Technical Report

TR-06-27

M3 version 3.0: Concepts, methods, and mathematical formulation

Javier B. Gómez, Department of Earth Sciences,
University of Zaragoza

Marcus Laaksoharju, Geopoint AB

Erik Skårman, Abscondo

Ioana Gurban, 3D-Terra

July 2006

Svensk Kärnbränslehantering AB
Swedish Nuclear Fuel
and Waste Management Co
Box 5864
SE-102 40 Stockholm Sweden
Tel 08-459 84 00
+46 8 459 84 00
Fax 08-661 57 19
+46 8 661 57 19

M3 version 3.0: Concepts, methods, and mathematical formulation

Javier B. Gómez, Department of Earth Sciences,
University of Zaragoza

Marcus Laaksoharju, Geopoint AB

Erik Skårman, Abscondo

Ioana Gurban, 3D-Terra

July 2006

Abstract

Hydrochemical evaluation is a complex type of work, carried out by specialists. The outcome of this work is generally presented as qualitative models and process descriptions of a site. To support and help quantify the processes in an objective way, a multivariate mathematical tool named M3 (Multivariate Mixing and Mass balance calculations) has been constructed. The computer code can be used to trace the origin of the groundwater and calculate the mixing portions and mass balances even from ambiguous groundwater data. The groundwater composition used traditionally to describe the reactions taking place in the bedrock can now be used to trace the present and past groundwater flow with increased accuracy.

The M3 code is a groundwater response model, which means that the changes in the groundwater chemistry in terms of sources and sinks are traced in relation to an ideal mixing model. The complexity of the measured groundwater data determines the configuration of the ideal mixing model. Deviations or similarities with the ideal mixing model are interpreted as being due to mixing or reactions. Assumptions concerning important mineral phases altering the groundwater or uncertainties associated with thermodynamic constants do not affect the modelling because the calculations are solely based on the measured groundwater composition.

M3 uses the opposite approach to that of many standard hydrochemical models. In M3 mixing is evaluated and calculated first. The constituents that cannot be described by mixing are described by reactions. The M3 model consists of three steps: the first is a standard principal component analysis, followed by mixing and finally mass balance calculations. The measured groundwater composition can be described in terms of mixing portions in % and the sink/sources of an element associated with reactions are reported in mg/l.

Summary

This report describes version 3.0 of the code M3 (**M**ultivariate **M**ixing and **M**ass balance calculations). This method and computer code were developed to trace the mixing and reaction processes in the groundwater. The aim of the M3 concept is to decode the complex information often hidden in the groundwater analytical data.

The present report (hereafter Report 1) will focus on the concepts, methods and mathematical formulation behind the M3 method. Two accompanying reports cover other aspects:

- Verification and Validation, /Gómez et al. 2006/ (hereafter Report 2) gathers a collection of validation and verification exercises, designed to test each part of M3 code and to build confidence in its methodology.
- User's Guide, /Laaksoharju et al. 2006/ (hereafter Report 3) includes detailed reference to the M3 program's capabilities, installation procedures and all functions and operations that the program can perform. It also describes sample cases of how the program is used to analyse a test data set.

The M3 method has been tested and modified over several years. The development work has been supported by the Swedish Nuclear Fuel and Waste Management Company (SKB). The main test site for the model was the underground Äspö Hard Rock Laboratory (HRL), but has been also extensively used in the Swedish sites of Forsmark, Simpevarp and Laxemar, and also in Canada, Jordan, Gabon and Finland.

The groundwater composition is a result of mixing processes and water-rock interaction. Standard groundwater models based on thermodynamic laws may not be applicable in a normal temperature groundwater system where equilibrium with many of the bedrock minerals is not reached and where biological processes seem to play a central role in the groundwater altering process. The major purpose of standard groundwater chemical codes is to describe the measured groundwater composition in terms of reactions. The constituents that cannot be described by reactions are described by mixing, possibly using a conservative tracer. The M3 model uses an opposite approach compared to the standard method. In M3 the mixing processes are evaluated and calculated first. This is possible due to the use of a multivariate technique (principal component analysis) to construct an ideal mixing model of a site. The constituents that cannot be described by mixing are described by reactions.

The M3 model consists of three steps; the first is a standard Principal Component Analysis (PCA), followed by mixing and finally mass balance calculations. In order to take into consideration as many relevant elements as possible, PCA is used to summarise and simplify the groundwater information. The M3 model compares the measured groundwater composition of each sample with known borehole-sampled waters or hypothetical extreme waters, called here simply *end-members*. All the measured groundwater samples at a site are compared to these end-members. The mixing calculations (i.e. mixing portions as a percentage of each selected end-member) determine how much of the observed groundwater composition is due to mixing from the selected end-member. The mass balance calculations (reported in terms of sinks/sources of groundwater constituents in mg/l or moles) determine how much of the measured groundwater constituents is the result of water-rock interaction. Since the calculations are relative to the selected end-members, modelling can only describe changes in terms of mixing and reactions taking place between the end-members. The results can be used to describe the groundwater characteristics both mathematically and quantitatively, rather than qualitatively as is often the case in a site description of the groundwater chemistry.

The M3 method can be used for tracing groundwater evolution, past-present dispersion and water-rock interactions. The outcome from the modelling can be reported in non-hydrochemical terms such as changes in mixing portions or in gain/losses due to mass balance reactions. The comparison with hydrogeological models is easier since the results from the M3 model can be compared to the results from the hydrodynamic models. The effect on the groundwater composition observed from biological reactions such as organic decomposition or sulphate reduction can be traced. In groundwater chemical response modelling such as M3 information concerning fracture mineralogy, thermodynamic data bases or groundwater flow directions are not included. The model concentrates solely on tracing changes in the measured groundwater composition which can be interpreted as a result of mixing and reactions.

A comparison of the M3 model's accuracy using two frequently used statistical tests, univariate and linear regression, shows that the prediction error (defined as the difference between the predicted concentration and the actual concentration, divided by the actual concentration) of the M3 calculations for water conservative constituents such as Cl and $\delta^{18}O$ is on average three times lower for the data from the Äspö site /Laaksoharju et al. 1999bc/. As with any groundwater models, the validity of the results must be examined carefully using expert knowledge in the areas of hydrochemistry, hydrogeology and alternative modelling.

The M3 computer program is a standalone program developed in the MATLAB 7.1 computation environment /MATLAB 2005/. The M3 toolbox calculates and displays the results both as graphs and numerical data. In addition, it offers the user several ways of examining and interpreting data.

Included with M3 is an online version of the reference manual, as well as Matlab libraries required to run the program. The M3 program has been tested on Windows 2000 and XP.

Contents

1 What is M3?

M3 /Laaksoharju et al. 1999c/ is a methodology and associated computer code to trace the mixing and reaction processes in a surface water-groundwater system. It was constructed to explain how and why we obtain the measured groundwater composition. The M3 method consists of 3 steps. The first is a standard principal component analysis, followed by mixing and finally by mass balance calculations.

1. A standard multivariate technique, called *Principal Component Analysis* (PCA) is used for cluster analyses of the data by using major elements (e.g. Cl, Ca, Na, Mg, K, SO_4, HCO_3, etc) in combination with stable isotopes (^2H, ^{18}O) and radiogenic isotopes (^3H). The chosen variables are generally relatively easy to analyse with a high analytical accuracy and they are known to describe the major groundwater features. Generally, depending on the analytical program, there may be other variables than those suggested which can also contain important information concerning the evolution of the groundwater at a site. Such variables should then be included in the PCA. Other cases, where isotopes are not analysed, may cause lower resolution in M3 modelling. The PCA aims to describe as much of the information as possible from the variables in the first principal component.

   The rest of the information is described by the second , third, etc principal components. Principal components are linear combinations of the original variables used to summarise most of the information in the data. The weights of the different variables in the equations are calculated automatically by the PCA. For the first two principal components an *x-y* scatter plot can be drawn (Figure 1-1). There, *x* is the equation for the first principal component and *y* is the equation for the second principal component. The plot is called the *PC-plot* and is used to visualise the clustering of the data as well as to identify end-members. Careful examination of the PC-plot can provide additional information as to the origin, complexity and relationship of the groundwater system. An *end-member* can in practice be any water; however, a real groundwater (sampled in a borehole) which resembles an assumed or modelled end-member is usually selected (Figure 1-1a and b). The number of selected end-members should be the minimum required to describe the observations.

2. **Mixing calculations** are used to calculate the mixing contribution of the end-members to the observed groundwater composition. The calculated mixing portion can be used to evaluate the mixing situation, the origin of the groundwater and possible flow paths (Figure 1-1c).

3. **Mass balance calculations (deviation calculations)** are used to define the sources and sinks for different elements which deviate from the ideal mixing model based on the mixing of end-members (Figure 1-1d). No deviation from the measured value indicates that mixing can explain the element behaviour, whereas source or sink is due to net reactions. The evolution caused by water-rock interactions can thus be described.

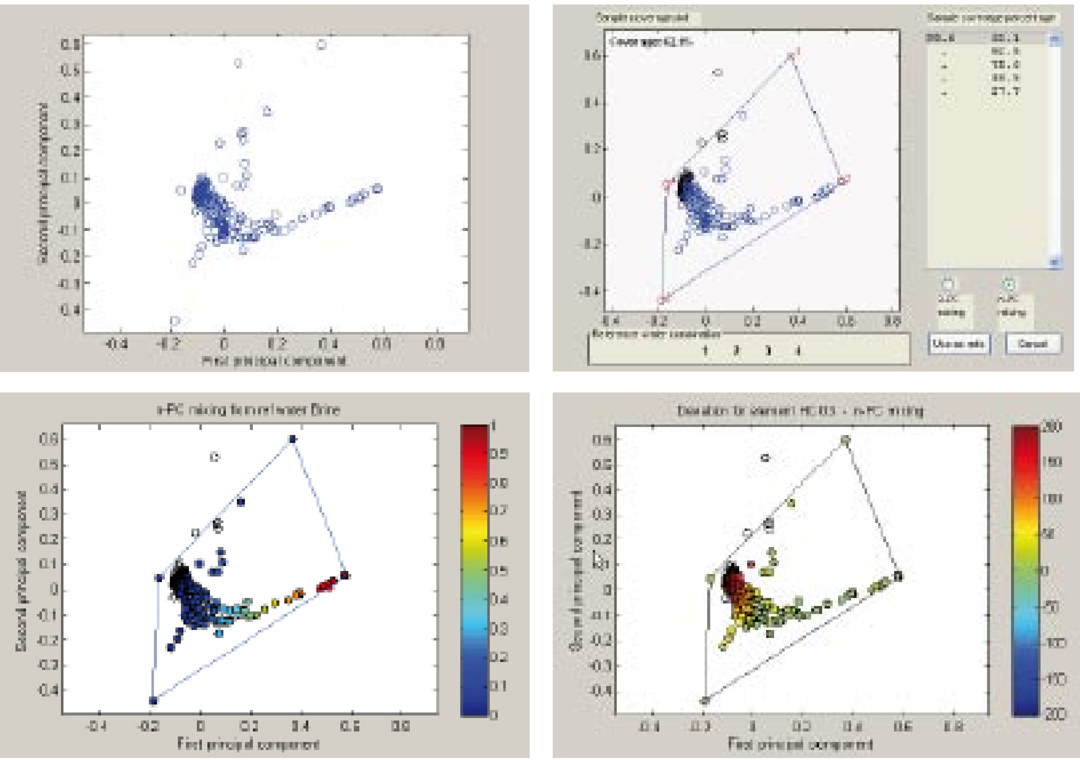


Figure 1-1. *Different steps in the M3 modelling; a) principal component analysis is used to obtain the maximum resolution of the data set; b) selection of end-members: the other groundwaters are compared to these; black dots are samples outside the mixing polyhedron and thus not explained by pure mixing of the selected end-members; c) mixing calculations: the distance in the PC space to the end-members e.g. the portions of brine end-member are shown in the figure, using the range from 0 (no brine end-member) to 1 (pure brine end-member); d) mass balance calculations: the sources and sinks of e.g. bicarbonate (HCO_3^-, mg/L) are shown which cannot be accounted for by mixing.*

1.1 What M3 can do

M3 is able to:

- Perform a principal component analysis on a dataset (each row a sample, each column a compositional variable) (Section 3.1). The benefits of using M3 increase when the size of the dataset increases.
- Plot the results in two and three dimensions using the first two or first three principal components as coordinate axes, respectively (PC-plots).
- Perform a mixing calculation using two or all the principal components once a set of end-members have been defined. This mixing calculation step gives the percentage of each end-member in every sample in the dataset (Section 3.2).
- Colour-code each sample in a PC-plot with respect to the contribution of each end-member to the mixing.
- Calculate, for each sample in the dataset, a “theoretical” composition from the above mixing proportions. This “theoretical” composition is that which the sample would have if mixing were the only process affecting its composition *and* if the end-members had been properly selected (for the meaning of “properly”, see Section 4.1).
- Perform a mass balance calculation by subtracting the real composition of a sample from the computed (“theoretical”) composition. Because the computed composition reflects pure mixing, the difference between the real and the computed compositions is interpreted by M3 as a mass balance, i.e. elements that have been enriched or depleted with respect to pure mixing (Section 3.3).

- Via the End-member Selection Module (Section 4.1), select the best combination of end-members for a given dataset. This selection procedure is semi-automatic but the user has first to produce a number of end-members (this number could be high) that are believed to have been involved in the genesis of every water sample in the dataset.
- Via the End-member Variability Module (Section 4.2), give mixing proportion ranges for a given sample from ranges in the chemical composition of the selected end-members. The ranges in the chemical composition of the end-members are supplied by the user.

1.2 What M3 can not do

M3 is not able to:

- Interpret in terms of reactions the mass balance calculation step. This is left to the user after taking into account a wide range of background information on the system being studied.
- Separate the effect of a chemical reaction from a pure mixing if the final result of both is identical. In other words, if a chemical reaction has the same effect on the composition of a water sample (considering only the compositional variables that have been used in M3) as a pure mixture of specific end-members, M3 would say that the mass balance is zero and therefore, that the sample can be "constructed" only by the mixing of the selected end-members. It is up to the users to assess the likelihood of one or the other outcome (again, after a careful geochemical and hydrological study of the system).
- Guarantee that the selected set of end-members is the best set in *absolute* terms. It can only assure that it is the best out of all the possibilities covered by the extended set of end-members included in the End-member Selection Module.
- Perform reaction path calculations as many geochemical codes like PHREEQC /Parkhurst and Appelo 1999/ or WATEQ4F /Ball and Nordstrom 2001/.
- Perform reactive transport calculations.

1.3 Alternatives to M3

M3 uses an specific methodology to compute mixing proportions. Nevertheless, other approaches has been proposed. Here we are interested in those already applied to mixing in groundwater systems. The purpose of this section is to let the potential M3 user know that alternative approaches to mixing can be tried with the same dataset in order to get a better understanding of the uncertainties involved in the calculations.

The chemometric algorithm of Cave and collaborators. This approach to computing mixing proportions /Cave and Harmon 1997, Cave and Wragg 1997, Dershowitz et al. 2000/ makes no initial assumptions about the nature of the end-members present and considers all the contributions to chemical variability in the groundwaters. In other words, end-members are not imposed from the beginning, but are computed along the way. This is the main difference with the M3 methodology and should be taken into consideration when end-members are not evident. It has been used in Task Force 5 at Äspö URL /Dershowitz et al. 2000/ where its results are compared to M3's.

The basic approach is illustrated in Figures 1-2 and 1-3. Here matrix **A** is the supplied groundwater data matrix (each column an input compositional variable, each row a different sample), and matrices **B** and **C** need to be found. The process for finding matrices **B** and **C** is carried out in four stages /Dershowitz et al. 2000/:

1. PCA and eigenvalue analysis are initially used in the same way as in M3.
2. The varimax-rotated loadings matrix /Reyment and Jvreskog 1996, Haan 2002/ from the PCA of matrix **A**, containing the initial groundwater compositions, is used to produce a first approximation of matrix **B**, which contains the mixing proportions.

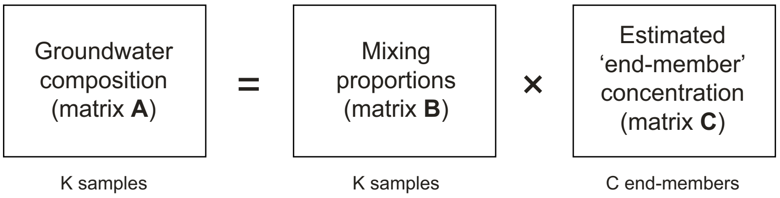


***Figure 1-2**. Relationships between matrices in the chemometric approach of Cave and collaborators /Cave and Harmon 1997, Cave and Wragg 1997, Dershowitz et al. 2000/.*

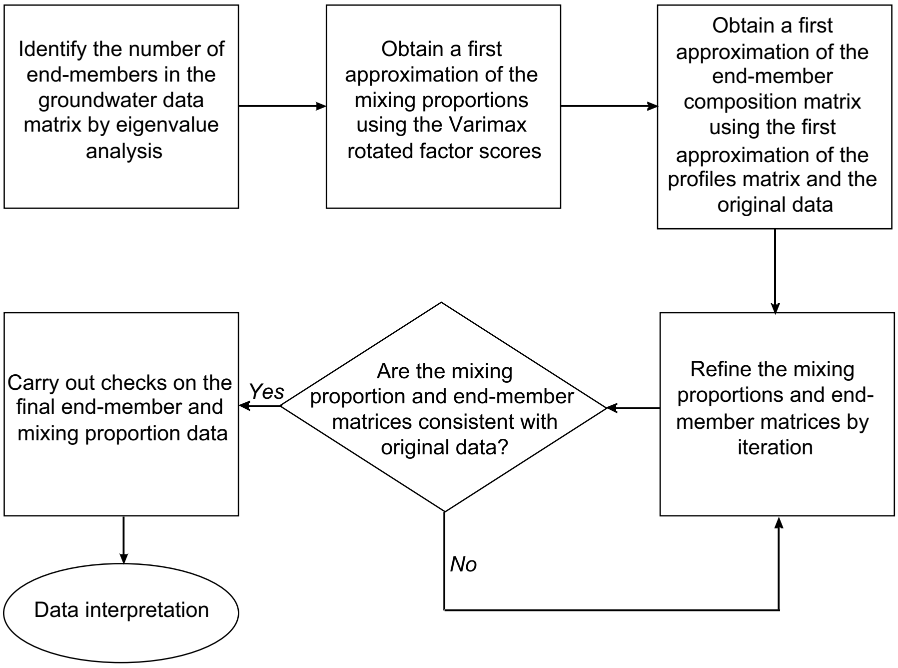


***Figure 1-3**. Summary of the procedure to compute mixing proportions under the chemometric approach.*

3. The pseudoinverse method for non-square matrices (Singular Value Decomposition, SVD, see footnote 3 in Section 3.1.2) is then applied to matrices **A** and **B** to produce a first approximation of matrix **C**, which contains chemical components that contribute to the chemical variability in the groundwaters, some of which should correspond approximately to end-members.
4. Matrices **B** and **C** are refined iteratively using the pseudoinverse method until both matrices contained estimates of mixing proportions and chemical components compositions that are consistent with the groundwater compositions of the original matrix **A**.

It is important to note that the chemical components obtained from the chemometric algorithm are not principal components (they have been rotated to make loadings simpler, see Section 3.1.2), but are derived from them.

The multi-parameter approach of /Rueedi et al. 2005/. /Rueedi et al. 2005/ have set up a method to compute mixing ratios based on the simultaneous use of several conservative parameters and applying a general mixing theory /e.g. Faure 2004/. No PCA is carried out and all non-conservative compositional parameters are excluded.

This approach has its weak point in the selection of end-members. As /Rueedi et al. 2005/ put it "this step always involves arbitrariness, which is why it is the least reproducible step of the presented approach. Therefore, the selection has to rely on a well-defined decision basis".

The approach to estimate mixing ratios and their uncertainties is based on the following procedure /Rueedi et al. 2005/:

1. Analysis of available data in terms of data quality, measurement techniques and precisions.
2. Exclusion of non-conservative parameters.
3. Selection of samples representing the end-members based on all available information (geology, hydrogeology and spatial distribution of hydrochemistry).
4. Estimation of end-member parameters and their uncertainties at each location.
5. Correction for other sources influencing the chemical and/or isotopic groundwater signature.
6. Calculation of the overall mixing ratios of a sample using the calculated mixing ratios of each groundwater parameter.

The method is applicable, in principle, for any number of end-members, but the mathematics could get quite involved when more than three end-members are used.

The method of /Carrera et al. 2004/ to compute mixing ratios with uncertain end-members. Most methods available for computing mixing ratios are based on assuming that end-member concentrations are perfectly known, which is rarely the case. Often, end-members cannot be sampled, and their concentrations vary in time and space. Still, much information about them is contained in the mixtures.

To take advantage of this information, /Carrera et al. 2004/ have developed a maximum likelihood method to estimate mixing ratios, while acknowledging uncertainty in end-member concentrations. Maximizing the likelihood of concentration measurements with respect to both mixing ratios and end-member concentrations leads to a general constrained optimisation problem /Carrera and Neuman 1986/. The method has been compared to more traditional methods like least squares and linear mixing, outperforming both of them. The method also yields improved estimates of end-member concentrations, thus enlarging the potential of mixing calculations.

The proposed algorithm consists of the following four steps /Carrera et al. 2004, paragraph [13]/:

1. Initialisation, consisting in the definition of initial mixing ratios by conventional least squares, *assuming that the composition of the end-members is fully known* (zero uncertainty).
2. Given the initial mixing ratios, maximise the log-likelihood function to estimate, *at the same time,* the expected values of the composition (i.e. concentrations) of the mixed waters *and* of the end-members.
3. Given the expected values of mixed water and end-member concentrations, maximise the log-likelihood to obtain the mixing ratios.
4. Repeat Steps 2 and 3 until convergence.

The log-likelihood function to be maximise is:

$$\ln L = \sum_{s=1}^{ns} \left[-\frac{1}{2} \left(\mathbf{z}_s - \boldsymbol{\mu}_s \right)^t \mathbf{A}_s^{-1} \left(\mathbf{z}_s - \boldsymbol{\mu}_s \right) \right],$$

where $\mathbf{z}_s$ are vectors containing the concentration of all species in *both* the samples and the end-members (there are *ns* such vectors, the sum of the number of samples plus the number of end-members), μ_s are vectors with the expected values of $\mathbf{z}_s$ and $\mathbf{A}^{-1}$ is the inverse of the covariance matrix. Of course, μ_s and $\mathbf{A}$ are unknown, and the procedure consists of finding those μ_s and $\mathbf{A}_s$ that maximise the above function. The resulting non-linear system of equations is solved iteratively by a Newton-Raphson method.

The authors test the method with two synthetic data sets, one consisting of two end-members and two species (species are the chemical variables used to trace mixing, like chlorine or deuterium), and the other consisting of three end-members and five species.

To simulate the uncertainty in the concentration of end-members, and to check the incidence of this uncertainty in the final mixing proportions, deviations from the “true” concentrations are introduced in the form of a “noise” term. Two level of noise are tested, a low variance and a high variance one.

The results are evaluated in terms of an “improvement index” for end-member concentrations, which evaluates the reduction in mean square error of end-member concentrations during estimation (an improvement index of 2 means that the error has been reduced by a factor of two). In the three-end-member, five-species case, improvement indices of 3 (in the case of only four samples) to 8 (in the case of 100 samples) have been obtained.

These results nicely show that redundancy in both the number of species and the number of samples greatly improve the quality of the estimations (this is the basic idea behind the M3 approach). This methodology has been applied to a real groundwater system by /Vázquez-Suñé et al. 1997/.

2 Historical development of M3

As already stated, the underlying hypothesis behind M3 is that *the origin and evolution of the groundwater can be described if the effects of mixing and reactions can be examined separately*. Starting with the mixing and letting those constituents that cannot be explained by mixing be described by reactions, rather than approaching reactions first and allowing mixing to explain the constituents, makes modelling simpler since less prerequisite information is needed. A simple two component mixing mass balance model was described by /Laaksoharju 1990, Smellie and Laaksoharju 1992/ and /Banwart et al. 1992/. The main aim of the model was to differentiate in the groundwater between what is due to mixing and what is due to water-rock reactions by using one variable e.g. Cl as a conservative tracer. A basic feature of this method is the identification of the original source water types that all others are mixed from. These original types of waters are called end-members. However, the model seemed to have limitations when tracing a more complex system involving several end-members. The real dynamics of the system may thus remain undetected and the processes behind the measured groundwater composition could be misinterpreted. Many variables are important for the understanding of the transport, mixing processes and reactions taking place in the groundwater system, such as Cl, $\delta^{18}O$ and $\delta^{2}H$. The information gathered in many variables can best be handled by using multivariate techniques.

A new method named M3 (Multivariate Mixing and Mass balance calculations) was developed in order to identify with a higher resolution the contribution of the observed groundwater composition from mixing and reactions /Laaksoharju et al. 1995/. The method is based on a Principal Component Analysis followed by a mixing and mass balance calculation once proper end-members have been identified and defined. The novel feature of this method is to use the result from the principal component analysis as a basis for calculating the effects of mixing and reactions.

The M3 method has been tested and modified over several years on contract from the Swedish Nuclear Fuel and Waste Management Company (SKB). The main test site for the model has been the underground Äspö Hard Rock Laboratory (HRL), but it has also been used in Canada /Smellie and Karlsson 1996/, Jordan /Waber et al. 1998/, Oklo in Gabon /Gurban et al. 1998/ and Palmottu in Finland /Laaksoharju et al. 1999a/, and most recently in SKB's Site Characterisation Programme /Laaksoharju et al. 2004ab, Laaksoharju 2004/.

The first version of the code (M3 v1.0) appeared in 1995 /Laaksoharju et al. 1995/ but it was not until 1999 when the first documented version of M3 (v2.0) was available as an SKB report /Laaksoharju et al. 1999b/ and as a paper in *Applied Geochemistry* /Laaksoharju et al. 1999c/.

The latest version of M3 (v3.0) has enough new features to deserve a thorough update of the 1999 manual. The added functionality of M3 v3.0 tries also to overcome some criticisms raised by specific methodological steps in previous versions, mainly those regarding the use of only two principal components to compute mixing proportions /e.g. Dershowitz et al. 2000, Bath and Jackson 2002/, and those regarding the selection of end-members /Bath and Jackson 2002/.

Besides the added capabilities, an extra reason to update the documentation is to comply with recent QA regulation to be imposed on all codes used in SKB's Site Investigation programme /SKB 2003, 2004, Hicks 2005/.

/Hicks 2005/ points out that "the results [of the review process of several codes] show that there are varying standard of code documentation and testing with some room for improvement in certain areas". Specifically, Hicks' conclusions for M3 state that "/Laaksoharju et al. 1999b/ provide a comprehensive description of how to run M3. However, no details of the implementation of Principal Component Analysis (PCA) theory and the mass balance calculation method have been identified in any of the references reviewed under this project." To overcome these shortcomings in the documentation and testing of M3 is another of the reasons for this revision.

3 Mathematical formulation

This section is dedicated to the detailed description of the three basic steps that form the backbone of M3. The emphasis here is on the mathematical formulation, but fundamental concepts (e.g. end-members) and methods (e.g. mixing and mass balance) are also introduced. The first subsection presents the multivariate statistical technique known as Principal Component Analysis from where the remaining computations are carried out. This is a rather technical topic but jargon and mathematical level have been kept to the bare minimum. The other two subsections are less demanding mathematically but more so geochemically. If the meaning of PCA is already known, skip Section 3.1 and go straight to Section 3.2 (mixing) or 3.3 (mass balance).

3.1 Principal component analysis

Principal Components Analysis (PCA), the core of M3, is a useful statistical technique that has found application in fields such as classification, pattern recognition, and image compression, and is a common technique for finding patterns in data of high dimension.

Before getting to a description of PCA, it is first useful to introduce the mathematical and statistical concepts that underlie PCA, mainly covariance, eigenvectors and eigenvalues. Further information can be found in many mathematics and statistics textbooks /e.g. Griffel 1989, Bevington and Robinson 1992, Anton 2000, Davis 2002/. What follows is based on the text “A tutorial on Principal Components Analysis” by Lindsay I Smith (http://csnet.otago.ac.nz/cosc453/student_tutorials/ principal_components.pdf).

3.1.1 Background mathematics

This section attempts to give some elementary mathematical background that will be required to understand the process of PCA. The topics are covered independently of each other, and examples given. First we deal with statistical concepts (standard deviation, variance and covariance) and then with that part of the algebra of matrices relevant to PCA (eigenvectors and eigenvalues). Readers familiar with the fundamentals of PCA can go directly to Section 3.1.2.

Covariace

Covariance is a generalization of the variance (a measure of the spread of the data around the mean value) to more than one variable, and defines the degree of correlation (and spread) between pairs of variables.

Variance (and standard deviation) only operates on one variable, so one can only calculate the variance for each variable of the data set independently of the other variables. However, it is useful to have a similar measure to find out how much the variables vary from the mean with respect to each other.

Covariance is such a measure. Covariance is always measured between two variables. If one calculates the covariance between one variable and itself, the variance is obtained. So, for a 3-dimensional data set (x, y, and z variables), one can measure the covariance between the x and y variables, the x and z variables, and the y and z variables. Measuring the covariance between x and x, or y and y, or z and z would give the variance of the x, y and z variables respectively.

Table 3-1. A two-variable dataset and covariance calculation. The dataset consists of ^{18}O and ^{2}H delta-values (permil deviations) of 12 groundwater samples from the Scandinavian Shield.

	Data		Covariance				
	δ^{18}O (‰ dev)	**δ^{2}H (‰ dev)**	**δ^{18}O**	**δ^{2}H**	$(\delta^{18}O_i - \delta^{18}\bar{O})$	$(\delta^{2}H_i - \delta^{2}\bar{H})$	$(\delta^{18}O_i - \delta^{18}\bar{O})(\delta^{2}H_i - \delta^{2}\bar{H})$
	–11.9	–83.4	–11.9	–83.4	–1.45	–4.73	6.85
	–9.7	–75.7	–9.7	–75.7	0.753	2.97	2.23
	–10.2	–78.4	–10.2	–78.4	0.253	0.267	0.0673
	–7.4	–61.5	–7.4	–61.5	3.05	17.2	52.4
	–13.6	–99.4	–13.6	–99.4	–3.18	–20.7	65.9
	–11.5	–89.4	–11.5	–89.4	–1.05	–10.7	11.2
	–10.0	–73.4	–10.0	–73.4	0.453	5.27	2.38
	–12.3	–90.6	–12.3	–90.6	–1.85	–11.9	22.0
	–11.1	–76.1	–11.1	–76.1	–0.647	2.57	–1.66
	–8.5	–64.8	–8.5	–64.8	1.95	13.9	27.1
	–10.6	–82.0	–10.6	–82.0	–0.147	–3.33	0.492
	–8.6	–69.3	–8.6	–69.3	1.85	9.37	17.4
Total	–125.4	–944.0					206.36
Mean	–10.45	–78.67					17.20

The formula for covariance is very similar to that for variance,

$$\mathrm{var}(x) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})}{n-1}, \tag{3-1}$$

the only difference being the replacement of the x's for the y's in the second set of brackets:

$$\mathrm{cov}(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}, \tag{3-2}$$

which can also be written in terms of random variables X and Y as

$$\mathrm{cov}(X,Y) = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle \tag{3-3}$$

In Equations (3-1) and (3-1) $\bar{x}$ is the mean of variable x and $\bar{y}$ is the mean of variable y, and in Equation (3-3) $\langle \cdot \rangle$ is the expected value of the random variable. From Equations (3-2) or (3-3) it is easily deduced that cov(x, y) = cov (y, x), a property that will be useful when the covariance matrix is introduced in the next section.

An example dataset would clarify the meaning of the covariance. Table 3-1 gives the deuterium and ^{18}O delta-values (‰) for ten groundwater samples from the Scandinavian Shield. There are two variables, δ^2H and δ^{18}O, and the lower part of Table 3-1 gives the steps needed to compute the covariance between both variables, cov (δ^{18}O,δ^2H) = 17.20.

What does this tell us? The exact value is not as important as its sign (ie. positive or negative). If the value is positive, as it is here, it indicates that both variables increase together, meaning that, in general, as the deuterium delta value increases, so does the ^{18}O delta-value. If the value is negative, then as one variable increases the other decreases. And if the covariance is zero, it indicates that the two variables are independent of each other.

The result that the delta-value of ^{18}O increases as the delta-value of ^{2}H increases can be easily seen by drawing a graph of the data, as in Figure 3-1. However, the advantage of being able to visualize data is only available in 2 and 3 dimensions (as it would be clear when dealing in M3 with multi-dimensional data). Since the covariance value can be calculated between any two variables in a data set, this technique is often used to find relationships between variables in high-dimensional data sets where visualisation is difficult.

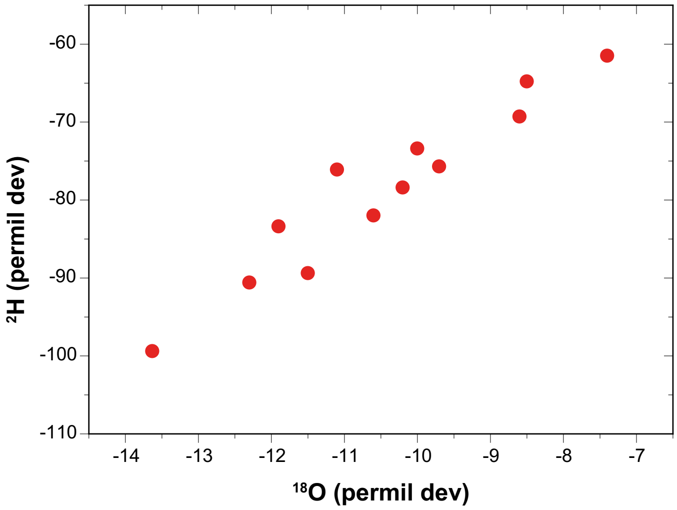


Figure 3-1*. Plot of the covariance data from Table 3-1 showing the positive relationship between* $\delta^{18}O$ *and* $\delta^{2}H$.

The covariance matrix

As already seen, covariance is always measured between two variables. If we have a dataset with more than two variables, there is more than one covariance measurement that can be calculated. For example, from a three-variable dataset (variables x, y, z) one can calculate $\mathrm{cov}(x, y)$, $\mathrm{cov}(x, z)$, and $\mathrm{cov}(y, z)$. In fact, for an n-variable dataset, one can calculate $n!/2(n-2)!$ different covariance values. These covariance values are usually arranged into a matrix, called the *covariance matrix*:

$$\mathrm{cov}(\mathbf{X}) = \langle (\mathbf{X}-\langle\mathbf{X}\rangle)(\mathbf{X}-\langle\mathbf{X}\rangle)^{\mathrm{T}} \rangle \tag{3-4}$$

where $\mathbf{X}$ is a column vector of n random variables. Covariance matrix $\mathrm{cov}(\mathbf{X})$ is an $n \times n$ matrix containing as elements the covariance of a pair of variables of the dataset. For example, for a dataset with three variables x, y and z, the covariance matrix will be written as:

$$\mathrm{cov}(\mathbf{X}) = \begin{pmatrix} \mathrm{cov}(x,x) & \mathrm{cov}(x,y) & \mathrm{cov}(x,z) \\ \mathrm{cov}(y,x) & \mathrm{cov}(y,y) & \mathrm{cov}(y,z) \\ \mathrm{cov}(z,x) & \mathrm{cov}(z,y) & \mathrm{cov}(z,z) \end{pmatrix}, \tag{3-5}$$

with $\mathbf{X} = (x,y,z)^{\mathrm{T}}$ and $\mathrm{cov}(a,b)$ computed from Equation (3-2).

Some points to note: Down the main diagonal, the covariance value is between one of the variables and itself, i.e. the diagonal contains the variance of the variables of the dataset. The other point is that since $\mathrm{cov}(a,b = \mathrm{cov}(b,a)$, the covariance matrix is a real square symmetric matrix.

Eigenvectors

A general result of matrix algebra is that one can multiply two matrices together, provided they have compatible sizes. Eigenvectors are a special case of this. Consider the following two multiplications between a matrix and a vector:

$$\begin{aligned} \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 3 \end{pmatrix} &= \begin{pmatrix} 11 \\ 5 \end{pmatrix} \\ \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} &= \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} \end{aligned} \tag{3-6}$$

In the first example, the resulting vector is not an integer multiple of the original vector, but in the second example, the resulting vector is exactly 4 times the starting vector. This is because a vector can be thought of as a point in 2-dimensional space. The vector $(3,2)^T$ (from the second example above)[1] represents an arrow pointing from the origin, $(0,0)^T$, to the point $(3,2)^T$. The other matrix, the square one, can be thought of as a *transformation matrix*. If one multiplies this matrix on the left of a vector, the answer is another vector that is transformed (rotated, rescaled, reflected) from its original position.

It is the nature of the transformation that the eigenvectors arise from. Imagine a transformation matrix that, when multiplied on the left, reflects a vector through the line $y = x$ (i.e. $y = x$ is the "mirror plane" of the transformation). It is easy to see that if a vector lies *on* the line $y = x$, it would be reflected onto itself. This vector (and all multiples of it, because it would not matter how long the vector is), would be an *eigenvector* (from the German *eigen*, itself) of the transformation matrix.

Eigenvector have several interesting properties. First of all, they can only be found for *square* matrices. But not every square matrix has eigenvectors. And, given an $n \times n$ matrix that does have eigenvectors, there are n of them. Given a 3×3 matrix, there are 3 eigenvectors.

Another of their properties is that even if an eigenvector is scaled by a certain amount before multiplying it, the same multiple of it is found as a result:

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

$$2 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \end{pmatrix} \tag{3-7}$$

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 6 \\ 4 \end{pmatrix} = \begin{pmatrix} 24 \\ 16 \end{pmatrix} = 4 \times \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

This is because if one scales a vector by a certain amount, the vector would be longer, but its direction would remain unchanged.

Lastly, all the eigenvectors of a matrix are orthogonal, i.e. at right angles to each other, no matter how many dimensions (variables) are involved. This is important because it means that one can express the data in terms of these orthogonal eigenvectors, instead of expressing them in terms of the original axes (one axis for each variable in the dataset). This is one of the key operations behind PCA.

Another important thing to know is that eigenvectors are always normalised to have a length of one. This is because the length of a vector does not affect whether it is an eigenvector or not, whereas the direction does. So, in order to keep eigenvectors standard, they are usually scaled to give them a length of one. Following the example above, $(3,2)^T$ is an eigenvector, and its length is

$$\sqrt{\left(3^2 + 2^2\right)} = \sqrt{13} \tag{3-8}$$

So we divide the original vector by $\sqrt{13}$ to give it a length of 1:

$$\begin{pmatrix} 3 \\ 2 \end{pmatrix} \div \sqrt{13} = \begin{pmatrix} 3/\sqrt{13} \\ 2/\sqrt{13} \end{pmatrix} = \begin{pmatrix} 0.832 \\ 0.555 \end{pmatrix} \tag{3-9}$$

---

[1] Interchanging rows with columns in a vector or a matrix gives the *transpose* of the vector or matrix. This operation is designated with the superindex T after the vector or matrix. The transpose of vector $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$ is vector $(3,2)^T$.

For small square matrices, eigenvectors (and eigenvalues; see below) are found by solving the *characteristic equation* $|\mathbf{C}-\lambda\mathbf{I}| = 0$, where **C** is the data matrix (the covariance matrix in our case), **I** is the unit matrix (a matrix with diagonal entries equal to one and off-diagonal entries all zero), and λ are the eigenvalues. This equation is a determinant that expands into a polynomial of degree *n* (the number of rows or columns of the data matrix). For bigger matrices, special numerical iterative methods are used. The most common and general technique is called *Singular Value Decomposition* (SVD), which consists in factorising the covariance matrix into three matrices: a diagonal matrix containing the square roots of the eigenvalues, and two orthogonal matrices containing the eigenvectors. See /Griffel 1989/ for a technical introduction to the subject and /Press et al. 1992/ or /Davis 2002/ for a more practical one. /Anton 2000/ also gives good advice on eigenvectors in general, how to find them, and orthogonality. There are also many ready-to-use numerical routines to find the eigenvectors of a matrix. A useful maths package, called newmat, is available at http://webnz.com/robert/.

Eigenvalues

Eigenvalues are closely related to eigenvectors; in fact, Equations (3-6) and (3-7) have an example of one eigenvalue. Notice how, in both examples, the amount by which the original vector was scaled after multiplication by the square matrix was the same. In this example, the value was 4. This number, 4, *is the eigenvalue associated with eigenvector* $(3/\sqrt{13}, 2\sqrt{13})^T$. No matter what multiple of the eigenvector we took before we multiplied it by the square matrix, we would always get 4 times the scaled vector as our result (as in Equation 3-7). So one can see that eigenvectors and eigenvalues always come in pairs. When one uses a programming library to calculate eigenvectors, the eigenvalues associated to each eigenvector are calculated as well.

3.1.2 Steps in PCA

PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analysing data.

The other main advantage of PCA (also exploited in M3) is that once these patterns in the data are found, one can "compress" the data by reducing the number of dimensions, without much loss of information. This is extremely useful when trying to graphically summarise the most important information.

Following /Lindsay 2000/, PCA can be decomposed into six basic steps:

Step 1: Get the data. To keep things simple, a two-dimensional example dataset will be used. Two dimensions give the opportunity of plotting the results to show what the PCA analysis is doing at each step. The data used are 1,550 analysis of deuterium and ^{18}O delta-values of groundwater samples from the Scandinavian Shield. They can be found in Appendix A and are plotted in Figure 3-2 colour-coded according to the Cl content of each sample (the Cl content is also listed in Appendix A). The two-variable data set can be arranged into a matrix **X** formed by two-columns: column 1 contains the $\delta^{10}O$ values and column 2 the δ^2H values.

Step 2: Normalise each variable of the dataset. For PCA to work properly, the variables should be normalised (or "reduced"), by subtracting the mean and dividing by the standard deviation. This operation is carried out separately for each variable in the dataset. If the dataset consists of two variables, x_i and y_i (in our example x is the ^{18}O delta-value and y is the ^2H delta-value, and i runs from 1 to 1,550) the reduced variables x' and y' are:

$$x_i' = \frac{x_i - \bar{x}}{\sigma_x} \qquad (3\text{-}10a)$$

and

$$y_i' = \frac{y_i - \overline{y}}{\sigma_y}, \qquad (3\text{-}10b)$$

where, as before, $\overline{x}$ and $\overline{y}$ are mean values and σ_x and σ_y are the standard deviation of each variable. The normalised or reduced variables are plotted in the right panel of Figure 3-2. Notice that now both variables have zero mean (values are centred around zero) and the axes limits have changed to reflect the one-standard deviation normalisation.

Step 3: Calculate the covariance matrix. This is done in exactly the same way as mentioned in section 3.1.1. Since the dataset is 2-dimensional, the covariance matrix is a 2×2 matrix. The result is

$$\mathrm{cov}(\delta^{18}\mathrm{O}, \delta^{2}\mathrm{H}) = \begin{pmatrix} 1.000 & 0.937 \\ 0.937 & 1.000 \end{pmatrix}. \qquad (3\text{-}11)$$

So, since the non-diagonal elements in this covariance matrix are positive, we should expect that both the x and y variables increase together. Notice also that the variance of both variables (diagonal terms) is one, as expected, because we are working with normalised variables, whose standard deviation (and therefore variance) is one. The off-diagonal value is the covariance of δ^{18}O and δ^{2}H. Because this value is close to one, it means that both variables are highly correlated.

Step 4: Calculate the eigenvectors and eigenvalues of the covariance matrix. Since the covariance matrix is square, we can calculate the eigenvectors and eigenvalues for this matrix (solving the *characteristic equation*, as explained in the previous section). These are rather important, as they give us useful information about the data (as will soon be clear). In the meantime, here are the eigenvectors and eigenvalues:

$$\mathbf{D} = \begin{pmatrix} 1.000 & 0 \\ 0 & 0.032 \end{pmatrix} \qquad (3\text{-}12)$$

$$\mathbf{V} = \begin{pmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{pmatrix} \qquad (3\text{-}13)$$

Matrix **D** is diagonal ($n \times n$) and has the eigenvalues in the main diagonal. Matrix **V** ($n \times n$) has each eigenvector as a column vector, i.e. $(0.7071, 0.7071)^{T}$ is the first eigenvector and $(0.7071, -0.7071)^{T}$ is the second eigenvector. **V** is usually called the *matrix of basis vectors*.

It is important to notice that these eigenvectors are both unit eigenvectors i.e. their lengths are both 1. This is very important for PCA; most maths packages, when asked for eigenvectors, give unit eigenvectors.

So what do they mean? Looking at the plot of the data in Figure 3-3 it can be seen that the dataset has quite a strong pattern. As expected from the covariance matrix, the two variables do indeed increase together. On top of the data, both eigenvectors have been plotted as diagonal dashed lines. As stated in the eigenvector section, they are perpendicular to each other. But, more importantly, they provide information about the patterns in the data. One of the eigenvectors (labelled “eigenvector 1” in Figure 3-3) goes through the middle of the elongated cluster of points, like drawing the best-fit line. This eigenvector shows how the two variables δ^{18}O and δ^{2}H are related along that line. The second eigenvector gives the other less important pattern in the data, that all the points roughly follow the main line (eigenvector 1), but have some dispersion above and below it. This dispersion is captured by the second eigenvector.

So, by this process of taking the eigenvectors of the covariance matrix, we are able to find lines that characterise the data.

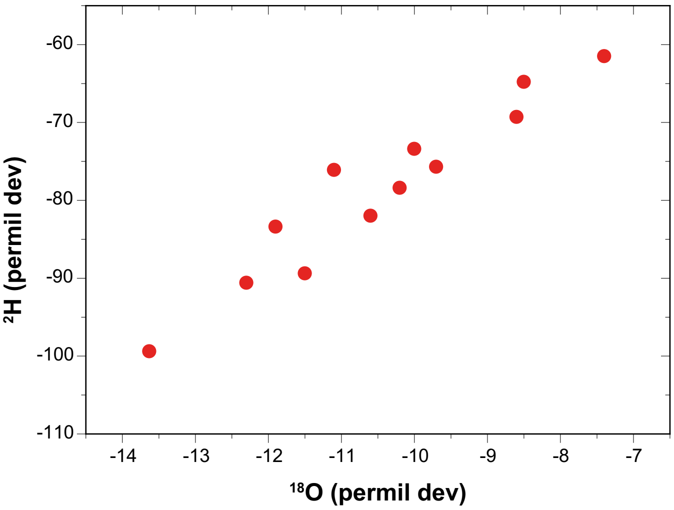


Figure 3-2. Plot of the data from Table 3-3. The plot on the left is for the original data and the plot on the right for the normalized data (zero mean and a standard deviation equal to one). Data points are colour-coded for Cl content as in Figure 3-1.

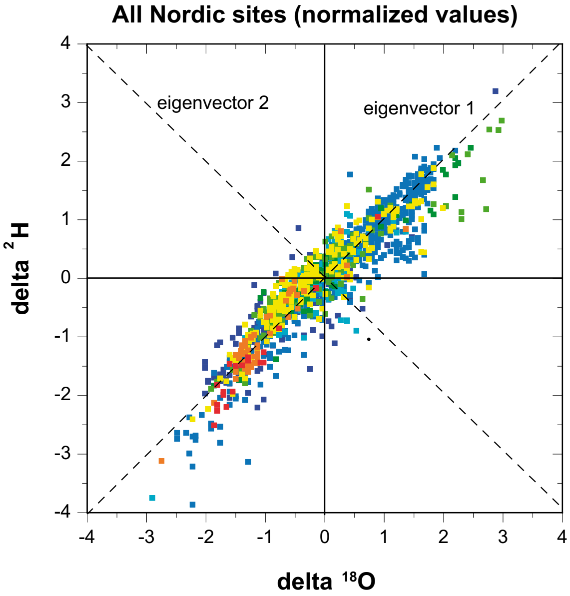


Figure 3-3. A plot of the normalised data with the eigenvectors of the covariance matrix overlaid on top (dashed lines). Data points are colour-coded for Cl content as in Figure 3-1.

Step 5: Choosing components and forming a reduced matrix of basis vectors. Here is where the notion of reduced dimensionality enters. If one looks at the eigenvectors and eigenvalues from the previous section, it will be noticed that they have quite different values. In fact, *it turns out that the eigenvector with the highest eigenvalue is the principal component of the dataset.* In the jargon of PCA the eigenvalues are called *principal components* (PCs) and organized in order of decreasing size of the corresponding eigenvalue: the eigenvector with the largest eigenvalue is called the first principal component (usually abbreviated PC1), the eigenvector with the second largest eigenvalue is called the second principal component (PC2), and so on.

In our example, the eigenvector with the largest eigenvalue (D_{11} = 1.000) was the one that pointed down the middle of the elongated cluster of points of the dataset. It is the most significant relationship between the variables of the dataset. In general, once eigenvectors are found from the covariance matrix, the next step is to order them by eigenvalue, highest to lowest. This gives the (principal) components in order of significance.

Now, one can decide to *ignore* the principal components of lesser significance. Some information is lost, but if the eigenvalues are small, the loss is not very big. If some components are left out, the final dataset will have fewer dimensions than the original. To be more precise, if one originally has n variables in the dataset, n eigenvectors and eigenvalues are obtained. If one then chooses to keep only the first p eigenvectors, the final dataset will have only p dimensions.

What needs to be done next is to form a *reduced matrix of basis vectors*, **W**, which is a subset of matrix **V**. This is constructed by taking the eigenvectors that one wants to keep from the matrix of eigenvectors, and forming a matrix with the eigenvectors as column vectors:

$$\mathbf{W} = (\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_p) \qquad (3\text{-}14)$$

where $\mathbf{V}_1$ is the first column (first eigenvector) in matrix **V**, and $\mathbf{V}_p$ is the last eigenvector form matrix **V** that is retained in the reduced matrix **W**. **W** is a $n \times p$ matrix.

Given the example set of data above, and the fact that it has 2 eigenvectors, there are two choices. We can either form a reduced matrix with both of the eigenvectors ($p = 2$),

$$\mathbf{W} = \begin{pmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{pmatrix}, \qquad (3\text{-}15)$$

or we can choose to leave out the smaller, less significant component (associated to eigenvalue $D_{22} = 0.032$) and only have a single column ($p = 1$):

$$\mathbf{W} = \begin{pmatrix} 0.7071 \\ 0.7071 \end{pmatrix}. \qquad (3\text{-}16)$$

We shall see the result of each of these in the next section.

Step 6: Deriving the new dataset with reduced dimensionality. This is the final step in PCA, and is also the easiest. Once the principal components (eigenvectors) that we wish to keep have been chosen and the reduced matrix of basis vectors formed, we simply take the transpose of this matrix and multiply it on the left of the matrix containing the original (normalised) dataset, also transposed:

$$\mathbf{Z}^{\mathrm{T}} = \mathbf{W}^{\mathrm{T}}\mathbf{X}'^{\mathrm{T}} \qquad (3\text{-}17)$$

where $\mathbf{W}^{\mathrm{T}}$ is the transpose of the matrix with the eigenvectors in the columns so that the eigenvectors are now in the rows, with the most significant eigenvector at the top, and $\mathbf{X}'^{\mathrm{T}}$ is the transpose of the normalised data matrix, i.e. the data items are in each column, with each row holding a separate variable. The final data are in matrix **Z**, with data items in rows and variables in columns (of course, $\mathbf{Z}^{\mathrm{T}}$, the result of the above matrix operation, has the variables in rows and data items in columns). Because **W** is an $n \times p$ matrix and $\mathbf{X}'$ an $m \times n$ matrix (where m is the number of data items in each variable of the dataset), **Z** will be an $m \times p$ matrix.

This operation gives us the original data solely in terms of the eigenvectors we have chosen. The original dataset has two axes, x and y, so the data were written in terms of them. It is possible to express data in terms of any two axes. If these axes are perpendicular the expression will be the most efficient. This is why it is so important that eigenvectors are always perpendicular to each other. We have changed the data from being in terms of the axes (variables) x and y, and now they are in terms of two eigenvectors. When the new dataset has a reduced dimensionality, i.e. some of the eigenvectors have been left out, the new dataset is expressed only in terms of the vectors that we decided to keep.

To show this on the example dataset, matrix operation (3-17) has been performed using both eigenvectors (Equation 3-15), and the result plotted in Figure 3-4 to show how the new data relate to the principal components. In this case, where we have kept all the eigenvectors, the plot is basically the original data, *rotated* so that the eigenvectors are the axes. This is understandable since we have lost no information in this operation.

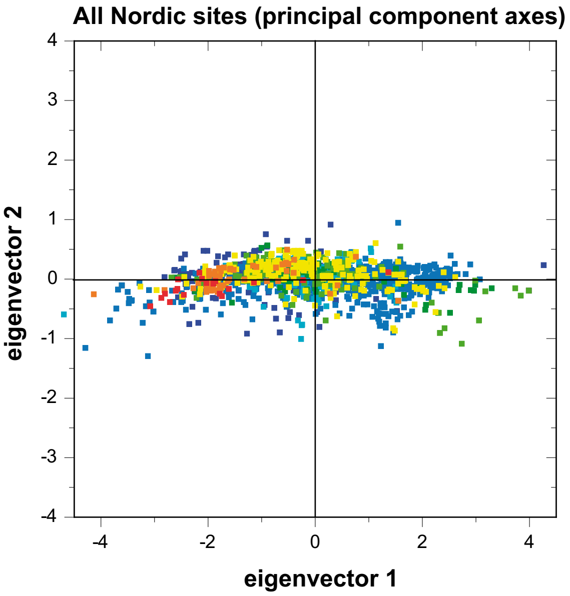


Figure 3-4. Plot of the new data points after rotation using eigenvectors. Data points are colour-coded for Cl content as in Figure 3-1.

The other transformation we can make is by taking only the eigenvector with the largest eigenvalue, Equation (3-16). The matrix resulting from this new operation has dimensions $m\times1$, i.e. is a column vector and is one-dimensional. If one compares this dataset with the one resulting from using both eigenvectors, it will be clear that this dataset is exactly the first column of the other. So, if we were to plot these data, they would be one-dimensional, and would be points on a line in exactly the x positions of the points in the plot in Figure 3-4. We have effectively thrown away the other axis entirely, which is the other eigenvector.

So what have we done up to here? Basically we have transformed our data so that they are expressed in terms of the *real* patterns between them, where the patterns are the lines that most closely describe the relationships between the data. This is helpful because the data points are now classified as a combination of the contributions from each of those lines (the eigenvectors). Initially we had the simple x (δ^{18}O) and y (δ^{2}H) axes. This is fine, but the x and y values of each data point don't really tell us exactly how that point relates to the rest of the data. Now, after the PCA, the values of the data points tell us exactly where they lie with respect to the trend lines (ie. above/below). In the case of the transformation using both eigenvectors, we have simply altered the data so that they are now expressed in terms of the eigenvectors instead of the usual axes. But the single-eigenvector decomposition has removed the contribution due to the smaller eigenvector and left us with data that are only in terms of the other.

3.1.3 Getting the old data back

Only if we take all the eigenvectors in our transformation (Equation 3-17) will we get back exactly the original data. If we reduce the number of eigenvectors in the final transformation, the retrieved data will lose some information. In the case of using all eigenvectors, we can recover the original data left-multiplying both sides of Equation (3-17) by the inverse of $\mathbf{W}^{T}$,

$$\mathbf{X}'^{T} = (\mathbf{W}^{T})^{-1}\,\mathbf{Z}^{T}. \qquad (3\text{-}18)$$

When we take all the eigenvectors in the reduced matrix of basis vectors **W**, it turns out that the inverse of the reduced matrix is actually equal to its transpose because in this case (and only in this case), matrix **W** is an orthogonal matrix (a square matrix with all its column vectors orthogonal and of unit length). This makes the return trip to the original data much easier, because the above equation then becomes

$$\mathbf{X}'^{\mathrm{T}} = \mathbf{W}\mathbf{Z}^{\mathrm{T}} \tag{3-19}$$

But, to actually get the original data back, we need to multiply each data item by the standard deviation of the variable and add on the mean. So, for completeness,

$$\mathbf{X}_i = \sigma_i \mathbf{X}_i' + \langle \mathbf{X}_i \rangle \mathbf{1}, \qquad i = 1, \ldots, n \tag{3-20}$$

where $\langle \mathbf{X}_i \rangle$ and σ_i are the mean and standard deviation of the original variable i, and $\mathbf{1}$ is an $m\times 1$ column matrix with all entries equal to one.

If **W** is a subset of **V** (i.e. not all the eigenvectors are used) the return trip is, *in theory,* not so easy because then **W** is not square and the inverse of a rectangular matrix is not defined. *In practice*, however, Equation (3-19) (*not* Equation 3-18!) can also be used in the general case to recover the original data. The derivation can be found in most books on PCA or matrix algebra under the *Karhunen-Loéve Transformation (KLT)* heading /e.g. Gershenfeld 1999/.

It is not very useful to perform the data re-creation using the complete reduced matrix **W**, because in this case **W** = **V** and the result is exactly the data we started with. However, it is very instructive to do this with the reduced matrix when not all eigenvectors are included to show how information has been lost. Figure 3-5 shows this plot for the example dataset. Comparing it to the original dataset as plotted in Figure 3-2 it becomes obvious how, while the variation along the principal eigenvector has been kept, the variation along the other component (the other eigenvector that has been left out) has gone.

3.1.4 The meaning of each principal component in a PCA

One problem with using PCA to replace n variables by p principal components, rather than the alternative strategy of replacing the n variables by a subset of p of the original variables, is interpretation.

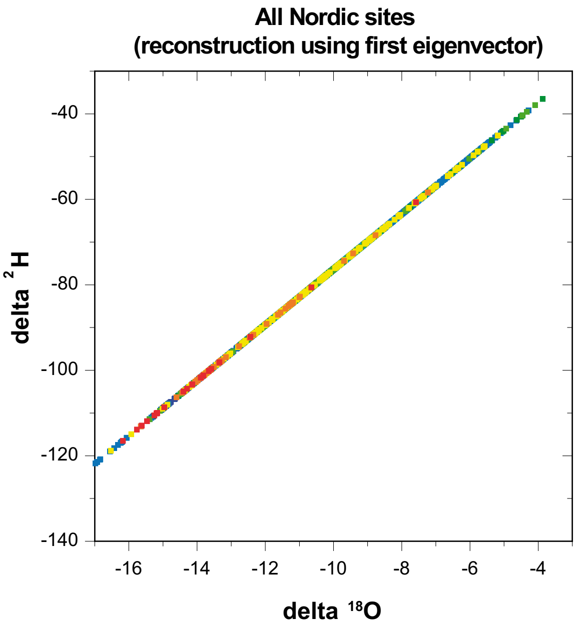


***Figure 3-5**. Reconstruction of the data using only the first eigenvector. Data points are colour-coded for Cl content as in Figure 3-1.*

A common question that arises when performing a principal component analysis is What is the meaning of each principal component? Or, if the results are graphically plotted (in one, two or three dimensions), What is the meaning of each axis?[2]

The answer is simple: each principal component (or eigenvector, or axis) is a *linear* combination of the original variables of the dataset. Because all principal components are orthogonal, this is the same as saying that the new axes (the eigenvectors) are obtained from the old axes (the original variables) by a simple rotation. For the example in the previous section the two principal components (PC) are:

$$\begin{aligned} \mathrm{PC}_1 &= 0.7071x + 0.7071y \\ \mathrm{PC}_2 &= 0.7071x - 0.7071y. \end{aligned} \tag{3-21}$$

As can be seen in Figure 3-3, this transformation is a *rotation* of the axes (dashed lines) with respect to the original ones. PC_1 and PC_2 are thus the coordinates of a point in the new coordinate system, expressed in terms of the old coordinates (x and y).

As each principal component is a linear combination of *all n* variables, to interpret a principal component we need to decide which original variables are important, and which are not, in defining that principal component. In PCA the elements of an eigenvector (its components in maths jargon) are called *weights* or *loadings*[3]. In Equation (3-21) the weights of the first principal component are 0.7071 and 0.7071, and those of the second principal component 0.7071 and −0.7071. If a weight of a PC is very small (compared to one, as they are normalized), it means that the contribution of that original variable to the PC is also small. In other words: the PC is independent (or almost) of that variable.

An important point to bear in mind is that a PCA does not change the relative position of the data points. The only transformation on the data is a rigid-body rotation, thus maintaining unchanged the *distances* (in the data space, of course, not in the real 3-dimensional space) between points. If the distance ratio between two points is 4.5 in the original coordinates, it would still be 4.5 in the rotated coordinate system (the one with each eigenvector as an axis).

3.1.5 How much variance is contained in each principal component?

As commented above, PCA is primarily a dimension-reduction technique, which takes observations on n correlated variables and replaces them by uncorrelated variables. These uncorrelated variables, the principal components, are linear combinations of the original variables, which successively account for as much as possible of the variation in the original variables. Typically a number p is chosen ($p < n$) such that by using only the first p PCs instead of the n variables will mean only a small loss of variation.

This “variation” is measured by the magnitude of the eigenvalue associated to each eigenvector. The magnitude of the eigenvalue gives the contribution of that eigenvector to the overall variance of the dataset. In our example, the first eigenvalue has a magnitude of 1 and the second of 0.032 (Equation 3-12), in other words, the second eigenvalue is about 30 times smaller than the first.

Another, more complicated, example will clarify this affirmation. Recall first that there are as many eigenvectors as original variables (because the covariance matrix is a $n \times n$ square matrix, where n is the number of variables).

---

[2] As we have already seen in a previous section, each eigenvector can be thought of as an orthogonal axis in a p-dimensional space.

[3] Do not confuse loadings with scores. Scores are the elements of the matrix $\mathbf{XV}$, i.e. the matrix version of Equation (3-21).

Table 3-6 summarises the results of a real example carried out with 1,080 water samples from the Scandinavian Shield /Laaksoharju 2005/. Ten compositional variables were used, so the covariance matrix has ten rows and ten columns, and ten is the number of eigenvectors. In its first column Table 3-6 gives the eigenvalues associated with the ten eigenvectors, the biggest first and the smallest last. The second column expresses the relative size as a percentage of the total. It is obtained by dividing each entry in column one by its sum total (10,839.771), and multiplying the result by 100 (e.g. first row is 4,589/10,839.771×100 = 42.33%). The last column in Table 3-6 is the cumulative variance of the first *p* eigenvalues.

If we decide to retain only the first *p* principal components, the variance explained once the dimension-reduction step has been taken is given by the cumulative percentage from the first up to the *p*th eigenvalue. If, in the example in Table 3-6, we decide to work only with the first two principal components, they would explain 66.1% of the overall variance of the dataset, the remaining 43.9% having been lost during the dimension-reduction procedure. The amount of variance explained by the retained principal components should always be indicated. This is important and will be put into geochemical perspective in Section 3.2.

3.2 Mixing calculations

Once the PCA is done, the real geochemical calculations start. The main goal of M3 is to know the contribution of several end-members (selected *a priori* on independent geochemical and hydrological grounds) to the chemical composition of each water sample in the dataset, based upon a number of compositional and/or physical-chemical variables (e.g. pH, Eh, concentration of selected elements and isotopes, etc). For this purpose, a key step in M3 methodology is the selection of a set of end-members that would act as end-members for mixing. In other words, *at that point M3 assumes that the sole process that has contributed to the final chemical composition of each water sample is mixing*. M3 tries to estimate mixing proportions by conservative mixing of all the input variables. In that respect, all compositional variables included in the PCA are treated equally, without distinguishing between conservative and non-conservative variables.

This is an important point to bear in mind, because all mass balance calculations (and hence all chemical reactions inferred from them) depend on differences between the composition predicted by mixing (all variables treated as conservative) and the real composition of the sample.

Table 3-6. Variance explained by each eigenvector in a real dataset with 1,080 water samples from the Scandinavian Shield.

Eigenvalue	As %	Cumulative %
4,589	42.33	42.33
2,577	23.77	66.10
1,225	11.30	77.41
1,089	10.05	87.46
776.3	7.161	94.62
305.8	2.821	97.44
156.5	1.443	98.88
61.92	0.5712	99.45
52.34	0.4829	99.94
6.911	0.6375E–1	100.0

Two different routines to compute mixing proportions are implemented in M3: (1) two-principal component mixing, and (2) hyperspace mixing. The first routine assumes that the first two principal components accumulate enough information to compute mixing proportions and thus all mathematical manipulations are performed in two dimensions. The second routine makes no assumption regarding how the variance is partitioned among the principal components and uses all of them to compute mixing proportions. Two-principal component mixing is simpler but can give significant errors (Section 3.2.2); hyperspace mixing is more complex but has a zero *a priori* error (Section 3.3.3).

3.2.1 End-members

The concept of *end-member* is a cornerstone in M3 methodology. How many end-members are to be used and their particular chemical and isotopic composition is something that in principle lies outside M3 methodology and should be decided (also in principle) by expert judgment after a careful geochemical and hydrological study of the system. However, this version of M3 includes a pseudo-automatic procedure to select the proper set of end-members for a given dataset which is based on a geometrical property that applies only to the hyperspace mixing routine described in section 3.2.3. The procedure is described in Section 4.1 under the heading "End-member Selection Module", to where the reader is referred for details.

All the measured groundwater composition is compared to some well-sampled and analysed groundwater of the site or to a hypothetical (modelled) extreme water. In both cases these waters are called in this report *end-members (*but see /Laaksoharju et al. 1999bc/ for a more detailed terminology)[4]. The M3 method compares the measured groundwater composition of each sample to the selected end-member composition. *The modelling is therefore always relative to the selected end-member composition* just as a measured altitude is relative to a chosen fixed point. As already stated, an end-member can be a modelled or sampled (extreme) groundwater e.g. glacial melt water, ancient sea water, brine water or rain water, dilute shallow groundwater or sea water.

As a general rule, the number and type of end-members to be selected in the modelling depend on the aim of the modelling and the complexity of the site. The groundwater data used in the modelling determines the minimum number of end-members needed to describe the observed groundwater composition.

Once the end-members have been chosen, they are included in the dataset and the PCA is performed. The inclusion of the end-members in the PC analysis could introduce a bias if (1) the end-members have a chemical composition radically different from the waters in the dataset; (2) there are very few water samples in the dataset; or (3) the statistical uncertainties affecting the samples of the dataset and of the end-members are different. Case (1) should never occur in a system dominated by mixing (if it were the case, it would mean that the end-members had been incorrectly selected); case (2) is not compatible with M3 philosophy, whose purpose is always the joint analysis of a large dataset (for very small datasets, other methods not based on multivariate statistics are more suitable); finally, case (3) is not realistic when both the end-members and the dataset belong to the same category (in this case, groundwaters). But if in a particular example one or more of the above-mentioned possibilities are expected, the PCA should be performed without the end-members and then compute the PC coordinates of the end-members (PC scores) by means of the eigenvectors, as explained in Section 3.1.2. Report 2 (Verification and Validation) includes a test case to quantify this bias.

---

[4] The name "end-member" has been preferred because it has no genetic connotation as how it has been defined or selected. Also, in this way other non-hydrological applications of M3 can be better described in a neutral manner.

3.2.2 Two-principal component mixing

This is the standard way mixing proportions were calculated in all previous versions of M3 /Laaksoharju et al. 1999bc/, and is kept here to be backward-compatible with previous versions of the code. The user should be well aware of the limitations of the two-principal component mixing algorithm when using M3 v3.0.

As its name suggests, only the first two principal components are used for such mixing calculation. The consequences of this dimension-reduction step are double: (1) on the good side, everything can be done on a two-dimensional plot, simplifying visualization and interpretations; (2) on the bad side, the computed mixing proportions are not unique (in the mathematical sense) when more than three end-members are used; for more than three end-members a trick has to be utilised (see below) that introduces an intrinsic error into the mixing proportions; the magnitude of that error depends on the position of the sample in the PC-plot, but can be high. It is up to the user to select which method, but this section and the next give some recommendations in that direction.

Let us use a specific example to explain how mixing proportions are computed with the two-principal component mixing routine. Figure 3-6 is a PC-plot for a real dataset where five end-members have been used: Litorina, Sea Sediment, Rain, Glacial, and Brine. These five end-members are joined by lines, forming a convex polygon.[5] If the polygon is not convex, one or more of the end-members have to be changed/deleted to make it convex in order to have the possibility of having complete binary mixing between any two end-members.

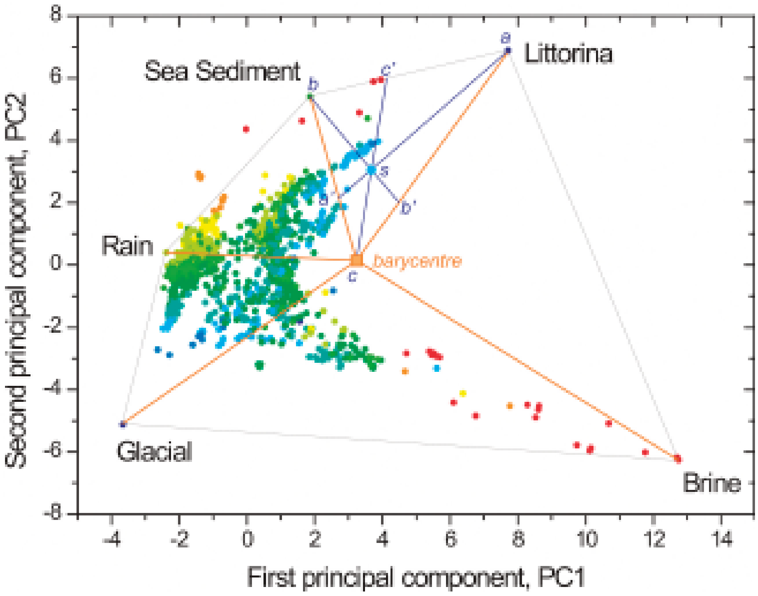


***Figure 3-6**. Procedure to compute mixing proportions in two-principal component mixing. End-members are joined by lines forming a polygon. Each water sample belongs to a triangle defined by two end-members and the barycentre of the polygon. The triangular coordinates of sample s give the mixing proportions in the triangle. To get the mixing proportions in the polygon, the contribution of the barycentre to the final mixing proportions has to be added. See text for details.*

---

[5] A planar polygon is *convex* if it contains all the line segments connecting any pair of its points. All the internal angles of a planar convex polygon are less than 180 degrees.

Next, the polygon is divided into triangles using its barycentre as a fictitious end-member. This is done because *on a plane three is the maximum number of coordinates that can uniquely define a point* (we shall comment on the importance of this limitation later). A point on a plane can be defined by two numbers using Cartesian coordinates and by three numbers using triangular coordinates. But it is not possible to use more than three numbers to define a point on a plane without ambiguity.

The coordinates of the barycentre are very easy to obtain. Denoting the PC coordinates by (x, y) the barycentre (BC) is the point satisfying

$$x_{BC} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad y_{BC} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{3-22}$$

where n is the number of end-members and (x_i, y_i) are the PC coordinates of the end-members. In our example $n = 5$.

Once the barycentre is known and the lines joining it with the end-members are drawn (see Figure 3-1), each sample in the dataset will belong to one and only one of the triangles that cover the polygon. Each triangle (there are n of them) has as vertices two real end-members *and* the fictitious barycentre end-member. The triangular coordinates of any sample will give us the mixing proportions we are looking for.

The selected sample in Figure 3-6 falls in the triangle Litorina-Sea Sediment-BC. Denoting by ξ_i' the *uncorrected* mixing proportions of the sample (i = Lit, Sed, BC), these are:

$$\begin{aligned} \xi_{\text{Lit}}' &= \frac{\overline{sa'}}{\overline{aa'}} = 0.2, \\ \xi_{\text{Sed}}' &= \frac{\overline{sb'}}{\overline{bb'}} = 0.3, \\ \xi_{\text{BC}}' &= \frac{\overline{sc'}}{\overline{cc'}} = 0.5, \end{aligned} \tag{3-23}$$

where the bar means the length of the corresponding segment. The three triangular coordinates give the mixing proportions *in the triangle*, but not *in the polygon*, which is what we want. To know the latter we should make a strong assumption regarding the mixing proportions of the fictitious barycentre end-member. What is the composition of this particular point, expressed as a percentage of each of the real end-members? There is no definite answer, and here lies the non-uniqueness of the mixing proportions computed with the two-principal component mixing algorithm. Only when there are three end-members is the answer unique, because in this case we do not even need the barycentre to compute the mixing proportions. But for four or more end-members a specific composition has to be assumed for the barycentre. The simplest ansatz is that *the barycentre is a mixture in equal proportions of all the end-members*:

$$\xi_{\text{Lit}}'' = \xi_{\text{Rain}}'' = \xi_{\text{Sed}}'' = \xi_{\text{Gl}}'' = \xi_{\text{Br}}'' = \frac{1}{5}\xi_{\text{BC}}', \tag{3-24}$$

where ξ_{BC}' is given by Equation (3-29). In our example $\xi_{\text{BC}}' = 0.5$, and so $\xi_{\text{Lit}}'' = \xi_{\text{Rain}}'' = \xi_{\text{Sed}}'' = \xi_{\text{Gl}}'' = \xi_{\text{Br}}'' = 0.1$. We use the notation ξ_i'' to indicate that this quantity is a *correction* to be added to ξ_i' (the mixing proportions *in the triangle*) to obtain the final mixing proportion *in the polygon*, i.e.

$$\xi_i = \xi_i' + \xi_i'', \tag{3-25}$$

where ξ_i are the final mixing proportions. The correction ξ_i'' is applied equally to all end-members. On the other hand, ξ_i' is non-zero only for the two end-members that form the triangle inside which the sample falls. For the other $n-2$ end-members ξ_i' is zero and so

$$\xi_i = \xi_i'' = \frac{1}{n}\xi_{\text{BC}}' \qquad \text{for } n-2 \text{ end-members.} \tag{3-26}$$

This last equation says that $n-2$ of the computed mixing proportions are numerically identical. In our example ξ_{Lit} and ξ_{Sed} have independent values, but ξ_{Rain}, ξ_{Gl} and ξ_{Br} all have the same value:

$$\begin{aligned}
&\xi_{\text{Lit}} = \xi'_{\text{Lit}} + \xi''_{\text{Lit}} = 0.2 + 0.1 = 0.3,\\
&\xi_{\text{Sed}} = \xi'_{\text{Sed}} + \xi''_{\text{Sed}} = 0.3 + 0.1 = 0.4,\\
&\xi_{\text{Rain}} = 1/5\xi'_{\text{BC}} = 0.1, \qquad\qquad (3\text{-}27)\\
&\xi_{\text{Gl}} = 1/5\xi'_{\text{BC}} = 0.1,\\
&\xi_{\text{Br}} = 1/5\xi'_{\text{BC}} = 0.1.
\end{aligned}$$

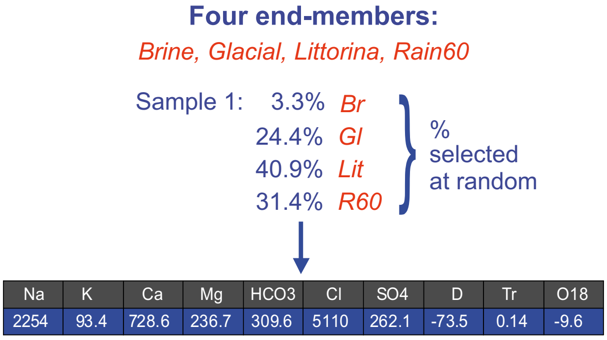
The fact that $n-2$ mixing proportions are identical is a general property for any sample when the two-principal component mixing routine is used. The reason, as already explained, is our assumption that the barycentre is an even mixture of all end-members. /Laaksoharju and Wallin 1997/ have tested an alternative way of computing the mixing proportions without a barycentre, but the results do not improve on the procedure explained here. The barycentre method is the one implemented in version 3 of the M3 code.

Table 3-7 shows part of the output of a mixing calculation using the two-principal component mixing routine. As in the example above the same five end-members have been used (Sea Sediment, Litorina, Brine, Glacial and Rain 1960). Note that each sample has three identical mixing proportions, shaded in grey in the table, and that the three identical end-members change depending on the position of the sample in the PCA polygon (e.g. the first sample has a 10.2% of Litorina, Brine and Glacial, Sample #5 has an 8% of Sea Sediment, Litorina and Brine, and Sample #8 has a 9.4% of Sea Sediment, Litorina and Rain60).

How does this non-uniqueness affect the computed mixing proportions? To assess the accuracy of the two-principal component mixing, a synthetic dataset has been created. 1,000 samples were generated by mixing a random amount of the four end-members Brine (Br), Glacial (Gl), Litorina (Lit), and Rain60 (R60), as shown in Figure 3-7.

Each end-member is defined by a set of compositional variables (Na, K, Ca, Mg, HCO_3, Cl, SO_4, deuterium, tritium, and ^{18}O in the example), as Table 3-8 summarises for this particular example.

To construct a synthetic sample, a randomly generated mixing proportion is assigned to it using a random number generator. A random number uniformly distributed between 0 and 1 is drawn for each end-member (four random numbers for the example above, one each for Brine, Glacial,



Na	K	Ca	Mg	HCO3	Cl	SO4	D	Tr	O18
2254	93.4	728.6	236.7	309.6	5110	262.1	-73.5	0.14	-9.6

Figure 3-7*. Procedure to generate synthetic samples with known proportions of each end-member. The example shows a case with four end-members. Each one has a particular chemical composition (Table 3-8) which is then multiplied by the randomly generated mixing proportion and summed up to build the chemical composition of the synthetic sample.*

Table 3-7. Output of the two-principal component mixing routine to show that three of the mixing proportions of each sample are identical (shaded cells), reflecting the assumption made on the composition of the barycentre.

Sample #	Sea Sed.	Litorina	Brine	Glacial	Rain60
1	0.507	0.102	0.102	0.102	0.187
2	0.121	0.057	0.057	0.057	0.708
3	0.176	0.034	0.034	0.034	0.723
4	0.080	0.080	0.080	0.170	0.589
5	0.122	0.122	0.122	0.500	0.134
6	0.094	0.094	0.257	0.462	0.094
7	0.112	0.112	0.211	0.453	0.112
8	0.060	0.060	0.060	0.124	0.695

Table 3-8. Composition of the four end-members waters used to generate the samples of the synthetic dataset.

End-member	Na (mg/l)	K (mg/l)	Ca (mg/l)	Mg (mg/l)	HCO_3 (mg/l)	Cl (mg/l)	SO_4 (mg/l)	D (dev)	Tritium (TU)	^{18}O (dev)
Brine	8,500	45.5	19,300	2.12	14.1	47,200	906	–44.9	0	–8.9
Glacial	0.17	0.4	0.18	0.1	0.12	0.5	0.5	–158	0	–21
Litorina	3,674	134	151	448	93	6,500	890	–38	0	–4.7
Rain 60	0.4	0.29	0.24	0.1	12.2	0.23	1.4	–80	2,000	–10.5

Litorina and Rain60). These are then added together and each one divided by the sum and multiplied by 100 to get a percentage. In the example shown in Figure 3-7 the mixing proportions are Brine = 3.3%, Glacial = 24.4%, Litorina = 40.9%, and Rain60 = 31.4%, obtained from the random numbers 0.05596, 0.40965, 0.68667, and 0.52717 (sum = 1.67939).

Once a mixing proportion has been assigned, the composition of the synthetic sample is computed by multiplying each end-member's compositional variable by the corresponding mixing proportion and summing up the contribution of all end-members. For example, the amount of Na in the synthetic sample of Figure 3-2 is

$$\mathrm{Na}_{\mathrm{Sample}} = \frac{\mathrm{Na}_{\mathrm{Br}} \times 3.3\% + \mathrm{Na}_{\mathrm{Gl}} \times 24.4\% + \mathrm{Na}_{\mathrm{Lit}} \times 40.9\% + \mathrm{Na}_{\mathrm{R60}} \times 31.4\%}{100} = 2254\ \mathrm{mg/L} \qquad (3\text{-}28)$$

The remaining compositional variables are obtained in the same way.

Finally, these synthetic samples are fed into M3's two-principal component mixing routine to compute the mixing proportions. Figures 3-8 and 3-9 plot the results for a 1,000 synthetic dataset. Each graph depicts the mixing proportion of a specific end-member (Brine, Glacial, Litorina and Rain60) colour-coded as a percentage of its proportion in every sample, from red (low mixing proportion) to blue (high mixing proportion). Figure 3-8 gives the real mixing proportions and Figure 3-9 the ones computed by M3.

Brine and Litorina mixing proportions are well resolved by M3 (compare the left hand side graphs in Figure 3-8 and 3-9), but the results for Glacial and Rain60 are less consistent. This is so because Glacial and Rain60 are the two end-members with lower concentrations of all the compositional variables and are thus most sensible to the real composition of the barycentre.

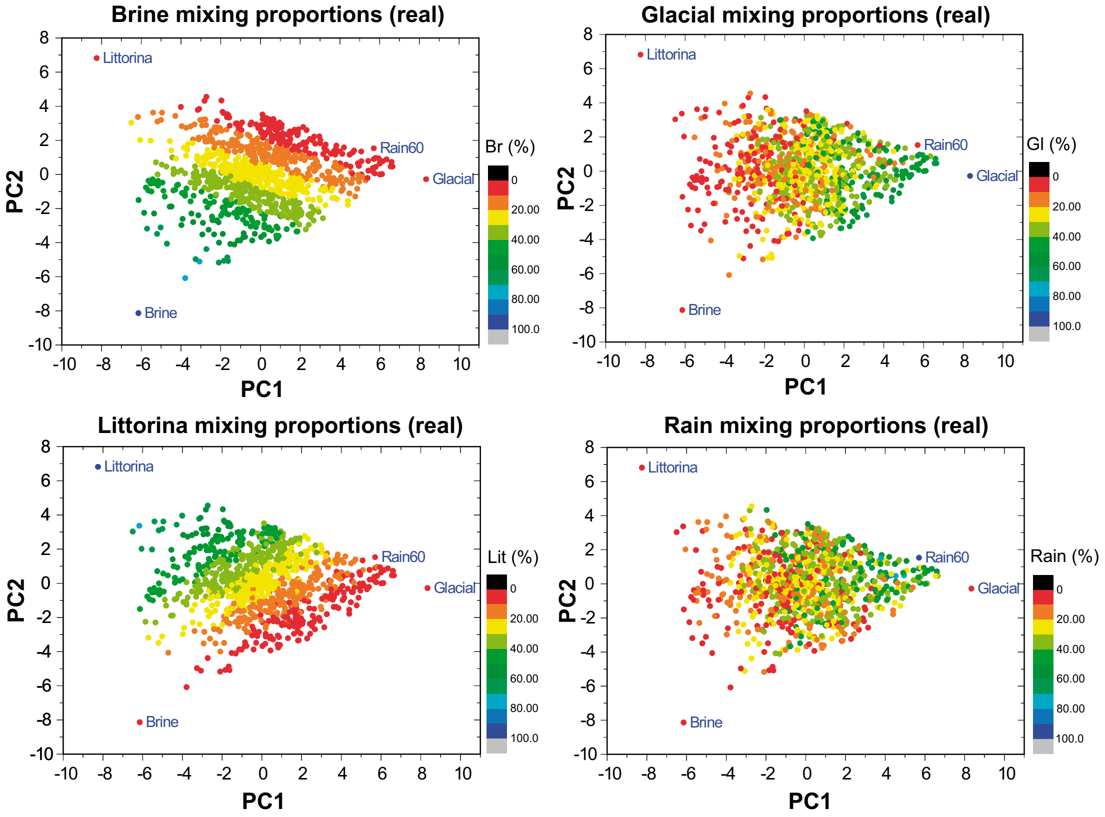


Figure 3-8. *1,000 synthetic samples on a PC-plot colour-coded with respect to the real mixing proportion of each end-member (Brine, Glacial, Litorina, and Rain60).*

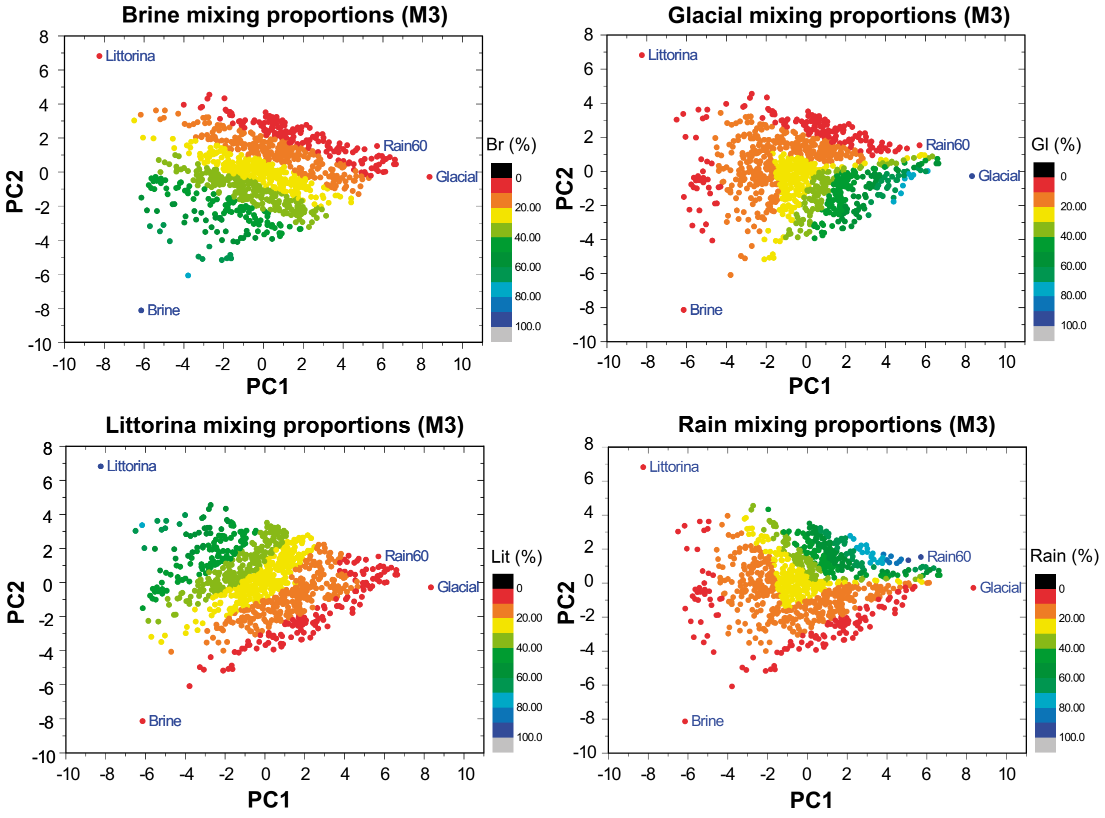


Figure 3-9. *The same synthetic samples as in Figure 3-8, but colour-coded with respect to the calculated proportion of each end-member. M3's two-principal component mixing routine has been used to compute the mixing ratios.*

A way to summarise the accuracy of the computed mixing proportions is by defining a generalised standard deviation between the real and computed mixing proportions:

$$\mathrm{StDev} = \sqrt{(\mathrm{Br}_{\mathrm{Real}} - \mathrm{Br}_{\mathrm{M3}})^2 + (\mathrm{Gl}_{\mathrm{Real}} - \mathrm{Gl}_{\mathrm{M3}})^2 + (\mathrm{Lit}_{\mathrm{Real}} - \mathrm{Lit}_{\mathrm{M3}})^2 + (\mathrm{R60}_{\mathrm{Real}} - \mathrm{R60}_{\mathrm{M3}})^2} \qquad (3\text{-}29)$$

In this expression $\mathrm{Br}_{\mathrm{Real}}$ refers to the known mixing proportion and $\mathrm{Br}_{\mathrm{M3}}$ to the one calculated by M3's two-principal component mixing routine. In Figure 3-10, where the results are graphically presented, each of the 1,000 samples is colour-coded with respect to the standard deviation. Maximum deviation is of the order of 85% and the mean standard deviation for the 1,000 samples is 17%, comparable to the variance not explained by the first two principal components used in the two-principal component mixing algorithm (the sum of the variance of the first two principal components in this example is 81%).

3.2.3 Hyperspace mixing

This new addition to M3 capabilities has a very simple rationale: instead of using only two of the principal components to calculate the mixing proportions, as with previous versions of M3, it calculates them by using all principal components. This is equivalent to saying that the mixing proportions are computed in a space with n-1 dimensions, where n is the number of end-members. Working in hyperspace has the advantage of avoiding the uncertainties derived from the projection of the coordinates onto a plane before calculating the mixing ratios. This uncertainty grows with the number of end-members and can be high when working with end-members which plot close together on the plane defined by the first two principal components (the plane used by the two-principal component mixing routine to perform the calculations), as has been made clear in Figures 3-8 and 3-10 in the previous section.

The mixing proportions are calculated from the principal component coordinates. The whole procedure is a simple coordinate transformation, from a n-1 Cartesian coordinate system (principal component coordinates) to a hyper-tetrahedral coordinate system (mixing proportion coordinates).

Figure 3-11 shows how to compute the mixing proportions from the principal component coordinates for a 2D example (three end-members). The mixing proportions are denoted (ξ_1, ξ_2, ξ_3), and the PCA coordinates as (x,y). In Figure 3-6 the red circle marks the location of the sample

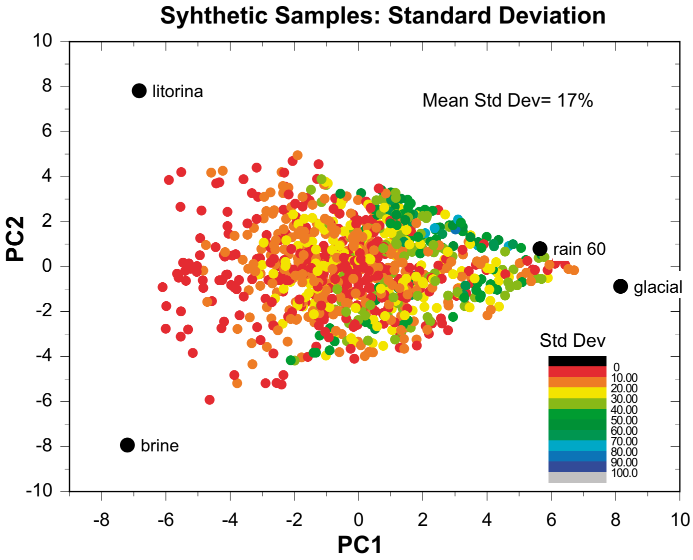


Figure 3-10*. Deviation of computed mixing proportions from real ones for 1,000 synthetic samples. M3's two-principal component mixing routine has been used for the calculations.*

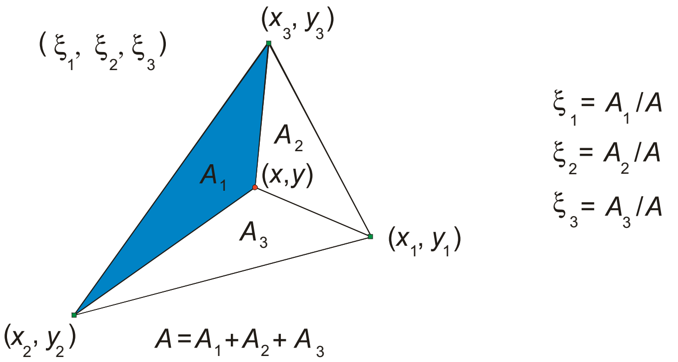


Figure 3-11. *Calculation of the mixing proportion coordinates from the Principal Component Analysis (PCA) coordinates.*

whose mixing proportions we want to compute. From this circle we draw straight lines to the three vertices of the triangle. This operation subdivides the total triangle into three sub-triangles (one coloured in blue) whose areas are A_1, A_2, A_3.

The area of the total triangle is $A = A_1+A_2+A_3$. From the knowledge of these areas it is straightforward to compute the mixing proportion coordinates, which are the ratio of the area of each sub-triangle to the total area (see the expressions in Figure 3-6). When coding this procedure, the required operations boil down to the computation of the determinant of a 3×3 coordinate matrix (one determinant for each coordinate):

$$\xi_1 = \frac{1}{2A}\begin{vmatrix} 1 & x & y \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}, \quad \xi_2 = \frac{1}{2A}\begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x & y \\ 1 & x_3 & y_3 \end{vmatrix}, \quad \xi_3 = \frac{1}{2A}\begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x & y \end{vmatrix} \qquad (3\text{-}30)$$

where ξ_1, ξ_2, ξ_3 are the mixing proportions, (x, y) are the PCA coordinates of the point whose mixing proportions are being calculated, and (x_i, y_i) are the PCA coordinates of the end-members.

For four end-members the polyhedron is a tetrahedron with four vertices, and the expressions for the mixing proportions are:

$$\xi_1 = \frac{1}{6A}\begin{vmatrix} 1 & x & y & z \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{vmatrix}, \quad \xi_2 = \frac{1}{6A}\begin{vmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x & y & z \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{vmatrix},$$

$$\xi_3 = \frac{1}{6A}\begin{vmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x & y & z \\ 1 & x_4 & y_4 & z_4 \end{vmatrix}, \quad \xi_4 = \frac{1}{6A}\begin{vmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x & y & z \end{vmatrix}. \qquad (3\text{-}31)$$

The generalization of this procedure to any number of end-members is straightforward and the matrix whose determinant is to be computed is in general a $d+1 \times d+1$ coordinate matrix, where d is the dimensionality of the PCA space and $d+1$ is the number of end-members:

$$\xi_1 = \frac{1}{d!A}\begin{vmatrix} 1 & x & y & z & \cdots & w \\ 1 & x_2 & y_2 & z_2 & \cdots & w_2 \\ 1 & x_3 & y_3 & z_3 & \cdots & w_3 \\ 1 & x_4 & y_4 & z_4 & \cdots & w_4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{d+1} & y_{d+1} & z_{d+1} & \cdots & w_{d+1} \end{vmatrix} \tag{3-32}$$

and similar expressions for coordinates ξ_2, ξ_3, ..., ξ_{d+1}. The meaning of each row and column of the determinant is explained in Figure 3-12.

To assess the ability of this routine to compute mixing proportions the same synthetic dataset employed to test the two-principal component mixing routine has been used. Figure 3-13 plots the results. This figure can be visually compared to Figure 3-8, which plots the real mixing proportions.

As was already done for the two-principal component mixing routine, we calculate the overall standard deviation using Equation (3-29).Maximum deviation for the hyperspace mixing routine is of the order of 2×10^{-5} percent and the mean standard deviation for the 1,000 samples is 7×10^{-6} percent (to be compared with maximum deviation of 85% and mean deviation of 17% for the two-principal component mixing routine), as Figure 3-14 shows. This value is equal to the precision of the calculation (six significant digits for a single-precision real number in its computer representation), so we can conclude that, as expected, *the n-principal component mixing routine has zero error when computing mixing proportions of synthetic samples*.

3.2.4 Allowance parameter

Any mixing calculation performed on a real, complex groundwater system, even in the ideal case of pure mixing (i.e. a groundwater system where the composition of the water is due only to dispersive mixing of several end-members), has several sources of uncertainty that contribute to the final accuracy of the computed mixing proportions. Among these sources, the most

$$\xi_1 = \frac{1}{d!V_d}\begin{array}{|cccccc|l} 1 & x & y & z & \cdots & w & \textbf{Sample's coord} \\ 1 & x_2 & y_2 & z_2 & \cdots & w_2 & \textbf{2}^{\textbf{nd}}\textbf{ end - member} \\ 1 & x_3 & y_3 & z_3 & \cdots & w_3 & \textbf{3}^{\textbf{rd}}\textbf{ end - member} \\ 1 & x_4 & y_4 & z_4 & \cdots & w_4 & \textbf{4}^{\textbf{th}}\textbf{ end - member} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \quad\vdots \\ 1 & x_{d+1} & y_{d+1} & z_{d+1} & \cdots & w_{d+1} & \textbf{n+1 end - member} \\ \hline & \textbf{PC1} & \textbf{PC2} & \textbf{PC3} & \bullet\bullet\bullet & \textbf{PC}\boldsymbol{n} & \end{array}$$

Figure 3-12. *Meaning of the rows and columns of the determinant used to compute the proportion of the first end-member in a sample. The proportions of the other end-members are computed by moving the sample's coordinates to the appropriate row (to the second row when computing the mixing proportion of the second end-member, to the third row when computing the mixing proportion of the third end-member, and so on).*

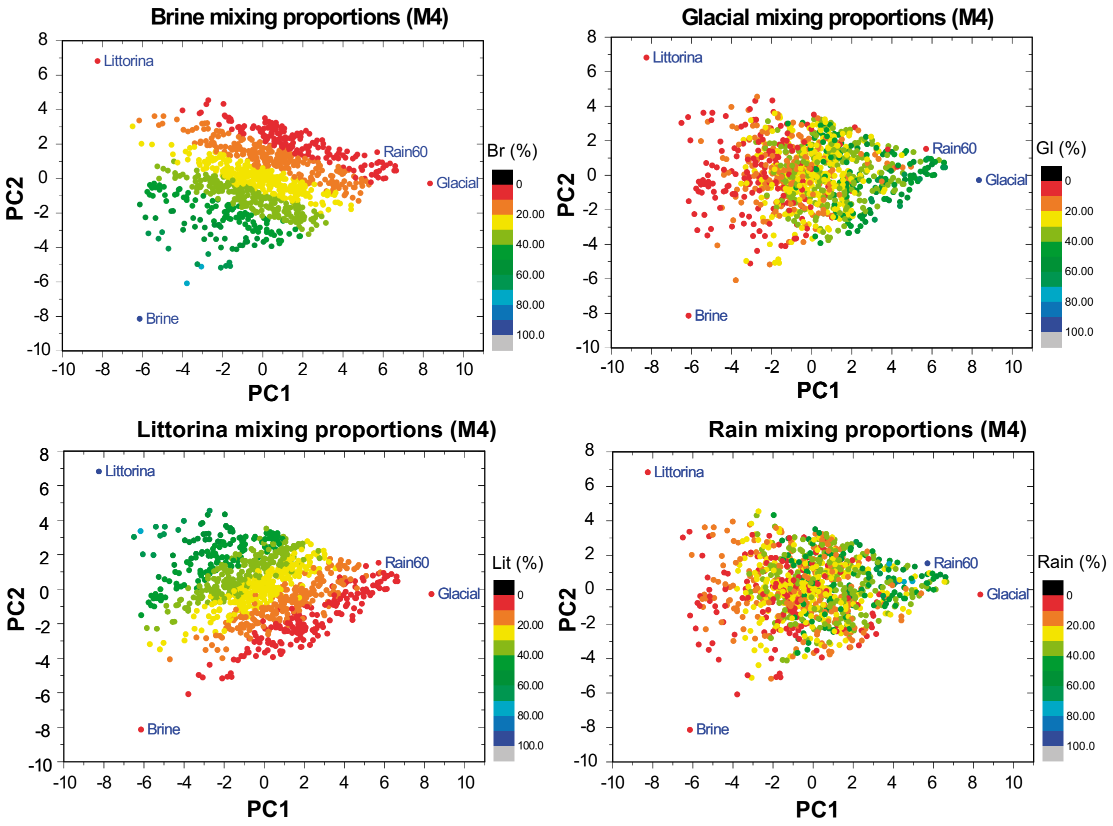


Figure 3-13. *The same synthetic samples as in Figure 3-8, but colour-coded with respect to the calculated proportion of each end-member. M3's hyperspace components mixing routine has been used to compute the mixing ratios.*

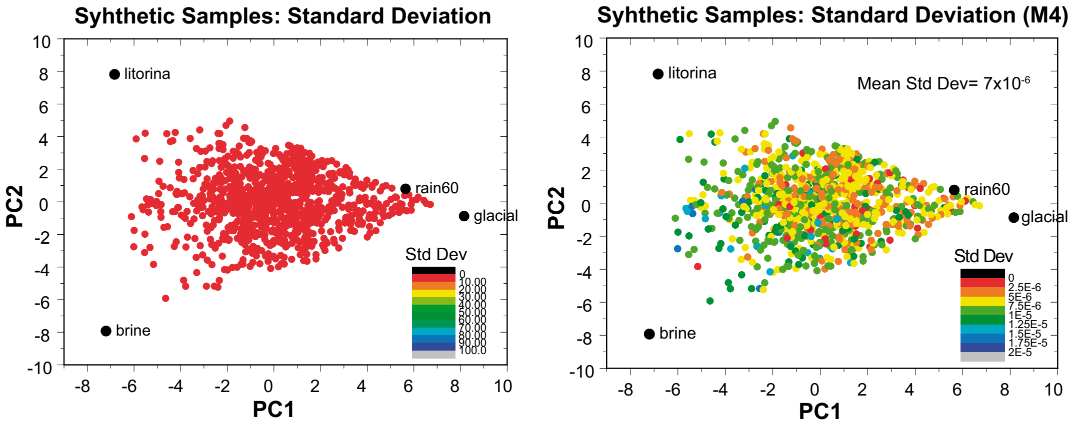


Figure 3-14. *Deviation of computed mixing proportions from real ones for 1,000 synthetic samples. M3's hyperspace mixing routine has been used for the calculations. The graph on the left uses the same 0–100% scale as Figure 3-10 to facilitate comparison. The graph on the right has a colour scale from 0 to 2×10*$^{-5}$ *percent to better appreciate how deviations are spread on the PCA plane.*

important are /Gómez et al. 2006/: (1) the number and composition of the water end-members that contribute to the chemical composition of the groundwater; (2) the spatial and/or temporal chemical variability of each end-member (end-members need not be spatially homogeneous or temporally constant); (3) the feasible chemical reactions that could contribute to the evolution of the groundwater system after each mixing event; (4) the conservative versus non-conservative behaviour of specific elements during mixing; and (5) analytical uncertainties.

When performing a mixing calculation (for which purpose a set of end-members has to be selected), some of the samples in the dataset could fall outside the mixing hyper-polyhedron, meaning that they cannot be explained by pure mixing of the chosen end-members for whichever of the 5 reasons pointed out in the previous paragraph. From this set of ”outsiders”, some samples will fall far from the ”walls” of the mixing hyper-polyhedron as they *simply* cannot be explained as a mixture of the selected end-members (case 1 above), but others will fall *just* outside the hyper-polyhedron, very close to any of its walls. These samples are *not* explained by mixing, but the reason in this case is almost certainly not the first one above, but any of the other uncertainties (2 to 5). In this case it is not unreasonable to “move” these samples to the nearest hyper-polyhedron wall and include them in the mixing calculations.

This procedure is implemented in M3 through the so-called *Allowance Parameter*. The allowance parameter, as its name suggests, allows for samples near the mixing hyper-polyhedron (but outside it) to be moved to the nearest wall. Essentially, the procedure is the following: Any sample inside the mixing hyper-polyhedron has all its PC coordinates positive; on the other hand, any sample outside the mixing hyper-polyhedron has at least one PC coordinate negative. If the negative PC coordinate is small compared to one, it means that its distance to a wall is small. So, setting the allowance parameter to, say, 0.05, would move to the nearest wall all the samples whose distance is less than 5% in terms of mixing proportions.

An example would clarify the meaning of the allowance parameter. Figure 3-15 shows a fictitious dataset (black dots) and four end-members (red circles), together with the mixing polygon projected onto the PC1-PC2 plane (shaded area). Four lines, each one parallel to one of the sides (projected wall) of the polygon, are drawn whose distance to the actual walls of the polygon is 5% in terms of mixing proportions. All the samples inside the blue extension of the polygon will be moved to the wall of the polygon if an allowance parameter of 0.05 is set in M3. Practically, this is carried out by setting to zero the negative mixing proportions of all samples with an absolute value equal or less than 0.05 and normalising to one the rest of the mixing proportions. For example, if a sample outside the mixing polyhedron has mixing proportions (0.500, 0.400, 0.120, −0.020), it would be reset to (0.490, 0.392, 0.118, 0.000). When a sample is moved to the wall of the mixing hyper-polyhedron by the allowance parameter, at least one of its mixing proportions is always zero.

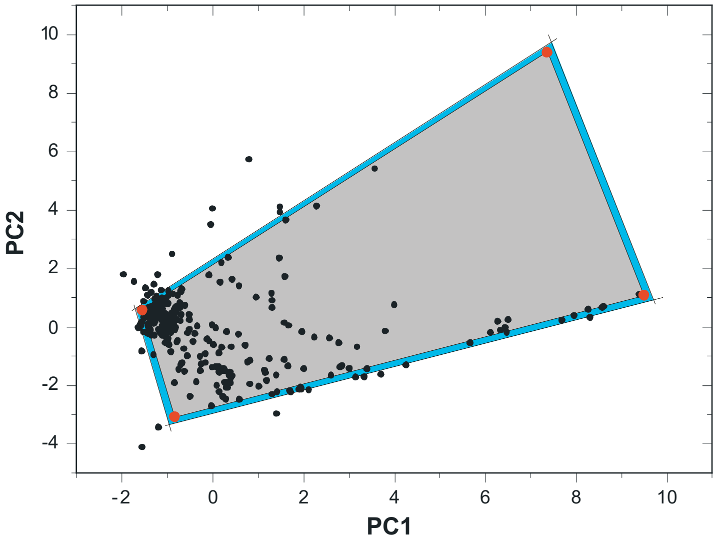


***Figure 3-15**. A fictitious dataset with four end-members projected onto the PC1-PC2 plane. The red circles are the end-members and define a polygon (shaded area) which includes all the samples than can be explained by the mixing of the four end-members. The blue extension of the polygon includes the samples that will be moved to the edges of the polygon if an allowance parameter different from zero is set in M3.*

M3 permits any value of the allowance parameter between 0 and 1, but allowance parameters greater than, say, 0.05 (5% in terms of mixing proportions) must be used with great care. A safe procedure is to repeat the calculations with a variable allowance parameter from zero to 5% in steps of 1% and check the coverage (i.e. the number of samples inside the mixing hyper-polyhedron) for each run. If the coverage differs only slightly between the runs with 0% and 5% allowance parameter, it would mean that most samples outside the mixing hyper-polyhedron are actually far from its walls. If, on the other hand, the coverage changes quickly when the allowance parameter is increased from 0 to 1%, and from 1 to 2%, but then the change becomes slower, it would mean that many samples are located very close to the walls of the mixing hyper-polyhedron and that they can safely be moved to the walls.

Finally, an allowance parameter of *exactly* zero is not recommended either, because rounding errors could translate during calculations into very small but negative mixing proportions for some samples. In summary, an allowance parameter in the range 0.01–0.05 (1–5%) is safest and therefore preferred.

3.3 Mass balance calculations

This section describes the calculations that M3 performs to estimate the compositional characteristics of each sample that can not be explained by pure conservative mixing.

In M3 mass balance calculations are always performed after mixing proportions have been computed. The meaning of this is the following: M3 assumes that mixing is the leading process (first order process) and tries to explain most of the chemical composition of the samples by conservative mixing. As already stated, all compositional variables included in the PCA are treated equally, without distinguishing between conservative and non-conservative variables. Once mixing proportions have been computed, M3 calculates the composition that a sample would have if mixing were the one and only process contributing to the chemical composition of the sample. This is done in the following straightforward way. For every sample:

- Take the mixing proportions computed in the previous step, ξ_i ($i = 1, ..., n$, where n is the number of end-members).
- Take the composition of each end-member (c_j^i, $i = 1, ..., n; j = 1, ..., m$; where n is the number of end-members and m the number of original variables). With this terminology $c_{\mathrm{Na}}^{\mathrm{Litt}}$ would be the Na content in the Litorina end-member and $c_{^{18}\mathrm{O}}^{\mathrm{Br}}$ the content of ^{18}O in the Brine end-member.
- Compute the expression

$$\hat{c}_j^{\mathrm{Sample}} = \sum_{i=1}^{n} c_j^i \times \xi_i. \qquad (3\text{-}33)$$

This gives the “theoretical” composition of each sample, The word “theoretical” should be understood as “the composition of the sample in the case of pure conservative mixing between the chosen end-members”. A particular case of the general expression (3-33) has already been used in Equation (3-28) to calculate the Na content of a synthetic sample.

- Subtract the real composition from the theoretical one, variable by variable, to get the absolute mass balance M:

$$M_j^{\mathrm{Sample}} = c_j^{\mathrm{Sample}} - \hat{c}_j^{\mathrm{Sample}}, \qquad (3\text{-}34)$$

  where c_j^{Sample} is the measured content of compositional variable j and $\hat{c}_j^{\mathrm{Sample}}$ is the computed one assuming pure conservative mixing.

Mass balance is carried out for each compositional variable *j*, not for the whole sample, i.e. a mass balance is computed for Na, another mass balance for Ca, and so on for all *j* compositional variables. Equation (3-34) gives the *absolute* mass balance, expressed in the same units as the measured variable (i.e. mg/L for concentrations of major elements), but M3 can also compute *relative* mass balances, *m*, just by dividing Equation (3-34) by the measured content of each compositional variable:

$$m_j^{\text{Sample}} = \frac{c_j^{\text{Sample}} - \hat{c}_j^{\text{Sample}}}{c_j^{\text{Sample}}} \times 100. \qquad (3\text{-}35)$$

The relative mass balance is expressed as a percentage deviation with respect to the actual value. For example, if a sample has a measured Na content of 2,500 mg/L and the predicted Na content is 2,000 mg/L, the absolute mass balance is +500 mg/L (the plus sign meaning that the actual Na content is *higher* than the computed one), and the relative mass balance +20%.

As in the previous section on mixing calculations, we have used the same synthetic dataset to compute absolute and relative mass balances from the mixing proportions delivered by the two-principal component and hyperspace mixing routines. Results for the absolute mass balance are shown in Figure 3-16. Chlorine has been selected because in real samples it should behave conservatively in most systems, and so its concentration depends only on mixing and not on reactions. In this way, mass balance should give values close to zero. For the synthetic dataset used here, mass balance should be exactly zero if mixing proportions have been correctly computed, as all the samples were created by pure conservative mixing of the four end-members.

As Figure 3-16 shows, this is correct for the mass balance calculated from the hyperspace mixing proportions, where the maximum absolute deviation is of the order of 0.008 mg/L, but not the case for the mass balance calculated by means of the two-principal component mixing proportions, which have maximum deviations of the order of 650 mg/L.

If we take a look at the *relative* deviations (Figure 3-17), the picture is similar, with a maximum relative deviation of 49% and an average of 2% for the two-principal component mixing (left-hand graph) and of 0.0001% (average of 2.2×10^{-6}) for hyperspace mixing (right-hand graph). The quoted deviations for the hyperspace mixing routine are below numerical error (i.e. differences are due to numerical rounding error when subtracting two numbers that are close to the machine precision).

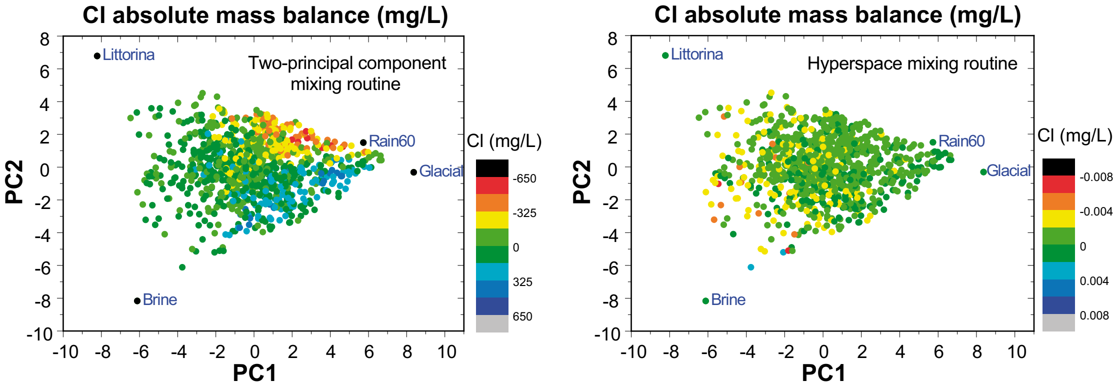


***Figure 3-16**. Absolute mass balance for Cl using the synthetic dataset. The graph on the left is for the two-principal component mixing routine, and the one on the right for the hyperspace mixing routine. Notice the change in colour bar scale from (–650 mg/L, +650 mg/L) in the left graph to (–0.008 mg/L, +0.008 mg/L) in the right graph, a difference of almost five orders of magnitude.*

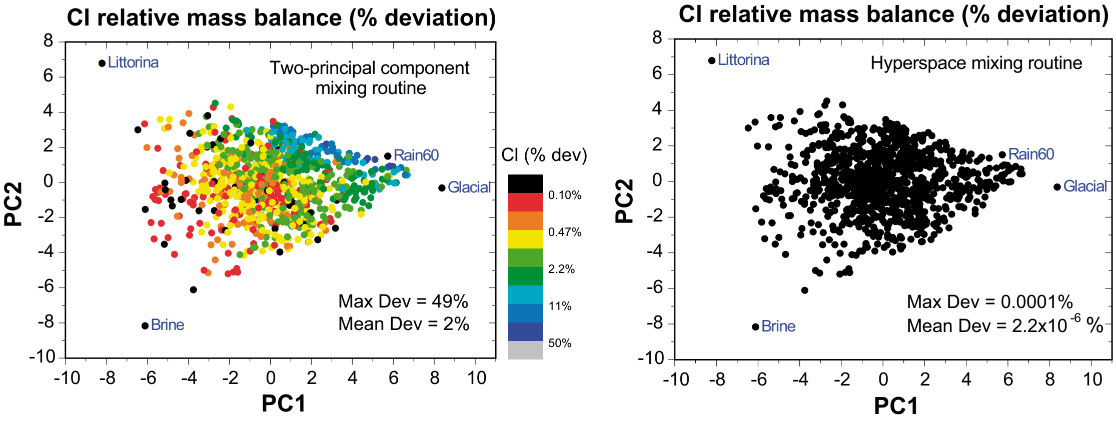


***Figure 3-17**. Relative mass balance for Cl using the synthetic dataset. The graph on the left is for the two-principal component mixing routine, and the one on the right for the hyperspace mixing routine. Scale bar is common to both graphs. Black dots are samples with a relative deviation of less than 0.1%.*

4 Modules

This section is dedicated to explain two extensions of M3 intended to facilitate two of the most critical steps in the whole M3 methodology, namely: (1) how to decide which end-members to use, given a dataset; and (2) what is the precision of the computed mixing proportions, given a correctly-selected but compositionally-variable set of end-members (i.e. the proper end-members have been selected but these are defined by a *compositional range* instead of by a specific composition).

The first question is addressed by the End-member Selection Module (ESM). The ESM (Section 4.1) accepts as an input a fairly large number (< 15 for practical purposes) of *potential end-members* and gives as an output the number of *explained* samples for each combination of potential end-members. The "best" combination of end-members is the one that can explain, only by mixing, the largest number of samples in the dataset. This module can be accessed through the Calculations/End-members/Select End-members menu, clicking on the "Test" button after having selected the end-members.

The second questions is taken by the End-member Variability Module (EVM), described in Section 4.2. The EVM accepts as an input a range of compositions for each end-member (the end-members have been already selected, either by expert judgment or with the help of the ESM) and construct from every range an input probability distribution (in terms of compositions), which is propagated through the PCA and converted into an output probability for the mixing proportions. This module is accessed directly via the EVM menu item in M3.

4.1 End-member selection module (ESM)

If the end-members have been correctly chosen for the given dataset, all samples should plot inside the $n-1$-dimensional hyper-volume defined by the n end-members. This is a generalization of the property of triangular coordinates: any point *inside* the triangle has all three triangular coordinates *positive*, and any point outside the triangle has, at least, one *negative* coordinate. With this rule in mind, it is very simple to know which samples are inside the hyper-volume defined by the n end-members: those with all their mixing proportions positive. Only these samples can be "explained" by pure conservative mixing of the selected end-members. The closer the number of explained samples is to the total number of samples, the better the combination is (in the sense that more samples can be explained as a mixing of the selected end-members). In other words, a set of end-members are properly selected for an specific dataset when most of the waters in the dataset are inside the mixing hyper-tetrahedron.

How many combinations of end-members are probed? M3 performs a systematic search of combinations, starting from two end-members and ending with the maximum number of end-members included as input potential end-members. The total number of combinations grows rapidly with this maximum number, and that is why 15 is a practical upper limit. From combinatorial theory, the number C of *combinations* of n elements taken r at a time is

$$C(n,r) \equiv \begin{pmatrix} n \\ r \end{pmatrix} = \frac{n!}{(n-r)!r!}. \qquad \text{(4-1)}$$

In our case, n is the maximum number of potential end-members, and r is the number of end-members of any particular combination, which goes from 2 to n. So, the total number of combinations of n end-members taken from two at a time to n at a time is then

$$C(n) = \sum_{r=2}^{n} \frac{n!}{(n-r)!r!}. \qquad \text{(4-2)}$$

Table 4-1 gives the number of combinations starting with 2 end-members and ending with 15. Each entry in the table gives the result of Equation (4-1) for the selected *n* (column) and *r* (row). For example, there are 56 combinations of 8 end-members taken 5 at a time, i.e, there are 56 combinations containing five end-members when the number of input potential end-members is 8. The "Total" entry in each column gives the result of Equation (4-2), i.e. the total number of combinations of *n* potential end-members when all the subsets from two to *n* end-members are included. This number of combinations (bottom row in Table 4-1) is the one M3 tries when selecting *n* potential end-members from the start.

Table 4-2 is an example of the output of the ESM for 119 samples from the Forsmark area /Laaksoharju 2005/. Each row is a combination of end-members selected from the set of potential end-members formed by Brine, Litorina, Glacial, Rain60, Lake Water, Sea Sediment, and Baltic Sea. In this particular example $n = 7$, so we have a total of 120 possible combinations: 21 of them are combinations of two end-members; 35 are combinations of 3 end-members; another 35 are combinations of 4 end-members; 21 more are combinations of 5 end-members; 7 are combinations of 6 end-members and finally there is just one combination of all the seven potential end-members. The last column of Table 4-2 is the total number of samples, out of 119, with all the mixing proportions positive (i.e. the number of samples that plot inside the hyper-tetrahedron in $n-1$-dimensional space; for the special case of two end-members, "inside" means "between the two end-members"; for example, the combination Brine + Glacial explains all 119 samples because all samples fall on the line segment (1-dimensional) between end-members Brine and Glacial and none to the left of Glacial or to the right of Brine; on the other hand, the combination Brine + Glacial + Litorina explains only 4 samples all 115 remaining waters are outside the two-dimensional triangle whose vertices are the end-members Brine, Glacial and Litorina; there is no contradiction in having fewer samples explained when one extra end-member is added, as the resulting geometrical construct "lives" in one more dimension).

Table 4-1. Number of combinations of end-members. Each column is for a maximum number of potential end-members and each row for the actual number of end-members included in the computation.

r	**n**													
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
2	1	3	6	10	15	21	28	36	45	55	66	78	91	105
3		1	4	10	20	35	56	84	120	165	220	286	364	455
4			1	5	15	35	70	126	210	330	495	715	1,001	1,365
5				1	6	21	56	126	252	462	792	1,287	2,002	3,003
6					1	7	28	84	210	462	924	1,716	3,003	5,005
7						1	8	36	120	330	792	1,716	3,432	6,435
8							1	9	45	165	495	1,287	3,003	6,435
9								1	10	55	220	715	2,002	5,005
10									1	11	66	286	1,001	3,003
11										1	12	78	364	1,365
12											1	13	91	455
13												1	14	105
14													1	15
15														1
Total	**1**	**4**	**11**	**26**	**57**	**120**	**247**	**502**	**1,013**	**2,036**	**4,083**	**8,178**	**16,369**	**32,752**

From the table is simple to select the "most promising" combinations, in the sense that they are the combinations of end-members that are able to explain, only by mixing, the highest percentage of samples. In the example below, combinations #38 (Glacial + Litorina + Lake Water) and #45 (Glacial + Lake Water + Baltic) are the best combinations with three end-members; and combinations #58 (Brine + Glacial + Litorina + Lake Water) and #65 (Brine + Glacial + Lake Water + Baltic) the best with four end-members. With five end-members only combination #101 (Brine + Glacial + Lake Water + Sea Sediment + Baltic) is rather good, explaining 95 out of 119 samples (80%). No combination with 6 or 7 end-members is able to explain more than 45% of the samples. Of course, the final selection of end-members will be based also on independent geochemical and hydrochemical arguments pertaining the specific site.

Table 4-2. Number of samples than can be "explained" by each possible combination of end-members, from two to seven end-members. Samples from Forsmark 1.2 Local Model, 119 samples /Laaksoharju 2005/.

1	Brine	Glacial	–	–	–	–	–	119
2	Brine	Litorina	–	–	–	–	–	0
3	Brine	Rain60	–	–	–	–	–	119
4	Brine	Lake Water	–	–	–	–	–	52
5	Brine	Sediment	–	–	–	–	–	31
6	Brine	Baltic	–	–	–	–	–	0
7	Glacial	Litorina	–	–	–	–	–	119
8	Glacial	Rain60	–	–	–	–	–	18
9	Glacial	Lake Water	–	–	–	–	–	86
10	Glacial	Sediment	–	–	–	–	–	115
11	Glacial	Baltic	–	–	–	–	–	119
12	Litorina	Rain60	–	–	–	–	–	119
13	Litorina	Lake Water	–	–	–	–	–	49
14	Litorina	Sediment	–	–	–	–	–	26
15	Litorina	Baltic	–	–	–	–	–	0
16	Rain60	Lake Water	–	–	–	–	–	72
17	Rain60	Sediment	–	–	–	–	–	115
18	Rain60	Baltic	–	–	–	–	–	119
19	Lake Water	Sediment	–	–	–	–	–	41
20	Lake Water	Baltic	–	–	–	–	–	49
21	Sediment	Baltic	–	–	–	–	–	29
22	Brine	Glacial	Litorina	–	–	–	–	4
23	Brine	Glacial	Rain60	–	–	–	–	48
24	Brine	Glacial	Lake Water	–	–	–	–	92
25	Brine	Glacial	Sediment	–	–	–	–	4
26	Brine	Glacial	Baltic	–	–	–	–	34
27	Brine	Litorina	Rain60	–	–	–	–	66
28	Brine	Litorina	Lake Water	–	–	–	–	30
29	Brine	Litorina	Sediment	–	–	–	–	4
30	Brine	Litorina	Baltic	–	–	–	–	0
31	Brine	Rain60	Lake Water	–	–	–	–	58
32	Brine	Rain60	Sediment	–	–	–	–	33
33	Brine	Rain60	Baltic	–	–	–	–	64
34	Brine	Lake Water	Sediment	–	–	–	–	5
35	Brine	Lake Water	Baltic	–	–	–	–	30
36	Brine	Sediment	Baltic	–	–	–	–	29

37	Glacial	Litorina	Rain60	–	–	–	–	84
38	Glacial	Litorina	Lake Water	–	–	–	–	115
39	Glacial	Litorina	Sediment	–	–	–	–	0
40	Glacial	Litorina	Baltic	–	–	–	–	25
41	Glacial	Rain60	Lake Water	–	–	–	–	54
42	Glacial	Rain60	Sediment	–	–	–	–	54
43	Glacial	Rain60	Baltic	–	–	–	–	84
47	Litorina	Rain60	Lake Water	–	–	–	–	36
48	Litorina	Rain60	Sediment	–	–	–	–	38
49	Litorina	Rain60	Baltic	–	–	–	–	1
50	Litorina	Lake Water	Sediment	–	–	–	–	30
51	Litorina	Lake Water	Baltic	–	–	–	–	1
52	Litorina	Sediment	Baltic	–	–	–	–	0
53	Rain60	Lake Water	Sediment	–	–	–	–	48
54	Rain60	Lake Water	Baltic	–	–	–	–	36
55	Rain60	Sediment	Baltic	–	–	–	–	37
56	Lake Water	Sediment	Baltic	–	–	–	–	30
57	Brine	Glacial	Litorina	Rain60	–	–	–	79
58	Brine	Glacial	Litorina	Lake Water	–	–	–	116
59	Brine	Glacial	Litorina	Sediment	–	–	–	1
60	Brine	Glacial	Litorina	Baltic	–	–	–	1
61	Brine	Glacial	Rain60	Lake Water	–	–	–	16
62	Brine	Glacial	Rain60	Sediment	–	–	–	50
63	Brine	Glacial	Rain60	Baltic	–	–	–	65
64	Brine	Glacial	Lake Water	Sediment	–	–	–	88
65	Brine	Glacial	Lake Water	Baltic	–	–	–	116
66	Brine	Glacial	Sediment	Baltic	–	–	–	2
67	Brine	Litorina	Rain60	Lake Water	–	–	–	47
68	Brine	Litorina	Rain60	Sediment	–	–	–	51
69	Brine	Litorina	Rain60	Baltic	–	–	–	3
70	Brine	Litorina	Lake Water	Sediment	–	–	–	34
71	Brine	Litorina	Lake Water	Baltic	–	–	–	1
72	Brine	Litorina	Sediment	Baltic	–	–	–	1
73	Brine	Rain60	Lake Water	Sediment	–	–	–	56
74	Brine	Rain60	Lake Water	Baltic	–	–	–	51
75	Brine	Rain60	Sediment	Baltic	–	–	–	45
76	Brine	Lake Water	Sediment	Baltic	–	–	–	34
77	Glacial	Litorina	Rain60	Lake Water	–	–	–	68
78	Glacial	Litorina	Rain60	Sediment	–	–	–	2
79	Glacial	Litorina	Rain60	Baltic	–	–	–	0
80	Glacial	Litorina	Lake Water	Sediment	–	–	–	18
81	Glacial	Litorina	Lake Water	Baltic	–	–	–	61
82	Glacial	Litorina	Sediment	Baltic	–	–	–	0
83	Glacial	Rain60	Lake Water	Sediment	–	–	–	40
84	Glacial	Rain60	Lake Water	Baltic	–	–	–	69
85	Glacial	Rain60	Sediment	Baltic	–	–	–	12
86	Glacial	Lake Water	Sediment	Baltic	–	–	–	58
87	Litorina	Rain60	Lake Water	Sediment	–	–	–	30
88	Litorina	Rain60	Lake Water	Baltic	–	–	–	1
89	Litorina	Rain60	Sediment	Baltic	–	–	–	0

90	Litorina	Lake Water	Sediment	Baltic	–	–	–	2
91	Rain60	Lake Water	Sediment	Baltic	–	–	–	29
92	Brine	Glacial	Litorina	Rain60	Lake Water	–	–	81
93	Brine	Glacial	Litorina	Rain60	Sediment	–	–	8
94	Brine	Glacial	Litorina	Rain60	Baltic	–	–	1
95	Brine	Glacial	Litorina	Lake Water	Sediment	–	–	29
96	Brine	Glacial	Litorina	Lake Water	Baltic	–	–	48
97	Brine	Glacial	Litorina	Sediment	Baltic	–	–	3
98	Brine	Glacial	Rain60	Lake Water	Sediment	–	–	40
99	Brine	Glacial	Rain60	Lake Water	Baltic	–	–	74
100	Brine	Glacial	Rain60	Sediment	Baltic	–	–	23
101	Brine	Glacial	Lake Water	Sediment	Baltic	–	–	95
102	Brine	Litorina	Rain60	Lake Water	Sediment	–	–	31
103	Brine	Litorina	Rain60	Lake Water	Baltic	–	–	1
104	Brine	Litorina	Rain60	Sediment	Baltic	–	–	0
105	Brine	Litorina	Lake Water	Sediment	Baltic	–	–	2
106	Brine	Rain60	Lake Water	Sediment	Baltic	–	–	30
107	Glacial	Litorina	Rain60	Lake Water	Sediment	–	–	33
108	Glacial	Litorina	Rain60	Lake Water	Baltic	–	–	27
109	Glacial	Litorina	Rain60	Sediment	Baltic	–	–	0
110	Glacial	Litorina	Lake Water	Sediment	Baltic	–	–	19
111	Glacial	Rain60	Lake Water	Sediment	Baltic	–	–	32
112	Litorina	Rain60	Lake Water	Sediment	Baltic	–	–	1
113	Brine	Glacial	Litorina	Rain60	Lake Water	Sediment	–	36
114	Brine	Glacial	Litorina	Rain60	Lake Water	Baltic	–	13
115	Brine	Glacial	Litorina	Rain60	Sediment	Baltic	–	2
116	Brine	Glacial	Litorina	Lake Water	Sediment	Baltic	–	31
117	Brine	Glacial	Rain60	Lake Water	Sediment	Baltic	–	57
118	Brine	Litorina	Rain60	Lake Water	Sediment	Baltic	–	1
119	Glacial	Litorina	Rain60	Lake Water	Sediment	Baltic	–	1
120	Brine	Glacial	Litorina	Rain60	Lake Water	Sediment	Baltic	2

4.2 End-member variability module (EVM)

A procedure has been developed an implemented in M3 v3.0 to assess the impact of the compositional variability of water end-members on the calculated mixing proportions. This section describes the procedure and the results obtained using Laxemar 1.2 data set (Local Model, 356 superficial and groundwaters, /Laaksoharju 2005/).

4.2.1 Motivation

The calculation of water mixing proportions by means of a PCA analysis is a well established and useful practice when dealing with a big number of samples /Laaksoharju and Skårman 1995ab, Smellie and Karlsson 1996, Laaksoharju and Wallin 1997, Gurban et al. 1998, Laaksoharju et al. 1995, 1999abcd, 2000, 2004, Laaksoharju 2004/. However, this type of analysis has the drawback of *a priori* selecting a set of end-members, whose number and composition are fixed in advance. A preliminary exploratory analysis can, in principle, identify the end-members to be used, but this selection, mainly based on expert judgment, is always tricky and, in many cases, difficult to justify /Bath and Jackson 2002, Svensson et al. 2002/.

The selection is even more critical when the mixing proportions coming out of the PCA analysis are to be used by hydrogeologists to "transport" them spatially through the system (using flow lines) or through time (to predict future changes in water composition).

4.2.2 General procedure

In order to overcome these difficulties, the EVM takes into account the intrinsic compositional variability of the end-members. The procedure starts from a pre-selected number of end-members, i.e. no attempt is made here to define *which* end-members to be use in the analysis (the selection could be made by expert judgment or using the ESM described in Section 4.1), and has the following steps (Figure 4-1):

1. Define the compositional variability of the end-members. This can be done by defining compositional *ranges* for each input compositional variable.
2. Construct a probability density function for each compositional variable. This probability density functions are called *input probabilities*.
3. Generate, according to the chosen input probabilities, a large number of end-member compositions.
4. For each run one of the calculated end-member composition is used to compute the mixing proportions of selected samples.
5. After all runs have been finished, mixing proportions for each sample are binned to construct the *output probability distributions*.

What follows is a brief summary of each step.

Definition of the compositional range for the selected end-members

Each end-member is characterized by two samples, real or synthetic, that represent the maximum variability expected for that end-member. The selection is made by expert judgment, taking into account all the geochemical and hydrological knowledge of the system. Note that this is not the same as selecting, by expert judgment, a fixed composition for the end-member, as the standard PCA analysis does. Here expert judgment selects a *range* of compositions, relaxing in this way the requirement of knowing the *exact* composition of each end-member to be used in the mixing calculations. Table 4-3 summarises the ranges that have been defined for the end-members in the modelling of the Local Model of Laxemar 1.2 dataset /Laaksoharju 2005/.

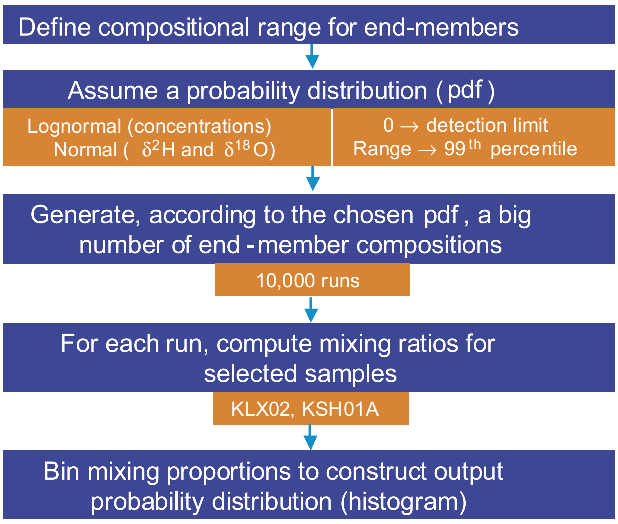


Figure 4-1. *Flowchart of the procedure implemented to assess the impact that the compositional variability of end-members has on mixing proportions.*

Table 4-3. Compositional ranges of the end-members used in Laxemar 1.2 PCA mixing modelling /Laaksoharju 2005/.

End-member	Na (mg/l)	K (mg/l)	Ca (mg/l)	Mg (mg/l)	HCO_3 (mg/l)	Cl (mg/l)	SO_4 (mg/l)	^2H (dev)	^3H (TU)	^{18}O (dev)
Brine 1	8,500	45.5	19,300	2.12	14.1	47,200	906	–44.9	0.00	–8.9
Brine 2	9,540	28	18,000	130	8.2	45,200	8.4	–49.5	0	–9.3
Glacial 1	0.17	0.4	0.18	0.1	0.12	0.5	0.5	–158	0.00	–21
Glacial 2	0.17	0.4	0.18	0.1	0.12	0.5	0.5	–125	0	–17
Litorina 1	3,674	134	151	448	93	6,500	890	–38	0.00	–4.7
Litorina 2	1,960	95	93.7	234	90	3,760	325	–53.3	0.00	–5.9
Rain 1	0	0	0	0	0	0	0	–125	0	–17
Rain 2	0	0	0	0	0	0	0	–44	168	–6.9
DGW 1	19.2	3	38.5	3.8	162	12	21.5	–68.4	11.913	–9.9
DGW 2	237	4	25	6	370	119	118	–73.8	0.775	–9.9

The extreme values for each end-member have been selected in the following way:

- **Brine 1** corresponds to the most saline sample found in Laxemar and is characterised by a very high sulphate content.
- **Brine 2** corresponds to the sample KRA/860/2 (from Finland) which is considered the most saline sample in Finland and is characterised by a low sulphate content.
- **Glacial 1** corresponds to a glacial melt water present in the system several thousand years ago.
- **Glacial 2** corresponds to a modern glacial melt water (different values for stable isotopes).
- **Litorina 1** corresponds to the theoretical composition of the Litorina Sea, so it represents an old Baltic sea water.
- **Litorina 2** corresponds to the present Baltic waters.
- **Rain 1** corresponds to an old winter rain with no tritium and very low values for the stable isotopes.
- **Rain 2** corresponds to the summer rain in the sixties with the maximum tritium and the highest values for the stable isotopes.
- **DGW 1** corresponds to a sample from the Bockholmen subarea (borehole HBH05) which is representative of a diluted granitic groundwater.
- **DGW 2** corresponds to a sample from the Äspö subarea (borehole HAS05) which is representative of a diluted granitic groundwater. Together with the DGW 1 cover the whole range of chemical and isotopic compositions of very shallow granitic groundwaters, below the overburden.

Mixing calculations have been carried out in Laxemar 1.2 with four end-members: Brine + Glacial + Litorina + Rain when dealing with the whole data set (superficial *and* groundwaters, 356 samples), and Brine + Glacial + Litorina + Dilute Groundwater when dealing only with the groundwaters (158 samples).

Construction of the input probabilities

For the definition of the probability density functions (pdfs) that characterize the compositional variation of each end-member the EVM adopts the following two assumptions: (1) all compositional variables follow a *log-normal distribution*; and (2) the ranges listed on Table 4-3 are equated to the *1st* and *99th percentiles* of the chosen probability function, which means that M3 allows for end-member compositions outside the reported range in Table 4-3 (with a probability of 1%). The log-normal distribution has been chosen because compositional variables are skewed non-negative physical quantities /Mishra 2002/. For example, Na in the Brine end-member (Table 4-3) has a lower Na content of 8,500 mg/L and an upper Na content of 9,540 mg/L. The value of 8,500 mg/L is equated to the 1st percentile and the value of 9,540 mg/L to the 99th percentile. The mean μ and standard deviation σ of the log-normally distributed Na content are /Mishra 2002/.

$$\mu = \exp\left(\alpha + \frac{\beta^2}{2}\right) \tag{4-3}$$

and

$$\sigma = \mu\sqrt{\exp(\beta^2) - 1}, \tag{4-4}$$

where α and β are the mean and standard deviation of the *log-transformed* variable; in other words, α and β are the mean and standard deviation of the normally distributed variable $\ln(x)$. Parameters α and β can be computed from the ranges in Table 4-2. For Na in the Brine end-member we have:

$$\alpha = \frac{\ln(9540) + \ln(8500)}{2},$$

$$\beta = \frac{\ln(9540) - \ln(8500)}{2 \times 2.576}.$$

In the latter equation the factor 2×2.576 is the number of standard deviations needed to include 99% of the area under a normal distribution, from the 1st to the 99th percentiles (i.e. 99% of the area under the normal distribution falls between $\alpha - 2.576\beta$ and $\alpha + 2.576\beta$). The area under a normal distribution is computed by means of the error function, $\text{erf}(x)$,

$$\text{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_{-x}^{x} e^{-t^2/2} dt, \tag{4-5}$$

and it can be shown that erf (2.576) = 0.9900. Figure 4-2 shows this in graphical form for a Gaussian distribution of zero mean and standard deviation of one.

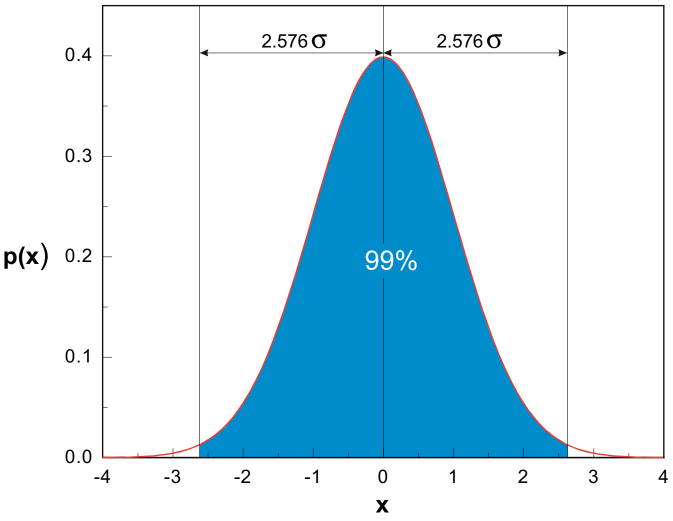


***Figure 4-2**. Normalised Gaussian distribution (zero mean and standard deviation of one) showing how the area under the curve is related to the distance from the mean, in unit of the standard deviation. 99% of the area is located between the 1st and the 99th percentiles (vertical lines). The first percentile is at a distance of 2.576σ to the left of the mean and the 99th percentile a distance of 2.576σ to the right of the mean.*

For the example given above, $\alpha = 9.1055$, $\beta = 0.022446$ and, therefore, $\mu = 9{,}007.0$ mg/L and $\sigma = 202.2$ mg/L (see the Brine pdf in the first panel of Figure 4-3). These are the two parameters that are needed to define the log-normal probability density function

$$f(x) = \frac{1}{x\beta\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{\ln x - \alpha}{\beta}\right]^2 \right), \qquad (4\text{-}6)$$

where x can have only positive values ($0 < x < +\infty$).

Once a probability function has been chosen and the statistical meaning of the empirical compositional range defined (in this case, it is equated to the 1st and 99th percentiles), the input probability functions are completely characterized. They are shown in graphical format in Figure 4-3. This figure plots the pdfs for all ten compositional variables used in the PCA mixing analysis: seven chemical species (Na, K, Ca, Mg, HCO_3, Cl, and SO_4), two stable isotopes (^2H and ^{18}O), and ^3H. The probability density functions have been constructed binning 10,000 values for each compositional variable and normalizing to ensure that the area under each curve is one. In the particular case of an end-member with identical upper and lower bounds for a compositional variable, the resulting probability density function is a delta function centred in the corresponding value (e.g. the $\delta^{18}O$ value for DGW in Table 4-3 is –9.9 for both the lower and upper limits; obviously, all realizations of this compositional variable will result in a $\delta^{18}O$ value of –9.9 for the end-member DGW).

The concentrations of Na, K, Ca, Mg, HCO_3, Cl, SO_4 and ^3H have been approximated by a pdf as explained above. However, δ^2H and $\delta^{18}O$ pose a particular problem because they are delta-values, calculated as the ratio of two concentrations and then normalized with respect to a reference concentration, which means that they can have both positive and negative values. This is incompatible with a lognormal distribution, which is only defined for positive real numbers. Instead, a Gaussian distribution has been used for δ^2H and $\delta^{18}O$, which is reasonable as their distribution is not skewed. The Gaussian or normal pdf has the form:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left[\frac{x - \mu}{\sigma}\right]^2 \right), \qquad (4\text{-}7)$$

where μ and σ are the mean and standard deviation of the normally-distributed variable x. Random variable x can take on any real value ($-\infty < x < +\infty$). For example, the mean and standard deviation of $\delta^{18}O$ in the Rain end-member are $\mu = (-17-6.9)/2 = -11.95$‰ and $\sigma = |-17-(-6.9)|/(2\times2.576) = 1.96$‰, and are shown in the last panel in Figure 4-3. Obviously, it is a symmetric distribution (to be compared to the positively-skewed log-normal distributions in the other panels).

Generation of end-member compositions

Once the input probability density functions are defined, the EVM generates a large number of compositions for each end-member. These compositions are randomly sampled from the corresponding pdf and fed into the PCA analysis. Figure 4-4 is a PC1-PC2 plot of the Laxemar 1.2 groundwater dataset (158 samples), where 10,000 compositions have been generated using end-members Brine + Glacial + Litorina + Dilute Groundwater. The end-member compositions are plotted in different colours while the position of the samples are indicated by a black dot. For the log-normal compositional variables (all input variables except ^2H and ^{18}O) the procedure to generate a log-normal random deviate is the following: (1) generate a random (Gaussian) deviate y with mean α and standard deviation β; (2) compute $x = e^y$; then x is a log-normal deviate with mean μ (Equation 4-3) and standard deviation σ (Equation 4-4). Basically, once each new realization of the composition of the end-members is produced, a principal component analysis of all the samples in the dataset (plus the end-members) is carried out and the PC coordinates of every water sample in the dataset saved into a file. Because the end-members are slightly different each time the PCA is carried out, the PC coordinates of the samples are also slightly different each time, and this is why each water sample is represented in Figure 4-4 not

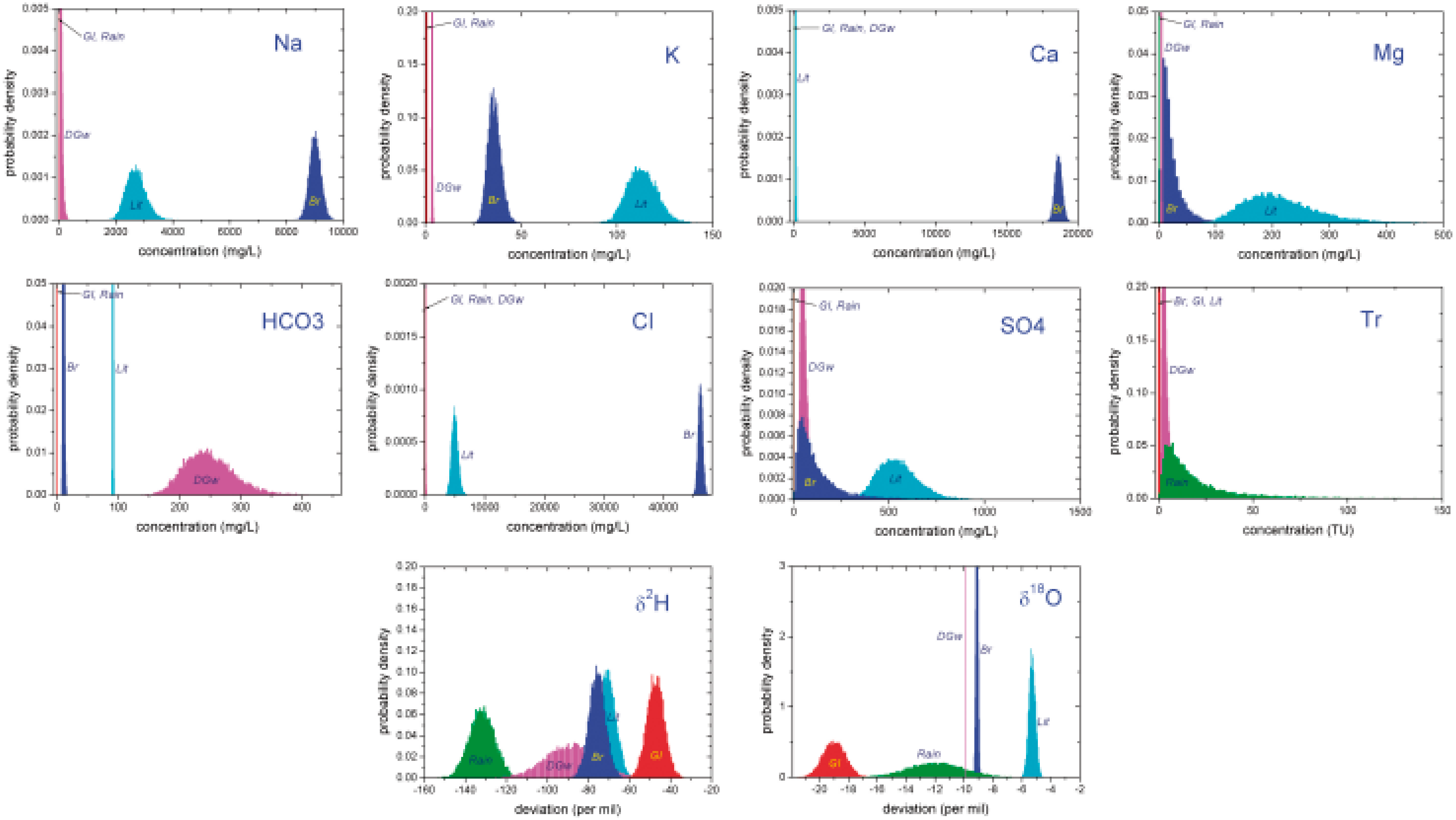


Figure 4-3. *Input probability density functions for all the chemical and isotopic variables considered in the PCA, as constructed from the compositional ranges of the end-members Brine (blue), Glacial (red), Litorina (cyan), Rain (green) and Dilute Groundwater (magenta).*

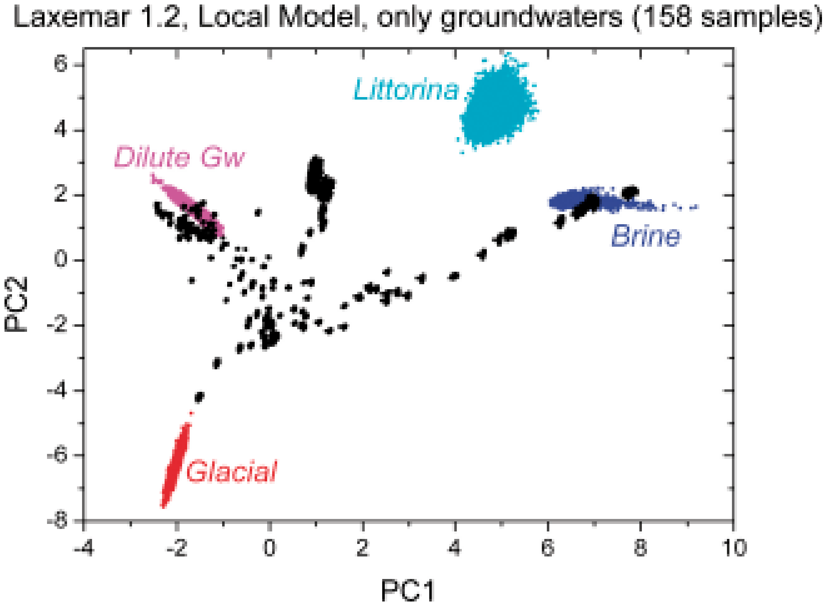


***Figure 4-4**. PC-plot for Laxemar 1.2 Local Model data set (only groundwaters, 158 samples). Each end-member is represented by 10,000 compositions (coloured dots) taken from a pdf defined by a predefined compositional range. Black dots correspond to the samples. These plots are like the superposition of 10,000 individual PC-plots, each computed with a different set of compositions for the end-members.*

by one dot, but by as many dots as different compositions of the end-members were generated (actually, in Figure 4-4 only one of every 10 dots is plotted so as not to clutter the graph with black dots; for the coloured dots, corresponding to the end-members, all 10,000 realizations are plotted).

The last point that has been taken into account to generate the random input compositional values for the end-members is the known correlation that exists between ^{2}H and ^{18}O delta-values in non-saline waters (Figure 4-5). In practice this means that M3 gives a random value to ^{18}O according to its input pdf and then computes the ^{2}H delta-value using the regression quoted in Figure 4-5. This regression is a weighted least-squared fit to 1,660 superficial and groundwater samples from the Baltic Shield. To the predicted deuterium value M3 adds a random Gaussian deviate of ±3.5 per mil, which is the dispersion of deuterium values around the best-fit line (inset in Figure 4-5). The EVM uses this method for fresh and brackish waters, but not for brines (Cl > 25,000 mg/L), as they do not follow the regression line but plot above it, inside the blue ellipse in Figure 4-5. For these waters the values of ^{2}H and ^{18}O coming out of the random sampling of the input pdfs are used directly.

Computation of mixing proportions

Once the PCA coordinates of each sample are known, the EVM computes their mixing proportions by means of one of the two mixing routines implemented in M3 (the two-principal component or the hyperspace mixing routines). The procedure has been described Section 3.2.

Construction of the output probability densities

Each M3 run gives, for each sample, a set of mixing proportions. For example, run #234 gives, say, the following mixing proportions for sample #15: Brine = 11.3%, Glacial = 58.6%, Litorina = 12.1%, and Dilute Groundwater = 18.0%. In other words, the 10,000 runs, *each with a different composition for the end-members,* give 10,000 different mixing proportions for sample #15 (and of course for any other selected sample). In this way M3 can assess how the variability

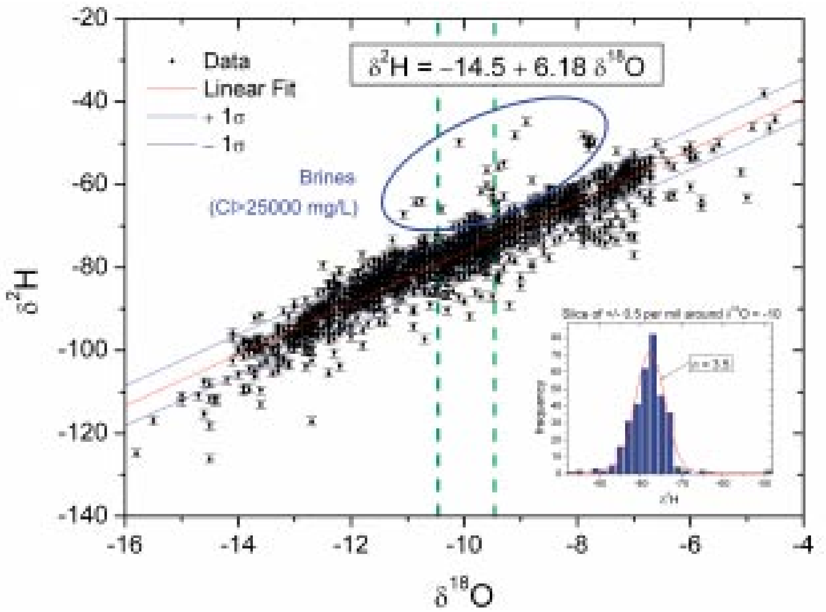


Figure 4-5. δ^2H–$\delta^{18}O$ *plot of 1,660 groundwater and superficial samples of the Baltic Shield. The red line is a weighted least square fit to the data assuming a* δ^2H *error equal to the detection limit (2‰). The blue lines are the ± 1σ bounds (68.3% of the data points fall between the blue lines). The inset shows, for a slice centred around* $\delta^{18}O = -10$ *(green vertical dashed lines), the dispersion of the data points, which is ± 3.5‰ in terms of the standard deviation. M3 assumes that* $\delta^{18}O$ *values are exact (no error in the horizontal axis). This correlation is applied to all samples with a Cl content less than 25,000 mg/L.*

in the composition of the end-members is propagated to the calculated mixing proportions. If the computed mixing proportions for a particular sample have a very broad variability (let's say, from 10% to 80% of the glacial end-member), it would mean that mixing proportions are very sensitive to changes in the composition of the end-members, casting serious doubts on the mixing results. If, on the other hand, mixing proportions for a sample concentrate around particular values, it would mean that they are not too sensitive to changes in end-member composition, thereby strengthening the case for a robust result.

Figure 4-6 shows the output pdfs for three selected groundwater samples from borehole KSH01A (Simpevarp area). As can immediately be appreciated, the range of mixing proportions for each of the selected samples is quite narrow, considering the *a priori* compositional variability of the end-members. This is a strong indication that the computed mixing proportions *are indeed a robust estimator of the mixing behaviour of the waters*.

Table 4-4 shows the mean and standard deviation of the mixing proportions for the three selected samples. The maximum deviation is ±3% (i.e. approximately 68% of the calculated mixing proportions are inside a bracket of width 6%) and the average deviation 2% (i.e. the average of 12 standard deviations: three samples times four end-members each).

The important conclusion that can be drawn from the above results is that, *once the number and type of end-members are known*, the inclusion of the compositional variability of the end-members in the PCA analysis gives a robust estimation of the mixing proportions, in the sense that the output probability functions are narrow, predicting mixing proportions tightly concentrated around a mean value. The bonus of this analysis, apart from the robustness itself, resides in the statistical bracketing of the variability of the mixing proportions, which is a fundamental issue when "exporting" these results for hydrogeological modelling.

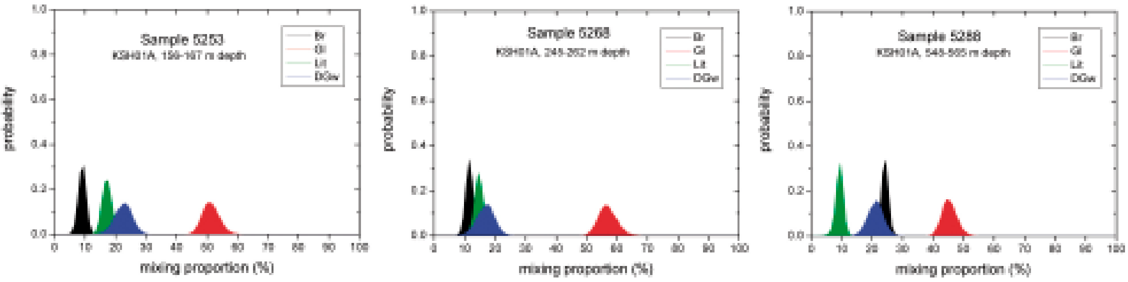


Figure 4-6. *Mixing proportions for three samples from borehole KSH01A (Simpevarp area). End-members used for the calculations are Brine + Glacial + Litorina + Dilute Groundwater. For the PCA analysis only groundwater samples from Laxemar 1.2 iteration were used (158 samples).*

Table 4-4. Mean and standard deviation for the mixing proportions of the selected samples from boreholes KLX02 and KSH01A.

		Brine (%)		Glacial (%)		Litorina (%)		Dilute Gw (%)	
	Sample	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
Simpevarp	5263	9.1	1.2	51.2	2.8	17.2	1.6	22.5	2.9
KSH01A	5268	11.5	1.2	57.0	3.0	14.7	1.4	16.8	2.9
	5288	24.2	1.2	45.4	2.5	9.2	1.4	21.2	2.6

5 Confidence building

The concept of confidence building is central to the reliable use of any piece of software, either very simple or extremely complex. All developers, users and decision makers will agree on this. But the agreement usually ends here, because the definition of "confidence building" is, to put it mildly, a thorny matter.

There is a huge amount of literature on verification, validation, and confidence building in connection with the Performance Assessment (PA) of deep geological repositories of high level radioactive wastes, including several PAs carried out by SKB or SKI like Project-90 /SKI 1991, NEA/OECD 1992/, SKB-91 /SKB 1992/, SITE-94 /SKI 1996ab/, and SR 97 /SKB 1999ab/. Although the opinions expressed there are diverse, they can be grouped into two extreme sides: (1) confidence building can demonstrate that a piece of software is sound and delivers what it says; and (2) confidence building can never demonstrate that a piece of software is sound and would deliver what it says. Of course, between these two opposite views, there are opinions that try to go straight to the point and which are much more useful from a practical viewpoint. As /Zuidema 1995/ put it in GEOVAL'94 (italics are ours):

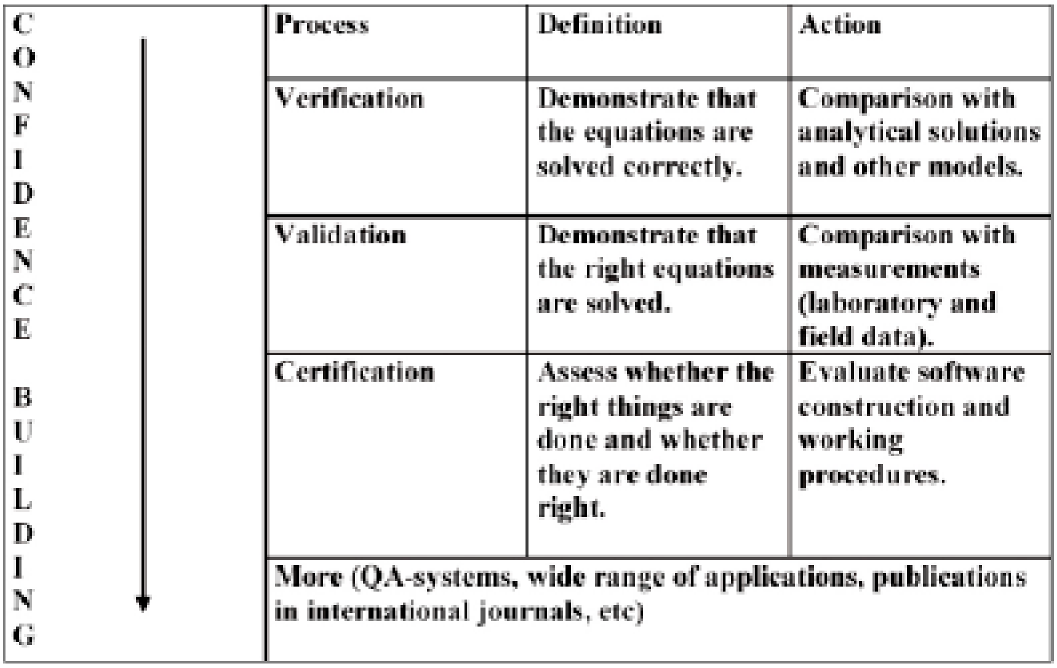
> "Absolute truth is not known, so we cannot of course provide a model which provides this. In practice, models will be used to support regulatory and legal decisions, and this will not change no matter how loudly and often it is proclaimed that it cannot be demonstrated that models represent the truth. Thus, our task should be to stop debating the impossibility of model validation in such an absolute sense, and to develop *procedures whereby all involved parties can be reasonably assured that models are appropriate and are being used correctly to meet the needs of the problem at hand*. Certainty cannot be achieved, we must and should be satisfied with engineering confidence – often it will be sufficient to provide confidence in our ability to bound the outcome of a specific phenomenon".

The very name "confidence building" is a later addition to the PA jargon which tries to acknowledge the intrinsic difficulties in demonstrating the soundness of any computer program. Before that, "validation" was the trigger word, and the discussion centred around the feasibility of validating a computer code /Greenwood 1989, Tsang 1991, Konikov and Bredehoeft 1992, Bredehoeft and Konikov 1993, McCombie and McKinley 1993, Bair 1994, Oreskes et al. 1994, Leijnse and Hassanizadeh 1994, Molnia 1996, Sargent 1999, Bredehoeft 2005/. The general conclusion was that a 100% validation was not possible, even recommendable, when the codes were asked to predict processes and whole-system behaviour in and around the repository in time spans of thousands of years. To soften up the implications that the word "validation" had, the term "confidence building" was coined for the process of gaining confidence in the workings of a computer program. Actually, the sentence written in italics in /Zuidema 1995/ citation could be used as a working definition of confidence building.

Figure 5-1 summarises the actions involved in confidence building, from the most basic step (verification) to more complex ones (validation, certification, etc). Some steps are intrinsic to the computer code, but others, the most critical, depend also on the specific application at hand and cannot be carried out without a complete specification of the system, the problem to be solved, and the time and space framework where the code should seek an answer.

Here and in Report 2, the first two steps in the confidence building process have been formalised in terms of a series of verification and validation experiments. No attempt has been made to devise certification procedures, but in Report 2 further details on the "More" step of the confidence building process are included.

In this report the verification and validation exercises are collected as two tables indicating the type of exercise and what the exercise tries to "verify" or "validate". The complete exercise can be found in Report 2.



CONFIDENCE BUILDING	Process	Definition	Action
	Verification	Demonstrate that the equations are solved correctly.	Comparison with analytical solutions and other models.
	Validation	Demonstrate that the right equations are solved.	Comparison with measurements (laboratory and field data).
	Certification	Assess whether the right things are done and whether they are done right.	Evaluate software construction and working procedures.
	More (QA-systems, wide range of applications, publications in international journals, etc)		

***Figure 5-1**. Processes and actions involved in confidence building /Svensson et al. 2004/.*

5.1 Verification of M3

The verification cases performed can be found in Report 2 and are summarised in Table 5-1. The cases tested show both “good results” and “bad results”, mainly to test when the two-principal component mixing routine (Section 3.2.2) should be used and when not. “Good result” cases are those where the comparison with the corresponding analytical solution, or another model study, is satisfactory; the potential M3 user is referred to Report 2 to study the details.

A few words may however be needed to explain the objectives when selecting the test cases and the way the comparisons have been carried out.

- The test cases should include all the parts of the M3 methodology (PCA, mixing, mass balance, ESM and EVM).
- Each test should focus on a particular algorithm or module.
- Synthetic datasets have been used in many tests as this is the best way to validate the results when dealing with mixing proportions and mass balance calculations.

Some verification studies are straightforward and test a specific algorithm, and other are more elaborate because they try to test longer pieces of the code.

Table 5-1. Summary of the verification tests that can be found in Report 2.

Group	Case	Comments
A. Principal Component Analysis	A1. Eigenvectors, eigenvalues and PC loadings.	Test of the mathematical routine that performs the Principal Component Analysis.
B. Mixing proportions	B1. Mixing proportions when end-members are fully known. B1.1. Mixing proportions with the two-principal component mixing routine. B1.2. Mixing proportions with the hyper-space mixing routine.	Test of the calculation of the mixing proportions from the PCA coordinates. Synthetic samples are used in all test cases. Both the two-principal component and hyper-space mixing routines are tested.
C. Mass balance	C1. Test of absolute and relative deviations using synthetic samples.	Test of the mass balance calculations. As these deviations depend on the mixing routine used, they are tested with the output of both mixing routines (see group B tests).
D. End-member Selection Module	D1. Test of the combinations generating routine. D2. Test of ESM using as end-members the same used to create the samples.	The main aim of this exercises is to test the routine that generates all the possible combinations of end-members.
E. End-member Variability Module	E1. Test of random number generator. E2. Construction of input probability distributions. E2.1. When lower and upper ranges are identical. E2.2. When lower and upper ranges are different.	This is a complex module that has to be tested by parts, starting with the random number generation routine, following with the input probability distributions and ending with the output probability distributions.
F. Comparison with other analytical and numerical solutions of pure mixing problems.	F1. Linear mixing (no redundancy). F2. Linear least squares (redundancy). F3. PHREEQC in the pure mixing mode.	This set of test will verify that M3 is able to solve pure mixing problems.

5.2 Validation of M3

Table 5-2. Summary of the validation tests that can be found in Report 2.

Group	Case		Comments
A. Stability of mixing proportions	A1.	Stability check 1: dependence of mixing proportions on the number of samples in the dataset.	Synthetic samples are used in A1.1 and A1.2, and a real data set with one synthetic sample inserted in cases A1.2 and A2.2.
	A1.1.	Only synthetic samples.	In case A2.3 only conservative elements are included among the variables. It is important to test the resolution of the method when only conservative elements are included.
	A1.2.	One synthetic sample inserted in a real-sample dataset.	
	A2.	Stability check 2: dependence of mixing proportions on the number of input variables.	Stability check 3 tries to quantify the error introduced in the mixing proportions when the end-members are included in the dataset prior to the principal component analysis.
	A2.1.	Only synthetic samples.	
	A2.2.	One synthetic sample inserted in a real-sample dataset.	
	A2.3.	Special case: Only conservative elements.	
	A3.	Stability check 3: dependence of mixing props on the inclusion/exclusion of end-members.	
B. End-member Variability Module	B1.	Output probability distributions.	This test will verify the effect of end-member compositional variability on the calculated mixing proportions. This variability could be intrinsic or due to analytical and/or sampling errors.
	B1.1.	When lower and upper ranges are identical.	
	B2.2.	When lower and upper ranges are different.	
	B2.	Testing mass balance in the case of variable end-member composition.	
C. End-member Selection Module	C1.	Test of ESM using as end-members others than the ones used to generate the samples.	This test will verify the stability of the mixing proportions against a misidentification of end-members: i.e. what would happen to the mixing proportions in the case of an erroneous selection of end-members (both in number and type).
D. Analysis of reactions	D1.	Test of absolute and relative deviations using one synthetic sample in a real dataset.	This tests will verify the meaning of the deviations between real and computed elemental concentrations and whether this deviations could be used to identify chemical reactions and in which circumstances.
	D1.1.	With conservative and non-conservative elements.	
	D1.2.	Only with conservative elements.	
E. Cross-check against other codes	E1.	Solve a mixing+reaction problem with M3 and PHREEQC.	This test will verify how the computed mixing proportions deviate from the real mixing proportions as the importance of chemical reactions is increased.
F. Cross-check against other methods to compute mixing ratios and reactions			This tests will verify the capability of M3 to solve real, complex groundwater mixing problems by comparing its results with those obtained by other authors using alternative approaches.
G. Ability to solve non-aqueous mixing problems			This tests will verify the capability of M3 to solve mixing problems outside the realm of water hydrogeochemistry.

6 An example with M3

To become familiar with the M3 method and to give an idea of the type of results M3 delivers, we shall use a real multi-dimensional dataset consisting of groundwater samples from the Scandinavian Shield. This is a dataset of 169 water samples from the Laxemar area, near Äspö, Sweden, used by /Laaksoharju et al. 2004a/ in iteration 1.2 of SKB's Site Investigation Program at Laxemar. A subset of the dataset is reproduced in Table 6-1 to see the type of information used in the input file.

As we are interested in groundwater mixing, only these type of water samples have been included in the dataset. All superficial waters (streams, lakes, seawater, precipitation) have been deleted. The input variables consist of ten compositional and isotopic parameters: the concentration in mg/L of Na, K, Ca, Mg, HCO_3, Cl, and SO_4; the concentration of ^3H in tritium units; and the deviation of ^2H and ^{18}O in ‰ (delta values). Notice that among these input variables are conservative and non-conservative ones, i.e. some would behave as conservative during mixing, but most of them will not. It is up to the user to decide which set of input variables best describe the system, but running M3 with different sets of input variables is a good idea to asses the impact of these on the mixing proportions and mass balances.

Table 6-1. Excerpt of the dataset (Laxemar 1.2 Regional Model, 169 samples).

Sample #	Na (mg/l)	K (mg/l)	Ca (mg/l)	Mg (mg/l)	HCO_3 (mg/l)	Cl (mg/l)	SO_4 (mg/l)	δ^2H ‰	^3H (TU)	δ^{18O} ‰
4116	498	13.8	60.4	16.9	440	529.8	201.81	–64.2	3.7	–9.5
4167	64.6	9.5	62	14	310	15.7	18.66	–79.6	12	–11.8
4399	169	6.68	27.6	6.9	390	72	44.65	–84.5	14.4	–11.7
4964	1,560	35.6	640	168	161	3,898.8	286.47	–77.3	2.2	–10.1
4464	447	19.1	92.9	27	418	585.8	155.56	–81.2	5.7	–11.1
4535	1,150	32.6	572	139	173	2,694.1	319.35	–72.8	1.2	–9.4
4522	2,210	67.8	754	287	115	5,421.7	533.97	–66.1	0.8	–8.4
8335	274	5.6	41.1	7.5	465	181	85.1	–80.6	12.1	–11.1
4965	1,610	21.6	945	173	161	4,466.4	410.22	–73.2	1.9	–9.5
8038	482	5.63	550	32.6	261	1,658.5	90.04	–91.4	2.5	–12.4
8020	783	6.34	779	63	176	2,617.1	173.5	–93.6	0.8	–12.6
8129	1,010	15.8	555	102	231	2,560	263	–77.4	1.1	–10.8
8095	276	9.27	176	25.5	332	545	104	–79.8	0.8	–11.5
8127	466	12.7	175	32.9	366	857	159	–77.5	1.9	–11.1
8164	282	5.93	32.7	8.1	450	178	106	–81.9	5.6	–10.8
8246	399	12.7	42.2	13.5	461	415	109	–80	7.8	–10.9
8247	1,870	48.4	893	242	137	4,950	543	–64.5	0.8	–8.5
4538	1,740	25.6	874	142	61	4,562.8	315.65	–88.2	0.8	–11.6
4724	2,000	29.2	934	204	99	5,329.5	546.97	–69.3	0.8	–8.8
4398	168	6.46	34.2	7.8	378	84.4	53.92	–82.1	11.2	–11.4
8100	366	9.74	139	31	353	642	89.6	–80.9	0.8	–11.6
8272	1,820	21.4	1,140	198	93	5,380	434	–75.6	0.8	–10.2
8016	2,040	34.2	934	226	125	5,410	498	–67	2.4	–8.8
8012	2,110	47.6	925	223	101	5,450	495	–69.4	2	–9.1
8017	2,070	26.8	985	202	91	5,430	472	–70.8	0.8	–9.4
8273	1,660	14.3	1,440	52.7	22	5,430	197	–84.3	0.8	–11.6

6.1 Selection of the end-members (ESM)

After deciding which and how many input variables to use, the next step is selecting the potential end-members to run the ESM. In our example, eight potential end-members were included in the input data file (Table 6-2), which means that a total of 247 combinations are possible. Because we know that at least three end-members are involved, we do not include the 28 combinations of only two end-members. That leaves 219 combinations of three (56), four (70), five (56), six (28), seven (8) and eight (1) end-members.

Table 6-3 collects all combinations able to explain more than 50% of the 169 samples. The rest of the combinations have been filtered out because they are not able to explain by mixing even half of the samples in the dataset.

In Table 6-3 the first column gives the combination number, the next eight columns give the end-members used in each particular run of M3. The second last column gives the actual number of samples inside the hyper-volume defined in *n*-1 dimensional space by the *n* selected end-members (where *n* runs from 3 to 8) and last column gives the percentage of samples explained. Samples inside this hyper-volume are the only ones that can be explained by pure conservative mixing of the given end-members. The more samples inside the hyper-volume, the more successful is the chosen combination in explaining the dataset. Combinations #61 (dark orange shading in Table 6-3) and #62 (light orange) are the most successful ones, able to explain 99.4% of the samples (i.e. all samples except one). We are interested in the mixing behaviour of the groundwaters. Combination #62 includes a superficial end-member (precipitation water) but Combination #61 includes none superficial end-member, so we decided to keep the latter in detriment of the former.

Notice that the best combination of three end-members with no precipitation end-member explains 98.2% of the samples (Br + Gl + DGw). This is a high percentage and we could have decided to retain that combination, but we know from the geological evolution of the site that some waters should have the signature of the Litorina Sea (the ancient Baltic Sea) and thus this end-member should be included. That is why we did not select the three-end-member combination Brine + Glacial + DiluteGW.

Among the five-end-members combinations none is able to explain more than 85.8% of the samples. And with more than five end-members, none explains more than 50% and are not shown in the filtered output (the best combination of 6 end-members explains 27.2% of the samples; the best one with seven end-members explains 6.5% of the samples, and the only eight-end-member combination explains just 2.4% of the samples). We see that the “quality” of the combination of end-members drops rapidly when more than 5 end-members are used, suggesting that five end-members is the likely maximum number of end-members involved in the mixing of the groundwaters in the Laxemar area. This datum in itself is not enough to rule out other combinations of end-members, but is a strong indication supporting combination Brine + Glacial + Litorina + DiluteGW as a good candidate to be used in all following M3 calculations.

The sole output of the ESM is of course insufficient to decide which set of end-members is to be used with M3, but combined with the geological, hydrological and geochemical knowledge of the system it could be converted into a powerful decision tool.

Table 6-2. Potential end-members.

	Na (mg/L)	K (mg/L)	Ca (mg/L)	Mg (mg/L)	HCO$_3$ (mg/L)	Cl (mg/L)	SO$_4$ (mg/L)	δ^{2}H ‰	^{3}H (TU)	δ^{18}O ‰
Brine	8,500	45.5	19,300	2.12	14.1	47,200	906	–44.9	0	–8.9
Glacial	0.17	0.4	0.18	0.1	0.12	0.5	0.5	–158	0	–21
Litorina Sea	3,674	134	151	448	93	6,500	890	–38	0	–4.7
Sea Sediment	2,144	91.8	103	258	793	3,383	53.1	–61	0	–7
Baltic Sea	1,960	95	234	93.7	90	3,760	325	–53.3	20	–5.9
Lake Water	21	3.21	30.3	5.9	110	45.8	16.18	–44.3	7.6	–4.5
Rain 1960*	0.4	0.29	0.24	0.1	12.2	0.23	1.4	–80	168	–10.5
Dilute GW	237	4	25	6	370	119	118	–73.8		–9.9

* Age corrected tritium to 2004 datum.

Table 6-3. Output of the End-member Selection Module. It has been filtered to show only those combinations explaining more than 50% of the samples in the dataset.

Comb #	EM1	EM2	EM3	EM4	EM5	EM6	EM7	EM8	Samples explained	%
2	Brine	Glacial	Rain60*	–	–	–	–	–	168	99.4%
3	Brine	Glacial	SeaSed	–	–	–	–	–	88	52.1%
4	Brine	Glacial	Baltic	–	–	–	–	–	88	52.1%
5	Brine	Glacial	LakeWater	–	–	–	–	–	141	83.4%
6	Brine	Glacial	DiluteGW	–	–	–	–	–	166	98.2%
22	Glacial	Litorina	Rain60*	–	–	–	–	–	139	82.2%
25	Glacial	Litorina	LakeWater	–	–	–	–	–	91	53.8%
26	Glacial	Litorina	DiluteGW	–	–	–	–	–	136	80.5%
28	Glacial	Rain60*	Baltic	–	–	–	–	–	133	78.7%
57	Brine	Glacial	Litorina	Rain60*	–	–	–	–	167	98.8%
60	Brine	Glacial	Litorina	LakeWater	–	–	–	–	144	85.2%
61	Brine	Glacial	Litorina	DiluteGW	–	–	–	–	168	99.4%
62	Brine	Glacial	Rain60*	SeaSed	–	–	–	–	168	99.4%
64	Brine	Glacial	Rain60*	LakeWater	–	–	–	–	86	50.9%
67	Brine	Glacial	SeaSed	LakeWater	–	–	–	–	127	75.1%
68	Brine	Glacial	SeaSed	DiluteGW	–	–	–	–	159	94.1%
69	Brine	Glacial	Baltic	LakeWater	–	–	–	–	118	69.8%
70	Brine	Glacial	Baltic	DiluteGW	–	–	–	–	141	83.4%
129	Brine	Glacial	Litorina	Rain60*	LakeWater	–	–	–	145	85.8%
130	Brine	Glacial	Litorina	Rain60*	DiluteGW	–	–	–	139	82.2%
134	Brine	Glacial	Litorina	Baltic	LakeWater	–	–	–	87	51.5%
138	Brine	Glacial	Rain60*	SeaSed	LakeWater	–	–	–	133	78.7%
139	Brine	Glacial	Rain60*	SeaSed	DiluteGW	–	–	–	129	76.3%
140	Brine	Glacial	Rain60*	Baltic	LakeWater	–	–	–	108	63.9%
141	Brine	Glacial	Rain60*	Baltic	DiluteGW	–	–	–	94	55.6%
143	Brine	Glacial	SeaSed	Baltic	LakeWater	–	–	–	98	58.0%

6.2 Principal components analysis and plot

Once the proper set of end-members have been selected, all remaining computations are carried out with it. For that purpose, the chosen end-members are included in the input data file and the PCA performed. Figure 6-1 shows the graphical visualization of the principal components and the position of each sample with respect to these coordinate axes. The upper graph is a 3D representation in the PC1-PC2-PC3 space, and the lower graph is the standard M3 PC-plot, with the first principal component (PC1) as the horizontal axis and the second principal component (PC2) as the vertical axis.

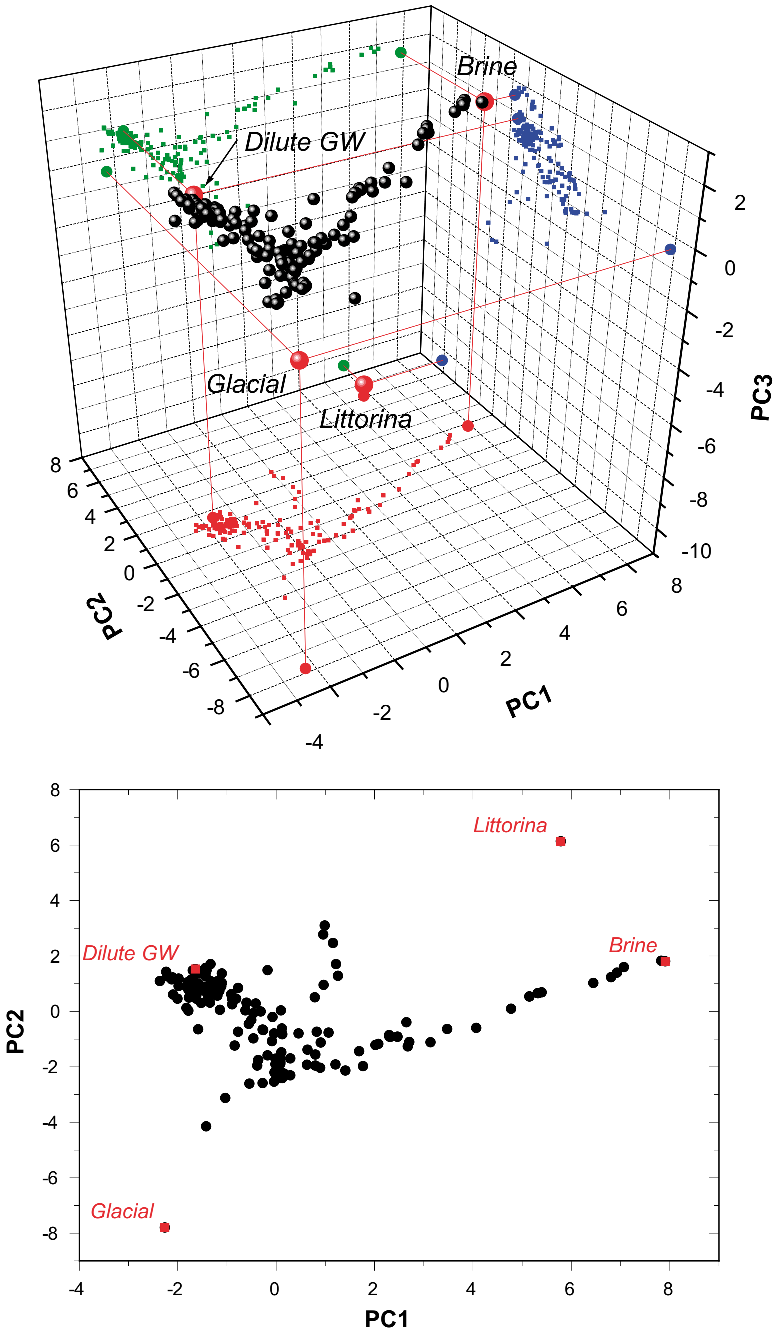


Figure 6-1. PC-plot of the Laxemar 1.2 dataset. The upper graph is a 3D visualization with the first three principal components as axes, and the lower graph is the standard M3 PC-plot with PC1 and PC2 as coordinate axes. Each sample is represented by a black dot and the end-members (Brine, Glacial, Litorina and Dilute GW) by a red dot. The 3D graph includes the projection of each sample onto the PC1-PC2, PC1-PC3 and PC2-PC3 planes (in red, blue, and green respectively). The PC1-PC2 projection is of course identical to the lower graph.

As we are using 10 input variables, the covariance matrix has dimension 10×10 and there are also ten eigenvectors (principal components) and ten eigenvalues (Section 3.1.2). Table 6-4 collects the first seven eigenvectors (principal components) and Table 6-5 all the eigenvalues in order of decreasing variance. As for the principal components, the first one, which contains 45.84% of the variance has the form:

$$\begin{aligned}PC1 = {} & 0.4502\times Na + 0.2848\times K + 0.4376\times Ca + 0.0638\times Mg - 0.3156\times HCO_3 + \\ & 0.4481\times Cl + 0.4253\times SO_4 + 0.1712\times {}^2H - 0.0974\times {}^3H + 0.0219\times Na^{18}O.\end{aligned}$$

It is not easy to interpret this principal components in terms of the original variables, as each one is a lineal combination of all the original input variables, some with positive loadings (e.g. Na, K, or deuterium for PC1), and some with negative loadings (HCO_3 and tritium for PC1). This is why in M3 methodology no attempt is made to undertake such an interpretation and in the graphs the axes are simply labelled PC1, PC2 and so forth.

The last column in Table 6-5 gives the cumulative variance (as a percentage) contained in the first *k* eigenvalues. The first two principal components contain 65.65% of the variance in the dataset. This is an important number to take into consideration when using the two-principal component mixing routine (Section 3.2.2). If the cumulative variance for the first two principal components is low (say, less than 70%), the amount of information lost during the compression is high and the computed mixing proportions could have significant errors. In this case, the hyperspace mixing routine is the only reliable way to compute the mixing proportions.

Table 6-4. First seven eigenvectors (principal components). Each column shows the loadings of an specific input variable.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Na	0.4502	–6.10E–2	–3.49E–2	8.03E–2	–8.77E–3	0.227	–0.1403
K	0.2848	0.5254	–6.36E–2	–6.54E–2	–0.1206	0.3179	0.2205
Ca	0.4376	–0.1428	0.1741	1.90E–2	9.76E–2	0.3169	2.07E–2
Mg	6.38E–2	0.5956	–0.3086	–9.50E–2	–0.448	–7.20E–2	–1.29E–2
HCO_3	–0.3156	6.48E–2	0.5372	–0.2369	–8.89E–2	0.3216	0.5894
Cl	0.4481	–0.1112	0.1177	2.91E–2	6.26E–2	0.26	1.23E–3
SO_4	0.4253	–9.34E–3	–1.12E–2	–6.19E–2	0.1768	–0.666	0.5804
^2H	0.1712	0.1804	0.6809	–0.2887	–0.2047	–0.3439	–0.4698
3H	–9.74E–2	0.5318	0.1148	0.1222	0.7946	1.79E–2	–0.1351
^{18}O	2.19E–2	–0.1144	–0.2998	–0.906	0.2425	9.31E–2	–8.88E–2

Table 6-5. Eigenvalues in order of decreasing variance.

Eigenvalue #	Variance	As percentage	Cumul. percentage
1	793.3	45.85	45.85
2	342.5	19.80	65.65
3	237.1	13.70	79.36
4	171.2	9.894	89.25
5	98.35	5.685	94.94
6	38.67	2.235	97.17
7	25.38	1.467	98.64
8	13.20	0.7633	99.40
9	9.677	0.5593	99.96
10	0.6971	0.4029E–1	100.0

6.3 Mixing proportions

The next step in the M3 methodology is the computation of the mixing proportions. In this step each sample is expressed in terms of the contribution of the chosen end-members to its composition. As we are using four end-members (Brine, Glacial, Litorina and Dilute GW), each sample will be characterize by four numbers: the percentage of Brine, Glacial, Litorina and Dilute GW necessary to reproduce its chemical composition *if pure conservative mixing were the only process affecting the chemical variability of the sample.*

In the previous section we saw that the two first principal components contain only 66% of the variance in the dataset. As this is a low figure, we have use the hyperspace mixing routine to compute the mixing proportions. Figure 6-2 shows the results. The graphs are PC-plots with PC1 and PC2 as coordinate axes. Each one plots the mixing proportions of a end-member colour-coded from 0% (red) to 100% (dark blue). Black circles are for samples outside the mixing hyper-volume (i.e. samples that cannot be explained by pure conservative mixing of the selected end-members). In this particular case there is only one non-explained sample, near the Dilute GW end-member. The end-members are marked by a black square (Brine, Glacial, Litorina, and Dilute GW).

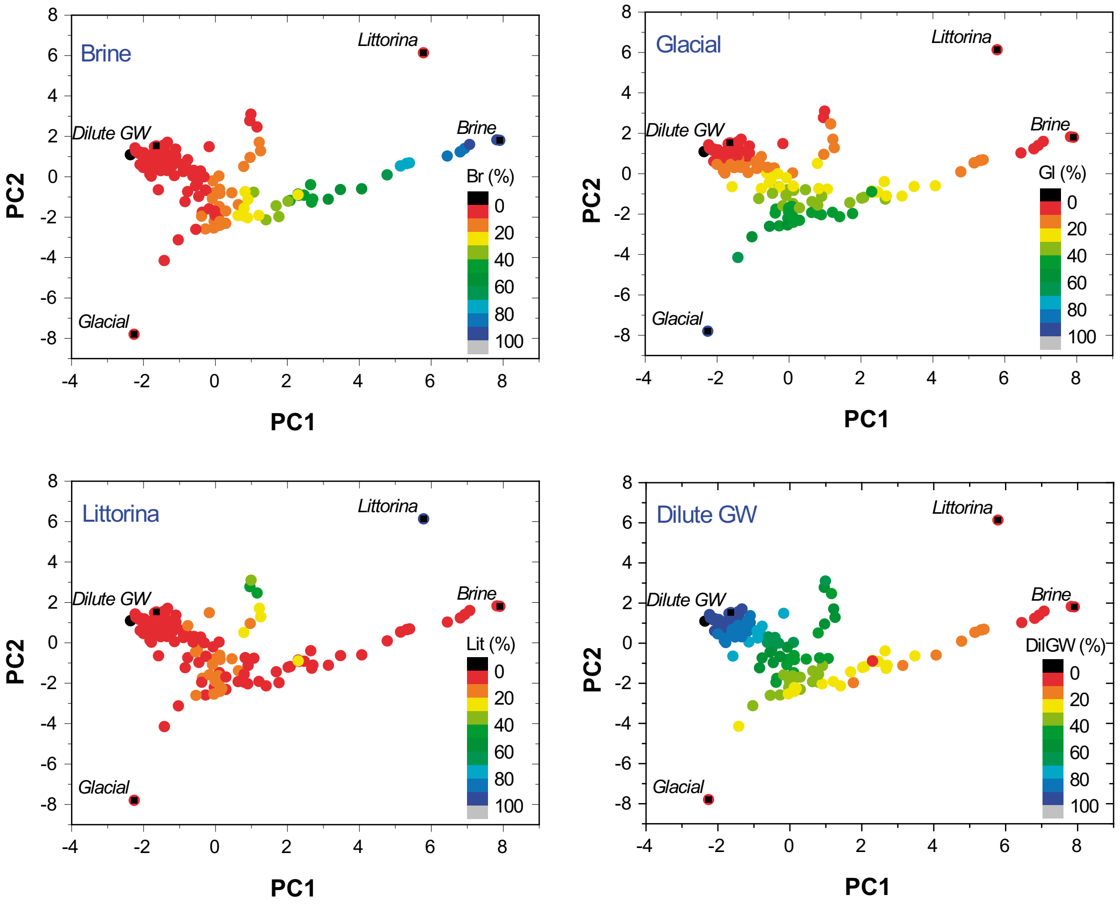


Figure 6-2. Mixing proportions for the four end-members (Brine, Glacial, Litorina and Dilute GW) computed with the hyperspace mixing routine. Labelled black squares identify the end-members and black circles those samples not explained by mixing (there is only one, near the Dilute GW end-member).

6.4 Mass balance

After the mixing proportions have been computed we are in the position of reconstructing the chemical composition of each sample ("theoretical composition" in Section 3.3). From this theoretical composition and the measured one M3 computes a mass balance for each input variable (and each sample) as formalised in Equations (3-40) and (3-41). Again, the best way to visualise the results for the whole dataset is by means of PC-plots. In this case the colour code indicates the difference (in relative percentage, see Equation 3-41) between the measured and theoretical concentrations of an specific input variable. Figure 6-3 shows the results for three conservative input variables: chlorine, oxygen-18 and deuterium. Deviation greater than 100% are coloured in grey. Again, a black circle marks a sample outside the mixing hyper-volume.

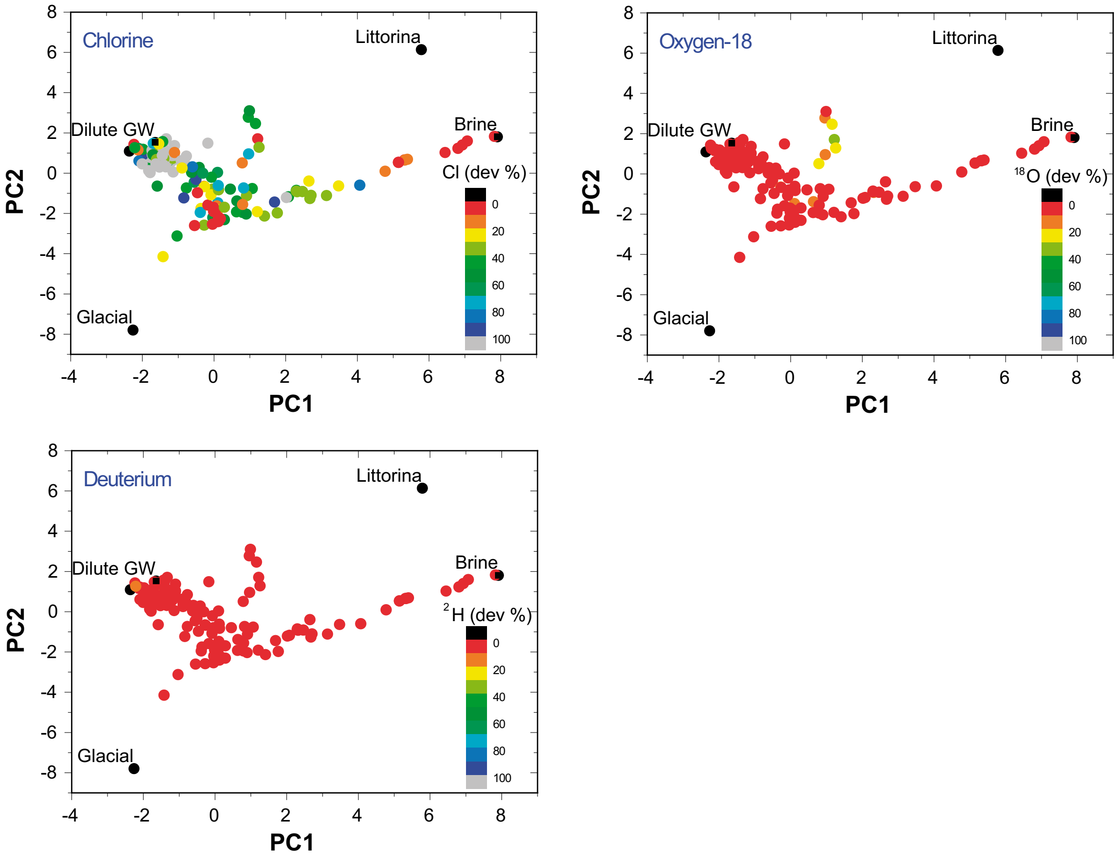


***Figure 6-3**. Mass balance expressed as a relative deviation from the measured composition for three conservative input variables: chlorine, ^{18}O and ^{2}H. Mass balance has been calculated from theoretical compositions computed with the hyperspace mixing routine. Black squares identify the end-members; black circles those samples outside the mixing hyper-volume; and grey circles samples with a relative deviation greater than 100%.*

6.5 End-member variability: uncertainty in mixing ratios

Up to now, all M3 calculations have been carried out assuming that end-members are fully known, both in number and in chemical composition. But this is a rather strong assumption in most real cases. The *number* of end-members that contribute to the mixing must commonly be deduced by independent geological, hydrological and geochemical arguments. But the *composition* of each end-member can vary in time and/or space, and this variability would certainly affect the computed mixing proportions and hence mass balance calculations. It is therefore very important to assess how the intrinsic compositional variability of the chosen end-members affects the mixing proportions. For that purpose the End-member Variability Module (see Section 4.2 for a review of this module) was implemented in M3 v3.0.

The EVM accepts as input a range of compositions for each end-member and construct from every range an *input probability distribution* (in terms of compositions), which is propagated through the PCA and converted into an *output probability* for the mixing proportions. In other words, the PCA and mixing calculation routines are run a large number of times, each one with a different composition for the chosen end-members. The composition of the end-members is selected in each run at random from the input probability distribution. This input probability distribution has been previously constructed from the compositional range of each input variable in each end-member (i.e. if we are using ten input variables and four end-members, we will have 40 input probability distributions: one for the Na content in Brine end-member, another for the Ca content in Litorina, and so on).

Each M3 run will give a set of mixing proportions for every sample. The EVM then constructs a histogram from the *m* runs (where *m* is a large number, of the order of 10^3–10^6), which summarises the variability in the mixing proportions due to the initial variability in the composition of the end-members. In that way the user can measure in a quantitative way how uncertainties in composition propagate through the different calculations and are converted into uncertainties in the computed mixing proportions.

Figure 4-2 in Section 4.2 shows the input probability distributions for the Laxemar example. Table 4-7 collects the compositional ranges that characterise each end-member (Brine, Glacial, Litorina, and Dilute GW), and Figure 4-6 plots the output probability distributions in term of mixing proportions. Also, the upper graph in Figure 4-4 shows how the compositional variability of the end-members affects its position on a PC-plot and the position of each sample.

As was commented on in Section 4.2, Figure 4-6 shows that the mixing proportion probabilities are in all cases very narrow, indicating that the whole procedure is indeed quite robust with regard to initial compositional uncertainties.

7 Conclusions

M3 version 3 is a *PCA computer code* to calculate mixing proportions of a *large* groundwater dataset from input compositional variables and a group of end-members. Once the mixing proportions have been computed, deviations from the actual composition of each sample are calculated by mass balance. These deviations can be interpreted in terms of chemical reactions if additional knowledge of the system under consideration is obtained.

Mixing proportions in M3 v.3 can be calculated by two different methods, called 2-principal component mixing and *n*-principal component mixing or hyperspace mixing. T*wo-principal component mixing* is the method implemented in all previous versions of M3 and only uses the information contained in the first two principal components to compute the mixing proportions. When more than three end-members are used, the 2-principal component mixing method does not give a unique set of mixing proportions for each sample, and an approximation has to be used in order to select one set among the infinite number of sets. This limits the accuracy of the computed mixing proportions.

Hyperspace mixing (also called *n*-principal component mixing) is new to this version of M3 and utilises all the principal components to compute the mixing proportions. For each sample, the computed mixing proportions are unique and so is the preferred method for more than three end-members.

M3 v.3 has two built-in modules aimed at reducing uncertainties in the computed mixing proportions. *The End-member Selection Module* (ESM) can facilitate the selection of end-members for a given groundwater system. The ESM gives the percentage (*coverage*) of samples in a dataset that can be explained by pure mixing of the chosen end-members. The bigger the coverage, the better is the selected set of end-members in explaining, by pure mixing, the chemistry of the samples. At any rate, the output of the ESM must be backed up by a detailed hydrological and hydro-geochemical study of the groundwater system under study.

T*he End-member Variability Module* (EVM) tries to take into account the intrinsic compositional variability that the chosen end-members can have. In a given groundwater system, even if the only important process shaping it is mixing, the end-members that mix to give a final mixed water can vary in space and/or time. The EVM accepts as input a *range* of compositions for each end-member (instead of a fixed composition, as in the standard M3 calculations) and calculates that range of mixing proportions for each sample in the dataset that is compatible with the input variability. The range of mixing proportions is converted at the output into a probability distribution of mixing proportions (one for each sample).

Before using M3 it is essential to acknowledge its aptitudes and limitations. There are things M3 is good at, and things that are beyond its capabilities. Showing these capabilities and limitations is the main goal of this Report. But it is only one of three volumes that together serve as the basic reference to M3 version 3. Report 2 concentrates on verification and validation issues, and Report 3 is the User's Guide. Both of them should be consulted before running M3. No computer code should be used without first knowing what the code actually does, but in the case of M3 this is specially true, as the outputs of M3 are not straightforward to interpret without a detailed geochemical knowledge of the studied groundwater system. The four basic requirements that must be fulfilled for the successful use of M3 are:

1. A large number of groundwater samples (> 100).
2. Mixing must be the first order process controlling the chemistry of the groundwater system.
3. Chemical reactions must account for only a small fraction of the chemical variability.
4. End-members must be deduced first by independent methods.

8 References

Anton H, 2000. Elementary Linear Algebra 8e, John Wiley & Sons Inc., 608 pp.

Bair E S, 1994. Model (in)validation: a view from the courtroom. Ground Water, 43(4), 530–531.

Ball J W, Nordstrom D K, 2001. 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. (Open File Report 91-183), USGS, USA.

Banwart S, Laaksoharju M, Nilsson A-C, Tullborg E-L, Wallin B, 1992. The large-scale Redox experiment: Initial characterization of the fracture zone. SKB Progress Report PR 25-92-04, Svensk Kärnbränslehantering AB.

Bath A H, Jackson C P, 2002. Äspö Hard Rock Laboratory: Task Force on modelling of groundwater flow and transport of solutes. Review of Task 5. SKB IPR-03-10, Svensk Kärnbränslehantering AB.

Bevington P R, Robinson D K, 1992. Data Reduction and Error Analysis for the Physical Sciences, 2e, WCB/McGraw Hill, 328 pp.

Bredehoeft J D, Konikov L F, 1993. Ground-water models: validate or invalidate. Ground Water, 31(2), 178–179.

Bredehoeft J D, 2005. The conceptualization model problem – surprise. Hydrogeol. J., 13, 37–46.

Carrera J, Neuman S P, 1986. Estimation of aquifer parameters under steady-state and transient conditions: 1. Background and statistical framework. Water Resour. Res., 22, 199–210.

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, W12101, doi:10.1029/2003WR002263.

Cave M, Harmon K, 1997. Determination of trace metal distributions in the iron oxide phases of Red Bed Sandstones by chemometric analysis of whole rock and selective leachate data. Analyst, 122, 501–512.

Cave M, Wragg J, 1997. Measurement of trace element distributions in soils and sediments using sequential leach data and a non-specific extraction system with chemometric data processing. Analyst, 122, 1211–1221.

Davis J C, 2002. Statistics and Data Analysis in Geology, 3rd edition, John Wiley and Sons, 638 p.

Dershowitz B, Shuttle D, Klise K, Uchida M, Metcalfe R, Cave M, 2000. Äspö Hard Rock Laboratory: Fracman modelling of geochemical end-member transport pathways, Äspö HRL, Äspö, Sweden. SKB IPR-02-37, Svensk Kärnbränslehantering AB.

Faure G, 2004. Isotopes: Principles and Applications. Wiley, 3rd edition, 928 pp.

Gershenfeld N, 1999. The Nature of Mathematical Modelling, Cambridge University Press, 345 p.

Gómez J B, Auqué L F, Gimeno M J, 2006. Sensitivity and uncertainty analysis of mixing and mass balance calculations with standard and PCA-based geochemical codes. To be published in an special issue of Applied Geochemistry.

Greenwood H J, 1989. On models and modeling, The Canadian Mineralogists, 27, 1–14.

Griffel D H, 1989. Linear Algebra and its Applications, 2 vols, Ellis Horwood Series in Mathematics and Its Applications.

Gurban I, Laaksoharju M, Ledoux E, Madé B, Salignac A L, 1998. Indications of uranium transport around the reactor zone at Bangombé (Oklo). Technical Report SKB 98-06, Svensk Kärnbränslehantering AB.

Haan C T, 2002. Statistical Methods in Hydrology. Iowa State Press, 2nd edition, 378 pp.

Hicks T W, 2005. Review of SKB's code documentation and testing, SKI Report 2005:05, Swedish Nuclear Power Inspectorate, Stockholm, Sweden, 68 p.

Konikov L F, Bredehoeft J D, 1992. Ground-water models cannot be validated. Adv. Water Resour., 15, 75–83.

Laaksoharju M, 1990. Measured and predicted groundwater chemistry at Äspö. Royal Institute of Technology, Stockholm. SKB Progress Report PR 25-90-13, Svensk Kärnbränslehantering AB.

Laaksoharju M, Smellie JAT, Nilsson A-C, Skårman C, 1995. Groundwater sampling and chemical characterisation of the Laxemar deep borehole KLX02. Technical Report SKB 95-05, Svensk Kärnbränslehantering AB.

Laaksoharju M, Skårman C, 1995a. Groundwater sampling and chemical characterization of the HRL tunnel at Äspö, Sweden. Progress Report SKB PR 25-95-29, Svensk Kärnbränslehantering AB.

Laaksoharju M, Skårman C, 1995b. Multivariate mixing and mass balance calculations of the Cigar Lake groundwaters. Internal Report, Svensk Kärnbränslehantering AB.

Laaksoharju M, Wallin B (ed), 1997. Evolution of the groundwater chemistry at the Äspö Hard Rock Laboratory. Proceedings of the second Äspö International Geochemistry Workshop, Äspö, Sweden, June 6–7, 1995. Svensk Kärnbränslehantering AB.

Laaksoharju M, Gurban I, Andersson C, 1999a. Indications of the origin and evolution of the groundwater at Palmottu. The Palmottu Analogue Project. (Technical Report 99-03), EC-NST, Luxembourg, Luxembourg.

Laaksoharju M, Skårman C, Skårman E, 1999b. Multivariate mixing and mass balance calculations (M3)-a new concept and computer program for decoding hydrochemical information. SKB Technical Report TS 00-01, Stockholm, Sweden.

Laaksoharju M, Skårman C, Skårman E, 1999c. Multivariate Mixing and Mass-balance (M3) calculations, a new tool for decoding hydrogeochemical information. Applied Geochemistry 14(7), 861–871.

Laaksoharju M, Tullborg E L, Wikberg P, Wallin B, Smellie J A T, 1999d. Hydrogeochemical conditions and evolution at the Äspö HRL, Sweden. Applied Geochemistry, 14(7), 835–860.

Laaksoharju M, Andersson C, Gurban I, Gascoyne M, 2000. Demonstration of M3 modelling of the Canadian Whiteshell Research Area (WRA) hydrogeochemical data. SKB TR-01-37, Svensk Kärnbränslehantering AB.

Laaksoharju M (ed), 2004. Hydrogeochemical evaluation of the Simpevarp area, model version 1.2. Preliminary site description of the Simpevarp area. SKB R-04-74, 463 p, Svensk Kärnbränslehantering AB.

Laaksoharju M, Gimeno M J, Auqué L, Gómez J B, Smellie J A T, Tullborg E-L, Gurban I, 2004a. Hydrogeochemical evaluation of the Forsmark site, model version 1.1. SKB R-04-05, 342 p, Svensk Kärnbränslehantering AB.

Laaksoharju M, Smellie J A T, Gimeno M J, Auqué L, Gómez J B, Tullborg E-L, Gurban I, 2004b. Hydrogeochemical evaluation of the Simpevarp area, model version 1.1. SKB R-04-16, 398 p, Svensk Kärnbränslehantering AB.

Laaksoharju M (ed), 2005. Hydrogeochemical evaluation of the Forsmark site, model version 1.2. Preliminary site description of the Forsmark area. SKB R-05-17, Svensk Kärnbränslehantering AB.

Laaksoharju M, Skårman E, Gurban I, Gómez J B, 2007. M3 version 3: User's Manual. SKB report (in preparation).

Leijnse A, Hassanizadeh M, 1994. Model definition and model validation. Advances in Water Resources, 17, 197–200.

Lindsay I S, 2000. "A tutorial on Principal Components Analysis", http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

MATLAB, 2005. MATLAB Programming version 7, MathWorks Inc, Natick, MA, USA, 835 p. (Online pdf version: www.mathworks.com/access/helpdesk/help/pdf_doc/matlab/matlab_prog.pdf).

McCombie C, McKinley I, 1993. Validation: another perspective. Ground Water, 31(4), 530–531.

Mishra S, 2002. Assigning probability distributions to input parameters of performance assessment models. SKB TR-02-11), Svensk Kärnbränslehantering AB.

Molnia B F, 1996. Modeling Geology: the ideal world vs. the real world. GSA Today, 6(5) 8–14.

NEA/OECD, 1992. SKI Project-90: A review carried out by an OECD/NEA team of experts. (Technical Report), NEA/OECD, Paris, France.

Oreskes N, Shrader-Frechette K, Belitz K, 1994. Verification, validation, and confirmation of numerical models in the Earth Sciences. Science, 263, 641–646.

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. (Science Report WRRIR 99-4259), USGS, 312 p.

Press W, Teukolsky S, Vetterling W, Flannery B, 1992. Numerical Recipes in C (2nd Edition). Cambridge University Press, Cambridge, U.K.

Reyment R A, Jvreskog K G, 1996. Applied Factor Analysis in the Natural Sciences. Cambridge University Press, 2^{nd} edition, 383 pp.

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.

Sargent R G, 1999. Validation and verification of simulation models. In: P.A. Farrington, H.B. Nembhard, D.T. Sturrock, and G.W. Evans (Eds.), Proceedings of the 1999 Winter Simulation Conference.

SKB, 1992. SKB 91-Final disposal of spent nuclear fuel. Importance of the bedrock for safety. Technical Report SKB TR 92-20, Svensk Kärnbränslehantering AB.

SKB, 1999a. SR 97: Processes in the repository evolution. Background report to SR 97. Technical Report SKB TR 99-07, Svensk Kärnbränslehantering AB.

SKB, 1999b. Deep repository for spent nuclear fuel. SR 97. Post-closure safety. Technical Report SKB TR 99-06, Svensk Kärnbränslehantering AB.

SKB, 2003. Planning Report for the Safety Assessment SR-Can. SKB TR-03-08, Svensk Kärnbränslehantering AB.

SKB, 2004. Interim Main Report of the Safety Assessment SR-Can. SKB TR-04-11, Svensk Kärnbränslehantering AB.

SKI, 1991. SKI Project-90. Technical Report SKI TR 91/23, SKI, Stockholm, Sweden.

SKI, 1996a. SKI SITE-94. Deep repository performance assessment project. 2 vols. Technical Report SKI R 96/36, SKI, Stockholm, Sweden, pp. 1–304 (Vol. 1) and 305–660 (Vol. 2).

SKI, 1996b. SKI SITE-94. Deep repository performance assessment project. Summary. Technical Report SKI R 97/5, SKI, Stockholm, Sweden, 90 p.

Smellie J A T, Laaksoharju M, 1992. The Äspö hard rock laboratory: Final evaluation of the hydrogeochemical pre-investigations in relation to existing geological and hydraulic conditions. SKB TR 92-31, Svensk Kärnbränslehantering AB.

Smellie J A T, Karlsson F, 1996. A reappraisal of some Cigar-Lake issues of importance to performance assessment. SKB Technical Report TR-96-08, 93 p, Svensk Kärnbränslehantering AB.

Svensson U, Laaksoharju M, Gurban I, 2002. Äspö Hard Rock Laboratory: Impact of the tunnel construction on the groundwater system at Äspö. Task 5. ÄspöTask Force on groundwater flow and transport of solutes. SKB IPR-02-45, Svensk Kärnbränslehantering AB.

Svensson U, Kuylenstierna H-O, Ferry M, 2004. DarcyTools, Version 2.1: Concepts, methods, equations and demo simulations. SKB R-04-19, Svensk Kärnbränslehantering AB.

Tsang C-F, 1991. The modeling process and model validation. Ground Water, 29, 825, 830.

Vázquez-Suñé E, Sánchez-Vila J, Carrera J, Marizza M, Arandes R, Gutierrez L A, 1997. Rising of groundwater level in Barcelona: evolution and effects on urban structures, In: J Chilton (ed), Groundwater in the Urban Environment, A.A. Balkema, pp. 267–271.

Waber H N, Clark I D, Salameh E, Savage D, 1998. Hydrogeochemistry of the Maqarin area, In: JAT Smellie (Ed), MAQARIN natural analogue study: Phase III. SKB TR-98-04, pp. 181–236, Svensk Kärnbränslehantering AB.

Zuidema P, 1995. Validation: demonstration of disposal safety requires a practicable approach. In: GEOVAL'94, Validation through model testing, Proceedings of a NEA/SKI Symposium, Paris, 11–14 October, 1994. OECD, EUR 16001, pp. 35–42.

Appendix A

Stable isotopes of 1,550 groundwater samples from the Scandinavian Shield

$\delta^{18}O$	$\delta^{2}H$	Cl (mg/L)
–11.07	–67.2	23,800
–9.0	–70.3	22,000
–9.0	–70.3	22,000
–12.0	–85.3	16,140
–12.3	–87.5	15,870
–12.3	–88.1	15,860
–12.2	–94.1	15,700
–12.2	–87.2	15,680
–12.4	–87.6	15,400
–12.5	–89.0	15,400
–11.4	–83.2	15,130
–11.4	–83.2	15,130
–11.1	–76.7	14,973.6
–12.2	–81.1	14,920.4
–12.4	–83.8	14,820
–11.4	–83.2	14,810
–10.7	–72.9	14,800
–10.7	–72.9	14,800
–10.7	–73.3	14,800
–12.3	–82.9	14,551.1
–11.9	–86.4	14,500
–11.4	–83.5	14,220
–12.2	–88.1	13,970
–12.2	–89.1	13,920
–11.81	–76.9	13,500
–10.4	–73.3	13,400
–12.7	–90.8	13,300
–12.2	–103.8	13,300
–12.4	–84.0	13,265.6
–12.0	–86.5	13,140
–10.4	–75.8	13,070
–11.0	–82.9	12,930
–12.6	–86.1	12,883
–13.0	–89.6	12,804.7
–11.9	–85.7	12,800
–10.4	–75.2	12,590
–11.7	–82.6	12,500
–10.3	–90.3	12,470
–10.5	–75.3	12,460
–12.5	–91.2	12,450
–11.5	–84.5	12,430
–10.2	–75.9	12,400
–10.5	–75.6	12,390
–10.5	–75.9	12,380
–10.5	–75.4	12,360
–13.0	–88.6	12,351.8
–10.5	–75.9	12,320
–10.4	–75.1	12,320
–12.7	–96.4	12,300
–12.7	–96.4	12,300
–11.27	–73.43	12,300
–11.6	–91.0	12,300
–12.4	–84.9	12,279.7
–13.6	–91.6	12,160
–13.2	–90.6	12,083.3
–12.5	–88.5	12,054
–10.0	–73.0	12,000
–12.8	–91.9	11,720
–8.7	–64.4	11,500
–12.5	–91.1	11,480
–11.0	–82.2	11,480
–11.0	–82.2	11,480
–12.8	–94.8	11,460
–12.9	–95.2	11,430
–12.8	–94.7	11,400
–12.8	–92.3	11,360
–12.7	–102.4	11,300
–13.0	–95.3	11,283.1
–13.1	–96.8	11,100
–12.4	–89.5	11,100
–13.1	–96.8	11,100
–13.0	–96.8	11,100
–13.0	–96.8	11,100
–12.8	–95.3	11,097.8
–12.4	–97.7	10,730
–11.9	–89.3	10,600
–13.7	–92.2	10,565
–13.1	–91.9	10,425
–13.6	–99.7	10,200
–9.1	–69.5	10,000
–14.1	–97.0	10,000
–14.1	–95.8	9,990
–13.6	–98.4	9,950
–13.6	–99.9	9,890
–13.6	–99.1	9,890
–13.5	–98.6	9,740
–13.5	–97.4	9,720
–13.6	–98.5	9,690
–11.8	–89.3	9,500
–11.8	–89.3	9,500
–11.4	–80.6	9,350
–12.7	–92.1	8,875.7
–10.9	–75.8	8,800
–12.6	–91.2	8,755.1
–11.9	–86.2	8,749.8
–13.3	–95.8	8,662.9
–12.7	–95.1	8,630
–12.0	–86.1	8,590
–11.2	–81.2	8,500
–11.1	–80.4	8,400
–12.3	–85.4	8,335
–12.9	–93.4	8,262.3
–11.1	–87.6	8,130
–12.5	–91.5	8,130
–13.0	–99.7	8,080
–13.0	–99.7	8,080
–11.3	–81.5	8,040
–12.27	–84.5	8,030
–12.1	–89.0	7,996.4
–11.7	–82.8	7,979.3
–12.4	–93.3	7,900
–12.4	–105.5	7,860
–12.8	–93.8	7,803.2
–10.9	–77.5	7,771.3
–11.2	–78.1	7,741.2
–11.1	–81.2	7,580
–11.0	–78.0	7,530
–11.5	–86.9	7,510
–12.9	–95.6	7,290
–12.5	–90.3	7,270
–11.7	–86.2	7,230
–10.4	–78.9	7,130
–10.1	–76.7	7,080
–11.2	–81.3	7,040
–10.3	–74.6	7,030
–10.5	–83.2	7,010
–12.4	–89.1	6,960
–12.4	–90.2	6,950
–10.6	–80.3	6,940
–10.1	–78.1	6,930
–10.9	–78.3	6,915.1
–10.3	–82.1	6,900

δ18O	δ2H	Cl (mg/L)
–10.9	–78.4	6,819.4
–10.5	–81.8	6,710
–10.1	–87.1	6,690
–12.5	–79.4	6,670
–10.1	–79.2	6,630
–11.2	–85.1	6,610
–10.3	–75.1	6,580
–10.6	–82.0	6,560
–9.7	–75.7	6,550
–10.1	–77.9	6,550
–10.1	–77.3	6,540
–4.7	–38.0	6,500
–12.6	–93.4	6,425.9
–9.9	–91.8	6,420
–12.3	–97.2	6,410
–11.5	–84.5	6,403.7
–12.1	–91.2	6,390
–10.1	–78.6	6,390
–13.3	–98.6	6,370
–13.3	–100.4	6,350
–13.8	–99.1	6,342.5
–14.0	–99.8	6,335.5
–13.4	–100.2	6322.2
–10.8	–84.3	6,300
–14.0	–100.0	6,298.2
–9.9	–78.0	6,296.5
–12.1	–89.8	6,290
–11.3	–84.8	6,270
–10.3	–79.0	6,240
–13.4	–99.5	6,230
–10.5	–85.7	6,220
–10.4	–80.4	6,210
–11.5	–85.5	6,207.3
–10.6	–89.0	6,200
–9.8	–76.1	6,190
–9.8	–77.8	6,180
–8.2	–70.8	6,150
–8.2	–70.8	6,150
–10.3	–79.9	6,110
–10.8	–78.9	6,110
–13.4	–100.2	6,100
–11.7	–87.2	6,064.5
–11.9	–89.2	6,062.5
–12.1	–94.2	6,060
–9.8	–77.5	6,050
–9.4	–76.4	6,000.8
–9.6	–74.4	6,000
–10.0	–78.1	5,990
–10.5	–81.2	5,990
–10.2	–78.2	5,990
–10.8	–78.5	5,985.4
–14.1	–102.5	5,982.3
–7.4	–69.2	5,970
–9.7	–75.9	5,970
–7.4	–69.2	5,970
–13.8	–98.9	5,970
–9.9	–76.4	5,950
–9.4	–74.3	5,940
–9.4	–82.5	5,930
–10.7	–77.6	5,921.5
–9.5	–74.0	5,920.6
–9.6	–74.8	5,910
–10.9	–90.1	5,890
–13.3	–103.4	5,880
–13.3	–103.4	5,880
–9.8	–85.9	5,870
–11.7	–86.9	5,865
–11.7	–86.9	5,865
–10.6	–78.5	5,854.2
–10.3	–75.9	5,849.7
–11.9	–92.3	5,840
–11.9	–92.3	5,840
–9.3	–74.1	5,830
–9.5	–74.1	5,830
–10.5	–78.3	5,815.5
–11.1	–85.1	5,790
–10.2	–74.9	5,751.4
–10.4	–75.8	5,750.5
–10.7	–97.3	5,750
–9.7	–74.7	5,750
–11.0	–84.7	5,730
–9.6	–74.3	5,690
–10.1	–73.9	5,687.5
–10.2	–73.7	5,684
–9.2	–77.8	5,680
–9.2	–77.8	5,680
–9.7	–75.7	5,680
–9.1	–68.3	5,680
–9.2	–77.7	5,670
–9.6	–74.7	5,660
–11.5	–84.9	5,653
–12.2	–88	5,650
–10.0	–88.2	5,640
–13.01	–98.2	5,590.1
–9.0	–84.4	5,580
–10.8	–83.5	5,570
–9.5	–76.2	5,570
–9.6	–81.0	5,570
–9.5	–74.5	5,560
–9.1	–70.6	5,550
–10.0	–73.4	5,547.5
–9.5	–73.4	5,540
–9.3	–71.1	5,540
–9.6	–75.1	5540
–9.5	–73.4	5,540
–9.6	–72.7	5,530
–9.7	–73.5	5,530
–9.0	–70.7	5,510
–13.0	–98.0	5,507.6
–11.81	–88.7	5,500
–11.1	–82.8	5,498.8
–9.6	–70.4	5,490
–13.0	–97.8	5,489.9
–11.6	–85.7	5,489
–9.0	–70.9	5,480
–11.8	–84.3	5,470
–12.4	–91.6	5,470
–7.5	–57.6	5,470
–9.1	–69.4	5,450
–12.8	–99.9	5,440
–12.8	–99.9	5,440
–11.7	–85.5	5,440
–9.2	–69.1	5,440
–13.8	–98.1	5,439.4
–13.0	–97.5	5,436.7
–9.4	–70.8	5,430
–11.6	–84.3	5,430
–8.4	–66.1	5,421.7
–9.1	–69.5	5,420
–9.4	–74.7	5,420
–10.1	–77.7	5,420
–9.4	–81.5	5,410
–8.8	–67.0	5,410
–8.8	–79.0	5,410
–9.0	–70.9	5,410
–9.6	–76.9	5,400
–11.7	–89.3	5,400
–11.7	–89.3	5,400
–8.7	–66.5	5,400
–8.8	–66.5	5,400
–9.8	–70.5	5,380
–8.5	–71.3	5,380
–11.6	–84.4	5,380
–11.6	–83.4	5,380
–10.2	–75.6	5,380
–12.7	–99.8	5,360
–12.7	–99.8	5,360
–10.7	–82.6	5,360
–9.2	–71.7	5,360
–9.1	–72.9	5,350

δ¹⁸O	δ²H	Cl (mg/L)
–9.1	–72.9	5,350
–12.3	–100.6	5,340
–12.3	–100.6	5,340
–9.9	–72.6	5,336.6
–12.9	–93.6	5,333.9
–9.1	–78.4	5,330
–8.8	–69.3	5,329.5
–9.1	–71.0	5,310
–8.5	–65.2	5,282.5
–8.8	–69.2	5,270
–8.8	–69.1	5,258.4
–9.5	–75.8	5,250
–8.8	–69.6	5,220.5
–10.4	–82.7	5,220
–11.75	–89.0	5,200
–8.8	–69.4	5,197.4
–9.6	–72.3	5,190
–8.8	–70.2	5,186.8
–13.3	–104.9	5,180
–8.9	–72.1	5,180
–13.3	–104.9	5,180
–12.7	–95.1	5,166.4
–8.4	–65.6	5,162.8
–7.8	–63.7	5,160
–13.1	–103.1	5,160
–8.8	–69.3	5,153.1
–10.3	–81.4	5,150
–8.8	–69.3	5,140.7
–8.9	–75.4	5,140
–11.9	–83.4	5,138
–10.2	–78.4	5,119.4
–12.34	–87.14	5,100
–11.1	–87.4	5,100
–8.8	–73.0	5,091.1
–10.2	–78.4	5,091.1
–10.2	–76.9	5,084
–9.2	–73.7	5,080
–9.2	–68.8	5,075.8
–7.2	–69.3	5,070
–9.1	–72.6	5,060
–8.9	–68.8	5,041.4
–9.7	–76.5	5,010
–9.7	–76.5	5,010
–9.0	–70.0	5,000
–12.0	–84.4	4,986.5
–10.1	–78.8	4,970
–10.3	–83.2	4,960.5
–8.5	–64.5	4,950
–9.6	–71.9	4,940
–9.2	–73.9	4,940
–11.2	–83.5	4,940
–9.9	–77.0	4,920.9
–11.1	–81.9	4,910
–8.8	–70.2	4,890
–11.3	–85.9	4,890
–9.5	–70.8	4,888.1
–9.5	–71.8	4,880
–9.2	–72.1	4,880
–8.6	–76.3	4,880
–10.0	–75.5	4,878.3
–8.7	–69.7	4,870
–13.3	–102.1	4,870
–8.9	–69.4	4,860
–8.7	–73.5	4,830
–11.2	–85.6	4,828
–9.1	–74.1	4,810
–9.2	–73.5	4,784.7
–9.4	–68.7	4,779.1
–9.6	–79.7	4,770
–9.6	–79.7	4,770
–7.5	–59.2	4,770
–9.5	–70.5	4,767.5
–8.5	–70.4	4,760
–9.2	–71.0	4,746.4
–9.4	–68.4	4,744.5
–10.9	–79.1	4,730
–11.6	–85.1	4,730
–10.3	–86.8	4,720
–9.6	–71.6	4,706.4
–11.3	–84.4	4,700
–11.3	–84.4	4,700
–9.3	–74.3	4,690
–11.8	–98.8	4,680
–10.2	–78.2	4,652.5
–11.8	–88.0	4,650
–8.5	–69.1	4,620
–11.2	–86.6	4,610
–9.2	–67.5	4,608.5
–8.5	–67.2	4,600
–10.6	–86.1	4,600
–9.2	–69.7	4,600
–13.6	–109.6	4,600
–9.2	–69.7	4,600
–13.6	–109.6	4,600
–7.7	–58.2	4,600
–7.6	–57.5	4,600
–8.7	–67.6	4,590
–8.6	–65.1	4,580
–11.6	–88.9	4,567.2
–11.6	–88.2	4,562.8
–11.5	–86.7	4,546.8
–11.5	–87.3	4,541.2
–11.2	–90.1	4,520
–10.0	–80	4,510
–9.6	–74.9	4,505
–10.2	–78.1	4,503.4
–10.0	–73.5	4,500
–9.5	–72.4	4,500
–7.9	–63.4	4,500
–13.3	–100.3	4,500
–9.5	–72.4	4,500
–11.6	–88.9	4,494.6
–8.6	–69.3	4,480
–9.4	–70.2	4,470
–9.5	–73.2	4,466.4
–8.8	–64.4	4,462.1
–8.6	–74.4	4,460
–8.7	–73.2	4,460
–8.6	–72.4	4,460
–8.7	–71.8	4,460
–10.0	–81.9	4,430
–9.3	–70.4	4,424.5
–9.3	–68.4	4,389.1
–7.3	–52.0	4,350
–7.3	–57.6	4,350
–10.3	–75.0	4,340.3
–8.9	–69.0	4,340
–9.1	–67.3	4,325.3
–12.3	–87.8	4,311.1
–10.3	–77.3	4,300
–11.1	–81.2	4,300
–11.7	–86.2	4,300
–10.3	–77.3	4,300
–11.1	–81.2	4,300
–8.3	–66.8	4,290
–8.7	–63.5	4,288
–8.1	–65.0	4,270
–9.1	–67.9	4,260
–14.6	–107.6	4,260
–11.7	–93.6	4,250
–8.8	–63.7	4,250
–9.4	–83.0	4,240
–8.4	–69.7	4,240
–8.4	–69.7	4,240
–14.2	–107.1	4,230
–8.7	–63.1	4,227.8
–7.7	–64.5	4,220
–7.7	–64.5	4,220
–11.4	–90.7	4,220
–8.0	–66.4	4,210

δ18O	δ2H	Cl (mg/L)
–9.4	–73.9	4,210
–8.0	–66.4	4,210
–12.4	–90.8	4,200
–8.8	–64.8	4,200
–12.4	–90.8	4,200
–7.9	–72.4	4,200
–9.7	–76.3	4,180
–9.7	–77.1	4,170
–9.4	–78.1	4,170
–12.0	–96.8	4,160
–11.2	–86.1	4,158.6
–9.9	–79.4	4,150
–7.9	–67.7	4,150
–7.5	–66.0	4,150
–8.5	–68.8	4,140
–12.0	–95.0	4,140
–6.6	–50.1	4,140
–8.0	–70.7	4,140
–11.1	–84.8	4,125.6
–7.6	–62.9	4,110
–7.6	–62.9	4,110
–8.9	–66.8	4,100
–7.5	–56.0	4,087.9
–6.7	–54.6	4,084.2
–8.3	–68.0	4,080
–11.5	–89.2	4,078.2
–9.3	–69.6	4,072.7
–8.0	–67.4	4,070
–8.0	–67.4	4,070
–14.7	–110.9	4,060
–8.0	–65.3	4,060
–9.0	–67.5	4,050.5
–10.3	–86.4	4,040
–9.2	–74.3	4,020
–8.1	–65.1	4,020
–8.1	–64.2	4,020
–7.9	–70.1	4,020
–8.9	–65.5	4,011.9
–11.6	–89.1	4,000
–11.6	–89.1	4,000
–12.8	–103.9	3,990
–7.3	–58.1	3,984.1
–7.9	–71.3	3,980
–8.8	–64.3	3,977.8
–7.9	–71.2	3,970
–8.7	–66.0	3,960
–7.8	–70.6	3,940
–8.2	–65.7	3,930
–11.7	–86.1	3,922.9
–13.1	–103.6	3,920
–7.4	–65.3	3,920
–9.4	–55.8	3,919.3
–8.8	–66	3,910
–11.5	–93.2	3,910
–7.9	–71.1	3,910
–7.8	–72.5	3,910
–9.6	–78.0	3,900
–12.2	–91.5	3,900
–8.5	–68.5	3,900
–12.2	–91.5	3,900
–10.1	–77.3	3,898.8
–8.1	–68.7	3,890
–8.3	–73.0	3,890
–7.4	–64.6	3,890
–12.7	–117.1	3,880
–8.7	–64.4	3,876.1
–10.8	–80.9	3,870
–10.8	–80.9	3,870
–6.8	–54.5	3,850.2
–8.1	–65.9	3,850
–8.6	–64.2	3,841.3
–7.7	–71.9	3,840
–6.9	–55.2	3,830.7
–8.1	–72.2	3,830
–7.7	–73.4	3,830
–11.4	–80.1	3,826.3
–13.9	–108.9	3,820
–8.2	–65.3	3,820
–13.9	–108.9	3,820
–7.4	–61.9	3,820
–9.2	–68.9	3,810
–8.9	–72.9	3,809.8
–12.8	–98.3	3,800
–10.0	–74.0	3,800
–10.0	–74.0	3,800
–6.8	–55.4	3,792.6
–7.4	–64.4	3,790
–12.4	–88.6	3,781.9
–7.0	–57.2	3,780
–8.8	–69.8	3,780
–12.7	–100.1	3,780
–8.2	–65.2	3770
–8.8	–66.2	3770
–7.6	–73.7	3760
–5.9	–53.3	3760
–6.8	–56.8	3757
–6.9	–57.2	3746.3
–10.3	–79.2	3744.2
–7.4	–72.8	3,730
–7.6	–63	3,730
–7.1	–56.1	3,727.5
–7.4	–61.5	3,720
–6.9	–57.0	3,706.3
–7.1	–57.3	3,690.3
–8.6	–64.9	3,690
–7.1	–56.2	3,689.2
–7.1	–55.6	3,682.1
–6.7	–53.6	3,680
–8.9	–73.7	3,680
–7.1	–57.9	3,674
–9.1	–72.0	3,670
–7.4	–69.5	3,670
–7.2	–56.9	3,665.1
–7.6	–71.4	3,650
–6.7	–56.9	3,650
–7.1	–55.3	3,642.8
–10.2	–78.1	3,640.7
–8.3	–65.0	3,640
–10.4	–75.9	3,640
–7.5	–70.0	3,640
–7.3	–60.2	3,630
–10.9	–94.3	3,630
–10.9	–94.3	3,630
–6.7	–58.7	3,630
–7.0	–77.0	3,630
–8.5	–64.2	3,630
–6.5	–52.3	3,620
–13.6	–113.0	3,620
–7.2	–55.7	3,618.7
–7.4	–59.0	3,610
–7.0	–54.1	3,610
–7.3	–59.6	3,570
–8.5	–64.2	3,570
–9.1	–73.0	3,570
–7.4	–59.2	3,560
–9.7	–79.2	3,560
–7.1	–56.2	3,541.8
–7.3	–61.5	3,538.4
–7.0	–54.9	3,532.5
–9.6	–77.5	3,530
–7.5	–74.2	3,530
–9.7	–80.1	3,530
–7.0	–52.5	3,530
–6.9	–54.8	3,529.2
–8.2	–63.0	3,528.3
–9.0	–75.2	3,510
–7.2	–56.9	3,497.8
–7.1	–55.2	3,491.8
–7.2	–55.8	3,491.2
–6.8	–57.0	3,490

$\delta^{18}O$	$\delta^{2}H$	Cl (mg/L)
–6.9	–59.0	3,490
–7.5	–71.3	3,490
–7.2	–56.7	3,484.7
–7.7	–64.3	3,480
–7.4	–59.6	3,480
–8.5	–63.0	3,474.4
–8.2	–65.8	3,470
–7.4	–60.4	3,470
–7.5	–59.8	3,450
–7.4	–63.7	3,450
–10.9	–86.0	3,450
–7.2	–56.2	3,449.2
–11.7	–83.0	3,445.3
–6.8	–56.8	3,441.4
–7.2	–57.6	3,434.5
–7.5	–62.4	3,420
–12.4	–92.1	3,420
–10.1	–82.5	3,420
–8.2	–64.6	3,420
–12.4	–92.1	3,420
–7.5	–59.0	3,419.9
–7.0	–67.8	3,410
–7.9	–60.5	3,410
–12.1	–90.6	3,400
–10.5	–86.0	3,400
–7.9	–62.8	3,400
–12.1	–90.6	3,400
–7.7	–63.1	3,390
–7.0	–61.0	3,383
–6.0	–50.7	3,380
–7.2	–59.5	3,380
–7.5	–60.9	3,380
–6.9	–55.4	3,369.7
–8.0	–64.6	3,360
–7.3	–60.7	3,350
–7.5	–60.7	3,340
–13.9	–110.0	3,331
–6.8	–54.4	3,325.1
–7.1	–54.9	3,313.8
–6.9	–59.1	3,300.3
–9.8	–75.7	3,300
–9.5	–75.1	3,300
–8.2	–66.0	3,290
–6.9	–57.7	3,286.8
–8.6	–63.8	3,281.5
–6.9	–56.9	3,263.8
–8.5	–66.3	3,260
–13.3	–94.9	3,250
–8.3	–63.4	3,250
–7.2	–62.8	3,250
–8.6	–64.3	3,250
–8.2	–62.6	3,250
–7.1	–56.3	3,247.1
–7.4	–59.9	3,246.4
–7.3	–61.6	3,240
–7.5	–61.4	3,230
–7.3	–67.5	3,230
–10.6	–78.5	3,200
–7.6	–61.1	3,200
–8.7	–68.7	3,180
–7.1	–54.5	3,169.5
–8.3	–68.1	3,160
–8.1	–61.5	3,160
–8.3	–63.3	3,150
–7.1	–59.8	3,140
–7.4	–63.6	3,120
–7.2	–58.6	3,120
–7.6	–61.2	3,110
–8.6	–66.9	3,110
–7.2	–58.1	3,110
–7.7	–68	3,100
–7.3	–60.3	3,100
–10.3	–79.6	3,093.6
–9.2	–68.8	3,086.2
–7.0	–59.5	3,080
–7.8	–63.6	3,080
–6.9	–62	3,080
–8.5	–65.9	3,070
–8.1	–63.8	3,060
–7.4	–60.7	3,060
–7.3	–58.1	3,050
–13.0	–99.6	3,030
–7.6	–60.7	3,030
–7.6	–57.4	3,030
–13.0	–99.6	3,030
–7.1	–58.8	3,030
–7.7	–58	3,020.2
–7.5	–60.9	3,020
–7.0	–56.3	3,020
–7.0	–59.4	3,020
–7.0	–59.6	3,015.6
–7.1	–58.5	3,010
–7.9	–74.1	3,003.6
–8.1	–62.1	3,000
–13.0	–96.7	3,000
–7.1	–56.9	2,991.5
–7.1	–57.7	2,990
–7.1	–57.8	2,980
–7.3	–57.5	2,963.9
–11.1	–81.5	2,960
–7.2	–58.5	2,960
–8.9	–71.9	2,960
–14.5	–118.1	2,950
–14.5	–118.1	2,950
–8.3	–63.9	2,946.1
–7.0	–58.6	2,940
–8.7	–70.3	2,940
–14.4	–111.6	2,940
–7.0	–72.9	2,930
–7.9	–58.9	2,919.2
–7.8	–59.8	2,915.3
–7.0	–57.9	2,910
–8.3	–64.8	2,910
–8.6	–67.0	2,910
–7.6	–56.9	2,901.1
–8.0	–66.4	2,900
–7.6	–57.9	2,893.7
–12.3	–92.5	2,880
–7.0	–58.9	2,870
–11.2	–79.1	2,865.7
–6.9	–58.4	2,860
–7.5	–59.1	2,860
–7.1	–57.1	2,860
–14.6	–115.3	2,850
–10.79	–77.63	2,850
–14.6	–115.3	2,850
–7.2	–57.3	2,850
–8.2	–64.5	2,840.7
–8.3	–68.2	2,840
–7.4	–56.1	2,833.4
–8.1	–64.8	2,808.5
–11.59	–84.0	2,800
–12.1	–92.1	2,800
–12.1	–92.1	2,800
–8.2	–68.1	2,800
–8.0	–63.4	2,800
–10.8	–83.8	2,790
–10.8	–76.5	2,790
–8.4	–73.6	2,790
–8.2	–65.1	2,786.8
–7.3	–73.5	2,780
–7.0	–66.7	2,780
–8.5	–61.8	2,770
–8.1	–63.6	2,770
–8.2	–65.6	2,770
–7.5	–56.3	2,763.9
–13.6	–103.4	2,760
–8.1	–62.4	2,760
–12.8	–98.7	2,760
–13.13	–96.9	2,760

δ¹⁸O	δ²H	Cl (mg/L)
–8.0	–62.1	2,757.2
–8.2	–65.1	2,751.9
–8.1	–61.0	2,740.5
–12.0	–90.2	2,730
–8.5	–65.8	2,730
–14.4	–111.5	2,730
–8.5	–65.8	2,730
–7.3	–58.9	2,730
–8.5	–65.1	2,723.5
–8.0	–61.1	2,720
–14.5	–111.7	2,720
–7.1	–57.1	2,720
–7.6	–58.2	2,720
–7.2	–59.1	2,720
–8.3	–65.6	2,710
–12.8	–94.9	2,700
–12.8	–94.9	2,700
–15.0	–112.2	2,700
–9.4	–72.8	2,694.1
–7.3	–60.9	2,690
–14.4	–112.4	2,690
–8.3	–64.1	2,685.6
–15.0	–110.9	2,681
–6.9	–57.2	2,680
–7.6	–58.3	2,676.7
–7.1	–57.0	2,674.9
–8.5	–77.7	2,670
–8.1	–64.1	2,670
–7.1	–60.0	2,669.3
–8.3	–64.1	2,667.1
–8.1	–65.3	2,660
–8.3	–66.4	2,660
–8.2	–61.6	2,656.8
–7.7	–58.3	2,650
–8.1	–61.8	2,636
–7.9	–65.9	2,630
–9.9	–75.5	2,630
–8.5	–66.4	2,630
–14.5	–126.2	2,630
–8.2	–61.3	2,627.6
–7.8	–65.1	2,627.4
–12.6	–93.7	2,622.1
–12.6	–93.6	2,617.1
–8.1	–66.1	2,615.6
–8.5	–60.9	2,613.2
–8.0	–65.7	2,612.9
–8.1	–61.6	2,610.2
–8.4	–61.4	2,610
–8.3	–65.8	2,610
–8.4	–62.6	2,591.5
–8.3	–61.2	2,590
–7.8	–60.1	2,582.4
–8.2	–61.6	2,573.1
–10.8	–77.4	2,560
–8.2	–62.5	2,555.1
–8.6	–67.2	2,540
–7.2	–70.3	2,530
–7.6	–64.1	2,522.8
–8.1	–60.9	2,520.9
–8.3	–67.5	2,520.4
–8.2	–64.0	2,520
–8.4	–67.1	2,510
–7.9	–72.5	2,509.5
–7.6	–63.7	2,507.5
–11.4	–93.3	2,500
–8.0	–60.2	2,494.6
–8.5	–62.4	2,476.4
–8.5	–64.1	2,464.7
–11.2	–87.1	2,460
–11.2	–87.1	2,460
–12.0	–94.0	2,450
–13.39	–96.9	2,450
–8.3	–61.6	2,449.8
–8.2	–67.5	2,437.6
–8.5	–63.5	2,427.3
–12.0	–90.3	2,400
–8.7	–65.2	2,365
–9.7	–78.5	2,340
–8.7	–67.5	2,340
–8.7	–67.5	2,340
–10.6	–76.4	2,330
–10.9	–77.6	2,300
–9.6	–75.1	2,290
–10.5	–78.9	2,290
–9.5	–77.3	2,239.9
–8.7	–64.8	2,204.5
–9.6	–78.2	2,192.2
–10.0	–74.8	2,180
–11.9	–92.2	2,160
–10.4	–80.1	2,150
–10.4	–80.0	2,130
–10.3	–79.8	2,110
–8.4	–63.0	2,100.9
–8.0	–61.5	2,098.5
–11.5	–89.9	2,050
–10.3	–80.6	2,040
–10.3	–80	2,020
–10.4	–81.3	1,990
–10.9	–80.3	1,970
–8.1	–60.7	1,952.9
–9.9	–81.1	1,920
–8.8	–68.0	1,917.3
–8.8	–68.2	1,916.1
–11.8	–87.9	1,900
–10.0	–80.0	1,900
–9.9	–81.9	1,880
–12.2	–87.5	1,850
–10.28	–74.79	1,850
–9.9	–80.2	1,840
–10.9	–77.0	1,830.1
–9.4	–73.3	1,810
–9.8	–80.1	1,800
–10.1	–73.2	1,800
–9.9	–72.7	1,790
–9.5	–73.5	1,778
–10.8	–81.0	1,776.9
–9.4	–66.6	1,760
–9.7	–80.0	1,750
–11.4	–83.4	1,740
–9.9	–73.2	1,740
–9.9	–74.6	1,740
–9.8	–72.2	1,720
–10.0	–75.9	1,710
–9.7	–75.5	1,701.6
–12.2	–94.5	1,700
–9.9	–72.8	1,690
–9.9	–67.9	1,690
–9.9	–74.6	1,690
–9.9	–74.6	1,680
–12.4	–91.4	1,658.5
–10.0	–70.8	1,650
–11.2	–81.2	1,650
–9.8	–77.8	1,634.2
–11.7	–89.9	1,598.3
–10.7	–85.5	1,560
–11.0	–84.0	1,560
–11.9	–89.2	1,500
–11.9	–89.2	1,500
–11.72	–85.7	1,500
–12.3	–90.3	1,421.7
–10.8	–80.1	1,300
–10.6	–77.8	1,298.6
–9.5	–68.0	1,280
–10.1	–76.3	1,270
–11.0	–76.8	1,270
–9.8	–76.1	1,260
–10.6	–82.3	1,250
–15.8	–124.8	1,220
–15.8	–124.8	1,220
–10.2	–80.6	1,170

δ18O	δ2H	Cl (mg/L)
–8.8	–67.5	1,170
–11.78	–85.37	1,160
–10.4	–77.4	1,160
–9.3	–71.3	1,145.3
–10.0	–76.0	1,140
–10.0	–76.0	1,140
–10.4	–79.7	1,130
–10.2	–77.5	1,130
–10.9	–81.3	1,130
–10.7	–82.6	1,120
–10.6	–82.2	1,120
–10.7	–82.4	1,120
–11.1	–80.5	1,114.1
–10.3	–81.1	1,110
–10.6	–79.9	1,108.3
–10.3	–78.3	1,080
–10.4	–80.2	1,070
–9.6	–73.4	1,060.6
–10.3	–75.8	1,056
–10.2	–81.9	1,040
–7.9	–68.2	1,040
–10.4	–80.5	1,040
–9.3	–72.5	1,040
–10.9	–77.6	1,030
–10.7	–75.6	1,028.1
–10.5	–78.9	1,020
–9.9	–79.4	1,020
–10.4	–77.3	1,010
–9.8	–81.1	1,010
–9.8	–80.3	1,000
–10.5	–85.3	1,000
–9.9	–79.9	970
–9.5	–86.9	947
–9.5	–87.0	938
–9.7	–78.5	934
–9.8	–72.4	932
–10.7	–79.3	932
–10.3	–81.0	924
–9.7	–69.0	922.8
–11.5	–89.4	918
–11.5	–83.2	915
–10.1	–71.3	910.6
–10.7	–77.2	910
–10.7	–77.2	910
–10.1	–75.2	901
–9.4	–74.7	900
–10.1	–76.4	895
–10.1	–75.8	895
–10.7	–81.4	895
–9.6	–75.5	890
–9.1	–72.3	888
–11.2	–78.1	887
–10.2	–76.7	878
–10.1	–82.3	876
–9.7	–72.4	876
–9.2	–89.1	874
–10.3	–74.8	870
–11.1	–77.5	857
–9.4	–72.3	856.7
–10.2	–77.8	843
–9.8	–71.9	840
–9.6	–77.0	840
–9.8	–72.9	835
–10.0	–72.4	830
–9.6	–80.8	827
–9.8	–72.9	823
–9.3	–80.1	818
–9.8	–76.9	812
–10.1	–77.6	805
–10.9	–80.3	797.4
–10.1	–69.6	792
–9.7	–69.2	789
–11.3	–78.8	785
–10.95	–76.64	780
–9.9	–68.1	780
–10.1	–72.9	760
–9.7	–73.8	739
–9.6	–76.7	737
–9.6	–75.0	733.5
–10.1	–73.7	729
–9.9	–72.0	710
–11.1	–77.4	697
–11.4	–85.6	690
–9.6	–74.1	665
–11.7	–90.0	665
–9.9	–75.1	662
–9.7	–72.4	654
–9.4	–64.6	651.1
–9.9	–73.9	646
–9.9	–74.1	645
–9.9	–73.1	644.4
–9.7	–86.9	643
–11.6	–80.9	642
–9.7	–80.5	628
–11.7	–82.2	612
–9.9	–70.1	610
–10.9	–80.5	608
–11.1	–81.2	585.8
–11.7	–82.1	584
–10.6	–78.6	575
–10.8	–76.4	574
–9.8	–75.3	560
–9.9	–78.3	560
–11.9	–89.0	555
–11.2	–83.9	555
–11.9	–89.0	555
–10.9	–78.7	548
–11.5	–79.8	545
–11.1	–83.4	543
–10.5	–75.5	530
–9.5	–64.2	529.8
–9.9	–71.0	519
–9.8	–74.3	510
–10.7	–76.2	509
–11.0	–84.8	508
–11.0	–84.8	508
–11.6	–81.1	503.3
–9.8	–71.4	484
–11.5	–82.1	482
–9.9	–75.1	476
–11.2	–82.6	476
–10.4	–73.2	474.5
–10.0	–73.5	440
–10.8	–78.0	440
–10.7	–82.7	438.4
–12.8	–97.2	435
–10.8	–81.1	431.7
–10.9	–79.2	430
–11.1	–78.4	430
–10.9	–80.0	415
–9.9	–74.1	415
–11.1	–78.4	413
–12.8	–90.7	396
–10.8	–76.3	392.4
–11.7	–84.6	389.4
–11.8	–88.0	380
–11.1	–80.3	371.1
–10.5	–85.6	369
–7.6	–67.4	332.7
–10.0	–73.9	329.3
–8.0	–65.7	307
–7.6	–68.0	303.6
–10.9	–90.6	300.6
–7.6	–65.6	290
–11.0	–82.4	289.4
–10.2	–73.3	280
–10.7	–79.0	280
–10.7	–79.0	280
–10.7	–79.0	280
–7.6	–67.5	276.7

δ18O	δ2H	Cl (mg/L)
–10.9	–77.7	275
–12.2	–87.3	266
–10.7	–79.0	260
–13.81	–100.8	260
–11.1	–84.0	259.4
–10.8	–79.3	250
–9.9	–75.3	250
–10.2	–77.1	230
–10.1	–76.8	230
–10.1	–76.5	229
–10.0	–77.2	228
–10.2	–77.8	227
–10.0	–76.3	225
–10.0	–76.5	225
–10.1	–77.3	224
–10.2	–76.9	221
–10.0	–73.0	220
–10.0	–76.5	215
–10.5	–76.0	215
–11.2	–82.9	215
–11.0	–83.5	211.6
–10.0	–76.8	211
–10.8	–79.7	210
–10.8	–79.7	210
–12.3	–82.1	198
–11.0	–80.9	189.1
–12.2	–80.9	187
–10.8	–82.9	184.8
–10.8	–82.8	182
–11.1	–80.6	181
–10.8	–82.8	178.8
–10.8	–81.9	178
–13.63	–99.4	178
–10.7	–84.8	176.7
–11.8	–87.0	173
–10.2	–73.9	168
–10.8	–76.6	162.5
–11.7	–84.8	160
–10.5	–75.1	146
–11.0	–80.7	145.7
–6.0	–60.6	136.2
–10.2	–75.0	133
–10.5	–78.6	133
–11.3	–83.0	133
–6.3	–62.6	127.2
–11.9	–83.7	127
–11.8	–95.2	126.2
–11.4	–83.0	124
–10.3	–76.3	123
–9.8	–68.7	123
–9.9	–73.8	119
–8.4	–63.0	117.8
–6.8	–63.5	116.5
–6.7	–64.3	116.1
–11.9	–83.5	113.1
–5.5	–50.1	111.7
–10.4	–76.1	109.0
–9.3	–69.0	109.0
–13.0	–99.6	109.0
–5.9	–53.4	108.9
–10.1	–73.5	108.0
–9.2	–70.6	107.7
–9.9	–79.7	106.0
–6.0	–53.8	101.6
–11.9	–84.0	100.3
–12.2	–84.7	99.8
–6.0	–52.3	98.1
–12.0	–87.0	96.7
–7.9	–64.4	95.9
–7.4	–62.2	90.8
–6.3	–55.8	88.9
–11.4	–82.1	84.4
–8.4	–66.9	84.2
–10.0	–79.3	83.4
–6.4	–54.8	83.2
–11.7	–84.6	81.4
–10.8	–77.4	78.8
–11.7	–85.4	76.4
–11.3	–85.0	75.0
–11.5	–85.0	75.0
–11.4	–85.0	75.0
–11.5	–81.0	74.0
–10.2	–73.3	73.0
–10.3	–73.4	73.0
–10.9	–85.0	72.0
–11.7	–84.5	72.0
–11.7	–83.9	71.4
–10.19	–80.8	69.4
–10.8	–73.8	68.3
–12.7	–92.4	68.1
–6.7	–62.1	66.4
–11.52	–83.0	66
–11.9	–89.4	63.3
–9.9	–71.2	63
–11.4	–83.0	63
–12.3	–86.2	62.3
–11.9	–86.5	61.9
–11.97	–88.2	61
–11.46	–82.1	60
–11.69	–83.9	60
–11.9	–89.6	59.7
–10.3	–73.4	59.3
–12.0	–85.9	57.5
–11.7	–79.0	57.1
–12.7	–92.5	56.4
–10.7	–76.1	55.1
–10.6	–78.0	53.1
–11.8	–82.3	51.5
–12.0	–82.2	50.3
–5.1	–57.0	49.4
–7.2	–61.9	49.3
–12.1	–85.2	48.6
–12.0	–88.2	48.4
–13.9	–101.5	48.0
–10.5	–73.4	47.5
–12.0	–89.2	47.0
–10.4	–83.0	47.0
–4.9	–46.2	46.8
–8.4	–68.4	46.1
–10.0	–77.4	46.0
–4.5	–44.3	45.8
–6.1	–51.7	45.7
–10.4	–83.0	45.0
–10.4	–83.0	44.0
–10.4	–83.0	44.0
–12.2	–84.4	42.5
–10.8	–79.0	42.3
–9.9	–71.0	41.0
–10.9	–79.0	40.9
–12.5	–85.8	40.7
–11.7	–87.4	40.6
–12.1	–88.5	40.6
–5.8	–63.8	38.9
–9.7	–71.0	38.0
–5.6	–51.5	37.7
–11.6	–87.0	37.0
–10.0	–69.0	37.0
–10.2	–70.1	36.0
–10.0	–77.8	35.4
–10.0	–82.2	35.4
–10.1	–81.2	35.4
–10.4	–80.3	34.6
–5.8	–65.3	34.1
–9.3	–75.1	33.6
–10.1	–75.9	33.5
–6.3	–62.5	33.5
–11.83	–85.3	33.0
–10.2	–77.0	33.0
–12.0	–83.3	32.5
–11.6	–85.6	32.1

δ¹⁸O	δ²H	Cl (mg/L)
–9.4	–71.1	30.1
–10.7	–85.9	30.0
–10.7	–85.9	30.0
–8.1	–62.4	29.9
–13.0	–97.4	29.0
–9.2	–74.0	28.8
–12.6	–89.1	28.1
–8.9	–70.2	27.9
–4.6	–46.3	27.8
–10.7	–73.7	26.9
–5.0	–63.2	26.6
–9.6	–71.5	26.2
–10.1	–78.5	26.0
–11.8	–86.2	25.9
–9.8	–72.3	25.8
–10.9	–75.7	25.7
–9.9	–71.1	25.6
–10.2	–73.9	25.5
–10.4	–77.0	25.0
–11.87	–85.0	25.0
–10.4	–77.0	25.0
–10.4	–77.0	25.0
–13.2	–94.9	24.1
–11.3	–79.8	24.0
–11.9	–81.4	23.5
–13.2	–97.5	23.0
–10.4	–77.0	23.0
–11.6	–87.0	23.0
–10.7	–76.3	22.6
–13.7	–100.0	22.2
–10.0	–73.4	22.0
–12.8	–100.3	22.0
–10.7	–80.4	21.2
–10.8	–78.1	21.0
–10.2	–85.5	20.5
–9.8	–71.5	20.4
–8.8	–59.0	20.0
–8.8	–65.1	19.9
–9.7	–71.8	19.8
–11.1	–83.7	19.7
–10.6	–75.9	19.6
–10.0	–74.9	19.2
–8.7	–67.7	19.2
–9.3	–69.7	19.1
–9.7	–72.9	19.1
–9.7	–79.1	19.0
–12.2	–90.1	18.9
–11.9	–85.2	18.7
–9.0	–82.3	18.6
–9.0	–72.8	18.6
–11.9	–88.3	18.4
–12.3	–90.6	18.3
–10.9	–81.0	18.1
–11.7	–82.5	18.1
–11.6	–88.0	18.0
–9.4	–72.3	18.0
–9.7	–76.3	17.9
–12.9	–96.0	17.8
–11.0	–80.2	17.8
–11.8	–85.1	17.8
–11.5	–84.9	17.7
–10.1	–78.7	17.7
–11.4	–79.8	17.7
–9.5	–74.4	17.5
–12.3	–86.3	17.1
–11.8	–85.2	17.1
–9.3	–68.7	17.0
–10.52	–77.1	17.0
–9.4	–72.8	16.3
–11.63	–82.1	16.0
–10.4	–77.7	15.8
–13.2	–102.6	15.8
–11.8	–79.6	15.7
–10.8	–77.5	15.7
–11.2	–77.4	15.5
–11.3	–82.9	15.5
–11.3	–82.9	15.5
–9.3	–69.5	15.3
–10.0	–72.4	15.2
–10.5	–76.4	15.1
–10.3	–78.0	15.0
–11.3	–84.1	15.0
–10.3	–78.0	15.0
–9.8	–67.1	15.0
–10.3	–78.0	15.0
–10.3	–78.0	15.0
–10.5	–77.4	14.9
–10.6	–76.4	14.9
–8.2	–63.5	14.6
–6.7	–54.4	14.4
–8.5	–64.9	14.2
–11.6	–83.5	14.1
–7.1	–55.7	14.1
–11.4	–81.9	14.0
–9.8	–71.8	14.0
–10.1	–72.2	13.9
–14.5	–108.0	13.8
–10.9	–80.8	13.7
–11.3	–80.3	13.6
–11.6	–81.7	13.5
–8.8	–79.9	13.5
–9.7	–70.4	13.5
–10.9	–78.0	13.4
–12.0	–85.1	13.4
–11.4	–77.6	13.3
–10.4	–77.8	13.2
–12.1	–91.4	13.2
–11.9	–86.1	13.1
–11.88	–81.7	13.0
–10.8	–77.1	13.0
–10.0	–71.3	12.9
–9.4	–71.7	12.8
–9.8	–75.5	12.8
–9.9	–74.9	12.7
–10.4	–73.3	12.7
–9.6	–72.7	12.6
–9.4	–78.2	12.5
–11.5	–85.0	12.5
–9.5	–62.5	12.4
–7.5	–62.6	12.3
–7.5	–62.6	12.3
–11.8	–83.1	12.2
–10.6	–77.0	12.1
–10.7	–74.2	12.0
–10.7	–83.5	12.0
–9.0	–69.0	12.0
–9.9	–68.4	12.0
–8.9	–69.3	12.0
–9.4	–77.0	12.0
–10.8	–75.8	11.8
–11.9	–86.3	11.8
–11.9	–85.2	11.8
–10.9	–77.5	11.8
–11.9	–86.5	11.7
–9.5	–75.8	11.7
–8.7	–65.4	11.6
–11.2	–77.6	11.5
–10.1	–76.2	11.4
–8.8	–67.3	11.3
–9.2	–70.5	11.3
–9.6	–75.3	11.2
–8.3	–64.2	11.2
–11.3	–76.5	11.2
–11.0	–80.1	11.2
–10.8	–80.0	11.0
–12.8	–93.1	11.0
–11.0	–79.0	11.0
–11.6	–88.0	11.0
–9.3	–72.0	10.9
–11.9	–86.0	10.9

δ18O	δ2H	Cl (mg/L)
–8.1	–62.9	10.8
–10.0	–76.5	10.7
–9.3	–72.3	10.7
–8.0	–64.9	10.6
–8.0	–63.9	10.6
–10.5	–75.5	10.6
–10.8	–78.6	10.5
–8.1	–67.8	10.5
–9.8	–68.5	10.4
–7.9	–63.6	10.4
–9.7	–68.6	10.3
–12.7	–90.4	10.3
–12.7	–90.4	10.3
–9.2	–68.4	10.3
–8.1	–67.3	10.2
–12.3	–84.8	10.2
–12.6	–94.8	10.0
–9.2	–74.0	9.9
–9.2	–66.0	9.9
–7.3	–61.5	9.8
–13.2	–94.9	9.7
–9.1	–69.8	9.6
–8.9	–70.8	9.6
–11.9	–88.6	9.6
–11.7	–88.8	9.5
–12.0	–90.0	9.4
–7.2	–61.8	9.3
–11.0	–82.5	9.3
–9.1	–65.6	9.2
–7.3	–62.0	9.2
–11.57	–82.6	9.0
–13.2	–95.3	9.0
–10.6	–76.3	8.9
–12.7	–93.8	8.9
–11.6	–81.4	8.9
–10.7	–75.1	8.8
–9.6	–81.9	8.8
–11.4	–79.3	8.8
–11.79	–83.9	8.7
–10.7	–73.7	8.7
–11.5	–78.1	8.5
–11.0	–83.1	8.5
–11.3	–81.6	8.5
–11.74	–83.4	8.4
–12.2	–87.3	8.3
–8.8	–73.8	8.3
–8.5	–61.6	8.3
–12.9	–93.1	8.2
–7.9	–63.2	8.2
–12.1	–88.3	8.2
–7.9	–63.3	8.2
–12.4	–94.6	8.17
–11.2	–76.0	8.1
–13.0	–95.0	8.0
–11.69	–84.7	8.0
–10.1	–72.3	8.0
–12.55	–89.7	7.9
–10.0	–75.6	7.8
–13.6	–99.3	7.8
–10.0	–76.0	7.7
–12.8	–99.3	7.6
–11.3	–80.4	7.6
–12.7	–90.1	7.6
–11.4	–84.6	7.5
–11.3	–77.2	7.5
–13.3	–94.2	7.4
–12.0	–84.3	7.4
–12.8	–87.8	7.4
–10.5	–72.1	7.4
–9.9	–72.9	7.3
–11.6	–78.8	7.3
–12.9	–98.4	7.2
–12.1	–87.5	7.2
–12.3	–85.9	7.2
–11.5	–81.5	7.0
–10.7	–78.6	7.0
–11.2	–86.2	7.0
–13.1	–95.4	7.0
–11.3	–82.0	6.6
–10.4	–86.6	6.56
–11.6	–83.5	6.5
–11.4	–77.2	6.4
–11.48	–81.9	6.4
–10.9	–78.8	6.3
–11.8	–87.9	6.2
–9.6	–77.8	6.2
–13.6	–99.3	6.2
–9.8	–69.7	6.1
–11.5	–81.8	6.0
–11.2	–76.7	6.0
–13.4	–99.9	6.0
–9.7	–68.9	6.0
–11.0	–78.3	6.0
–9.6	–74.4	5.9
–11.9	–89.2	5.9
–8.0	–66.9	5.9
–10.9	–80.0	5.8
–11.2	–77.8	5.8
–14.0	–105.8	5.8
–11.4	–81.8	5.8
–11.5	–79.3	5.7
–10.6	–77.0	5.7
–12.7	–90.0	5.7
–10.7	–77.8	5.7
–10.8	–81.2	5.7
–11.3	–80.4	5.6
–7.1	–61.5	5.5
–12.68	–91.4	5.4
–7.4	–62.6	5.4
–11.2	–79.4	5.3
–10.3	–84.4	5.2
–11.1	–76.1	5.2
–11.9	–86.0	5.1
–10.7	–75.8	5.1
–11.5	–79.1	5.1
–9.8	–77.3	5.1
–10.0	–75.6	5.0
–11.45	–81.6	5.0
–11.79	–84.0	5.0
–12.44	–89.5	5.0
–10.0	–75.6	4.8
–11.9	–86.8	4.8
–12.55	–90.9	4.8
–12.8	–87.8	4.8
–10.6	–81.6	4.7
–12.57	–90.1	4.7
–13.0	–94.9	4.68
–12.73	–92.7	4.6
–11.88	–84.4	4.6
–12.4	–93.2	4.5
–13.1	–96.1	4.5
–10.2	–81.3	4.5
–6.4	–62.9	4.4
–11.57	–83.8	4.4
–12.7	–94.4	4.3
–6.7	–57.9	4.3
–7.0	–72.4	4.2
–7.1	–65.3	4.2
–11.9	–81.3	4.2
–7.1	–72.3	4.1
–6.9	–58.5	4.1
–12.62	–90.4	4.1
–8.7	–71.5	4.1
–11.22	–79.3	4.0
–10.94	–77.8	4.0
–11.36	–80.3	4.0
–9.8	–77.5	3.9
–8.8	–71.9	3.9
–11.7	–86.7	3.8
–11.47	–82.7	3.7

δ¹⁸O	δ²H	Cl (mg/L)
–11.6	–78.1	3.6
–9.5	–75.2	3.6
–12.4	–92.2	3.53
–12.7	–95.3	3.5
–12.0	–89.7	3.5
–13.1	–99.4	3.3
–12.1	–86.2	3.2
–13.8	–104.5	3.1
–9.5	–77.7	3.1
–9.7	–67.9	3.1
–11.39	–81.2	3.0
–11.33	–80.3	3.0
–12.0	–85.9	3.0
–11.23	–80.2	3.0
–11.12	–81.5	3.0
–7.6	–67.4	3.0
–12.9	–92.2	2.4
–11.4	–86.2	2.3
–12.5	–90.0	2.3
–12.4	–94.0	2.24
–12.4	–94.9	2.22
–12.94	–94.1	2.2
–12.8	–94.1	2.2
–13.0	–95.0	2.1
–13.1	–94.3	2.1
–12.6	–90.6	2.1
–12.6	–90.4	2.05
–13.1	–97.9	2.0
–12.6	–96.9	1.93
–12.6	–96.4	1.77
–12.7	–97.9	1.77
–13.0	–94.9	1.7
–12.8	–91.3	1.68
–12.93	–93.4	1.5
–11.2	–78.5	1.5
–12.8	–97.2	1.5
–12.4	–96.0	1.48
–12.5	–96.0	1.48
–12.4	–92.7	1.46
–12.7	–96.5	1.4
–12.8	–92.4	1.38
–11.1	–84.4	1.3
–12.8	–93.9	1.21
–12.5	–94.6	1.2
–15.5	–116.9	1.2
–12.8	–95.8	1.14
–12.7	–93.3	1.09
–10.9	–80.6	1.0
–11.24	–80.1	1.0
–11.34	–80.7	1.0
–12.9	–96.4	0.88
–12.2	–95.9	0.86
–13.8	–109.3	0.8
–13.1	–96.6	0.77
–12.2	–93.7	0.76
–13.2	–102.1	0.75
–13.7	–100.7	0.75
–13.5	–102.8	0.74
–12.9	–96.5	0.74
–8.5	–64.8	0.7
–11.7	–88.6	0.7
–13.5	–102.2	0.65
–12.7	–95.5	0.59
–13.7	–104.9	0.58
–13.7	–106.2	0.54
–13.3	–96.2	0.54
–13.4	–105.4	0.53
–12.7	–94.8	0.47
–12.3	–95.9	0.43
–10.5	–80.0	0.23
–10.5	–80.0	0.23