EXTENDED QMODEL—Objective Definition of External End Members in the Analysis of Mixtures

William E. Full, Robert Ehrlich, and J. E. Klovan

Analysis of empirical data considered to be mixtures of a finite number of end members has been a topic of increasing interest recently. The algorithms EXTENDED CABFAC and QMODEL by Klovan and Miesch (1976) represent a satisfactory solution to this problem if pure end members are captured within the data set or if the composition of "true" end members are known a priori. Where neither condition is satisfied, the composition of "external" end members can, under certain conditions, be deduced from the structure of the data. Described herein is an algorithm termed EXTENDED QMODEL which defines feasible end members which are "closest" to the data envelope.

KEY WORDS: mixtures, end members, algorithm, Q-mode factor analysis.

INTRODUCTION

The growing trend toward quantification in the earth sciences has resulted in the generation of large data sets of compositional information. These data sets generally involve measurements of several compositional variables on a large number of geologic objects. One way of viewing such a data set, although by no means the only way, is to consider individual samples as mixtures of a relatively small number of end members. That is, by combining end member compositions in various proportions, the composition of individual objects may be reproduced. In many geological problems, the number of end members and their composition may be unknown and the problem becomes one of unmixing the given compositions into a minimum number of end members whose composition is mathematically and geologically "reasonable." Q-mode factor analysis is one analytical approach to this general class of problem.

In geology, Q-mode factor analysis has been used as a means to "unmix"
heavy mineral assemblages (Imbrie and Van Andel, 1964), grain-size distribution data (Klovan, 1966), geochemical and petrologic data (Miesch, 1976b), and for unmixing of detrital particles via shape analysis (Mazzullo and Ehrlich, 1980; Brown, Ehrlich, and Colquhoun, 1980; Hudson and Ehrlich, 1980). One accepted method of "unmixing" has been developed by Klovan and Miesch (Miesch, 1976a; Klovan and Miesch, 1976). In this method, reference axes derived from Q-mode factors are interpreted as end members and the data points are expressed as mixtures of these components. Klovan and Miesch's computer programs, EXTENDED CABFAC and QMODEL (Klovan and Miesch, 1976), allow, via a number of options, some latitude on how these component fractions, or end members, can be chosen. Several of these options "assume" that the true end members are contained within the data set, in which case the most divergent samples are used as end members. In geology many instances exist where true end members are not captured within a data set, and in other cases, possibly cannot exist for physical or chemical reasons. Miesch (1976b) foresaw this possibility and devised options which allow use of end member compositions not contained within the data set. Such external end members may be based on a strong a priori hypothesis or on an intuitive appreciation of the geological situation. On the other hand, the structure of the data itself may indicate something about the existence and composition of external end members. Where a priori external end members are known, Klovan and Miesch's (1976) solution involving manual input of these end members seems reasonable. A problem arises in the more general case of empirically determining external end members—especially where the data set is large and four or more end members may exist.

A method for the solution of this general problem is proposed herein. The implementation of the new solution has been designed as an extension of program QMODEL. Before discussing this new method, a brief synopsis of the end member concept is presented.

**DEFINITION OF ENDMEMBERS**

Let \( N \) be the number of data points and \( n \) the number of measurements comprising each data point. Furthermore, let \( X_i \) represent a typical data point consisting of \( x_{i1}, \ldots, x_{ij}, \ldots, x_{in} \), where \( x_{ij} \) represents the \( j \)th measurement on the \( i \)th sample which satisfy the following

\[
\begin{align*}
(1) & \quad x_{ij} \geq 0 \quad \text{for } j = 1, \ldots, n \\
(2) & \quad \sum_{j=1}^{n} x_{ij} = K, \quad \text{(constant for all } X_i) 
\end{align*}
\]

The set of data points \( X_i \) represents points in a variable space that, due to constraint (2) above, are in a space of dimensionality \( n \) but on a \( n - 1 \) dimensional