: critical pressure of the gas - T_c: critical temperature of the gas
RATES	refer to Parkhurst and Appelo (2013)

The reactions in the file RXN must be consistent with those in *phreeqc.dat*. The molar volume *Vm* of both solution species and phases is needed for pressure correction. A limited number of *Vm* is available in *phreeqc.dat*. This study provides a more complete list of these values by linking it to SUPCRTBL. The RATES block defines rate expressions for kinetic reactions. The mineral names must be identical to the names in PHASES.

Table 3 lists the five data blocks of interest in the framework of *llnl.dat*. There is an additional LLNL_AQUEOUS_MODEL_PARAMETERS block not present in *phreeqc.dat*. This data block defines the parameters for the *B*-dot equation, including Debye-Hückel parameters *A* and *B*, the $\dot{B}$ (*B*-dot) parameter, as well as parameters for activity coefficient function of CO_2(aq) (co2_coefs). In addition, the SOLUTION_SPECIES block of *llnl.dat* also defines the ion size parameter of aqueous species $\mathring{a}_i$ for the *B*-dot equation. There are no parameters for pressure corrections of log *K* and Peng-Robinson equation of state in *llnl.dat*. *llnl.dat* uses the *B*-dot equation to calculate activity coefficients. The ideal gas law is used for gases.

Both *phreeqc.dat* and *llnl.dat* dataset formats use the following polynomial analytical expression for the temperature dependence of log *K* (Parkhurst and Appelo, 2013)

Table 3
The framework of the *llnl.dat* dataset.

Data block	Parameters
LLNL_AQUEOUS_MODEL_PARAMETERS	- temperatures: temperature grid - dh_a: Debye-Hückel *A* parameter - dh_b: Debye-Hückel *B* parameter - bdot: Debye-Hückel *B*-dot parameter - co2_coefs: parameters for activity coefficient function of CO_2(aq)
SOLUTION_MASTER_SPECIES	element name, master species, alkalinity, formula, gram formula weight of element
SOLUTION_SPECIES	association reaction - llnl_gamma: hard-core diameter in *B*-dot equation - analytic: coefficients for an analytical expression of log *K* - log_k: log *K* at 25 °C - delta_h: enthalpy of reaction at 25 °C. - CO2_llnl_gamma: indicate activity coefficient for CO_2(aq) is used for this uncharged species
PHASES	phase name dissolution reaction - analytic: coefficients for an analytical expression of log *K*
RATES	refer to Parkhurst and Appelo (2013)

$$\log_{10}K = A_1 + A_2T + \frac{A_3}{T} + A_4\log_{10}T + \frac{A_5}{T^2} + A_6T^2 \quad (1)$$

where *T* is the temperature in kelvin; A_1, A_2, A_3, A_4, A_5, and A_6 are fitting parameters. Without these parameters available in the datasets, PHREEQC uses the reaction enthalpy value (-delta_h) and van't Hoff equation for extrapolating the log *K* to elevated temperatures. Lacking both parameters results in automatic calculations at 25 °C and 1 bar without warning to the user. To ensure that all species listed in the dataset have the requisite A_1, A_2, A_3, A_4, A_5, and A_6 fitting parameters was a major effort in this study.

2.2. SUPCRTBL input and output formats

SUPCRTBL produces a list of log *K* values at the specified *T-P* formatted as a plain text file. These log *K* values must be translated into the PHREEQC formats mentioned above. To produce the output file with the desired log *K*s, a series of files are necessary to ensure consistency with PHREEQC. The detailed input and output formats for SUPCRTBL are described in Zimmer et al. (2016) and Johnson et al. (1992). Here, we provide only a brief introduction. The input files include the *CON* file, *RXN* file, and direct-access SUPCRT database file (*dpronsbl.dat*). The *CON* file defines a set of *T-P* grids for calculations of thermodynamic properties and is automatically generated by SUPPHREEQC for the user-specified *T-P* range. The *RXN* file defines the chemical reactions of interest. A default *RXN* file is provided, which covers almost all inorganic aqueous species, minerals, and gases, as well as several organic aqueous species from the SUPCRTBL database, *dpronsbl.dat*. *dpronsbl.dat* contains the standard molal thermodynamic data at 25 °C and 1 bar, and parameters needed to calculate temperature and pressure dependence of these properties. Note that *dpronsbl.dat* is in direct-access format, meaning that it can be directly read by SUPCRTBL, but is unfriendly for users to read and modify. Therefore, a sequential-access counterpart, *spronsbl.dat*, is available for human utilization, which can then be converted to *dpronsbl.dat* (using CPRONSBL). The outputs of SUPCRTBL are the calculated standard molal thermodynamic properties of the reactions, including log *K* values, defined in the *RXN* file over the range of conditions specified in the *CON* file. These log *K* values are used for generating PHREEQC datasets.

2.3. The workflow of SUPPHREEQC

SUPPHREEQC is written in C++11 and contains seven functional modules (Fig. 1): (1) A reading module that prompts users for the desired dataset framework and *T-P* values. (2) A *CON* file generation module that, using the read *T*(s) and *P*, generates a SUPCRT-format *CON* file (Johnson et al., 1992) to specify the *T-P* conditions. (3) A format conversion module that calls the CPRONSBL program to convert the sequential-access SUPCRT database (*spronsbl.dat*) to direct-access format (*dpronsbl.dat*). CPRONSBL is a derivative of CPRONS92 modified to support SUPCRTBL (Zimmer et al., 2016). (4) A log *K*-calculation module that calls the SUPCRTBL program to calculate log *K* values using a *CON* file, a reaction (*RXN*) file in SUPCRTBL-format, and the *dpronsbl.dat* file. (5) A log *K* analytical expression generation module that regresses the coefficients for the polynomial analytical expression used in PHREEQC for the temperature dependence of log *K* based on the 11 discrete log *K* values from the step 4. (6) An activity coefficient parameter module that calculates the *T*- and *P*-dependent values of the Debye-Hückel *A*, *B*, and *B*-dot parameters for the *B*-dot equation (Parkhurst and Appelo, 2013). This module is only available for datasets following the *llnl.dat* framework. (7) An output module that produces the desired PHREEQC dataset following the *phreeqc.dat* or *llnl.dat* framework by integrating parameters for the analytical expression of log *K*, reactions extracted from the *RXN* file, and a set of customized master species. BASIC scripts for rate equations and kinetic parameters for minerals, the activity coefficient parameters (only for *llnl.*dat type datasets), and an accompanying

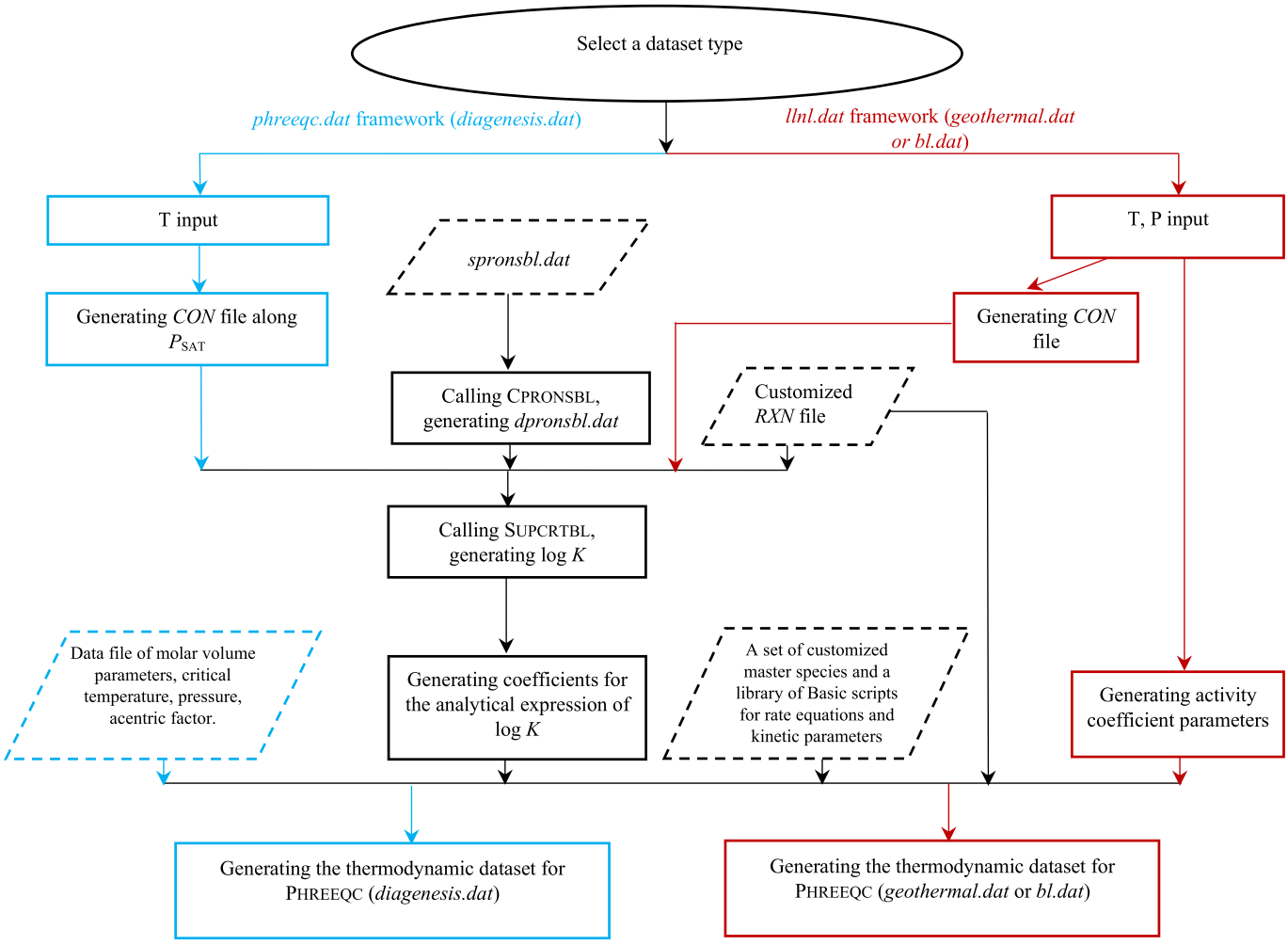

Fig. 1. A flow chart illustrating the generation of a thermodynamic dataset for PHREEQC using SUPPHREEQC. Dashed parallelograms indicate the accompanying datasets used with SUPPHREEQC. Blue and red items represent the data flow for generating *phreeqc.dat* and *llnl.dat* framework datasets, respectively. Black items indicate data flow shared by both frameworks of datasets. Names in parenthesis (*diagenesis.dat*, *geothermal.dat*, *bl-1kb.dat*) are example datasets generated using SUPPHREEQC and discussed in Section 4. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

dataset (only for *phreeqc.dat* type datasets) of molar volume parameters, critical temperatures, critical pressures, and acentric factors are embedded in SUPPHREEQC.

3. Embedded files, equations, and parameters in SUPPHREEQC

3.1. Master species file and reaction file

A set of master species in the "master species file" and an *RXN* file are customized to be consistent with *spronsbl.dat*. Currently, the master species and *RXN* file contain almost all the inorganic aqueous species, minerals, and gases found in the SUPCRTBL database as well as several organic aqueous species (e.g., CH_4, $Acetate^-$, HAcetate). The master species are subdivided into primary and secondary species, following Parkhurst and Appelo (2013).

In the *RXN* file, all aqueous complexes (master species and secondary species) are written in the format of association reactions, and all solid and gas phases are written in the format of dissolution reactions, following Parkhurst and Appelo (2013). The reactant in the association reaction for primary master species is the primary master species itself and their log *K*s are always zero. The secondary master species are defined via redox reactions, which include the related primary master species and O_2. O_2 is also a secondary master species.

Generally, the reactions of secondary aqueous species, minerals, and gases in the *RXN* file are written in a non-redox form, using the primary and secondary master species as reactants.

However, if the valence state of an element in the secondary aqueous species or phase is not in the list of master species (e.g. S_2^{2-}, polysulfides, and polythionates), then the redox-form reactions are individually specified and their valances are defined following the conventions in *llnl.dat* (Table 4).

3.2. Activity coefficient model for databases with the llnl.dat framework

3.2.1. Activity coefficient equations

In datasets generated by SUPPHREEQC that follow the *llnl.dat* framework, the "LLNL_AQUEOUS_MODEL_PARAMETERS" keyword is included, which means that the *B*-dot equation is employed to calculate the activity coefficient for aqueous species (Parkhurst and Appelo, 2013):

$$\log\gamma_i = \frac{-Az_i^2 I^{0.5}}{1 + \mathring{a}_i B I^{0.5}} + \dot{B} I \quad (2)$$

where γ_i denotes the activity coefficient for species *i*; *A*, *B* are Debye-Hückel parameters; z_i is the charge of species *i*; *I* is the ionic strength; $\mathring{a}_i$ is the ion size parameter of species *i*; $\dot{B}$ (*B*-dot) is the correction parameter. The $\mathring{a}_i$ values for the aqueous species in SUPPHREEQC are mostly taken from the EQ3/6 database (*data0.com.V8.R6*). However, if the species is not

Table 4
Special species association and dissolution reactions in *RXN* file.

Species or phases	Corresponding association or dissolution reactions
S_2^{2-}	$2\ HS^- + 0.5\ O_2 = S_2^{2-} + H_2O$
$S_2O_6^{2-}$	$2\ SO_3^{2-} + 0.5\ O_2 + 2\ H^+ = S_2O_6^{2-} + H_2O$
S_3^{2-}	$H^+ + O_2 + 3\ HS^- = S_3^{2-} + 2\ H_2O$
$S_3O_6^{2-}$	$3\ SO_3^{2-} + 4\ H^+ = S_3O_6^{2-} + 0.5\ O_2 + 2\ H_2O$
S_4^{2-}	$1.5\ O_2 + 2\ H^+ + 4\ HS^- = S_4^{2-} + 3\ H_2O$
$S_4O_6^{2-}$	$4\ SO_3^{2-} + 6\ H^+ = S_4O_6^{2-} + 1.5\ O_2 + 3\ H_2O$
S_5^{2-}	$2\ O_2 + 3\ H^+ + 5\ HS^- = S_5^{2-} + 4\ H_2O$
$S_5O_6^{2-}$	$5\ SO_3^{2-} + 8\ H^+ = S_5O_6^{2-} + 2.5\ O_2 + 4\ H_2O$
Pyrite	$FeS_2 + H_2O = Fe^{2+} + 2\ HS^- + 0.5\ O_2$
Copper	$Cu + 0.5\ O_2 + 2\ H^+ = Cu^{2+} + H_2O$
Diamond	$C + O_2 + H_2O = HCO_3^- + H^+$
Graphite	$C + O_2 + H_2O = HCO_3^- + H^+$
Sulfur	$S + H_2O = HS^- + H^+ + 0.5\ O_2$
Pyrrhotite(trov)	$Fe_{0.875}S + 0.75\ H^+ + 0.125\ H_2O = 0.875\ Fe^{2+} + HS^- + 0.0625\ O_2$
Iron	$Fe + 0.5\ O_2 + 2\ H^+ = Fe^{2+} + H_2O$
Nickel	$Ni + 0.5\ O_2 + 2\ H^+ = Ni^{2+} + H_2O$
S_2,g	$S_2 + 2\ H_2O = 2.5\ H^+ + 1.5\ HS^- + 0.5\ SO_4^{2-}$

listed in the EQ3/6 database, the following rules (compatible with the EQ3/6 database) are applied:

(1) $\mathring{a}_i$ for anions is 4.0
(2) $\mathring{a}_i$ for neutral species is 3.0
(3) $\mathring{a}_i$ for cations with $z \leq 4$ is $3.5 + 0.5z$, where z is the charge of the species
(4) $\mathring{a}_i$ for cations with $z > 4$ is 6.0.

3.2.2. Debye-Hückel and B-dot parameters

The Debye-Hückel parameters A and B are calculated following Helgeson (1969),

$$A = \frac{1.8246 \times 10^6 \rho_{H_2O}^{0.5}}{(\varepsilon_{H_2O} T)^{1.5}} \tag{3}$$

$$B = \frac{50.29 \times 10^8 \rho_{H_2O}^{0.5}}{(\varepsilon_{H_2O} T)^{0.5}} \tag{4}$$

where ρ_{H_2O} is the density of water (g/cm^3), and ε_{H_2O} is the dielectric constant of water.

In SUPPHREEQC, ρ_{H_2O} values are calculated with SUPCRTBL, which uses the equation of state for water in Johnson and Norton (1991). The calculations for ρ_{H_2O} in the critical region are different from those outside of the region (Johnson and Norton, 1991; Johnson et al., 1992). The dielectric constant is calculated via (Johnson and Norton, 1991; Johnson et al., 1992),

$$\varepsilon_{H_2O} = 1 + \frac{a_1}{T^*}\rho_{H_2O}^* + \left(\frac{a_2}{T^*} + a_3 + a_4 T^*\right)\rho_{H_2O}^{*2} + \left(\frac{a_5}{T^*} + a_6 T^* + a_7 T^{*2}\right)\rho_{H_2O}^{*3} + \left(\frac{a_8}{T^{*2}} + \frac{a_9}{T^*} + a_{10}\right)\rho_{H_2O}^{*4} \tag{5}$$

where T is expressed in K, $T^* = T/298.15$ K, ρ_{H_2O} in g/cm^3, $\rho_{H_2O}^* = \rho_{H_2O}/(1\ g/cm^3)$, and $a_{1\ldots10}$ are parameters.

In SUPPHREEQC, the extended-term parameter (*B-gamma*) for calculating the activity coefficient of NaCl in the activity coefficient equation of Oelkers and Helgeson (1990) is employed to represent $\dot{B}$. SUPPHREEQC estimates the value of $\dot{B}$ through linear interpolation for the values of *B-gamma* of NaCl (up to 1000 °C and 5000 bars) in Oelkers and Helgeson (1990)'s Table A-2. Note that this function misbehaves outside a narrow pressure-temperature region.

3.2.3. Activity coefficient of aqueous carbon dioxide

The activity coefficient of aqueous carbon dioxide CO_2^0(aq) in datasets following the *llnl.dat* framework is defined as a function of temperature and ionic strength (Drummond, 1981; Parkhurst and Appelo, 2013),

$$\log\gamma_{CO_2} = \left(C + FT + \frac{G}{T}\right)I - (E + HT)\left(\frac{I}{I+1}\right) \tag{6}$$

where T is the temperature (K); I is ionic strength; C, F, G, E, and H are species-dependent parameters. The stated application range of this equation is 0–400 °C, 0–500 bars, and 0–6 molal NaCl solutions. Outside this P-T-X range, the uncertainty is not known. Activity coefficients for H_2(aq) and O_2(aq) are calculated following Eq. (6).

3.3. Other files

Other files in SUPPHREEQC provide molar volume parameters, critical temperatures, critical pressures, and acentric factors. The molar volume parameters for most minerals and aqueous species are from *spronsbl.dat*. The critical temperatures, critical pressures, and acentric factors for most gases are from *phreeqc.dat*. However, the related data for CO(g) are from Voigt et al. (2018), and are unavailable for S_2(g). Another dataset contains a library of callable BASIC scripts for dissolution and precipitation rate equations and kinetic parameters for minerals, which takes from Zhang et al. (2019) but omits phases unavailable in SUPCRTBL.

4. Example datasets and applications

4.1. Dataset types and examples

SUPPHREEQC can generate three types of datasets: one adopts the framework of *phreeqc.dat* and the other two use the framework of *llnl.dat*. For the latter, their *T-P* ranges are either 0.01–100 °C at 1 bar and 100–300 °C at P_{SAT} or up to 1000 °C at constant pressure up to 5000 bars (the limits set by SUPCRTBL, Zimmer et al., 2016. See also Sverjensky et al., 2014).

Users of SUPPHREEQC can generate their own datasets at any *T-P* range within the limits stated above. To illustrate the three types of datasets, we produced an example of each: *diagenesis.dat*, *geothermal.dat*, and *bl.dat* (Table 1). The *diagenesis.dat* dataset follows the framework of *phreeqc.dat*. It includes (1) the temperature polynomials for the log K of aqueous species, minerals, and gases reactions covering the *T-P* range of 0.01–100 °C at 1 bar and 100–200 °C at P_{SAT}; (2) molar volume parameters for solids and aqueous species for estimating the effect of pressure on log K (suitable up to ~1000 bar and 200 °C; Appelo et al., 2014); and (3) Peng-Robinson equation parameters, critical temperatures, pressures, and acentric factors (Parkhurst and Appelo, 2013) for calculating gas fugacities. The Davies equation for aqueous species activity coefficients is used. Note that the list of species with molar volume parameters was greatly expanded in *diagenesis.dat* compared to that from *phreeqc.dat*. Also, in *diagenesis.dat*, almost all species have the parameters for the temperature polynomials of Eq. (1).

The *geothermal.dat* dataset follows the framework of *llnl.dat* and

includes (1) the temperature polynomial for the log K of aqueous species, minerals, and gases reactions covering the T-P range of either 0.01–100 °C at 1 bar and 100–300 °C at P_{SAT}; and (2) B-dot equation parameters for computing activity coefficients of aqueous species, which are also a function of temperature and pressure. *geothermal.dat* does not include the pressure modification parameters used by PHREEQC and therefore the issue of double accounting the pressure correction does not apply (see below). Again, parameters for the temperature polynomials of Eq. (1) are available for almost all species.

The *bl.dat* dataset follows the framework of *llnl.dat* and includes (1) the temperature polynomial for the log K of aqueous species, minerals, and gases reactions covering the T-P range up to 1000 °C and 5000 bars, respectively; and (2) B-dot equation parameters for computing activity coefficients of aqueous species, which are also a function of temperature and pressure. T-P grids inputted in SUPPHREEQC for generating *bl.dat* should be in the applicable range of SUPCRTBL (Fig. 2c), which is 0–1000 °C and 1–5000 bars. For users' convenience, we have generated *bl.dat* datasets at 0.5, 1, 2, and 5 kb (*bl-0.5 kb dat, bl-1kb.dat, bl-2kb.dat, bl-5kb.dat*). *bl.dat* datasets also do not include the pressure modification parameters used by the *phreeqc.dat* framework and therefore the issue of double accounting the pressure correction also does not apply (see below).

The coefficients for the analytical expression of log K (conforming to Eq. (1)) in *diagenesis.dat*, *geothermal.dat*, and *bl.dat* were calculated. Eleven (11) discrete log K values at the specified T-P points were automatically created using SUPCRTBL. For *diagenesis.dat*, the lower and upper-temperature points were 0.01 and 200 °C, respectively, and nine other temperature points evenly distributed between 0.01 and 200 °C along the liquid-vapor saturation curve. The log K values at these 11 temperature points were used to regress the six parameters in Eq. (1). For *geothermal.dat*, the lower and upper-temperature points were 0.01 and 300 °C, respectively, and nine other points evenly distributed between 0.01 and 300 °C along the liquid-vapor saturation curve. For *bl.dat*, the points are isobaric (up to 5 kb) with variable temperatures: minimum T,

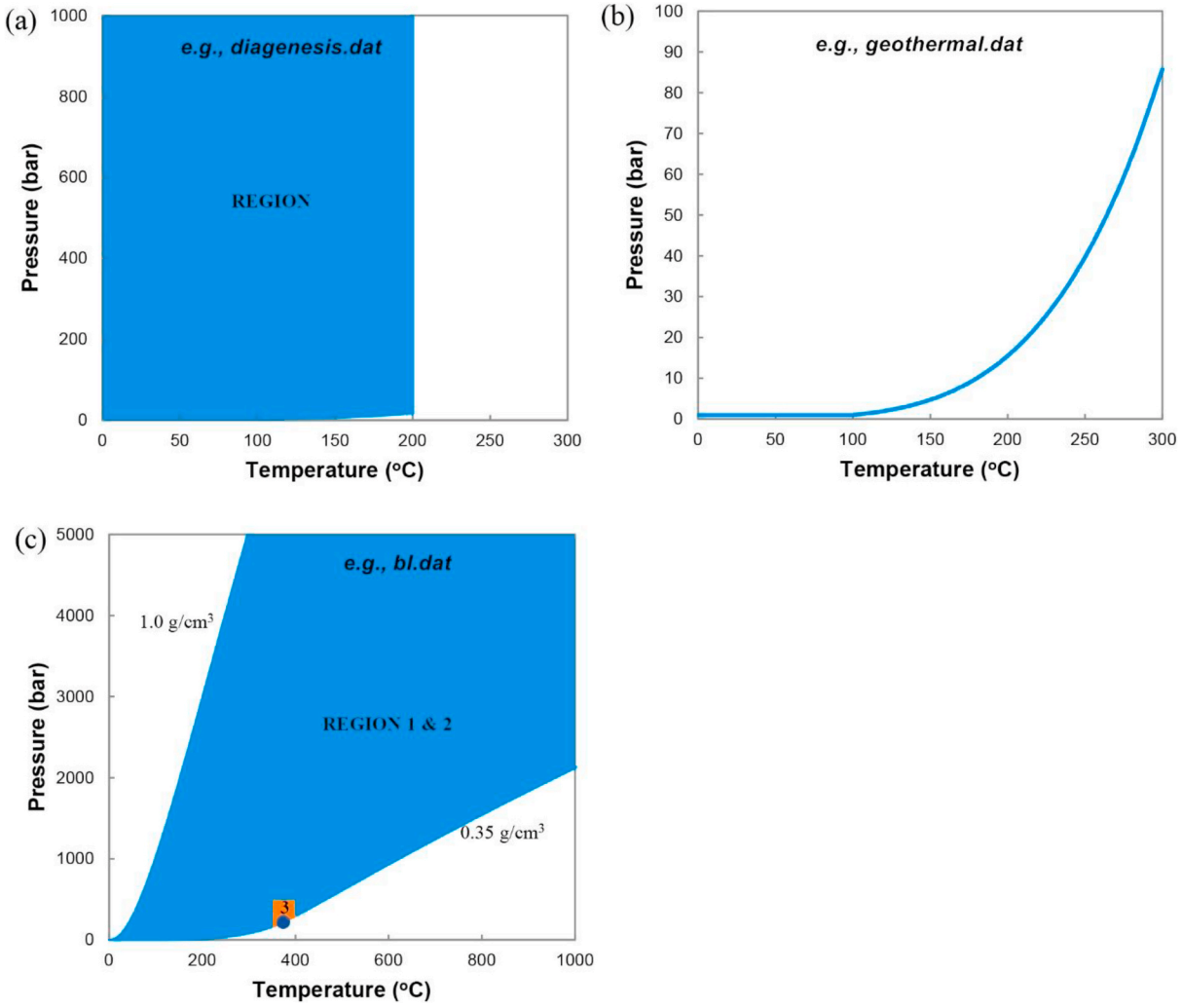

Fig. 2. The blue regions or lines represent applicable T-P ranges for the three types of datasets generated by SUPPHREEQC. REGION 1, 2, and 3 correspond to the definitions of Johnson et al. (1992). (a) The T-P range for the dataset following the *phreeqc.dat* framework (e.g., *diagenesis.dat*), which is bound by 0.01 and 200 °C isotherms, vaporization boundary, and 1000 bar isobar. (b) T-P range for one of the two datasets following the *llnl.dat* framework (i.e., *geothermal.dat*), which is 0.01–100 °C at 1 bar and 100–300 °C at P_{SAT}. (c) REGION 1 & 2 cover the T-P range for the other dataset following the *llnl.dat* framework (e.g., *bl.dat*), which is bound by 1.0 and 0.35 g/cm^3 isochores, 5000 bar isobar, 1000 °C isotherm, vaporization boundary, and analogous perimeter of REGION 3. REGION 3 shows the T-P range not applicable, which is bound by 350 and 400 °C isotherms, 500 bar isobar, 0.35 g/cm^3 isochore, and vaporization boundary. Note that REGION 1 and 2 cover the range of the apparent standard molal Gibbs free energy and equilibrium constant calculations of SUPCRTBL. However, the applicability to other thermodynamic properties of aqueous species (enthalpy, entropy, volume, and heat capacity) depends on whether the species is neutral or charged (Johnson et al., 1992). The dot labeled (3) in (c) indicates the critical point of water (373.917 °C and 220.46 bar). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

maximum T and nine other points evenly distributed between minimum T and maximum T. Subsequently, non-linear fitting of Eq. (1) with the discrete log K values is conducted in order to generate the coefficients (A_1 to A_6).

A comparison of these three SupPhreeqc generated datasets with the default Phreeqc datasets is shown in Table 1 and the applicable T-P ranges for the three datasets generated by SupPhreeqc are also shown in Fig. 2.

4.2. Applications of the SupPhreeqc datasets

Three examples below demonstrate the applications of the datasets generated by SupPhreeqc.

4.2.1. Saturation state of geothermal waters

Stefánsson and Arnórsson (2000) provided geochemical analyses of natural geothermal waters (listed in **Appendix 2** of this study). The samples were taken from rivers, springs, and drill hole discharges with a temperature range from 4 to 300 °C. Therefore, they are suitable examples to demonstrate the applications of *geothermal.dat*. We did not consider pressures higher than P_{SAT} in this case because these waters were predominantly sampled from the Earth's surface or near-surface.

Speciation and solubility calculations were conducted with three datasets (Table 1) for comparison: one by SupPhreeqc (*geothermal.dat*) and two that accompany the Phreeqc software distribution (*llnl.dat* and *phreeqc.dat*). In this case, polynomial analytical expressions of log K in *geothermal.dat* and *llnl.dat* cover a T-P range of 0.01–100 °C at 1 bar and 100–300 °C along P_{SAT}. Saturation indices (*SI*) for relevant minerals were calculated from temperatures and chemical compositions of the water samples. *SI* is defined as log(Q/K), where Q denotes the reaction quotient and K is the equilibrium constant (Zhu and Anderson, 2002).

Stefánsson and Arnórsson (2000) concluded that natural waters above 200 °C were near equilibrium with microcline and low-albite when they are in contact with these feldspars. Fig. 3a and b show that the modeling results using *geothermal.dat* and *llnl.dat* are consistent with this conclusion (note that the use of *llnl.dat* derived datasets, *core10.dat* (Neveu et al., 2017) and *carbfix.dat* (Voigt et al., 2018) produced results very similar to those from *llnl.dat*). However, *SI* values calculated from *phreeqc.dat* are about three units lower for albite and microcline.

llnl.dat uses the standard state Gibbs free energy of formation ($\Delta_f G^o$) for albite and microcline from Helgeson et al. (1978) while *geothermal.dat* uses those from Holland and Powell (2011). In *llnl.dat*, $\Delta_f G^o$ for SiO_2(aq) came from Supcrt92 while the $\Delta_f G^o$ in *geothermal.dat* is based on the quartz solubility data of Rimstidt (1997). As a result, there are differences between the *SI* values calculated with *llnl.dat* and *geothermal.dat* (Fig. 3c and d), yet the discrepancies are within ±0.5 log unit.

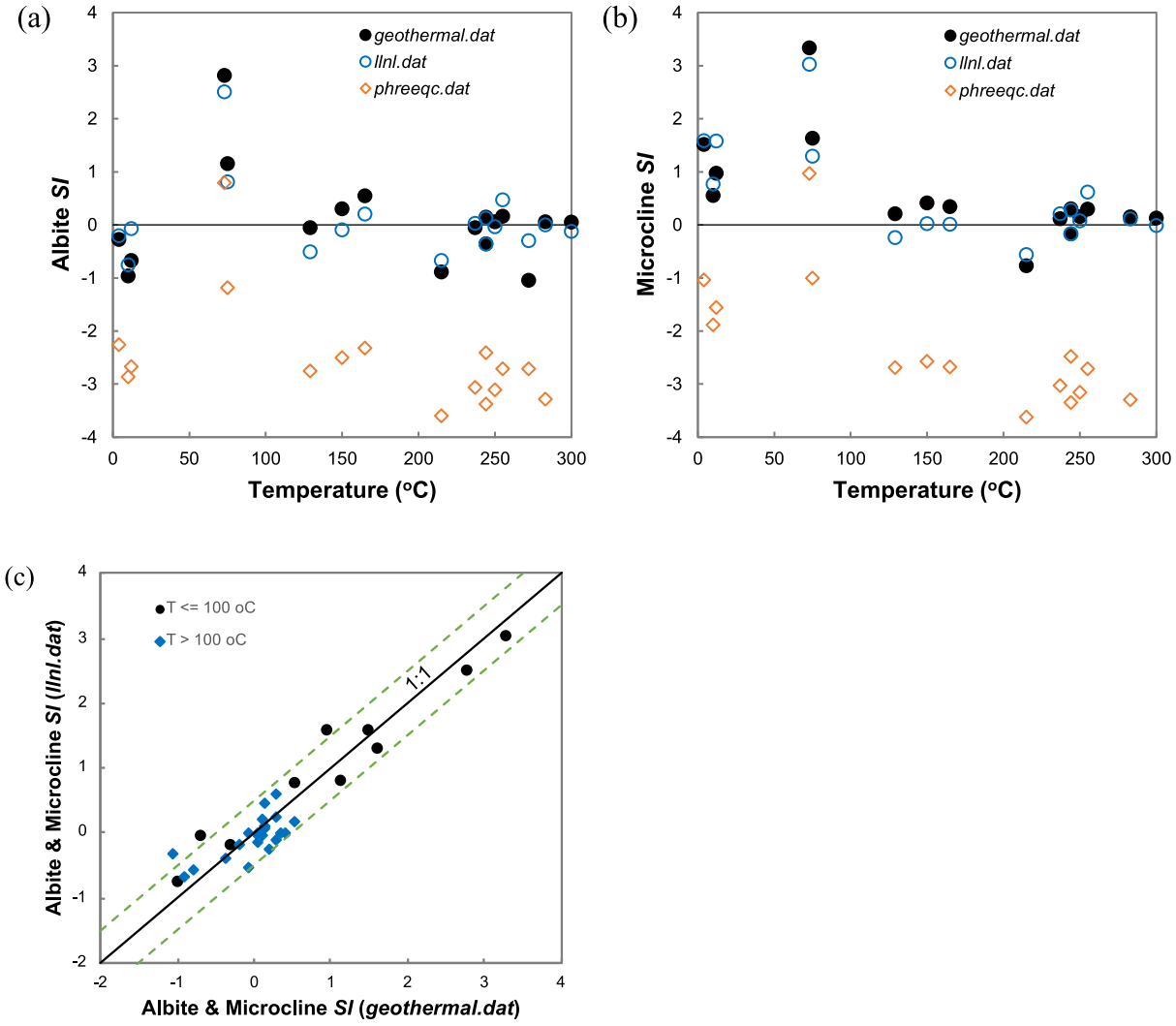

Fig. 3. Mineral saturation indices (*SI*) of albite and microcline for geothermal waters analyzed by Stefánsson and Arnórsson (2000) (their Table 1) calculated using Phreeqc (a)–(b): *SI* vs. temperature; (c): Cross plot albite and microcline *SI* calculated with *llnl.dat* and *geothermal.dat*. The dotted lines in (c) represent a deviation of ±0.5 log units.

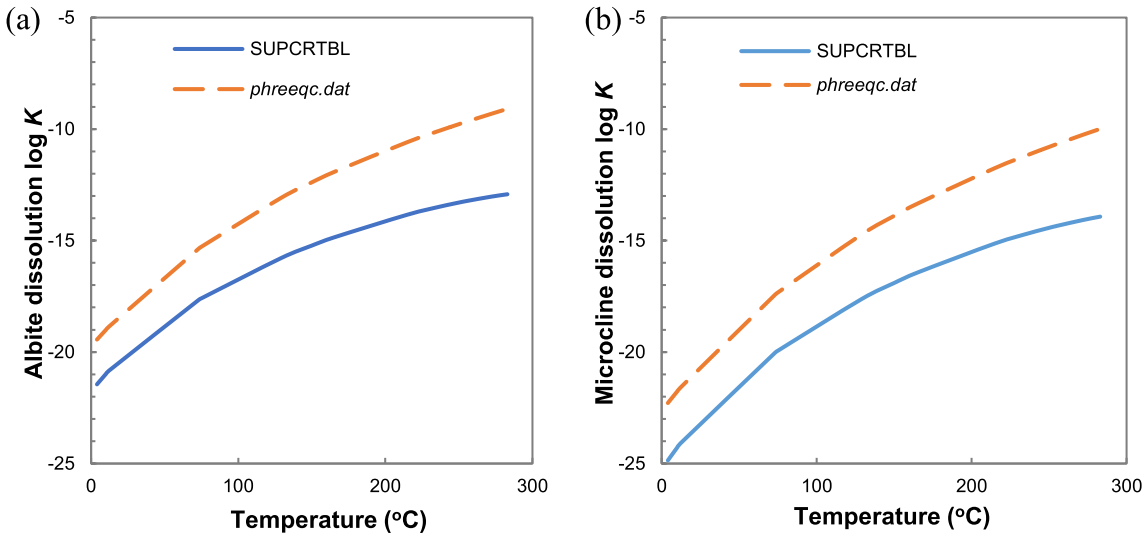

Fig. 4. Comparison of log *K*s for albite and microcline dissolution reactions calculated from using *phreeqc.dat* and *geothermal.dat*.

The differences in calculated *SI* between *phreeqc.dat* from *geothermal.dat* (Fig. 3a and b) are large, mainly due to the differences of log *K*s for microcline and albite dissolution reactions in *phreeqc.dat* and *geothermal.dat* (Fig. 4). The log *K*s at 25 °C in the two datasets differ about 2 and 2.5 units for albite and microcline, respectively, but the temperature dependence also differs. Because *phreeqc.dat* uses the van't Hoff equation to extrapolate log *K* at 25 °C to higher temperatures, the differences between log *K*s in *phreeqc.dat* and *geothermal.dat* increase with temperature. This example illustrates that *geothermal.dat* from SupPhreeqc is more reliable for modeling geothermal waters.

4.2.2. Example of quartz solubility at elevated T and P

Experimental data of quartz solubility are available for a wide range of *T* and *P* and they are recently evaluated critically (Sverjensky et al., 2014; Miron et al., 2016), which allows us to test the functionality of SupPhreeqc and the datasets that SupPhreeqc generated with Phreeqc calculations. Towards this goal, speciation and solubility calculations of quartz in pure water were carried out with Phreeqc using datasets generated in this study (*bl.dat, diagenesis.dat, geothermal.dat*) and those distributed along with the Phreeqc program (*llnl.dat* and *phreeqc.dat*). The speciation-solubility simulation results using these datasets are shown in Fig. 5 and compared to experimental data. In order to better compare with experiments at elevated temperatures and pressures, we replaced the $\Delta_f G^o$ at 298.15 K and 1 bar for aqueous species SiO_2^o (aq) in Supcrtbl (Zimmer et al., 2016) with that of Miron et al. (2016). Miron et al. (2016) adopted the entropy S^o, the HKF parameters a1-a4, c1-c2, ω from those of Shock et al. (1989). Note that the $\Delta_f G^o$ at 298.15 K values in recent literature are all close to each other and all relied on the experimental data from Rimstidt (1997). This update in Supcrtbl was mainly for HKF parameters that are suitable for high *T* and *P* calculations. A thorough analysis of quartz solubility and aqueous Si species is also provided in Sverjensky et al. (2014). Here, our objective is to verify computer program functions of Supcrtbl – SupPhreeqc – Phreeqc.

For the temperature and pressure range of 25–300 °C and 1 bar to P_{sat}, the calculations from *geothermal.dat* compared well with the experimental data that Miron et al. (2016) used for optimizing their $\Delta_f G^o$ at 298.15 K and 1 bar for aqueous species SiO_2^o (aq) (Fig. 5a). These experimental data included those at from Rimstidt (1997), Crerar and Anderson, 1971, Siever (1962), Morey et al. (1962), and Lier et al. (1960). Note that our Phreeqc calculations included speciation – solubility simulations. The agreement between quartz solubility calculated using *geothermal.dat* and experimental data verifies the computer program functions in SupPhreeqc and Phreeqc. However, while quartz solubility calculated using *geothermal.dat* matched the experimental data well, the calculated values from *llnl.dat* did not match (Fig. 5a) because of the known issue of the $\Delta_f G^o$ at 298.15 (Miron et al., 2016; Zimmer et al., 2016).

For the 1 kb experimental data, Rimstidt (1997) and Miron et al. (2016) convincingly argued that the experimental values for quartz solubility at temperatures below 130 °C from Morey et al. (1962) are incorrect because equilibrium was not attained. Therefore, only comparisons of their 130–300 °C data from Morey et al. (1962) are meaningful. The experimental data from Walther and Orville (1983) were included in the optimization of Miron et al. (2016). The agreement of calculated quartz solubility using *bl-1kb.dat* just verified the correct functioning of computer programs. Quartz solubility calculated with *bl.dat* at 1 kb also matched well with experimental data at 1 kb not included in Miron et al. (2016), namely Hemley et al. (1980) and Weill and Fyfe (1964). In Fig. 5b, the line calculated using *phreeqc.dat* also matched well with the experimental data, which illustrates the effectiveness of the pressure correction function implemented in Phreeqc by Appelo et al. (2014). In contrast, quartz solubility calculated with *llnl.dat*, for which the pressure is P_{SAT}, did not match well with experimental data at 1 kb.

We also generated *bl.dat* datasets from SupPhreeqc at 2 and 5 kb, respectively. Quartz solubility calculated from *bl.dat* at 2 kb matched well with Weill and Fyfe (1964), Hemley et al. (1980), and Walther and Orville (1983); *bl.dat* at 5 kb matched well with Manning (1994). These calculations show that SupPhreeqc functioned well in generating datasets up to 1000 °C and 5 kb for Phreeqc. It should be noted that the pressure effect on log *K* values is already taken account of in the Supcrtbl calculations.

SupPhreeqc's ability to generate datasets at temperature up to 1000 °C and pressure up to 5000 bars and the ability of Phreeqc to calculate solubility and speciation up to these *T* and *P* do not mean that the programs will produce accurate solubility and speciation calculations. The above example of quartz solubility is a simple chemical system. Users are responsible for whether these calculations can match with experimental data and whether they can trust the calculations.

4.2.3. An application to clastic diagenesis—Albitization

During burial diagenesis, albitization of detrital plagioclase and K-feldspar in sandstones has been observed in sedimentary basins worldwide, and has been well documented since the 1960s (e.g., Aagaard et al., 1990; Boles, 1982; Dickinson et al., 1969; Gold, 1987; Land and Milliken, 1981; Middleton, 1972; Morad et al., 1990). Albitization of plagioclase is an important reaction during diagenesis of clastic sediments because it regulates the Ca and Si concentrations in the pore water

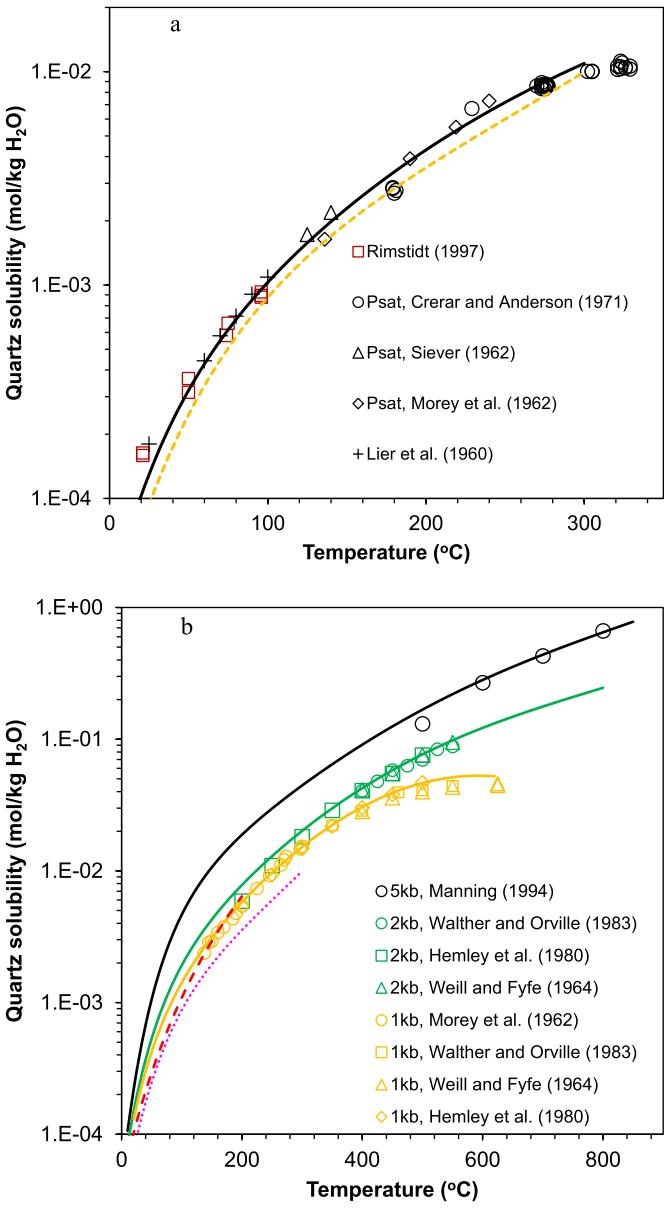

Fig. 5. Quartz solubility (mol/kg H_2O) in pure water. Lines represent quartz solubility from speciation modeling with PHREEQC using assuming the monomer of SiO_2^o(aq) is the dominant species. Symbols denote the experimental data. (a) 25–300 °C at 1 bar or $P_{sat.}$ The black and the orange dashed lines were calculated with *geothermal.dat* and *llnl.dat*, respectively; (b) 1, 2, and 5 kb and high temperatures. Solid Lines are calculated with *bl.dat* at fixed pressures of 1, 2, and 5 kb, respectively. The red dashed line and purple dotted line are calculations with *phreeqc.dat* at 1 kb and llnl.dat at P_{sat}, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

and thus also the cementation of calcite and quartz (Baccar et al., 1993). Albitization of K-feldspar is a critical control of the Na^+ and K^+ concentrations in the formation waters because this reaction consumes Na^+ and releases K^+ (Eq. (7)). Its application, Na/K geothermometer, is a widely used tool for estimating reservoir temperatures of hydrothermal systems (Can, 2002; Fournier and Potter, 1979; Fournier and Truesdell, 1973; Nicholson, 2012). Albitization is potentially aided by NaCl-dominated formation waters, which have a high activity of Na^+.

The reaction of albitization of K-feldspar can be expressed as:

$$Na^+ + KAlSi_3O_8\ (\text{K-feldspar}) = NaAlSi_3O_8\ (\text{albite}) + K^+ \quad (7)$$

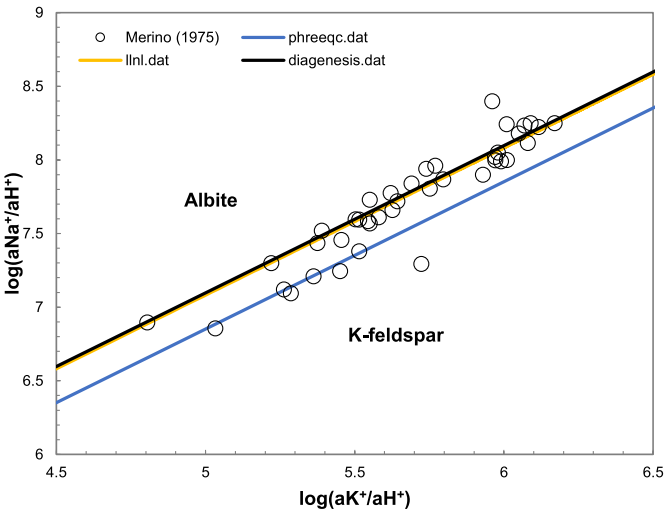

Fig. 6. Stability diagram of equilibrium between K-feldspar, albite, and an aqueous solution calculated at 100 °C and 150 bars using different PHREEQC datasets. Field data (brines from Kettleman North Dome) are from Merino (1975). The pressure is corrected to 150 bars in *diagenesis.dat,* and *phreeqc.dat*, while P_{SAT} is used in *llnl.dat*.

If anorthite is used as a proxy for plagioclase, albitization of plagioclase can be expressed as.

Al and Si are conserved in this reaction. Fig. 6 shows the calculated log(aNa^+/aH^+) vs. log (aK^+/aH^+) at 100 °C and 150 bars using different PHREEQC datasets. The calculated activity ratios are also compared with field data from Merino (1975) that define the boundary of stability fields of albite and K-feldspar. Calculated lines using *diagenesis.dat* and *llnl.dat* fit the field data quite well. The line calculated with *phreeqc.dat* is ~0.3 log units lower than those calculated with *diagenesis.dat* and *llnl.dat*. The effects of pressure on the calculated activity ratios are small (*cf.* the orange line vs. the black line).

$$2\ Na^+ + CaAl_2Si_2O_8\ (\text{anorthite}) + 4\ SiO_2^o = 2\ NaAlSi_3O_8\ (\text{albite}) + Ca^{2+} \quad (8)$$

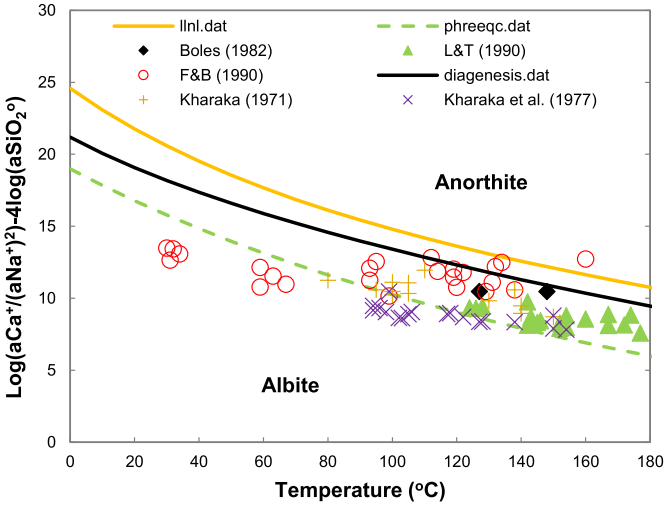

Fig. 7. Calculated values of log($aCa^+/(aNa^+)^2$)-4log(a $SiO_{2(aq)}^o$) as a function of temperature from formation water chemistry data from the literature projected on a stability diagram of equilibrium between anorthite, albite, and an aqueous solution. The boundary of the anorthite to albite stability fields was calculated with different thermodynamic datasets for PHREEQC. Hydrostatic pressure was used in calculations using *diagenesis.dat* and *phreeqc.dat*, while P_{SAT} or 1 bar at $T < 100$ °C was used in *llnl.dat*. L&T (1990): Lundegard and Trevena (1990); F&B (1990): Fisher and Boles (1990).

For the conversion of 1 mol anorthite to 2 mol albite, 4 mol of aqueous SiO_2 are required. This reaction will reduce the activity of $SiO^0_{2(aq)}$ in the formation water and thus reduce the potential of quartz cementation. However, the addition of Ca^+ may promote calcite precipitation, provided HCO_3^- is available in the system.

Fig. 7 depicts the relationship of $\log(aCa^+/(aNa^+)^2)$-$4\log(aSiO^0_{2(aq)})$ as a function of temperature. The $\log(aCa^+/(aNa^+)^2)$-$4\log(aSiO^0_{2(aq)})$ values decrease as a function of temperature, which makes the stability field larger for albite at lower temperatures and anorthite at higher temperatures. The boundaries of anorthite to albite stability fields were calculated with different thermodynamic datasets for comparison.

Because albitization is a common alteration reaction occurring in the subsurface, geopressure should be considered when conducting speciation-solubility calculations for the formation waters. In this study, the pressure is assumed to be the same as the hydrostatic pressure. The normal hydrostatic gradient for seawater is about 10.45 kPa/m (Pashin, 2008). The *diagenesis.dat* dataset generated by SupPhreeqc can be applied for calculations at pressures up to ~1000 bars. Hydrostatic pressure is used in *diagenesis.dat* and *phreeqc.dat*, while P_{SAT} or 1 bar at $T < 100$ °C is used in *llnl.dat*.

The calculated differences between the sizes of anorthite and albite stability fields are large (Fig. 7). This is mainly due to the different thermodynamic data of $SiO^0_{2(aq)}$ in different datasets. The source of thermodynamic data of $SiO^0_{2(aq)}$ in *phreeqc.dat* is unknown. The data in *llnl.dat* is from Supcrt92, and the data in *diagenesis.dat* is from Rimstidt (1997).

Plagioclase is missing in the field site of Lundegard and Trevena (1990) but co-exists with albite in the sandstones studied by Boles (1982) and Fisher and Boles (1990). The calculated lines should pass the data points from Boles (1982) and Fisher and Boles (1990). Most data points from Lundegard and Trevena (1990) are expected to be within the albite stability field. The offset likely resulted from the use of anorthite in calculations rather than a proper component of the plagioclase solid solution. However, for *phreeqc.dat*, most data from Lundegard and Trevena (1990) lie in the anorthite stability field, which is problematic. In *phreeqc.dat*, modeling at temperatures higher than 25 °C is done through the van't Hoff approximation, provided the enthalpy of reaction ($\Delta_r H^o$) is available. This method may have large uncertainty if the temperature range is large (Anderson and Crerar, 1993).

5. Discussion

5.1. Limitations

Note that other geochemical modeling software packages also can perform calculations at a wide range of temperatures and pressures, provided there is an appropriate accompanying dataset. These software packages include, but are not limited to, Eq3/6 (Wolery, 1992), Chess (Van der Lee and De Windt, 2002), The Geochemist's Workbench (Bethke and Yeakel, 2014), Chim-xpt (Reed, 1998), Gems-psi (Kulik et al., 2004), and Reaktoro (Leal et al., 2017). Several allow users to use the Peng-Robinson model for gases in combination with the extended Debye-Hückel activity model.

However, a more fundamental limitation is the lack of internal consistency in the thermodynamic data behind the calculated log *K* values. While in *spronsbl.dat* the thermodynamic properties for minerals and aqueous species are respectively internally consistent, they may not be consistent with each other. Solubilities calculated from combining the two may not be consistent with the experimental data (Miron et al., 2016; Tutolo et al., 2014).

5.2. Errors due to pressure over-correcting and constant mineral molar volume

Because the Phreeqc program uses 1 bar as the reference pressure to correct the pressure effect on log *K*s with molar volumes of solids and aqueous species (Appelo et al., 2014), *diagenesis.dat* accounted for the pressure effects twice: once in Supcrtbl and again in Phreeqc for the temperature range of 100–200 °C, where log *K*s are calculated at P_{SAT} rather than at 1 bar. This is shown in the equation below from Appelo et al. (2014),

$$\log K_{P,T} = \log K_{P=P_{ref},T} - \Delta V_r \frac{P-1}{2.303RT} \quad (9)$$

where log $K_{P,T}$ denotes the equilibrium constant at pressure *P* and temperature *T*; $\log K_{P=P_{ref},T}$ is the equilibrium constant at temperature *T* and reference pressure P_{ref} ($P_{ref} = 1$ bar); ΔV_r is the standard volume of the reaction; *R* is the universal gas constant. When log *K*s are calculated at P_{SAT} at $T > 100$ °C, double accounting of the pressure effect occurs. Other datasets that follow the same *phreeqc.dat* approach (*core10.dat* and *carbfix.dat*) have the same issue of double accounting.

This study has investigated the effects of pressure over-correction on log *K*. Note that Appelo et al. (2014) assumed that mineral molar volume is invariable with *T* and *P*. We followed this assumption to calculate the errors in log *K* of dissolution reactions of 30 minerals due to pressure over-correction, based on **Eq. (9)**. The maximum error is only ~0.045 at 200 °C and 1 kb. Therefore, this type of error can be ignored at $T < 200$ °C. Appelo et al. (2014) advised their approach (the *phreeqc.dat* framework) should be used below 200 °C, which should be followed.

However, in reality, mineral molar volume changes with *T* and *P*. We compared the log *K* of dissolution reactions of 30 minerals from *diagenesis.dat* (mineral molar volume invariable with *T* and *P*) with those from Supcrtbl (mineral molar volume changes with *T* and *P*). The maximum difference is ~0.82 in log *K* at 200 °C and 1 kb. As a result, the errors from the assumption of invariable mineral molar volumes are more significant than those from pressure over-correcting.

6. Conclusions

This paper introduces SupPhreeqc, a computer program written in C++11, which links Phreeqc and Supcrtbl to facilitate geochemical modeling of mineral-water-gas interactions at elevated *T-P*. SupPhreeqc generates thermodynamic datasets suitable for use with Phreeqc by translating log *K* values from Supcrtbl (Zimmer et al., 2016) to Phreeqc-readable datasets. This link allows utilization of the high-quality thermodynamic data for minerals (Holland and Powell, 2011; Helgeson et al., 1978) and a large number of inorganic and organic aqueous species with the Helgeson-Kirkham-Flowers equation of state parameters (e.g., Sverjensky et al. (1997) and other HKF related publications) for speciation and solubility modeling at elevated *T-P*.

As examples, we generated three Phreeqc datasets (*diagenesis.dat*, *geothermal.dat*, and *bl.dat*) to represent the three types of datasets that SupPhreeqc can generate. *diagenesis.dat* follows the framework of *phreeqc.dat* and includes the *T-P* range of 0.01–100 °C at 1 bar and 100–200 °C at P_{SAT}, as well as the pressure correction and Peng-Robinson equation parameters for gas fugacity from Appelo et al. (2014). In addition to different sources of thermodynamic data, the difference between the *phreeqc.dat* and the *diagenesis.dat* datasets (and other datasets generated from SupPhreeqc) is their reliability for modeling at elevated temperatures. Only a small set of reactions included in *phreeqc.dat* have the necessary parameters for calculations up to 200 °C and 1 kb. In contrast, in *diagenesis.dat* and other SupPhreeqc-generated datasets that follow the *phreeqc.dat* format, nearly all species have temperature dependence via the temperature polynomial approach.

The second example is *geothermal.dat,* which follows the framework of *llnl.dat*. Its *T-P* range is 0.01–100 °C at 1 bar and 100–300 °C at P_{SAT}. The default *llnl.dat* and *geothermal.dat* datasets differ in their thermodynamic database sources (for minerals Helgeson et al., 1978 vs Holland and Powell, 2011; critical aqueous Si and Al species, and other species outlined in Zimmer et al. (2016)). Again, in *geothermal.dat* and other

SupPhreeqc-generated datasets that follow the *llnl.dat* format and structure, nearly all species have temperature dependence via the temperature polynomial approach. However, the default *llnl.dat* dataset has many species that do not have log *K* values other than those at 25 °C and 1 bar.

The third example, *bl.dat*, follows the framework of *llnl.dat* and generates datasets that cover a range of temperatures (up to 1000 °C) at a constant pressure (up to 5000 bars). These *T* and *P* limits correspond to the *T-P* limits of Supcrtbl. Additionally, this study has supplied the parameters for calculating activity coefficients using the B-dot equation up to 1000 °C and 5000 bars for speciation calculations. As examples, we have generated in this study of *bl.dat* series datasets at 0.5 kb (65–300 °C), 1 kb (110–600 °C), 2 kb (180–940 °C), and 5 kb (300–1000 °C).

This study has also developed an online version of Phreeqc, which is available at https://models.earth.indiana.edu. The three datasets above appear as options in the pulldown menu for dataset selection. Users may also use their own custom dataset generated from SupPhreeqc with Phreeqc.

Computer code availability

The SupPhreeqc C++ source code and executables are available at http://hdl.handle.net/2022/23355 (https://doi.org/10.5967/yjxn-5548) and https://github.com/HydrogeoIU/SupPHREEQC. The executable programs run with Windows and Linux systems. An online version of SupPhreeqc, along with an online version of Phreeqc, is available at https://models.earth.indiana.edu. Contact Chen Zhu at chenzhu@indiana.edu or by phone: 001 812 856 1884 to report any issues. The programs have become available in the year 2020.

Author contribution

Guanru Zhang coded the SupPhreeqc program. Yilun Zhang helped with thermodynamic calculations. Kevin Tu coded and maintains the online versions of SupPhreeqc and Phreeqc. Peng Lu Writing - original draft, designed and supervised this project and wrote part of the manuscript. and Chen Zhu Writing - original draft, designed and supervised this project and wrote part of the manuscript. Peng Lu developed the application examples.

Declaration 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.

Acknowledgments

This work was partially supported by the U.S. National Science Foundation grant EAR-1926734 to CZ, the Faculty Research Support Program through the office of the Vice Provost for Research at Indiana University, the Haydn Murray chair endowment, the American Chemical Society Petroleum Research Fund (PRF#57727-ND2 to CZ), and the Natural Science Foundation of China (#41702125 to GRZ "Effects of Hydrocarbon Acidic Byproducts on Local Reservoir Porosities and Secondary Migration of Natural Gas: Insights from Reactive Transport Modeling"). Although the work was partly sponsored by an agency of the United States Government, the views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. We would like to thank David Parkhurst and Tony Appelo for their PHREEQC program and discussions. We thank all reviewers and editors for their time. Their comments have improved the quality and clarity of this paper. Editing of the manuscript by Anne Hereford and proofreading by Lei Gong is greatly appreciated.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cageo.2020.104560.

References

Aagaard, P., Egeberg, P.K., Saigal, G.C., Morad, S., Bjorlykke, K., 1990. Diagenetic albitization of detrital K-feldspars in Jurassic, Lower Cretaceous and Tertiary clastic reservoir rocks from offshore Norway; II, Formation water chemistry and kinetic considerations. J. Sediment. Res. 60, 575–581.

Anderson, G.M., Crerar, D.A., 1993. Thermodynamics in Geochemistry: the Equilibrium Model. Oxford University Press, New York.

Appelo, C.A.J., Parkhurst, D.L., Post, V.E.A., 2014. Equations for calculating hydrogeochemical reactions of minerals and gases such as CO_2 at high pressures and temperatures. Geochem. Cosmochim. Acta 125, 49–67.

Baccar, M.B., Fritz, B., Made, B., 1993. Diagenetic albitization of K-feldspar and plagioclase in sandstone reservoirs; thermodynamic and kinetic modeling. J. Sediment. Res. 63, 1100–1109.

Ball, J.W., Nordstrom, D.K., 1991. User's manual for WATEQ4F, with revised thermodynamic data base and text cases for calculating speciation of major, trace, and redox elements in natural waters. U.S. Geological Survey 1–198. Menlo Park.

Bethke, C.M., Yeakel, S., 2014. The Geochemist's Workbench User's Guides. version 10.0. Aqueous Solutions LLC, Champaign.

Boles, J.R., 1982. Active albitization of plagioclase, Gulf coast tertiary. Am. J. Sci. 282, 165–180.

Can, I., 2002. A new improved Na/K geothermometer by artificial neural networks. Geothermics 31, 751–760.

Crerar, D.A., Anderson, G.M., 1971. Solubility and solvation reactions of quartz in dilute hydrothermal solutions. Chem. Geol. 8, 107–115.

Dickinson, W.R., Ojakangas, R.W., Stewart, R.J., 1969. Burial metamorphism of the late mesozoic Great valley sequence, cache creek, California. Geol. Soc. Am. Bull. 80, 519–526.

Drummond, S.E., 1981. Boiling and Mixing of Hydrothermal Fluids-Chemical Effects on Mineral Precipitation. Pennsylvania State University, p. 760.

Fisher, J.B., Boles, J.R., 1990. Water—rock interaction in Tertiary sandstones, San Joaquin basin, California, U.S.A.: Diagenetic controls on water composition. Chem. Geol. 82, 83–101.

Fournier, R.O., Potter, R.W., 1979. Magnesium correction to the Na K Ca chemical geothermometer. Geochem. Cosmochim. Acta 43, 1543–1550.

Fournier, R.O., Truesdell, A.H., 1973. An empirical Na K Ca geothermometer for natural waters. Geochem. Cosmochim. Acta 37, 1255–1275.

Gold, P.B., 1987. Textures and geochemistry of authigenic albite from Miocene sandstones, Louisiana Gulf Coast. J. Sediment. Res. 57, 353–362.

Helgeson, H.C., 1969. Thermodynamics of hydrothermal systems at elevated temperatures and pressures. Am. J. Sci. 267, 729-804.

Helgeson, H.C., Delany, J.M., Nesbitt, H.W., Bird, D.K., 1978. Summary and critique of the thermodynamic properties of rock-forming minerals. Am. J. Sci. 278, 1–229.

Hemley, J.J., Montoya, J.W., Marinenko, J.W., Luce, R.W., 1980. Equilibria in the system Al_2O_3-SiO_2-H_2O and some general implications for alteration/mineralization processes. Econ. Geol. 75, 210–228.

Holland, T.J.B., Powell, R., 2011. An improved and extended internally consistent thermodynamic dataset for phases of petrological interest, involving a new equation of state for solids. J. Metamorph. Geol. 29, 333–383.

Johnson, J.W., Norton, D., 1991. Critical phenomena in hydrothermal systems; state, thermodynamic, electrostatic, and transport properties of H_2O in the critical region. Am. J. Sci. 291, 541-648.

Johnson, J.W., Oelkers, E.H., Helgeson, H.C., 1992. SUPCRT92: A software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species, and reactions from 1 to 5000 bar and 0 to 1000°C. Comput. Geosci. 18, 899–947.

Kulik, D., Berner, U., Curti, E., 2004. Modelling Chemical Equilibrium Partitioning with the GEMS-PSI Code. Paul Scherrer Institut.

Land, L.S., Milliken, K.L., 1981. Feldspar diagenesis in the Frio Formation, Brazoria County. Texas Gulf Coast. Geology 9, 314–318.

Langmuir, D., Mahoney, J., Rowson, J., 2006. Solubility products of amorphous ferric arsenate and crystalline scorodite ($FeAsO_4{\cdot}2H_2O$) and their application to arsenic behavior in buried mine tailings. Geochem. Cosmochim. Acta 70, 2942–2956.

Leal, A.M.M., Kulik, D.A., Smith, W.R., Saar, M.O., 2017. An overview of computational methods for chemical equilibrium and kinetic calculations for geochemical and reactive transport modeling. Pure Appl. Chem. 89, 597–643.

Lier, J.A.V., Bruyn, P.L.D., Overbeek, J.T.G., 1960. The solubility of quartz. J. Phys. Chem. 64, 1675–1682.

Lundegard, P.D., Trevena, A.S., 1990. Sandstone diagenesis in the Pattani Basin (Gulf of Thailand): history of water-rock interaction and comparison with the Gulf of Mexico. Appl. Geochem. 5, 669–685.

Manning, C.E., 1994. The solubility of quartz in H_2O in the lower crust and upper mantle. Geochem. Cosmochim. Acta 58, 4831–4839.

Marini, L., Accornero, M., 2010. Erratum to: Prediction of the thermodynamic properties of metal–arsenate and metal–arsenite aqueous complexes to high temperatures and pressures and some geological consequences. Environ. Earth Sci. 59, 1601–1606.

Merino, E., 1975. Diagenesis in tertiary sandstones from Kettleman North Dome. California—II. Interstitial solutions: Distribution of aqueous species at 100°C and

chemical relation to the diagenetic mineralogy. Geochem. Cosmochim. Acta 39, 1629–1645.
Middleton, G.V., 1972. Albite of secondary origin in Charny sandstones, Quebec. J. Sediment. Res. 42, 341–349.
Miron, G.D., Wagner, T., Kulik, D.A., Heinrich, C.A., 2016. Internally consistent thermodynamic data for aqueous species in the system Na–K–Al–Si–O–H–Cl. Geochem. Cosmochim. Acta 187, 41–78.
Morad, S., Bergan, M., Knarud, R., Nystuen, J.P., 1990. Albitization of detrital plagioclase in Triassic reservoir sandstones from the Snorre Field, Norwegian North Sea. J. Sediment. Res. 60, 411–425.
Morey, G.W., Fournier, R.O., Rowe, J.J., 1962. The solubility of quartz in water in the temperature interval from 25 to 300 °C. Geochem. Cosmochim. Acta 26, 1029–1040.
Neveu, M., Desch, S.J., Castillo-Rogez, J.C., 2017. Aqueous geochemistry in icy world interiors: Equilibrium fluid, rock, and gas compositions, and fate of antifreezes and radionuclides. Geochem. Cosmochim. Acta 212, 324–371.
Nicholson, K., 2012. Geothermal Fluids: Chemistry and Exploration Techniques. Springer-Verlag Berlin.
Nordstrom, D.K., Archer, D.G., 2003. Arsenic thermodynamic data and environmental geochemistry. In: Welch, A.H., Stollenwerk, K.G. (Eds.), Arsenic in Ground Water: Geochemistry and Occurrence. Springer, Boston, pp. 1–25.
Oelkers, E.H., Helgeson, H.C., 1990. Triple-ion anions and polynuclear complexing in supercritical electrolyte solutions. Geochem. Cosmochim. Acta 54, 727–738.
Parkhurst, D.L., Appelo, C.A.J., 1999. User's guide to PHREEQC (Version 2): A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. US Geo. Survey 312.
Parkhurst, D.L., Appelo, C.A.J., 2013. Description of input and examples for PHREEQC Version 3–A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. U.S. Geological Survey 497.
Pashin, J.C., 2008. Coal as a petroleum source rock and reservoir rock. In: Suárez-Ruiz, I., Crelling, J.C. (Eds.), Applied Coal Petrology. Elsevier, Burlington, pp. 227–262.
Reed, M., 1998. Calculation of simultaneous chemical equilibria in aqueous-mineral-gas systems and its application to modeling hydrothermal processes. Tech. Hydro Ore Deposits Geo. 10, 109–124.
Rimstidt, J.D., 1997. Quartz solubility at low temperatures. Geochem. Cosmochim. Acta 61, 2553–2558.
Siever, R., 1962. Silica solubility, 0°-200° C., and the diagenesis of siliceous sediments. J. Geol. 70, 127–150.
Stefánsson, A., 2001. Dissolution of primary minerals of basalt in natural waters: I. Calculation of mineral solubilities from 0°C to 350°C. Chem. Geol. 172, 225–250.
Stefánsson, A., Arnórsson, S., 2000. Feldspar saturation state in natural waters. Geochem. Cosmochim. Acta 64, 2567–2584.
Sverjensky, D.A., Harrison, B., Azzolini, D., 2014. Water in the deep Earth: The dielectric constant and the solubilities of quartz and corundum to 60kb and 1200°C. Geochem. Cosmochim. Acta 129, 125–145.
Sverjensky, D.A., Shock, E.L., Helgeson, H.C., 1997. Prediction of the thermodynamic properties of aqueous metal complexes to 1000°C and 5 kb. Geochem. Cosmochim. Acta 61, 1359–1412.
Tagirov, B., Schott, J., 2001. Aluminum speciation in crustal fluids revisited. Geochem. Cosmochim. Acta 65, 3965–3992.
Tutolo, B.M., Kong, X.-Z., Seyfried, Saar, M.O., 2014. Internal consistency in aqueous geochemical data revisited: Applications to the aluminum system. Geochimica et Cosmochimica Acta 133, 216–234.
Van der Lee, J., De Windt, L., 2002. CHESS Tutorial and Cookbook, User's Manuals. Ecole des Mines de Paris, Fontainebleau.
Voigt, M., Marieni, C., Clark, D.E., Gíslason, S.R., Oelkers, E.H., 2018. Evaluation and refinement of thermodynamic databases for mineral carbonation. Energy Procedia 146, 81–91.
Walther, J.V., Orville, P.M., 1983. The extraction-quench technique for determination of the thermodynamic properties of solute complexes; application to quartz solubility in fluid mixtures. Am. Mineral. 68, 731–741.
Weill, D.F., Fyfe, W.S., 1964. The solubility of quartz in H_2O in the range 1000–4000 bars and 400–550°C. Geochem. Cosmochim. Acta 28, 1243–1255.
Wolery, T.J., 1992. EQ3NR, a Computer Program for Geochemical Aqueous Speciation-Solubility Calculations: Theoretical Manual, Users Guide, and Related Documentation. Lawrence Livermore National Lab, pp. 1–246, Version 7.0.
Zhang, Y., Hu, B., Teng, Y., Tu, K., Zhu, C., 2019. A library of BASIC scripts of reaction rates for geochemical modeling using PHREEQC. Comput. Geosci. 133, 104316.
Zhu, C., Anderson, G., 2002. Environmental Applications of Geochemical Modeling. Cambridge University Press.
Zhu, Y., Zhang, X., Xie, Q., Chen, Y., Wang, D., Liang, Y., Lu, J., 2005. Solubility and stability of barium arsenate and barium hydrogen arsenate at 25°C. J. Hazard Mater. 120, 37–44.
Zimmer, K., Zhang, Y., Lu, P., Chen, Y., Zhang, G., Dalkilic, M., Zhu, C., 2016. SUPCRTBL: A revised and extended thermodynamic dataset and software package of SUPCRT92. Comput. Geosci. 90, 97–111.