Application of a mathematical model to determine the source of produced water in an oil field

Nguyen Minh Quy
Luong Van Huan
Le Thi Thu Huong
Vietnam Petroleum Institute

1. Introduction

The *XY*, an oil field in Southern offshore Vietnam, has produced oil from a basement reservoir since 2003. In order to maintain reservoir pressure, water injection has been started from Dec 2004. Water was first appeared in produced fluid from the well X-1 in May 2004. Water encountered in other wells started to increase in late 2005. Hundreds of water samples were taken and analyzed. Analytical results indicated that the chemical compositions of produced waters vary from well to well and even from time to time in some wells. For monitoring and optimizing production performance, determining the source of the produced water was required, and this was set as the main objective of this study.

A mathematical model, the so-called the Linear Mixing Model was developed, mainly based on the statistical assessment of variation of conservative chemical species in available produced water analytical results, to identify all possible sources and the contribution of each source to the produced water. The results of the model indicate that the produced water is a mixture of three sources: formation water, injected water and drilling fluid. Among these sources, formation water is the dominant component in almost produced water samples.

This paper presents the mathematical model which was successfully applied to determine the source of produced water in the XY oil field.

2. The linear mixing model

2.1. The Linear Mixing Approach

In many geochemical related observations, compositional variation among a series of specimens (e.g., rock, sediment or water samples) may be attributed to physical mixing or mathematically linear mixing. Datasets which conform to a linear mixing model can be expressed as mixtures of a fixed number of end members. The end members represent a series of fixed compositions (or compositional profiles), which can be regarded as distinct contribution sources to the geological body for which the datasets are being analyzed [1]. In our case, a water body is assumed to be supported from mixing *p* independent water sources, *m* water samples are taken and concentrations of *n* soluble chemical species are those of interest.

The fundamental principle of the linear mixing model is that mass conservation can be assumed and a mass balance analysis can be used to identify and apportion contribution sources. Mass balance equation can be written to account for all *n* soluble chemical species in the *m* samples as contributions from *p* independent water sources:

$$y_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} \quad (1)$$

Where y_{ij} is the j^{th} elemental concentration (mg/l or meq/l) measured in the i^{th} sample, g_{ik} is the contribution proportion of the k^{th} water source to the i^{th} sample, and f_{kj} is concentration (mg/l or meq/l) of the j^{th} soluble chemical constituent in water from the k^{th} source.

When all the measurements y_{ij}'s of *n* chemical species in *m* samples are populated in a *m-by-n* matrix *Y*, then equation (1) can be written in the matrix form as:

$$Y = G \times F \quad (2)$$

Where *G* is a *m-by-p* matrix of source proportions and *F* is a *p-by-n* matrix of source compositions (or source profiles).

In fact, measurements in matrix *Y*, of course, are likely to include some noise and/or analytic, as well as systematic errors. So equation (2) should additionally

include an error term *E* (a *m-by-n* matrix), then equation (2) can be rewritten as:

$$Y = G \times F + E \quad (3)$$

There exist a set of natural physical constraints on the solution that must be considered in developing any model for identifying and apportioning the sources of water contribution. The fundamental, natural physical constraints that must be obeyed are:

- The original data must be reproduced by the model; this means the error term *E* must be minimized and values in the matrix *E* would be distributed in certain and explainable patterns.

$$minimize \left\|E\right\|^2 \quad (4)$$

- All values in matrices *G* and *F* must be non-negative; a water source cannot have a negative concentration of chemical species or a water source cannot contribute negative proportions to any water sample.

$$G \geq 0 \text{ and } F \geq 0 \quad (5)$$

- When all possible water sources are taken into account, the sum of source proportion contributions to each water sample must be constant (e.g. equal to unit or a hundred percent).

$$sum(G) = 100\% \quad (6)$$

It is assumed that the concentrations of a series of chemical species have been measured for a set of samples from the water body so that the matrix *Y* is always known. If the number of sources *p* that contribute to those water samples can be identified and their compositional profiles measured, then only the contributions of the sources to each sample need to be determined. These calculations are generally made without much difficulty, using standard linear equation or more effective alternatives, such as non-negative least-square techniques [2].

There is situation in which the chemical composition of the water body is believed to have been produced by mixing from some water sources, but the number of water sources and their chemical composition are unknown. In this case, the objective of the linear mixing modeling is to determine the number of water sources *p*, the chemical profile of each water source and the proportion that each of the *p* sources contributes to each water sample. Recasting the chemical compositions of water samples into a linear mixing model in the absence of *a priori* knowledge about the water sources requires a solution of the bilinear (or explicit) mixing problem. The multivariate data analysis methods that are used to solve this problem are generally referred to as factor analysis.

2.2. Principal Component Analysis (PCA)

The conventional approach to solve the bilinear mixing problem is the most common form of factor analysis named Principal Components Analysis (PCA). This method is generally calculated using an eigenvector analysis of a correlation matrix.

The matrix *Y* can always be defined in terms of the *singular value decomposition*.

$$Y = U \times S \times V' \quad (7)$$

Characteristics of singular value decomposition are that: *U* and *V* matrix are orthogonal, and singular values *S* are always ordered so that those with the largest variation come first. When only the first *p* columns of the *U* and *V* matrices and the first *p* values of *S* are take into account, which are denoted as $\overline{U}$, $\overline{V}$ and $\overline{S}$ respectively, and an error terms *E* is added, then equation (7) will be:

$$Y = \overline{U} \times \overline{S} \times \overline{V} + E \quad (8)$$

Error matrix *E* represents the part of the data variance un-modeled by the linear mixing model with *p* factors. It can be shown [2] that the first term on the right side of equation (8) estimates *Y* in the least-squares sense that it gives the lowest possible value for $\left\|E\right\|^2$ when the data matrix *Y* is approximated by the linear mixing model with *p* factors.

Equation (8) is a mathematically feasible solution for the bilinear mixing problem which was addressed in equation (3). The problem can be solved, but it does not produce an *unique* solution. It is always possible to include a transformation into the equation:

$$Y = G \times T \times T^{-1} \times F \quad (9)$$

where *T* is one of the potential infinity of transformation matrices. This transformation is called a rotation and is generally included in order to produce factors that appear to be closer to physically real source profiles.

In fact, *G* and *F* are usually consisting of many negative values. However, the rotation matrix T cannot, in most cases, eliminate all negativity in *G* and *F*, and constant-sum constraints (6) is hardly satisfied in customary PCA.

2.3. Matrix Factorization with Non-Negativity and Constant-Sum Constraints

There are various approaches available to impose nonnegativity constraints in factor analysis. One of the alternatives for positive matrix factorization is Lee and Seung's Euclidean Update algorithm which is preferably called Non-Negativity Matrix Factorization (NNMF). This algorithm is preferred because it is rather clear, simple easily computable, but more important is of its guarantee of convergence, although it is somehow expensive in CPU time [3].

This algorithm minimize *Euclidean distance* $\| X - GF \|^2$ with respect to *G* and *F*, subject to the constraints $G, F \geq 0$.

- *G* and *F* are initialized to be two random non-negative matrices or two roughly-estimated matrices.

- *G* and *F* are continuously kept updating until $\| X - GF \|^2$ converges. The multiplicative update rules are as the following:

$$F_{a\mu} = F_{a\mu} \frac{(G^T X)_{a\mu}}{(G^T GF)_{a\mu}}, \; G_{ia} = G_{ia} \frac{(XF^T)_{ia}}{(G^T FF^T)_{ia}} \quad (10)$$

This means that each element of *F* is multiplied by corresponding element of matrix G^TX then divided by corresponding element of matrix G^TGF.

During the above updates, *G* will be updated column-wise while *F* will be updated row-wise, and *G* and *F* should be "simultaneously" updated. This means, after updating one row of *F*, the corresponding column of *G* needs to be updated subsequently; so actually we update *F* and *G* alternately.

The whole algorithm scheme of this *NNMF* model is given out in Fig. 1. Updating elements of *G* and *F* in each iteration is carried out in the inner loop, while calculating *Euclidean distance* $\| X - GF \|^2$ and checking criteria of its convergence is carried out in the outer loop.

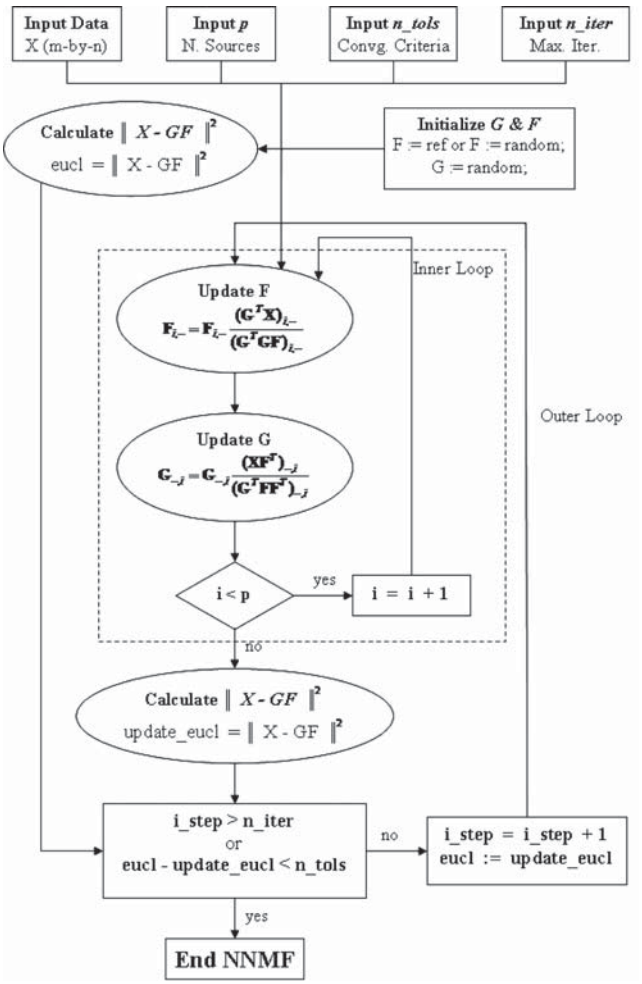

Fig. 1. *Algorithm Scheme of Lee and Seung's NNMF*

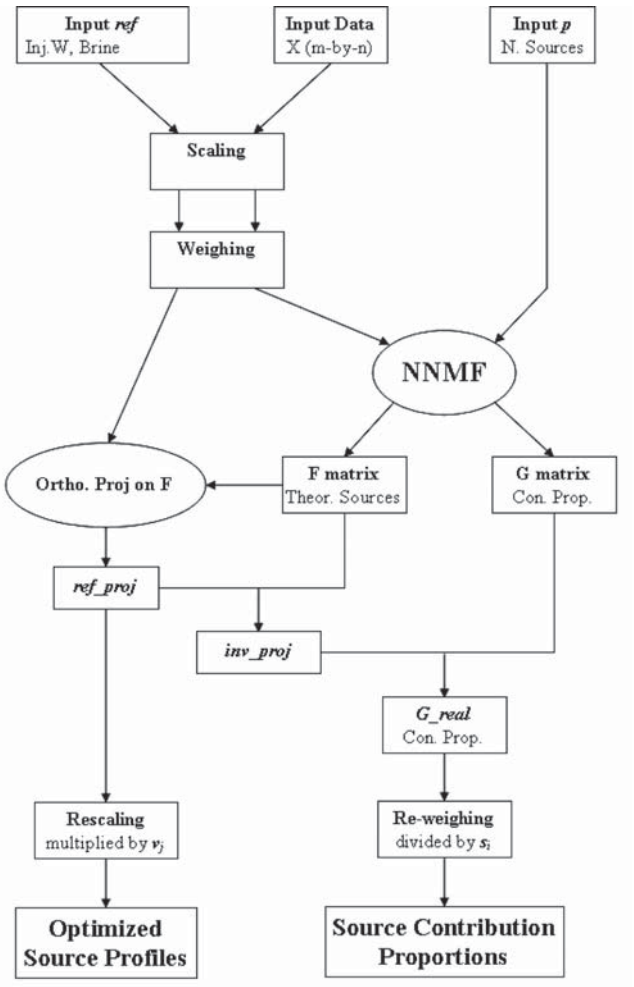

Fig. 2. *Outline of Source Unmixing Calculation*

3. Computations for produced water of XY field

3.1. Preparing Data Input

The water-rock physico-chemical interaction was conducted and the results showed that: there are 5 chemical components including bromide, chlorite, sulfate, sodium and total ion which are necessarily stable in the *XY* basement reservoir and are considered as conservative components or chemical "fingerprints" to clarify the contribution of each water source to produced water. Chemical data of produced waters are assembled into a matrix X, samples are arranged row-wise, and parameters are arranged column-wise. A total number of 177 produced water samples were taken in to account so data matrix will have 177 rows and 5 columns.

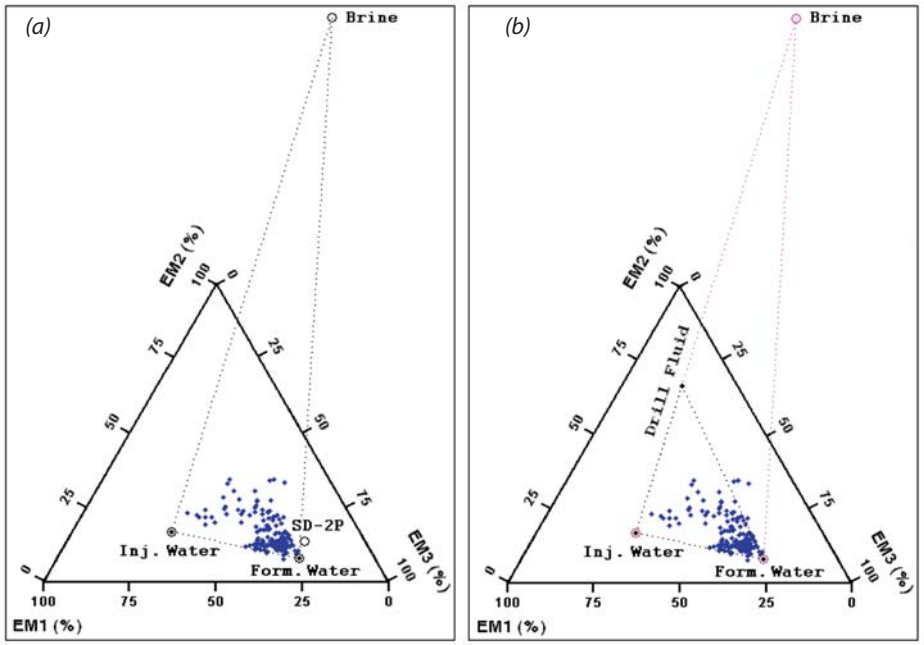

Fig. 3. *Expression of produced water as mixtures of mathematical EMs*

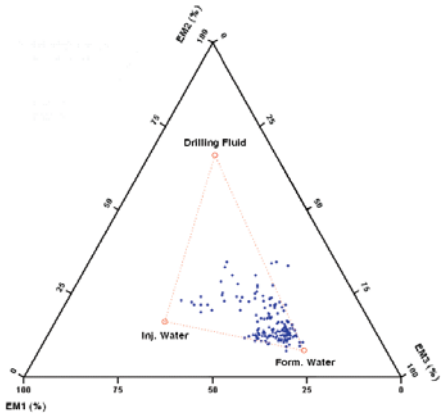

Fig. 4. *Positions of realistic end-members in space of mathematical EMs*

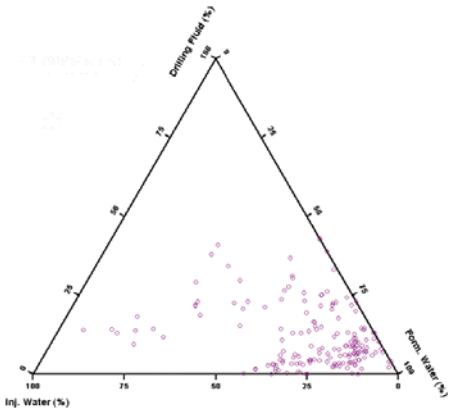

Fig. 5. *Expression of produced water as mixtures of water sources*

3.2. Computational Scheme

Input data, after eliminating extremely eliminating, scaling and/or weighting, are assembled in matrix *X* (177-by-5), including 177 produced water samples and 5 chemical parameters. This input matrix is trained in a computational process in which an outline of the computational scheme is given in Fig. 2.

3.3. Computational Output

In this study, the computation process was optimized with three water sources. The PMF computation produced three mathematical profiles (EM1-3), the expressions of all water samples, injected water, brine and formation water sample as mixtures of these 3 mathematical profiles are represented in Fig. 3b. The representations of produced water samples by these mathematical profiles show a clear acute angle at formation water. This clue indicates that all produced water samples are actually mixtures of 3 realistic water sources with unique chemical profiles.

Initially, it is believed that produced water is mixing from formation water, injected water and brine, but computational results show that no produced water sample is distributed in the large area spreading from the brine position (Fig. 3b). Moreover, there exists also a clear upper edge of the acute angle from the optimized position of formation water. This evidence allows the conclusion that produced water was mixed from an intermediate composition between brine and injected water (sea water) rather than directly from a pure brine composition. This intermediate composition, so-called *drilling fluid*, is positioned in the line from brine to injected water and its position, as shown in Fig. 3b, can be determined by

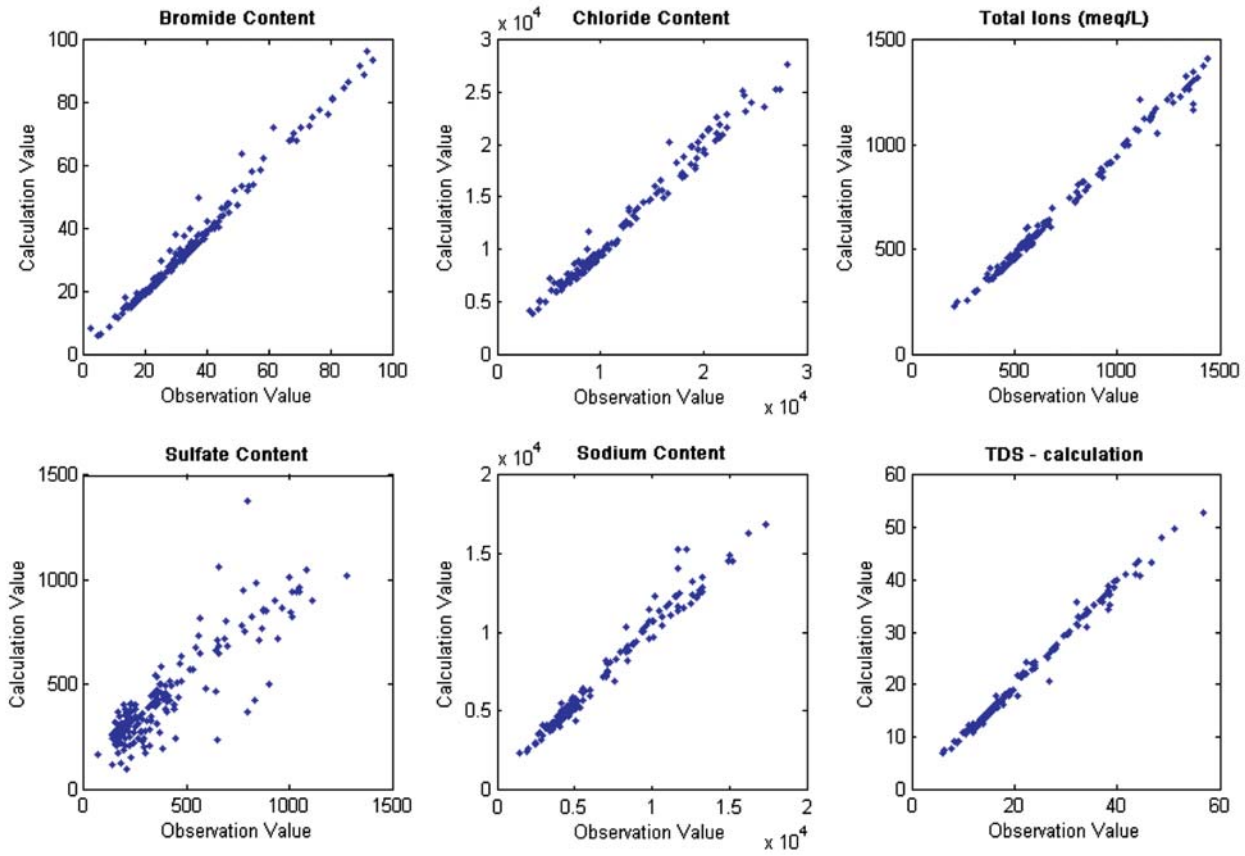

Fig. 6. *Calculation versus Observation of Chemical Components*

convexity optimization. The convexity optimization gives a proportion of 28.7% brine in drilling fluid. This value is agreeable with the proportion of about 30% brine in total mudlosses which include brine and seawater.

Finally, three realistic end-members which contribute to produced water are positioned in the mixing space of three mathematical end-members as shown in Fig. 4. It can be realized that all produced water samples and their natural trends, including acute angle and sharp edges, are enclosed well by three realistic end-members. A spatial base transformation or rotation to these realistic end-members will give the expressions of all produced water samples as mixtures of three realistic water sources as shown in Fig. 5.

In order to validate the model, an inverting model was performed. The recalculated values of chemical components of water samples obtained by the inverting model are in good agreement with the observation as shown in Fig. 6.

Conclusions

In summary, all computational results have definitely confirmed the appropriateness and accuracy of applying a linear mixing model to identify water sources and their contributions to produced water. The results of the model indicate that the produced water is a mixture of three sources: formation water, injected water and drilling fluid. Among these sources, formation water is the dominant component in almost all produced water samples.

The application of the mathematical models is the fundamental factor for the success of this study.

References

1. Weltje, G. J. *End-member modeling of compositional data: numerical-statistical algorithms for solving the explicit mixing problem*. Journal of Mathematical Geology. 1997; Vol. 29: p. 503 - 549.

2. Lawson, C.L. and Hanson, R.J. *Solving Least Squares Problems*. Prentice-Hall Press. 1974.

3. Lee, D.D. and Seung, H.S. *Algorithms for nonnegative matrix factorization, in Advances in Neural Information Processing 13*. MIT Press. 2001: p. 556 - 562.