Journal of Hydrology 550 (2017) 144–156

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Identifying geochemical processes using End Member Mixing Analysis to decouple chemical components for mixing ratio calculations

Flavia Pelizardi [a,*], Sergio A. Bea [a,b], Jesús Carrera [c], Luis Vives [a]

[a] Instituto de Hidrología de Llanuras "Dr. Eduardo J. Usunoff", Av. República de Italia 780, 7300 Azul, Buenos Aires, Argentina
[b] CONICET, Av. Rivadavia 1917, C1033AAJ Ciudad Autónoma de Buenos Aires, Argentina
[c] GHS UPC-CSIC, Instituto de diagnóstico Ambiental y Estudios del Agua (IDAEA), CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain

ARTICLE INFO

Article history:
Received 22 January 2017
Accepted 3 April 2017
Available online 8 April 2017
This manuscript was handled by Prof. P. Kitanidis, Editor-in-Chief, with the assistance of Massimo Rolle, Associate Editor

Keywords:
Mixing analysis
EMMA
Conservative components
Groundwater
Chemical reaction identification

ABSTRACT

Mixing calculations (i.e., the calculation of the proportions in which end-members are mixed in a sample) are essential for hydrological research and water management. However, they typically require the use of conservative species, a condition that may be difficult to meet due to chemical reactions. Mixing calculation also require identifying end-member waters, which is usually achieved through End Member Mixing Analysis (EMMA). We present a methodology to help in the identification of both end-members and such reactions, so as to improve mixing ratio calculations. The proposed approach consists of: (1) identifying the potential chemical reactions with the help of EMMA; (2) defining decoupled conservative chemical components consistent with those reactions; (3) repeat EMMA with the decoupled (i.e., conservative) components, so as to identify end-members waters; and (4) computing mixing ratios using the new set of components and end-members. The approach is illustrated by application to two synthetic mixing examples involving mineral dissolution and cation exchange reactions. Results confirm that the methodology can be successfully used to identify geochemical processes affecting the mixtures, thus improving the accuracy of mixing ratios calculations and relaxing the need for conservative species.

1. Introduction

The term "mixing calculation" refers here to the computation of the proportions (mixing ratios) in which two or more end-members are mixed in a sample. The calculation of mixing ratios is a key task in hydrological research and water resources management because it yields quantitative information about the origin of the water sources, which is one of the ultimate goals of hydrology.

Mixing ratios are usually computed from hydrochemical data such as major and trace ion chemistry and stable isotopes. The identification of mixing and chemical processes in a groundwater system typically involves computation of saturation indexes with respect to minerals that are suspected to be present in the aquifer, or dispersion plots of available data to help in establishing mixing lines and associating deviations caused by chemical reactions (e.g., Mazor et al., 1973, 1985; Mazor, 1990; Appelo and Postma, 2005). The approach is rich in that insights into the system behaviour are gained in the process, but it can be tedious and difficult when many end-members are present, which shadow chemical reactions. Christophersen and Hooper (1992) proposed a two-step procedure to systematize preliminary mixing analyses. The first step consists in identifying the end-members that mix in the water samples. This step is essentially conceptual, but it can be greatly aided by methodologies based on Principal Component Analysis, such as End Member Mixing Analysis, (EMMA), which consists of finding the minimum number of end-members that explain the observed variability within a dataset and helps in establishing the composition of these potential end-members (see Christophersen and Hooper, 1992; 1993, for a detailed description). The second step consists in calculating the mixing ratios for the identified end-members in the mixtures from the concentrations of conservative species. It is a mechanical step accomplished by constrained linear least squares (Christophersen et al., 1990), separately for each sample; by non-linear optimization, taking into account all mixtures, when end-member concentrations are uncertain (Carrera et al., 2004); or geometrically based on the information of two or n-principal component mixing (Laaksoharju et al., 2008). Combinations of these two steps have been broadly used in hydrology for specific tasks, such as hydrochemical characterization and conceptual models validation (Hooper et al., 1990; Christophersen et al., 1990; Suk and Lee, 1999; Laaksoharju et al., 2008; Long and Valder, 2011; Jiménez-Martínez et al., 2011; Valder et al., 2012; Gómez et al., 2014),

* Corresponding author.
E-mail address: pelizardi@faa.unicen.edu.ar (F. Pelizardi).

http://dx.doi.org/10.1016/j.jhydrol.2017.04.010
0022-1694/

runoff quantification (Burns et al., 2001), hydrograph separation (Hooper et al., 1990; Katsuyama et al., 2001; Soulsby et al., 2003; James and Roulet, 2006; Menció et al., 2014), urban groundwater recharge evaluation (Carrera et al., 2004; Vázquez-Suñé et al., 2010), and geochemical processes identification (Gómez et al., 2008; Tubau et al., 2014; Jurado et al., 2015).

Application of this methodology needs to satisfy two conditions: (1) a sufficient number of species must exhibit conservative behaviour, and (2) their concentrations must be sufficiently different (Christophersen and Hooper, 1992; Carrera et al., 2004; Barthold et al., 2011). These conditions are often met in surface waters, which motivated initial EMMA developments (Christophersen and Hooper, 1992), but they are difficult to achieve in groundwater systems because of reactions driven by interaction with mineral surfaces and gases, and long residence times (Parkhurst, 1997; Carrera et al., 2004; Rueedi et al., 2005). The problem of the non-conservative behaviour was identified by Hooper (2003) who also pointed out that time variability in end-member concentrations might affect EMMA. The issue of chemical reactions can be addressed by geochemical methods using speciation codes (e.g., NETPATH, Plummer et al., 1991; PHREEQC, Parkhurst and Appelo, 2013). However, they require that both end-members and potential chemical reactions be known a priori. Reactions can also be heuristically inferred by simple inspection of the elemental mass balance of those constituents that cannot be described by pure mixing supported by independent knowledge of the system (Gómez et al., 2008). Non-conservative behaviour is poorly addressed by multi-variate methods. Rueedi et al. (2005) suggested excluding species suspected to be affected by chemical reactions from mixing ratios calculation, whereas Carrera et al. (2004) suggested increasing their variance. The issue of time variability was addressed by Tubau et al. (2014) by treating the extreme concentrations of a water source as end-members. These authors and Jurado et al. (2015) went further and used mixing calculations to identify chemical reactions by examining the species that display the largest residual from conservative mixing calculations and attributing these residuals to reactions.

All approaches above evaluate the contribution of mixing from conservative tracers. The contributions from reactions are considered in subsequent steps. Nevertheless, in real chemical analyses, there are many species, with different and unknown sources and behaviors. Some species concentrations could be explained by conservative mixing, others by chemical reactions and others (often, most) by both processes. The challenge here is to quantify the mixing processes and chemical reactions taking into account the information contained in the whole dataset. It follows the need to have a methodology capable to identify and eliminate the impact of the chemical reactions on mixing calculations.

The concept of chemical components sounds appealing to overcome the difficulties and uncertainties induced by chemical reactions in EMMA and mixing calculations because a chemical component is defined as a linear combination of species that is not affected by chemical reactions (e.g., Yeh and Tripathi,1989; Rubin, 1983). This concept has been widely exploited to reduce the numerical demand in reactive transport problems (Saaltink et al., 2001; Molins et al., 2004; De Simoni et al., 2007; Kräutle and Knabner, 2005, 2007; Bea et al., 2009). The introduction of conservative components is convenient in reactive transport calculations because it allows eliminating contributions of reactions to the mass balance equations (nonlinear/coupling terms). The approach is quite elegant when all species follow the same transport processes because conservative components also do, which facilitates transport formulation (e.g., Rubin, 1983; Kirkner and Reeves, 1988; Steefel and Lasaga, 1994; Steefel and MacQuarrie, 1996; Zhang et al., 2005; Bea et al., 2010). However, the key issue in the use of components is not the resulting simplification on the transport equation, which is lost if transport processes are species dependent (e.g., different diffusion coefficients for each species). The key issue is the elimination of (fast) reactions sink-source terms. Since conservative mixing calculations also require elimination of reactions, it is natural to conjecture that the same concept can also be beneficial for mixing calculations.

This work is devoted to propose a methodology with a twofold purpose: (1) to assess whether EMMA can help in the identification of not only end-members, but also geochemical reactions occurring in the groundwater system, and (2) to test whether the performance of mixing ratio calculations can be improved by the use of conservative chemical components as a linear combination of non-conservative ones based on the identified geochemical system.. The proposed methodology is not so much to be an alternative to conventional hydro-geo-chemical tools, as a complementary tool that should be especially useful in the early stages of exploratory chemical data analysis.

2. Concepts

The proposed approach is based on the use of chemical components, EMMA and mixing calculations, which are described below.

2.1. Chemical systems, reactions and components

Chemical systems are defined in terms of the species of interest (i.e., those that have been measured and the modeller is interested in explaining), and the reactions among them. Both can be explained in terms of the stoichiometric matrix, **S**, a $n_r \times n_s$ matrix, whose n_r (number of reactions) rows contain the stoichiometric coefficients of each of the n_s species, aligned by columns, in each reaction. These concepts can be illustrated by means of two synthetic examples that will be used later.

The first example (Case A) involves gypsum ($CaSO_4{\cdot}2H_2O$) and magnesite ($MgCO_3$) dissolution or precipitation:

$$Ca^{+2} + SO_4^{-2} + 2H_2O \Longleftrightarrow CaSO_4 \cdot 2H_2O(s) \quad \text{(R1)}$$

$$Mg^{+2} + CO_3^{-2} \Longleftrightarrow MgCO_3(s) \quad \text{(R2)}$$

The coefficients in the stoichiometric matrix are the numbers multiplying each species in the chemical equation of the reaction, with its sign switched when they are in the left hand side. For instance, in Reaction (R1), the stoichiometric coefficient of Ca^{2+} is -1, that of gypsum is 1, and that of water is -2. However, we are only interested in the concentrations of aqueous species. Therefore, the coefficients of gypsum and water will not be included in the stoichiometric matrix. Moreover, we arrange the species and reactions so that: (1) the matrix can be split into two blocks of dimensions $n_r \times (n_s - n_r)$ and $n_r \times n_r$ and (2) the second block equals minus the identity matrix. This may require rearranging the reactions and it is helped by splitting the species in primary and secondary, a concept we will not develop here. Further details on the construction of stoichiometric matrices are given by Molins et al. (2004). The resulting matrix for a chemical system involving Reactions (R1) and (R2) is:

$$\mathbf{S} = (\mathbf{S}_1 \mid -\mathbf{I}) = \left(\begin{array}{c|cc|cc} & Ca^{2+} & Mg^{2+} & SO_4^{2-} & CO_3^{2-} \\ \text{Gyp. diss.} & -1 & 0 & -1 & 0 \\ \text{Mgs. diss.} & 0 & -1 & 0 & -1 \end{array}\right) \quad (1)$$

The second example (Case B) includes the reactions of Case A, and Na-Ca exchange:

$$0.5X_2Ca + Na^+ \Longleftrightarrow 0.5Ca^{+2} + XNa \quad \text{(R3)}$$

In this case, the corresponding stoichiometric matrix for the three reactions would be:

$$\mathbf{S} = (\mathbf{S}_1 \mid -\mathbf{I}) = \left(\begin{array}{c|ccc|cc} & Ca^{2+} & Mg^{2+} & SO_4^{2-} & CO_3^{2-} & Na^{+} \\ \text{Gyp. diss.} & -1 & 0 & -1 & 0 & 0 \\ \text{Mgs. diss.} & 0 & -1 & 0 & -1 & 0 \\ \text{Cat. exch} & 0.5 & 0 & 0 & 0 & -1 \end{array}\right) \quad (2)$$

where we have had to increase the number of reactions and species. The system may include other species, which are conservative because they do not participate in any reaction. We do not include such species in the stoichiometric matrix because their corresponding column would be a vector of zeros.

The definition of the stoichiometric matrix allows expressing reactions in a compact manner, which is useful to abstract and simplify speciation calculations (e.g., Bea et al., 2009), or reactive transport calculations (e.g., Saaltink et al., 1998; Molins et al., 2004; De Simoni et al., 2005; Kräutle and Knabner, 2005, 2007). Thus, the increments in the concentrations of all species (Δc) due to reactions can be written as:

$$\Delta \mathbf{c} = \mathbf{R} = \mathbf{S}^t \mathbf{r} \quad (3)$$

where $\mathbf{c}$ is the vector of concentrations of all species, $\mathbf{R}$ is the contribution of reactions to the mass of each species (per unit volume), and $\mathbf{r}$ is the vector of the extent of all reactions. For example, in R1, if r_1 moles (per m^3) evolve from left to right, the same amount of SO_4^{2-} and Ca^{2+} will be subtracted from their concentrations. In the framework of (1), this results from multiplying r_1 times the -1 stoichiometric coefficients of SO_4^{2-} and Ca^{2+} in the first row of $\mathbf{S}$

Eliminating the impact of reactions on mixing calculations requires introducing components, defined as linear combinations of species that remain unchanged by reactions. This can be achieved by computing components as $\mathbf{u} = \mathbf{U}\mathbf{c}$, where $\mathbf{U}$ is called the component matrix, a full-ranked $(n_s - n_r) \times n_s$ kernel matrix defined such that:

$$\mathbf{U} \cdot \mathbf{S}^t = 0 \quad (4)$$

The use of this matrix is interesting because it helps us to achieve the goal of eliminating reactions. Indeed, multiplying Eq. (3) by $\mathbf{U}$ yields:

$$\Delta \mathbf{u} = \mathbf{U}\mathbf{R} = \mathbf{U}\mathbf{S}^t \mathbf{r} = 0 \quad (5)$$

which expresses that components $\mathbf{u}$ are conservative (i.e., their concentration is not affected by reactions). Note that this definition of components should not be confused with that of compositional data analysis, when data are viewed as fractions constrained to add up to one (e.g., Tolosana-Delgado et al., 2005).

There are a number of ways to compute matrix $\mathbf{U}$. In fact, its definition is non-unique. This is not a problem because all definitions are equivalent, in the sense that the concentration subspace defined by the rows of $\mathbf{U}$ is always the same. It is the kernel (null space) of the reactions, which represents the concentration (changes) that will remain unaffected by reactions. In fact, one can take advantage of the different ways of defining $\mathbf{U}$ for specific purposes (see, e.g., Molins et al., 2004). In our case $\mathbf{U}$ can be simply built as consisting of two blocks. The first one is an $(n_s - n_r) \times (n_s - n_r)$ identity matrix. The second one is the $(n_s - n_r) \times n_r$ transpose of the first block of $\mathbf{S}$, $\mathbf{S_1}$. The resulting $\mathbf{U}$ for Case A is:

$$\mathbf{U} = (\mathbf{I}|\mathbf{S}_1^t) = \left(\begin{array}{c|cccc} & Ca^{2+} & Mg^{2+} & SO_4^{2-} & CO_3^{2-} \\ u_{\mathbf{Gyp}} & 1 & 0 & -1 & 0 \\ u_{\mathbf{Mgs}} & 0 & 1 & 0 & -1 \end{array}\right) \quad (6)$$

It is easy to check that this matrix satisfies Eq. (4). That is, $\mathbf{U}$ is orthogonal to $\mathbf{S}$, which simply means that $u_{\mathbf{Gyp}} = Ca^{+2} - SO_4^{-2}$ is not affected by any of the reactions. It is not affected by gypsum dissolution which will increase the concentrations of Ca^{+2} and SO_4^{-2} by the same amount, and it is not affected by magnesite dissolution, which does not contribute to either solute. The same can be said about $\mathbf{u}_{\mathbf{Mgs}}$Similarly, the components matrix for Case B is:

$$\mathbf{U} = (\mathbf{U}|\mathbf{S}_1^t) = \left(\begin{array}{c|ccccc} & Ca^{2+} & Mg^{2+} & SO_4^{2-} & CO_3^{2-} & Na^{+} \\ \mathbf{u}_{\mathbf{Gyp-Cat.Ex.}} & 1 & 0 & -1 & 0 & 0.5 \\ \mathbf{u}_{\mathbf{Mgs}} & 0 & 1 & 0 & -1 & 0 \end{array}\right) \quad (7)$$

Here, the second component is identical to that of Case A. The first component involves both cation exchange and gypsum dissolution. It is easy to see that $\mathbf{u}_{\mathbf{Gyp-Cat.Ex}} = Ca^{+2} - SO_4^{-2} + 0.5Na^{+}$ is not affected by gypsum dissolution or precipitation, or by cation exchange processes. Note that this new component involves two reactions because they share one species. In fact, this is always the case. All reactions sharing at least one species will lead to a simple component. This will reduce the number of components in complex cases, where numerous reactions affecting all species are present. In the limit, as the number of reactions increases beyond the number of analysed species, no conservative components might be available, so that the approach proposed here would be of no use.

2.2. EMMA and mixing calculations

The goal of EMMA is to identify the minimum number of end-members that are mixing in the system, so as to explain the chemical variability assuming conservative mixing. The method, which is clearly explained by Christophersen and Hooper (1992), and Hooper (2003), is basically motivated by the need to reduce the dimensionality of the analysis. Visualization and interpretation of a large number of chemical analyses of many species is hard. The problem is much easier to handle if, instead of studying a dataset potentially as large as the number of species, it would suffice to analyse their projections on a space of much smaller (say 2 or 3) dimensions. This is achieved by the following procedure:

(1) Eigenanalysis of the correlation matrix of data ($\mathbf{X}^{*t}\mathbf{X}^{*}$), where $\mathbf{X}^{*}$ is a matrix containing the standardised values of each concentration measurement (i.e., the value obtained after subtracting the mean and dividing by the standard deviation of each species). The resulting eigenvalues divided by the number of species, represent the fraction of the total variance explained by each eigenvector, which can be viewed as an axis of the much smaller dimension space in which to analyse the data.
(2) Projection of the concentration data onto the new space defined by the selected eigenvectors (i.e., by those that explain most of the variability). The reduced dimensions of this space imply that one can visualize the projections of the data in one (if only two eigenvectors are retained) or two (if three are required) graphs. In the context of mixing analysis, end-members should encircle the data, which facilitates the testing of candidates for end-members by simply projecting them on the reduced space and analysing if they indeed encircle the data (see Hooper, 2003, for details on the procedure).

If the concentrations are conservative and really result from mixing of n_e end-members, EMMA will be successful because, except for measurements errors, concentrations will belong to a

$(n_e - 1)$ dimensional space. The concept is illustrated in Fig. 1, for a 2 end-members problem, where concentrations lie on the straight line between the end-members (a 1-dimensional space). The eigenvector direction coincides with the line joining these two end-members. For the purpose of identifying mixing, it would suffice to know the relative positions of samples along that line, instead of working with all concentration data. This is only possible if the concentrations of all species are explained by conservative mixing.

The approach may fail in the case of reacting species, when reactions modify concentrations according to Eq. (3), (Fig. 1b). For this reason, the key issues in this work are the identification and subsequent elimination of the contribution of reactions. In cases in which both mixing and chemical reactions occur, concentrations would be given by:

$$\mathbf{c_i} = \underbrace{\mathbf{X_e}\lambda_\mathbf{i}}_{conservative\ mixing} + \underbrace{\mathbf{S^t r_i}}_{chemical\ reactions} \tag{8}$$

where $\mathbf{c_i}$ is the vector of concentrations in sample i, $\mathbf{X_e}$ is a $(n_s x n_e)$ matrix containing the concentration of all species at the end-members, $\lambda_\mathbf{i}$ is the n_e dimensional vector of mixing ratios and $\mathbf{r_i}$ is the vector of reaction extents for the $i-th$ sample. This equation shows that concentrations do not need to lie on the n_e dimensional space, but on a $n_e + n_r$ dimensional space, if the directions of reactions, given by the rows of $\mathbf{S}$, are independent of the mixing directions, given by the columns of $\mathbf{X_e}$.

Since reactions increase the dimension of the vector space explaining the variability of concentrations, they should show an additional eigenvector in an EMMA analysis. The example of Fig. 1b shows that reactions are not necessarily orthogonal to the mixing space, while eigenvectors are. Therefore, the identification of reactions needs not to be unequivocal. Still, the loadings of eigenvectors should reflect reactions somehow. This together with conceptual understanding of the site should help modellers in identifying the actual reactions.

Assuming that the reactions have been identified, the corresponding component matrix can be computed as explained in the previous section. This step eliminates the contribution of the chemical reactions in Eq. (8). In fact, multiplying Eq. (8) by $\mathbf{U}$, and using Eq. (4), yields:

$$\mathbf{u_i} = \mathbf{X_{ue}}\lambda_\mathbf{i} \tag{9}$$

where $\mathbf{u_i}$ is the vector $(n_s - n_r)$ of all conservative components at sample i and $\mathbf{X_{ue}}$ is the $((n_s - n_r) \times n_r)$ matrix of components concentrations at end-members. Note that, conservative components include not only those resulting from reactions (e.g., $\mathbf{u_{Gyp}}$ or $\mathbf{u_{Mgs}}$) in Eq. (6) but also the species that do not participate in reactions (e.g., Cl^-, which is often conservative).

Mixing computations would then proceed as usual for each sample, by minimizing the objective function:

$$J_{ui} = (\mathbf{u}_i - \mathbf{u}_i^*)^t \mathbf{V}_u^{-1} (\mathbf{u}_i - \mathbf{u}_i^*) \tag{10}$$

where $\mathbf{u}_i^*$ are the measured components and $\mathbf{V_u} = \mathbf{U^t V_c U}$ is the covariance matrix of errors in components, with $\mathbf{V}_c$ the covariance matrix of errors in species concentrations. This objective function is minimized with respect to $\lambda_\mathbf{i}$ subject to the constraint that mixing ratios add up to 1:

$$\lambda_i^t \mathbf{1}_u = 1 \tag{11}$$

where $\mathbf{1}_u$ is a vector of 1 s of dimensions $n_u = n_s - n_r$.

3. Methods

3.1. Approach procedure

The proposed approach consists of the following steps:

Step 1: **Chemical data processing**, data which includes all steps from sample collection to having a well-defined concentration data matrix.

Step 2: **Application of EMMA** to all data using the full set of chemical species

Step 3: **Identification of end-members and chemical reactions** in the system on the basis of EMMA results and conceptual understanding, and definition of new conservative components.

Step 4: **Application of EMMA** to the conservative components and species resulting from the identified reactions.

Step 5: **Assessing the previously identified end-members and chemical reactions**, which entails testing if chemical data, once projected on the subspace of the eigenvectors of Step 4, can be explained in terms of conservative mixing.

Step 6: **Mixing calculations** using the end-members identified in Step 5, and the final conservative components and species.

Step 7: **Calculations of reaction extents** from Eq. (4), or from the deviations from conservative mixing (e.g., Tubau et al., 2014).

Note that steps 1, 3 and 5 involve a good deal of conceptual understanding of the system. In fact, steps 4 and 5 may have to be repeated testing different combinations of reactions until a satisfactory solution is obtained. In this context, satisfactory means that a large percentage of variability is explained by a few

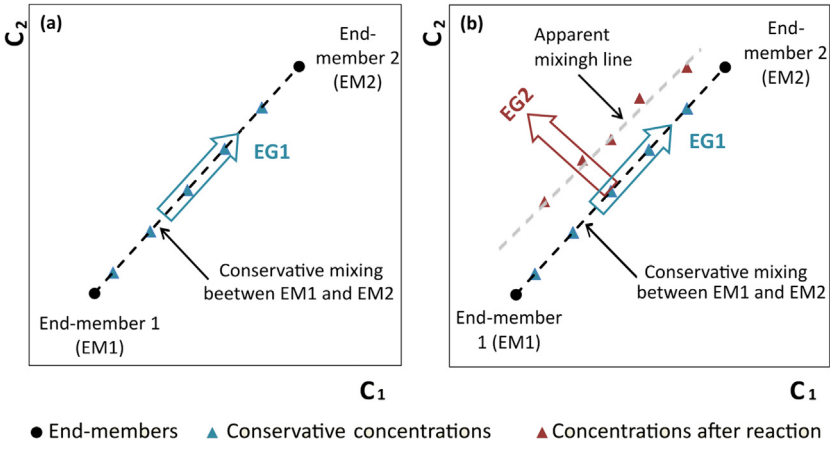

Fig. 1. Plot of species C1 over species C2 considering conservative mixing between two end-members (a) and considering the occurrence of chemical reactions (b). Note that the concentrations resulting from conservative mixing lie on straight line beetwen EM1 and EM2 (conservative mixing line in (a) and (b)). Chemical reacions cause deviation from the mixing line (concentration after reactions) originating a new apparent mixing line (b).

eigenvectors, and that these eigenvectors make hydrological sense (i.e., they can be used to identify meaningful end-members).

3.2. Generation of synthetic data

The application of the proposed methodology is illustrated using two synthetic datasets. The advantage of using synthetic data lies in the fact that the contributions of conservative mixing and reaction extents are known for every species in each sample. These datasets were generated by conservative mixing of two end-members and subsequently modified by mineral dissolution and cation exchange according to the following procedure:

(1) *Generation of end-members compositions*. A diluted (DIL) and a brackish (BRA) end-members were used. The chemical composition of DIL was obtained modifying a standard rainwater composition (e.g., see Appelo and Postma, 2005) by processes usually observed in the unsaturated zone: evaporation (the initial solution composition was evaporated 4 times), addition of sea salt deposition (1 mmol of NaCl), cation exchange (1 mmol of cationic exchange Mg/Ca and 1 mmol of cation exchange Na/Ca), and gypsum dissolution (1 mmol of $CaSO_4{\cdot}2H_2O$). BRA was obtained by mixing of standard seawater with rainwater in a 1:2 proportion. The concentrations of Cl^-, Na^+, SO_4^{2-}, Ca^{2+}, Mg^{2+}, K^+ and CO_3^{2-} for these two end-members are shown in Table 1.

(2) *Mixing*. The chemical compositions of the 20 mixtures were obtained by conservative mixing between these two end-members (DIL and BRA) using the mixing ratios shown in Table 2.

(3) *Reactions*. The resulting mixtures are undersaturated with respect to gypsum and magnesite (saturation indices are also shown in Table 2). The concentrations of conservative mixtures were modified considering two sets of chemical reactions: (1) Case A: Gypsum and magnesite dissolutions (Reactions 1 and 2), and (2) Case B: Gypsum and magnesite dissolutions and cation exchange Na/Ca (Reactions 1, 2 and 3). For Case A, we first computed the amount of mineral dissolution required to reach equilibrium using PHREEQC 3.0 (Parkhurst and Appelo, 2013) with the WATEQ4F thermodynamic database (Ball and Nordstrom, 1991) at 25 °C. A random fraction of this amount was added to the conservative mixtures (random to reflect that in natural systems, samples are often away from equilibrium). The actual dissolution extents are shown in Table 2. For Case B, the chemical composition of Case A was modified by adding Na/Ca exchange assuming that the Na-exchange complex was the dominant fraction on the total exchange capacity. The cation exchange extents are shown in Table 2. For both cases measurement errors were simulated by adding a random noise (between 0 and 5% of the concentration) to all species.

Fig. 2 shows Ca^{2+} vs. SO_4^{2-} concentrations (Fig. 2a) and Mg^{2+} vs. CO_3^{2-} concentrations (Fig. 2b) for conservative mixing and gypsum and magnesite dissolutions. Note that the deviations of SO_4^{2-} and Ca^{2+} from the mixing line are greater than those ones for Mg^{2+} and CO_3^{2-}. This is consistent with the reactions extents for both processes shown in Table 2.

Table 1
Cl^-, Na^+, SO_4^{2-}, Ca^{2+}, Mg^{2+}, K^+ and CO_3^{2-} concentrations for the DIL and BRA end-members.

End-members	Concentration (mmol/L)						
	Cl^-	Na^+	SO_4^{2-}	Ca^{2+}	Mg^{2+}	K^+	CO_3^{2-}
Diluted (DIL)	1.03	3.04	0.54	2.67	1.13	0.54	4.10
Brackish (BRA)	189.31	162.07	9.79	3.58	18.40	3.55	0.82

Table 2
Mixing ratios, saturation indexes for gypsum ($CaSO_4{\cdot}2H_2O$) and magnesite ($MgCO_3$) and reaction extents (mmol/L) for the samples obtained by mixing of DIL and BRA. Note that, prior to reactions, all the samples are undersaturated with respect to gypsum and magnesite.

Data	Mixing ratios		Sat Indexes Cons. Mixing		Reaction extent (mmol/L)		
	δ DIL	δ BRA	SI_Gyp	SI_Mgs	Gypsum	Magnesite	Cat. Exch
DIL	1	0	−1.99	−2.36	0.00	0.00	0.00
1	0.99	0.01	−1.95	−2.32	17.23	2.30	19.41
2	0.98	0.02	−1.92	−2.29	17.07	1.15	19.30
3	0.96	0.04	−1.86	−2.24	1.16	0.71	3.86
4	0.95	0.05	−1.84	−2.21	6.47	2.19	9.14
5	0.92	0.08	−1.79	−2.16	21.88	0.26	23.16
6	0.9	0.1	−1.76	−2.12	17.12	1.84	19.12
7	0.89	0.11	−1.75	−2.11	7.24	1.39	9.91
8	0.87	0.13	−1.73	−2.08	20.82	2.04	22.08
9	0.85	0.15	−1.71	−2.06	7.33	0.64	10.00
10	0.83	0.17	−1.69	−2.04	0.27	1.72	3.08
11	0.82	0.18	−1.68	−2.03	15.47	1.46	17.53
12	0.8	0.2	−1.67	−2.01	13.08	1.55	15.37
13	0.75	0.25	−1.63	−1.97	20.72	0.37	21.46
14	0.65	0.35	−1.59	−1.91	17.28	1.67	18.49
15	0.55	0.45	−1.55	−1.85	2.90	1.41	5.78
16	0.45	0.55	−1.53	−1.82	7.19	0.11	9.73
17	0.35	0.65	−1.49	−1.72	17.71	0.14	17.61
18	0.25	0.75	−1.46	−1.63	7.45	0.52	9.69
19	0.15	0.85	−1.44	−1.49	8.34	0.26	10.26
20	0.05	0.95	−1.42	−1.22	21.36	0.13	18.36
BRA	0	1	−1.41	−0.89	0.00	0.00	0.00

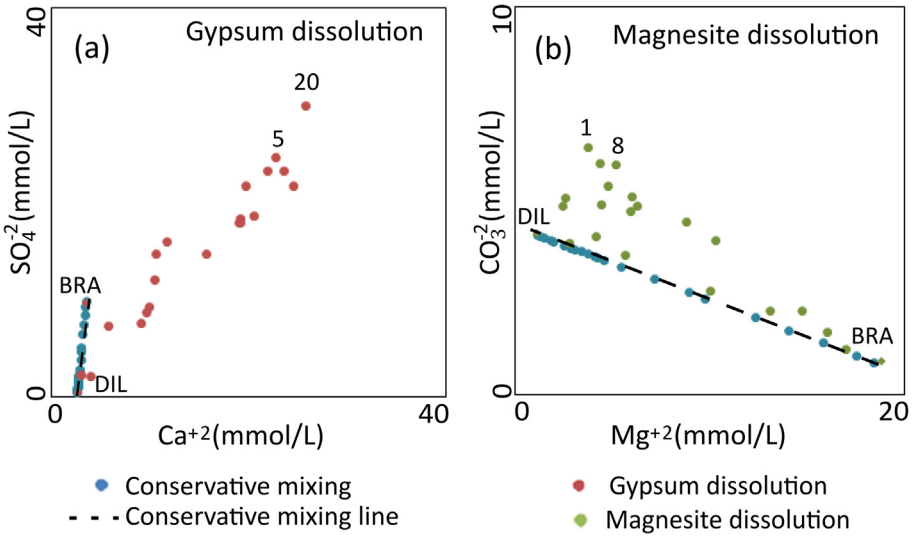

Fig. 2. Ca^{+2} vs. SO_4^{-2} (a) and Mg^{+2} vs. (b) for conservative mixing and after mineral dissolution. Conservative mixing values were obtained by linear combination of the two end-members (DIL and BRA). Deviations from the mixing line are due to gypsum and magnesite dissolutions. The samples that show most deviation from the mixing line are labelled.

3.3. Analyses procedure

In this section we describe the EMMA and MIX analyses procedure performed to illustrate the methodology.

3.3.1. EMMA analyses procedure

We assessed the impact of chemical reactions on end-members identification by comparing the projection of concentrations on the selected eigenvectors space for the two analyses that are described below:

(1) Using all-data (All): These EMMA analyses were performed with the datasets described above using the full set of species for Case A and Case B (termed here as A-All and B-All). These included conservative species (the ones whose concentration were obtained by conservative mixing) and reactive species (the ones affected by chemical reactions). Hence, these were performed to illustrate the identification of reactions (Step 3 in Section 3.1).

(2) Using conservative components (Cons): These EMMA analyses were performed with conservative species and components (Step 4) for both cases (termed as here as A-Cons and B-Cons). Decoupled conservative chemical components (**u**) were defined consistently with both reaction sets, following the methodology described in Section 2.1.

3.3.2. MIX analyses procedure

Mixing ratios were computed using MIX, the approach of Carrera et al. (2004) assuming that the end-member compositions are uncertain. The method minimizes:

$$J_u = \sum_i J_{ui} \tag{12}$$

where J_u is the global objective function and J_{ui} the contribution of each species i to the objective function. The eight analyses performed are described below:

(1) Using all data (All): performed with reactive and conservative species for both cases (A-All and B-All). The variance assigned to each datum is the square of its value.

(2) Using conservative components (Cons): performed with conservative species and conservative components (**u**) consistent with both reaction sets (A-Cons and B-Cons). The variance of the components was obtained as the sum of the variances of participating species multiplied by their components coefficients (i.e., $\mathbf{V_u} = \mathbf{U^t V_c U}$).

(3) Partial identification of reactions (Par): these analyses were performed assuming that the true chemical system could not be identified, and magnesite dissolution was not considered in the definition of chemical components for both cases (termed here as A-Par and B-Par).

(4) Using only conservative species (Csp): these analyses (termed here as A-Csp and B-Csp) were performed with those species that remained unaffected by reactions: Cl^-, Na^+ and K^+ for Case A and Cl^- and K^+ for Case B.

The results were assessed through the RMSE (Root mean square error) of computed mixing ratios compared to their true values. The overall fit in terms of concentrations was evaluated by means J_u (Eq. (12)) and the impact of each species through its contribution to the objective function J_{ui}/J_u (Eq. (10)).

4. Results and discussion

4.1. EMMA results

Case A: gypsum ($CaSO_4{\cdot}2H_2O$) and magnesite ($MgCO_3$) dissolutions

The EMMA results for A-All are shown in Fig. 3, which displays the percentage of variance explained by Eigenvectors 1, 2 and 3 (EG1, EG2, and EG3, respectively), and the corresponding contributions of each species. The first eigenvector (EG1) explains 70.62% of the total variance (Fig. 3a). It is clearly associated with mixing between the BRA and DIL end-members because: (1) the contributions are similar (approx. 0.45) for all species except Ca^{2+} and SO_4^{2-} (they should have been $\sqrt{1/7}$=0.38 if all species were equally weighted); and (2) they are all positive except for CO_3^{2-}, which is the only species with higher concentration in the DIL end-member than in the BRA end-member (Table 1).

The different contributions of the species to EG1 (Fig. 3b) could be explained as the result of the significance of mixing and reactive processes. Note that Mg^{2+} and CO_3^{2-} were minimally affected by magnesite dissolution (Table 2). Thus, their concentrations are mainly controlled by conservative mixing. In the same way, Ca^{2+} and SO_4^{2-} are the lesser contributors to this eigenvector because

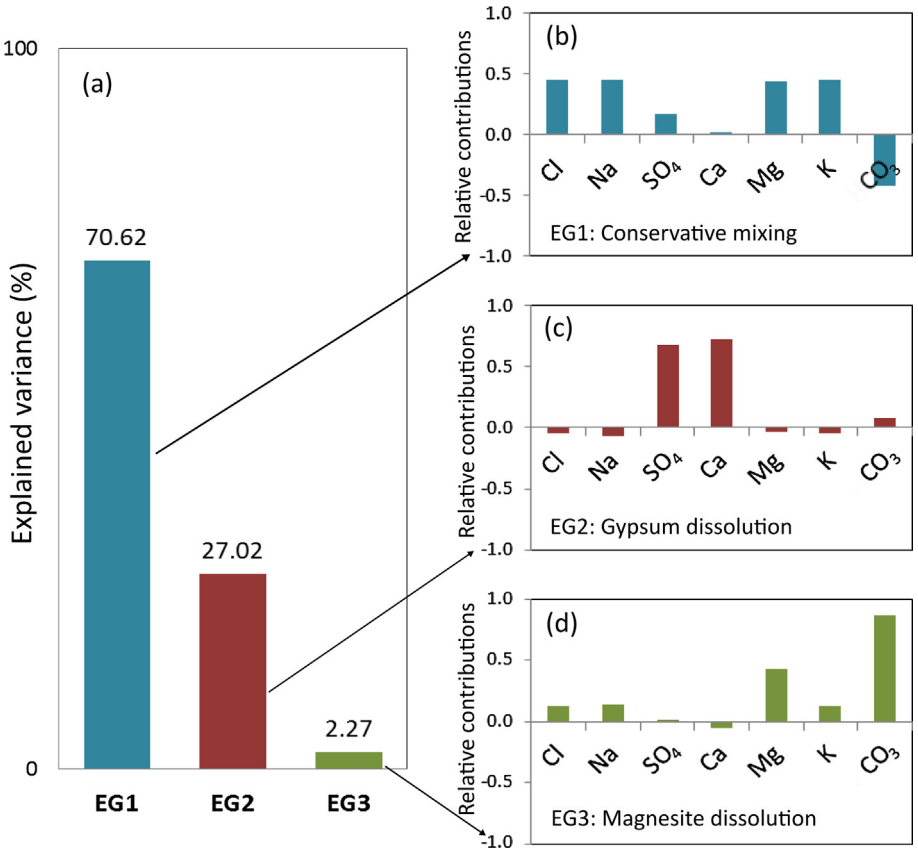

Fig. 3. Results of A-all: (a) Contributions of Eigenvectors 1, 2 and 3 to the total data variance. (b), (c) and (d) relative contributions of the species to each Eigenvector. Note that the eigenvectors are associated with a process due to the contribution of the chemical species to each one of them.

of the impact of gypsum dissolution in their concentrations (see reaction extents in Table 2 and Fig. 2a).

The second eigenvector (EG2) explains 27.02% of the variance (Fig. 3a). The main contributors are Ca^{2+} (0.73) and SO_4^{2-} (0.68), thus it is associated with gypsum dissolution (Fig. 3c). Note that the species that contribute least to the first eigenvector (EG1) are the ones with the largest loadings on the second.

Lastly, the third eigenvector (EG3) explains 2.27% of the total variance (Fig. 3a) and CO_3^{2-} (0.87) and Mg^{2+} (0.43) are the main contributors, whereas the contributions of the remaining species are below 0.14 (Fig. 3d). Hence, this eigenvector should be related to magnesite dissolution.

Summarizing, these three eigenvectors account for a 99.91% of the total variance, and can be associated with the three processes that had been imposed to obtain the chemical composition of the samples: (1, EG1) conservative mixing, (2, EG2) gypsum, and (3, EG3) magnesite dissolutions. The percentage of the total variance explained by each eigenvector can be related to the magnitude of these processes (Figs. 2 and 3). The remaining eigenvectors explained less than 0.1% of the data variance and are related to the random noise.

EMMA allows screening the potential end-members projecting the data into the lower dimensional space defined by the eigenvectors. The end-members should circumscribe the data and they should be the extreme points, outside the observed data, to be able to explain the mixture. However, the data dispersion induced by reactive processes could lead to a misinterpretation about the total number of end-members. The sample projection onto the first three eigenvectors for A-All is shown in Fig. 4a and c. The extremes values (potential end-members) onto EG1 (associated with conservative mixing) are DIL and sample 20. DIL is the end member with lowest concentrations and sample 20 is the one with highest contribution of BRA and input of SO_4^{2-}, Ca^{2+}, Mg^{2+} and CO_3^{2-} due to gypsum and magnesite dissolution (Table 2). The projections of DIL and BRA samples on the EG2 (related to gypsum dissolution) exhibit the lowest values because neither is affected by the gypsum dissolution. Oppositely, the highest values onto this eigenvector correspond to samples 20 and 5 (Fig. 4a), which are affected by the highest gypsum dissolution extents (Table 2). They might be interpreted as fictitious end-members (Fig. 4a).

A similar conclusion could be achieved analysing the projections on EG3, related to magnesite dissolution (Fig. 4c). Here, the DIL end-member, unaffected by the reaction, shows the lowest value. The positive extreme corresponds to sample 1 (Fig. 4c), which showed the largest magnesite dissolution extent (Table 2) and the highest deviation from the mixing line (Fig. 2b). Again, this sample might also be interpreted as an end-member.

Note that the application of the EMMA methodology considering all species dataset suggests that five, instead of two, water samples are candidates to end-members. This reflects that chemical reactions increase the dimension of the vector space explaining the variability of concentrations. In this case, there is an additional eigenvector in an EMMA analysis for each reaction (EG2 related to gypsum dissolution and EG3 related to magnesite dissolution). In general, this information should be taken as qualitative. It is the responsibility of the modeller to decide how to use it, but it appears natural to us that the potential for gypsum and magnesite dissolutions should be assessed. Also the modeller should check the conceptual likelihood of having five or two end-members.

In our case, the assumptions of gypsum and magnesite dissolutions (chemical reactions identified in the Step 3 described in Section 3.1) are considered. The following step is to perform EMMA of

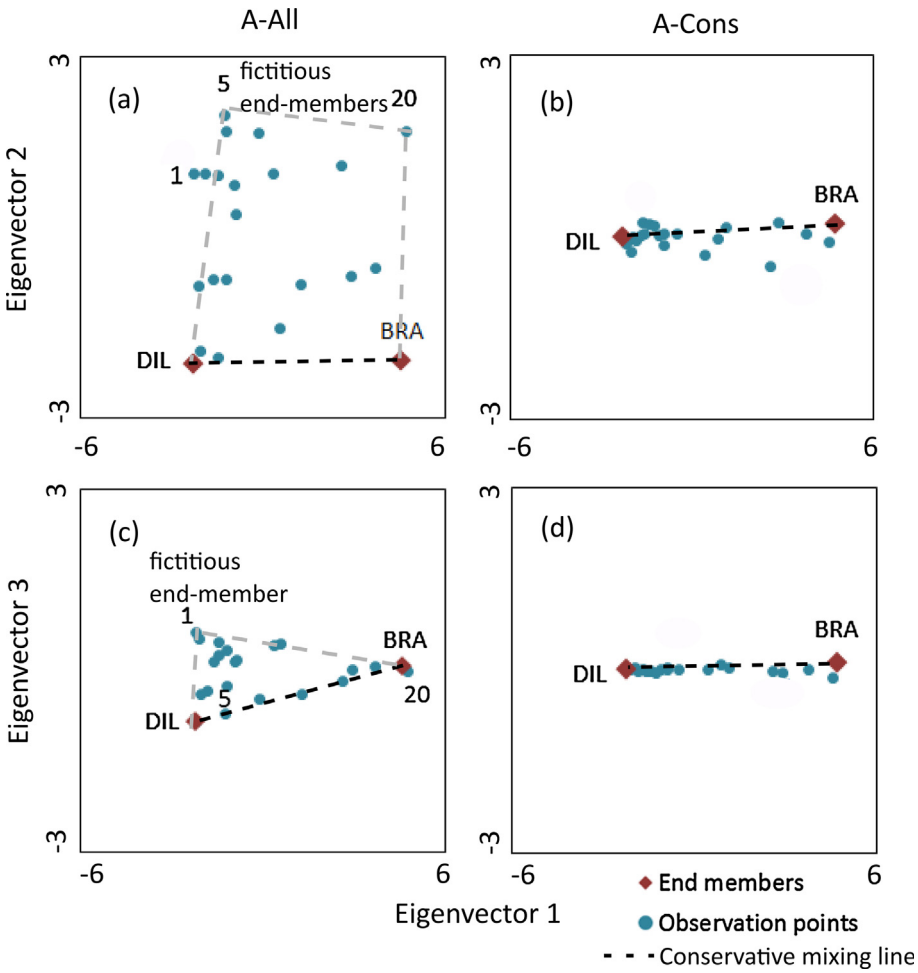

Fig. 4. EMMA projections of concentration data on the space defined by eigenvectors 1, 2 and 3 for Case A using all data (a) and (c), and using conservative components (b) and (d). Note that conservative components explain the composition of the samples as the mixture of the two end-members.

the conservative components and species resulting from the identified reactions. In this analysis, termed as A-Cons (Fig. 4b and d). EMMA was performed with five conservative components: Cl^-, Na^+, K^+ and the new decoupled components $\mathbf{u}_{Gyp} = Ca^{+2} - SO_4^{-2}$, and $\mathbf{u}_{Mgs} = Mg^{+2} - CO_3^{-2}$ (recall Section 2.1). In this analysis EG1 explains the 97.27% of the total variance, with similar contribution for all components (approximately $\sqrt{1/5} = 0.45$, results not shown). This eigenvector could be fully associated with conservative mixing. Note that it explains as much variance as the first and second eigenvectors together in the Analysis A-All. We attribute the fact that not all variance (say, 99%) is explained by this eigenvalue to the fact that errors in $\mathbf{u}_{Gyp}$ and $\mathbf{u}_{Mgs}$ have a larger variance than individual species and that the absolute values of $\mathbf{u}_{Gyp}$ are relatively small. The remaining eigenvectors explain less than 1% of the data variance and are mainly related to the random noise. Thus, the projection of concentrations in the new eigenvector space (Fig. 4b and d) clearly shows that data dispersion has been virtually eliminated, and the extreme waters in the first eigenvector are the real end-members DIL and BRA.

Case B: gypsum ($CaSO_4{\cdot}2H_2O$) and magnesite ($MgCO_3$) dissolution, and cation exchange Na/Ca

Fig. 5 summarizes the EMMA results for B-All. Now, the first two eigenvectors explain more than 92% of the total variance. As in the previous analysis, EG1 (77.89%) is also associated to mixing (Fig. 5a and b) and EG2 (14.55%) to gypsum dissolution (the main contributors are SO_4^{2-} and Ca^{2+} with loadings of 0.83 and 0.39, respectively, Fig. 5c). However, the association is not as clear as in Case A. EG3 (5.83%) is different; the main contributor is Ca^{2+} (0.80), followed by the negative contributions of SO_4^{2-}, CO_3^{2-} and Na^+ (Fig. 5d). This eigenvector can not be easily related to any specific reaction, though the opposite sign of the contributions of Ca^{2+} and Na^+ suggests association with cation exchange. The main contributors to EG4 (1.65%, Fig. 4e) are CO_3^{2-} (0.81), and Mg^{2+} (0.42), which suggest association to magnesite dissolution as in Case A.

The projections of the samples onto the first three eigenvectors are shown in Fig. 6. As in Analysis A-All, the data dispersion caused by reactions might lead modellers to overestimate the number of end-members. The extreme values for EG1 are sample 3 and sample 20 (Fig. 6a), whereas the extreme ones for EG2 are BRA and samples 20 and 8 (samples are pointed as fictitious end-members) affected with the maximum gypsum dissolution extents (Table 2). The DIL end-member and the sample 18 are the extreme values of EG3 (Fig. 6c).

According to the proposed approach, we performed analysis B-Cons with four conservative components: Cl^-, K^+, and the decoupled components (recall Section 2.1) $\mathbf{u}_{Gyp-Cat.Ex} = Ca^{+2} - SO_4^{-2} + 0.5Na^+$ (to eliminate data dispersion generated by gypsum dissolution and cation exchange Ca/Na), and $\mathbf{u}_{Mgs} = Mg^{+2} - CO_3^{-2}$ (to eliminate dispersion generated by magnesite dissolution). As in A-Cons, EG1 explained almost the total variance (99.84%) with the same contribution of all the components (0.5). The sample projections onto the eigenvectors space confirms

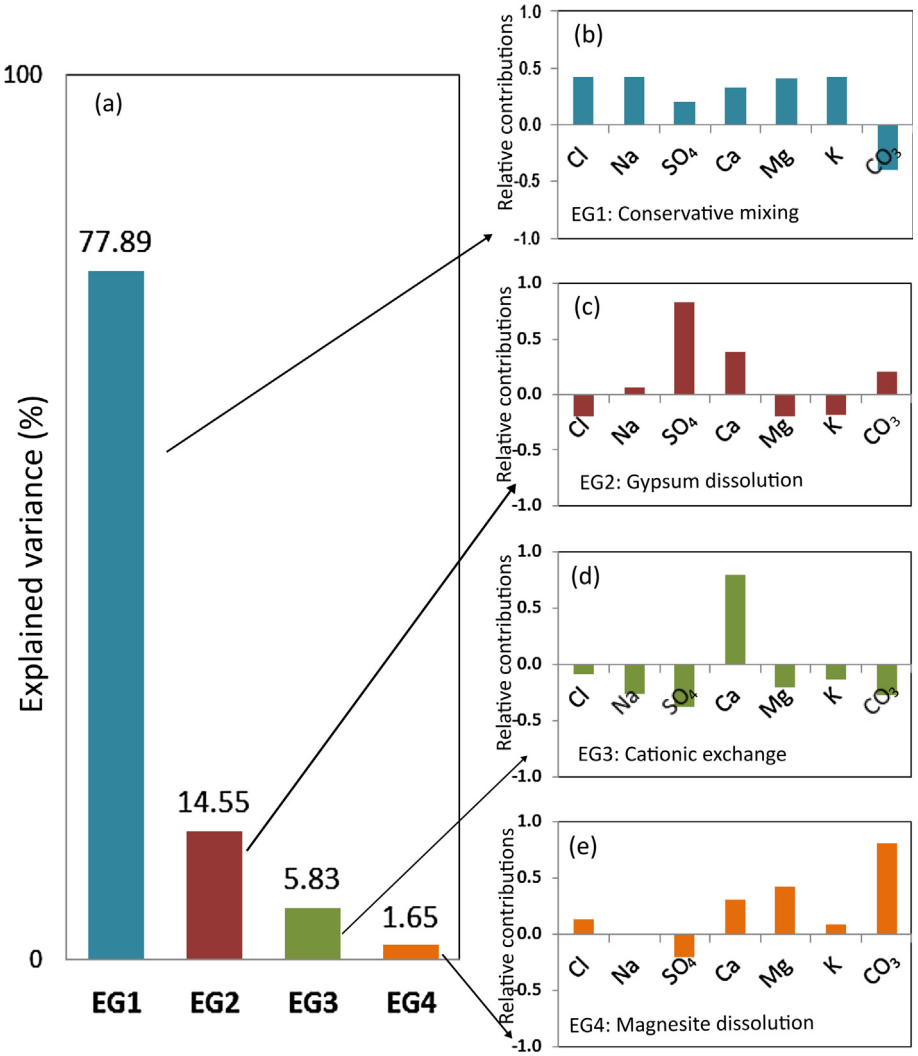

Fig. 5. Results Case B-all data: (a) Contributions of Eigenvectors 1, 2, 3 and 4 to the total data variance. (b), (c), (d) and (e) relative contributions of the species to each Eigenvector.

conservative mixing (Fig. 6b and d). Clearly, the data dispersion generated by chemical reactions was eliminated by the use of decoupled conservative components.

The conceptual interpretation of the eigenanalysis of Case B is not as well defined as that of Case A. In fact, there are numerous combinations of minerals that might cause deviations from the mixing line. This, coupled to the fact that the number of end-members is unknown, and their composition may evolve over time, leads us to assume that eigenanalyses will rarely be clear-cut. To assess the impact of this uncertainty, we analysed several combinations of reactions involving the actual reacting species. The reactions, the corresponding decoupled components, and the percentages of variance explained by the resulting eigenvectors are shown in Table 3. The true set of reacting minerals clearly outperforms the alternative sets in that the percentage of variance explained by the first eigenvector is far larger than any of the others. It confirms that the proposed approach can be a powerful tool to: (1) identify chemical reactions, (2) eliminate the contribution of the identified reactions and (3) identify end-members.

4.2. Assessing mixing calculations

Mixing ratios were computed using MIX (Carrera et al., 2004) for all the analyses described in section 3.3.2 in order to assess and quantify the impact of chemical reactions in their calculations. Mixing calculations were evaluated by: (1) comparing the real and estimated mixing ratios by means of the root-mean-square error (RMSE) and bivariate plots; and (2) assessing the objective function values.

Case A: gypsum ($CaSO_4{\cdot}2H_2O$) and magnesite ($MgCO_3$) dissolutions

Computed mixing ratios for A-All and A-Cons are compared with the real mixing ratios in Fig. 7a and b, respectively. In general, mixing proportions of the most saline end-member are overestimated in A-All, which suggests that the calculation is trying to compensate the effect of mineral dissolution by increasing the proportion of saline water (Fig. 7a). This effect is overcome by the use of conservative components (Fig. 7b) which yields results closest to the 1:1 fit (Fig. 7b).

Table 4 displays the global objective function values (Eq. (12)), and the contribution of each component (Eq. (10)). Note that the objective function decreases from −10.58 to −0.04 when the conservative component set is used. Further insight can be gained by evaluating the contributions of each species to the objective function (a high contribution indicates large differences between measured and computed concentrations). Results indicate that, when using all data, high contributions are associated to species affected by chemical reactions. This implies that the magnitude of the

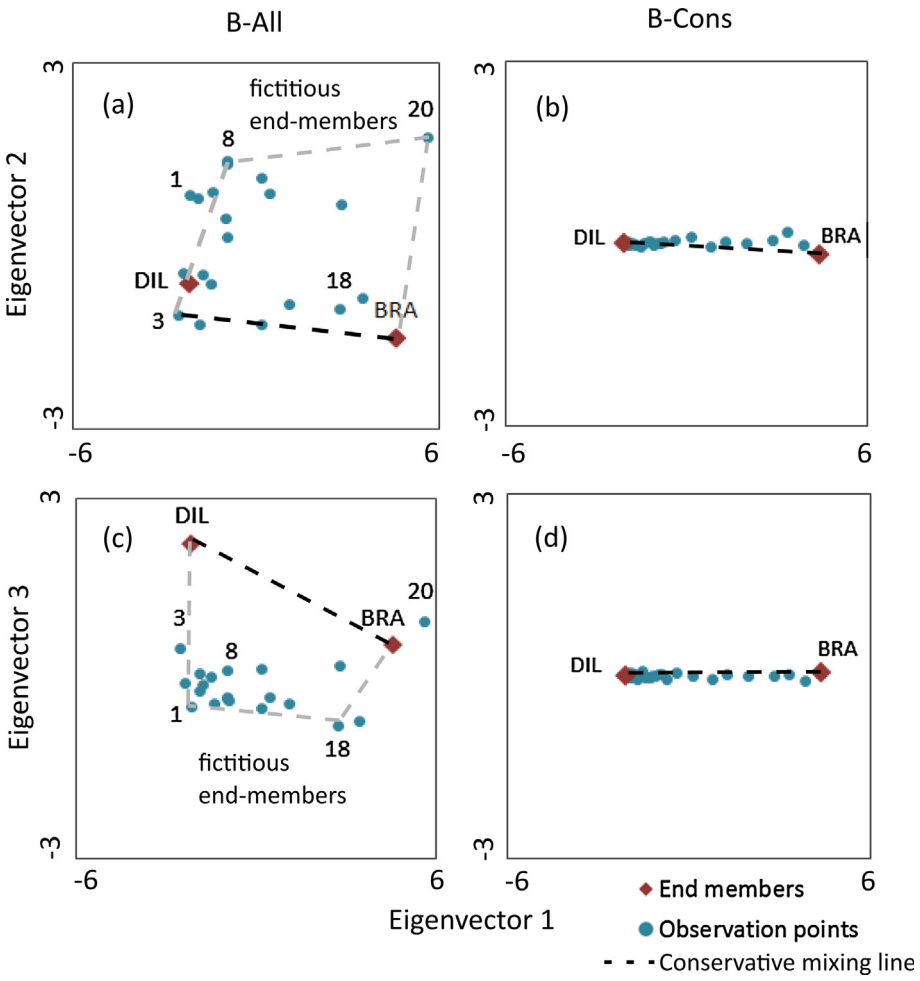

Fig. 6. EMMA projections of concentration data on the space defined by eigenvectors 1, 2, 3 and 4 for the case B using all data (a) and (c), and with conservative components (b) and (d).

Table 3
Sets of chemical reactions involving the reacting species, conservative components obtained for each system and percentages of variance explained by the eigenvectors resulting from processing each case with EMMA.

Chemical reactions	Components	Explained variance (%)		
		EG1	EG2	EG3
$Ca^{+2} + SO_4^{-2} + 2H_2O \Longleftrightarrow CaSO_4 \cdot 2H_2O(s)$ $Mg^{+2} + CO_3^{-2} \Longleftrightarrow MgCO_3(s)$ $0.5X_2Ca + Na^{+} \Longleftrightarrow 0.5Ca^{+2} + XNa(s)$	$Ca^{+2} - SO_4^{-2} + 0.5Na^{+}$ $Mg^{+2} - CO_3^{+2}$	99.84	0.11	0.03
$Na^{+} + 0.5CO_3^{-2} \Longleftrightarrow 0.5Na_2CO_3(s)$ $Mg^{+2} + SO_4^{-2} \Longleftrightarrow MgSO_4(s)$ $Ca^{+2} + CO_3^{-2} \Longleftrightarrow CaCO_3(s)$	$Ca^{+2} + 0.5Na^{+} - CO_3^{+2}$ $Mg^{+2} - SO_4^{-2}$	75.76	24.14	0.09
$Na^{+} + 0.5CO_3^{-2} \Longleftrightarrow 0.5Na_2CO_3(s)$ $Mg^{+2} + SO_4^{-2} \Longleftrightarrow MgSO_4(s)$ $Ca^{+2} + Mg^{+2} + 2CO_3^{+2} \Longleftrightarrow Ca \cdot Mg(CO_3)_2(s)$	$Ca^{+2} + 0.5Na^{+} - CO_3^{+2}$ $Ca^{+2} + SO_4^{-2} - Mg^{+2}$	73.99	25.91	0.09
$Na^{+} + 0.5CO_3^{-2} \Longleftrightarrow 0.5Na_2CO_3(s)$ $Ca^{+2} + Mg^{+2} + 2CO_3^{+2} \Longleftrightarrow Ca \cdot Mg(CO_3)_2(s)$ $Ca^{+2} + SO_4^{-2} + 2H_2O \Longleftrightarrow CaSO_4 \cdot 2H_2O(s)$	$2Mg^{+2} + 0.5Na^{+} - CO_3^{+2}$ $Mg^{+2} + SO_4^{-2} - Ca^{+2}$	89.65	10.26	0.08
$Na^{+} + 0.5CO_3^{-2} \Longleftrightarrow 0.5Na_2CO_3(s)$ $Mg^{+2} + CO_3^{-2} \Longleftrightarrow MgCO_3(s)$ $Ca^{+2} + SO_4^{-2} + 2H_2O \Longleftrightarrow CaSO_4 \cdot 2H_2O(s)$	$Mg^{+2} + 0.5Na^{+} - CO_3^{+2}$ $Mg^{+2} + SO_4^{-2} - Ca^{+2}$	77.16	22.74	0.08

contribution on the objective function may be used, instead of EMMA, to identity reacting species. However, the contribution is related to the mass that reactions added stoichiometrically to each species (displayed in Table 2). In our case Ca^{2+} and SO_4^{2-} contribute with almost 90% of the objective function while Mg^{2+} and CO_3^{2-} contribute with less than 8%. The contribution on the objective function for Mg^{2+} or CO_3^{2-} is larger than that of conservative species (Cl^{-}, Na^{+} and K^{+}). In A-Cons the main contributors to the objective function are the decoupled chemical components. It is important to highlight that these decoupled components are a linear

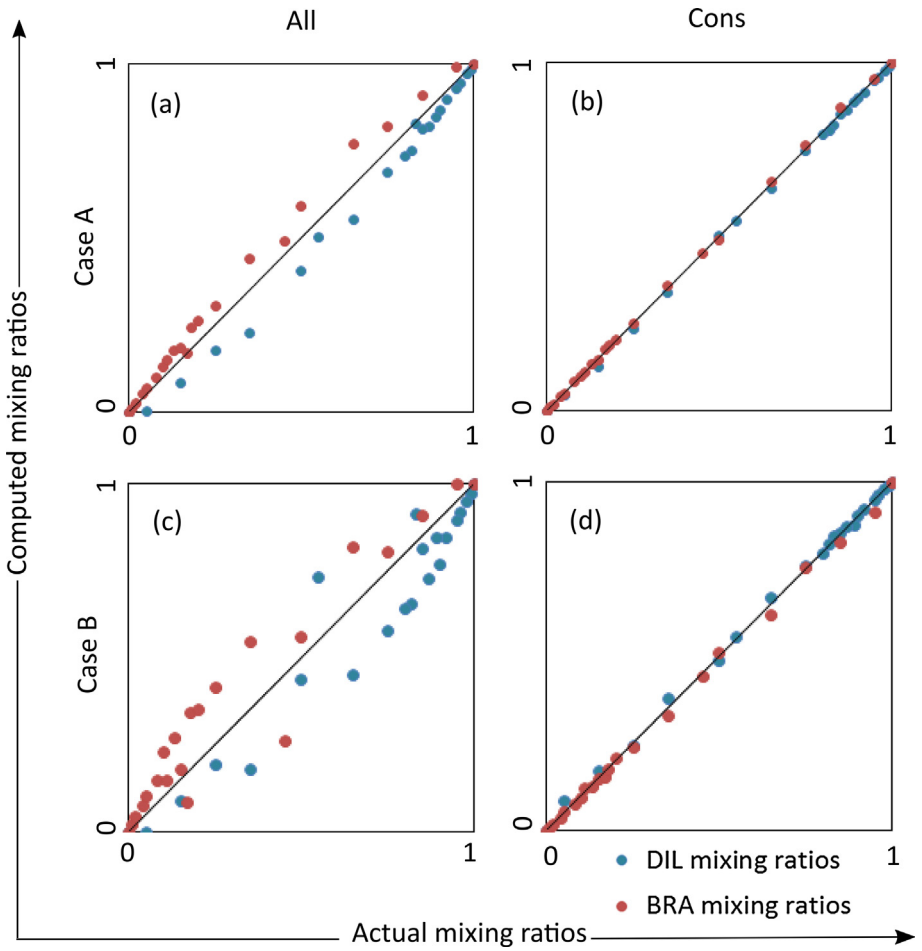

Fig. 7. Comparison between true and computed mixing ratios. (a) Case A using all data (A-all), and (b) using conservative components (A-cons). Graphs (c) and (d) display the same results of Case B. Note that the best estimations of DIL and BRA mixing ratios are obtained with the use of conservative components in both cases.

Table 4
MIX results obtained from processing the data of cases A and B in different analyses: Root-mean-square error (RMSE) of the mixing ratios in all the samples, values of the objective function (Eq. (10)) and percentages of the contributions of each species to the objective function.

Case	Analysis	RMSE (%)	Objective Function	Contribution to the Objective Function%									
				Cl^-	Na^+	SO_4^{2-}	Ca^{2+}	Mg^{2+}	K^+	CO_3^{2-}	u_{Gyp}	u_{Mgs}	$u_{gyp\text{-}Cat.Ex}$
A	A-All	5.67	−10.58	1.06	0.96	43.61	46.24	4.72	0.41	3.00	–	–	–
	A-cons	0.79	−0.04	13.18	14.14	–	–	–	24.22	–	22.66	25.80	–
	A-Par	1.07	−0.83	2.37	2.08	–	–	62.78	2.10	29.49	1.18	–	–
	A-Csp	1.75	−0.02	17.54	35.39	–	–	–	47.07	–	–	–	–
B	B-All	11.19	−12.73	9.56	8.44	31.75	38.85	5.67	2.54	3.19	–	–	–
	B-cons	1.12	−0.03	10.32	–	–	–	–	39.40	–	–	28.20	22.08
	B-Par	1.42	−0.79	4.21	–	–	–	61.72	1.73	31.11	–	–	1.24
	B-Csp	1.10	−0.01	21.18	–	–	–	–	78.82	–	–	–	–

combination of the reactive species and besides their high contribution to the objective function, its global value has decreased.

It is clear that the nice results of the mixing calculations reflect the synthetic nature of the example. In a real case, modellers might have neglected $MgCO_3$ (a relatively rare mineral) dissolution or might have preferred to work only with conservative species (Cl^-, Na^+ and K^+ in this example). To assess this alternative approach we performed two additional mixing calculations: A-Par and A-Csp. The first one includes only gypsum dissolution (i.e., it was performed with only one conservative component $\mathbf{u_{Gyp}}$), and the second one considering only the conservative species (Cl^-, Na^+ and K^+). Results are summarized in Table 4. The values of the root-mean-square error (RMSE) are highest for the A-All (5.67%). If mixing calculation are computed considering only gypsum dissolution (A-Par) the error decreases (RMSE 1.07%). When using only conservative species (A-Csp), the error decreases (1.75%) compared with A-All, but is higher than the obtained with A-Cons (0.79%). The results confirm that the best approach (best estimation of mixing ratios) in terms of minimum error is obtained when using both, $\mathbf{u_{Gyp}}$ and $\mathbf{u_{Mgs}}$ conservative components.

Case B: gypsum ($CaSO_4{\cdot}2H_2O$), magnesite ($MgCO_3$) dissolutions, and cation exchange Na/Ca

Computed and measured mixing ratios for B-All and B-Cons are plotted in Fig. 7c and d. Again, the differences between real and computed mixing ratios are smallest for the calculation with conservative components.

The analysis of objective function (Table 4) leads to similar conclusions as the previous analysis. The global objective function decreases from −12.73 to −0.03 with the use of decoupled conservative components. The main contributors to the objective function in B-All are the species affected by chemical reactions.

In the same way than Case A, we repeated the calculations using only conservative species (B-Csp), and assuming that magnesite dissolution was not identified (B-Par). Contrary to Case A, using B-Csp yields results that are comparable to those of B-Cons. We attribute this result to overfitting, because the RMSE obtained in B-Csp (Cl^- and K^+) should have been larger than the obtained in A-Csp (Cl^-, Na^+ and K^+) where three species were available. Still, the result points to the fact that there is always a random component in any mixing calculation (in fact, other realizations of the random error yield better results for the proposed approach), and that performing computations with conservative species when they are available should complement the proposed approach.

5. Conclusions

We presented a methodology devoted to the identification of end-members and chemical reactions by means of EMMA, and to improve the performance of mixing ratio calculations by defining decoupled conservative components based on the assumed reactions. We applied and validated this methodology with two synthetic datasets generated by imposing different chemical reactions using a geochemical calculation tool (PHREEQC). Results indicate that this methodology can be applied to: (1) identify chemical reactions, (2) improve the identification of the end-members that are mixing in the system, and (3) enhance mixing ratio calculations defining decoupled conservative components as linear combination of species consistent with the identified set of reactions. Regarding the identification of the chemical reactions, the proposed methodology presents some limitations. Chemical reactions could be clearly identified and related to an eigenvector in Case A, but the identification was not so evident in Case B, which probably reflects the presence of reactions that affect the same species (common ion effect). Nevertheless this methodology can clearly identify the involved chemical species in the reactions. Regarding the identification of end members, the approach improves with the use of conservative components in both cases, which reflects that conservative components eliminate the dispersion generated by the reactions. Regarding mixing calculations the use of conservative components decreases the differences between computed and measured mixing ratios and lead to reductions of the global objective function in both cases. While the approach appears powerful, it should be viewed a complementary tool to traditional geochemical interpretations. Specifically, the likelihood of the reactions should be ascertained by the saturation index calculations and the actual presence of the reacting minerals. Furthermore, the modeller should also verify independently the hydrological meaningfulness of the identified end-members.

Acknowledgements

Funding for this research was provided by the Authority of Matanza-Riachuelo Basin (ACUMAR) grant held by L. S. Vives. Also we are grateful for financial support from CONICET (National Council of Scientific and Technical Research).

References

Appelo, C.A.J., Postma, D., 2005. Geochemistry, Groundwater and Pollution. CRC Press.

Ball, J.W., Nordstrom, D.K., 1991. User's manual for WATEQ4F, with revised thermodynamic data base and test cases for calculating speciation of major, trace, and redox elements in natural waters.

Barthold, F.K., Tyralla, C., Schneider, K., Vaché, K.B., Frede, H.G., Breuer, L., 2011. How many tracers do we need for end member mixing analysis (EMMA). A sensitivity analysis. Water Resour. Res. 47 (8).

Bea, S.A., Carrera, J., Ayora, C., Batlle, F., 2010. Modeling of concentrated aqueous solutions: Efficient implementation of Pitzer equations in geochemical and reactive transport models. Comput. Geosci. 36 (4), 526–538.

Bea, S.A., Carrera, J., Ayora, C., Batlle, F., Saaltink, M.W., 2009. CHEPROO: a Fortran 90 object-oriented module to solve chemical processes in Earth Science models. Comput. Geosci. 35 (6), 1098–1112.

Burns, D.A., McDonnell, J.J., Hooper, R.P., Peters, N.E., Freer, J.E., Kendall, C., Beven, K., 2001. Quantifying contributions to storm runoff through end-member mixing analysis and hydrologic measurements at the Panola Mountain Research Watershed (Georgia, USA). Hydrol. Process. 15 (10), 1903–1924.

Carrera, J., Vázquez-Suñé, E., Castillo, O., Sánchez-Vila, X., 2004. A methodology to compute mixing ratios with uncertain end-members. Water Resour. Res. 40 (12). http://dx.doi.org/10.1029/2003WR002263.

Christophersen, N., Hooper, R.P., 1992. Multivariate analysis of stream water chemical data: The use of principal components analysis for the end-member mixing problem. Water Resour. Res. 28 (1), 99–107.

Christophersen, N., Hooper, R.P., 1993. Modelling the hydrochemistry of catchments: a challenge for the scientific method. J. Hydrol. 162 (1–4), 1–12.

Christophersen, N., Neal, C., Hooper, R.P., Vogt, R.D., Andersen, S., 1990. Modelling streamwater chemistry as a mixture of soilwater end-members – a step towards second-generation acidification models. J. Hydrol. 116, 307–320.

De Simoni, M., Carrera, J., Sanchez-Vila, X., Guadagnini, A., 2005. A procedure for the solution of multicomponent reactive transport problems. Water Resour. Res. 41, W11410.

De Simoni, M., Sanchez-Vila, X., Carrera, J., Saaltink, M.W., 2007. A mixing ratios-based formulation for multicomponent reactive transport. Water Resour. Res. 43 (7).

Gómez, J.B., Auqué, L.F., Gimeno, M.J., 2008. Sensitivity and uncertainty analysis of mixing and mass balance calculations with standard and PCA-based geochemical codes. Appl. Geochem. 23 (7), 1941–1956.

Gómez, J.B., Gimeno, M.J., Auqué, L.F., Acero, P., 2014. Characterisation and modelling of mixing processes in groundwaters of a potential geological repository for nuclear wastes in crystalline rocks of Sweden. Sci. Total Environ. 468, 791–803.

Hooper, R.P., Christophersen, N., Peters, N.E., 1990. Modelling streamwater chemistry as a mixture of soilwater end-members — an application to the Panola Mountain catchment, Georgia, U.S.A. J. Hydrol. 116 (1–4), 321–343. http://dx.doi.org/10.1016/0022-1694(90)90131-G.

Hooper, R.P., 2003. Diagnostic tools for mixing models of stream water chemistry. Water Resour. Res. 39 (3).

James, A.L., Roulet, N.T., 2006. Investigating the applicability of end-member mixing analysis (EMMA) across scale: a study of eight small, nested catchments in a temperate forested watershed. Water Resour. Res. 42, W08434. http://dx.doi.org/10.1029/2005WR004419.

Jiménez-Martínez, J., Aravena, R., Candela, L., 2011. The role of leaky boreholes in the contamination of a regional confined aquifer. A case study: the Campo de Cartagena region, Spain. Water Air Soil Pollut. 215 (1–4), 311–327.

Jurado, A., Vázquez-Suñé, E., Carrera, J., Tubau, I., Pujades, E., 2015. Quantifying chemical reactions by using mixing analysis. Sci. Total Environ. 502, 448–456.

Katsuyama, M., Ohte, N., Kobashi, S., 2001. A three-component end-member analysis of streamwater hydrochemistry in a small Japanese forested headwater catchment. Hydrol. Process. 15 (2), 249–260.

Krautle, S., Knabner, P., 2005. A new numerical reduction scheme for fully coupled multicomponent transport-reaction problems in porous media. Water Resour. Res. 41, W09414.

Krautle, S., Knabner, P., 2007. A reduction scheme for coupled multicomponent transport-reaction problems in porous media: Generalization to problems with heterogeneous equilibrium reactions. Water Resour. Res. 43, W03429.

Kirkner, D.J., Reeves, H., 1988. Multicomponent mass transport with homogeneous and heterogeneous chemical reactions: effect of the chemistry on the choice of numerical algorithm: 1. Theory. Water Resources Research 24 (10), 1719–1729.

Laaksoharju, M., Gascoyne, M., Gurban, I., 2008. Understanding groundwater chemistry using mixing models. Appl. Geochem. 23 (7), 1921–1940.

Long, A.J., Valder, J.F., 2011. Multivariate analyses with end-member mixing to characterize groundwater flow: wind cave and associated aquifers. J. Hydrol. 409 (1), 315–327.

Mazor, E., Kaufman, A., Carmi, I., 1973. Hammat Gader (Israel): geochemistry of a mixed thermal spring complex. J. Hydrol. 18 (3–4), 289–303.

Mazor, E., Vuataz, F.D., Jaffé, F.C., 1985. Tracing groundwater components by chemical, isotopic and physical parameters—example: Schinznach, Switzerland. Journal of Hydrology 76 (3), 233–246.

Mazor, E., 1990. Applied chemical and isotopic groundwater hydrology.

Menció, A., Galán, M., Boix, D., Mas-Pla, J., 2014. Analysis of stream-aquifer relationships: a comparison between mass balance and Darcy's law approaches. J. Hydrol. 517, 157–172.

Molins, S., Carrera, J., Ayora, C., Saaltink, M.W., 2004. A formulation for decoupling components in reactive transport problems. Water Resour. Res. 40 (10).

Parkhurst, D.L., 1997. Geochemical mole-balance modeling with uncertain data. Water Resour. Res. 33 (8), 1957–1970.

Parkhurst, D.L., Appelo C.A.J., 2013. PHREEQC (Version 3.0.4) – a computer program for speciation, batch speciation, one-dimensional transport, and inverse geochemical calculations U.S. Geological Survey Techniques and Methods, Book 6, Chapter A43 497 p., http://pubs.usgs.gov/tm/06/a43/.

Plummer, L.N., Prestemon, E.C., Parkhurst, D.L., 1991. An interactive code (NETPATH) for modeling net geochemical reactions along a flow path.
Rubin, J., 1983. Transport of reacting solutes in porous media: relation between mathematical nature of problem formulation and chemical nature of reactions. Water Resour. Res. 19 (5), 1231–1252.
Rueedi, J., Purtschert, R., Beyerle, U., Alberich, C., Kipfer, R., 2005. Estimating groundwater mixing ratios and their uncertainties using a statistical multi parameter approach. J. Hydrol. 305, 1–14.
Saaltink, M.W., Ayora, C., Carrera, J., 1998. A mathematical formulation for reactive transport that eliminates mineral concentrations. Water Resour. Res. 34 (7), 1649–1656.
Saaltink, M.W., Carrera, J., Ayora, C., 2001. On the behavior of approaches to simulate reactive transport. J. Contam. Hydrol. 48 (3), 213–235.
Soulsby, C., Petry, J., Brewer, M.J., Dunn, S.M., Ott, B., Malcolm, I.A., 2003. Identifying and assessing uncertainty in hydrological pathways: a novel approach to end member mixing in a Scottish agricultural catchment. J. Hydrol. 274 (1), 109–128.
Steefel, C.I., MacQuarrie, K.T., 1996. Approaches to modeling of reactive transport in porous media. Rev. Mineral. Geochem. 34 (1), 85–129.
Steefel, C.I., Lasaga, A.C., 1994. A coupled model for transport of multiple chemical species and kinetic precipitation/dissolution reactions with application to reactive flow in single phase hydrothermal systems. Am. J. Sci. 294 (5), 529–592.
Suk, H., Lee, K.K., 1999. Characterization of a ground water hydrochemical system through multivariate analysis: clustering into ground water zones. Groundwater 37 (3), 358–366.
Tolosana-Delgado, R.N., Otero, N., Pawlowsky-Glahn, V., Soler, A., 2005. Latent compositional factors in the Llobregat River Basin (Spain) Hydrochemistry. Math. Geol. 37, 683–706.
Valder, J.F., Long, A.J., Davis, A.D., Kenner, S.J., 2012. Multivariate statistical approach to estimate mixing proportions for unknown end members. J. Hydrol. 460, 65–76.
Vázquez-Suñé, E., Carrera, J., Tubau, I., Sánchez-Vila, X., Soler, A., 2010. An approach to identify urban groundwater recharge. Hydrol. Earth Syst. Sci. 14 (2085–2097), 2010. http://dx.doi.org/10.5194/hess-14-2085-2010.
Tubau, I., Vàzquez-Suñé, E., Jurado, A., Carrera, J., 2014. Using EMMA and MIX analysis to assess mixing ratios and to identify hydrochemical reactions in groundwater. Sci. Total Environ. 470–471, 1120–1131. http://dx.doi.org/10.1016/j.scitotenv.20130.121.
Yeh, G.T., Tripathi, V.S., 1989. A critical evaluation of recent developments in hydrogeochemical transport models of reactive multichemical components. Water Resour. Res. 25 (1), 93–108.
Zhang, G., Zheng, Z., Wan, J., 2005. Modeling reactive geochemical transport of concentrated aqueous solutions. Water Resour. Res. 41 (2).