A quantitative procedure for chemostratigraphy

Richard A. Reyment
Naturhistoriska Riksmuseet, Stockholm
richard.reyment@pal.uu.se

ABSTRACT: A method for condensing a sequence of geochemical determinations into a set of vector-lengths for plotting is proposed. Data from the eastern Chinese non-marine Upper Cretaceous are used for exemplifying the procedure. The effect of reducing the dimensionality by deleting minor elements is illustrated.

INTRODUCTION

The idea of using frequencies of species occurrences for making stratigraphically oriented curves or profiles is not a new one in palaeontology. Pollen frequency curves are commonplace in Quaternary geology just as time-oriented profiles are often used in foraminiferal studies. Reyment (1980) summarized several of these concepts and also discussed statistical procedures for their analysis. Another type of data is often available for sedimentary sequences, namely, determinations of major and minor elements and it is this category that is considered in this brief note.

Finding the most suitable method for producing a geochemical profile can pose a problem. The approach that might seem to lie closest to hand, that of a multivariate analytical technique of some kind for condensing the dimensionality of the array, is not as simple as it may seem at first sight. The most obvious choice, principal component analysis, has attractive properties, but there are two hurdles to be cleared.

Firstly, the spectral decomposition of a principal component analysis leads to a set of several scores. It is often quite difficult to identify which of these sets is stratigraphically informative. Secondly, tables of geochemical determinations have a constant sum and hence lie in simplex space. This signifies that the application of standard principal component analysis is not advisable, but rather the appropriate version of the technique for compositional data should be used (Aitchison, 1986). However, this still does not solve the problem of indeterminancy with respect to constructing a profile. An alternative, and very simple approach, is suggested. I propose here to use vector lengths as a measure, computed for the log-ratio transformed array of geochemical determinations. This transformation places the observations in simplex space where the restriction imposed by the constant row-sum closure no longer applies.

The significance of simplex space

A restricted part of real space, the simplex, constitutes the basic concept for the study of compositional data. The significance of simplex space can be appreciated by considering the nature of the unit-sum constraint expressed by equation (1):

$$\mathrm{x} = x_1 + x_2 + \ldots + x_D = 1 \qquad (1)$$

where x is the vector of D proportions. The sum of the elements is always the same. Here it is 1 because we are thinking in terms of proportions. In the case of percentage frequences, the sum would be 100. These vector elements are not variables in the true statistical sense, but ingredients, or parts. If one of the parts is deleted, the constant-sum condition must be reinstated. This is the principle of subcompositional coherence. This is a necessary condition for without it, any form of multivariate computation lacks interpretability. This is because the parts constituting a compositional vector are bound to each other. If one or more parts are changed or removed the values of all the other parts in the rows of the array of frequencies will be affected and the rows of the array will not sum to the same value. No such restriction applies for full space variables, for example distances measured between diagnostic points on a fossil.

Aitchison (1986) showed that the ratio of two parts remains unchanged for full compositions in relation to subcompositions. Consequently subcompositions s_i and full compositions x_i maintain the fixed relationship

$$s_i / s_j = x_i / x_j$$

Although the vector x defined in (1) consists of D parts, the composition it represents is completely specified by the d components of a d-part subvector, where $d = D - 1$.

Hence, $x_D = 1 - x_1 - \ldots \; x_d$

The log-ratio used here for constructing the vectorial lengths is that defined by the common division of one of the parts

$$y_i = \log(x_i / x_D) \qquad (2)$$

where ($i = 1, \ldots, d$) .

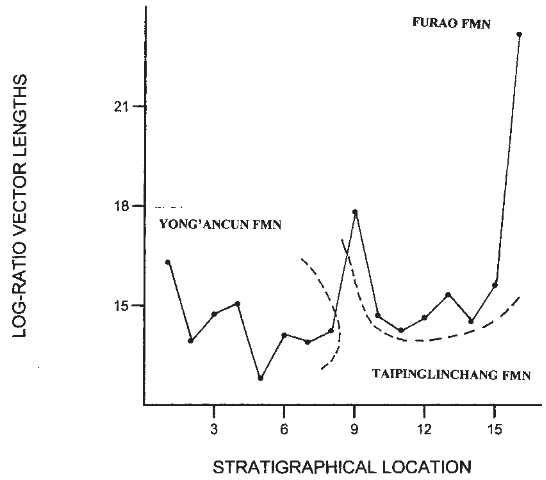
The common divisor x_D can be any one of the parts in the data array. One might wonder whether the choice of component divisor makes a difference to an analysis. Aitchison (1986, p. 78) reported that a statistical procedures involving parts of a composition asymmetrically are invariant under the group of permutations.

The vector length method

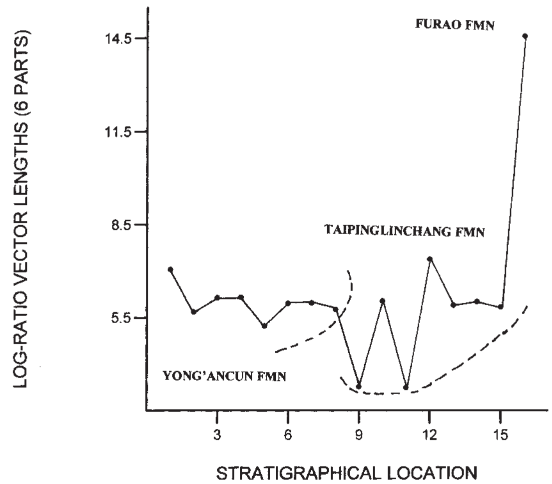
The length of the vector of log-ratios of parts is readily found as

$$\mathrm{v} = (\mathrm{y}_1^2 + \mathrm{y}_2^2 + \ldots + \mathrm{y}_d^2)^{1/2} \qquad (3)$$

The stratigraphical profile is constructed by plotting the lengths of the vectors against stratigraphical location. A geologically oriented reference for vector operations in geology is given in the book by Reyment and Jöreskog (1993).



TEXT-FIGURE 1
Plot of the 12-part vector lengths against stratigraphical location for the Heilongjang data.



TEXT-FIGURE 2
Plot of the 6 vector lengths (major elements) against stratigraphical location for the Heilongjiang data.

THE DATA

The data used for exemplifying the method for chemostratigraphy were taken from a recent publication by Hirano *et al.* (2003). These data are not particularly remarkable nor extensive although they are typical of what is found in chemostratigraphy. My choice was governed by my current interest in the non-marine Cretaceous of China. The array of observations consists of chemical analyses (major and minor elements) of a composite stratigraphical sequence in the non-marine Upper Cretaceous of the Songli Basin of far-eastern China. The sediments sampled lie in the Jiayin Group exposed along the Heilongjiang River. The sequence consists of the middle and upper 86 m of the Yong'ancun Formation and the lower 66 m of the Taipinglinchang Formation. A total of 15 mudstone samples from the two formations and one siltstone sample from the Furao Formation were analysed. The sampled section of the Yong'ancun is assigned tentatively to the Albian-Cenomanian. The section through the Taipinglinchang Formation is considered to be Turonian in age (Chen 1996). The age assigned to the Furao Formation is latest Cretaceous.

The sediments were analysed by X-ray fluorescence for 12 major and minor elements (Hirano et al. 2003, p. 18). The following elements were determined:
SiO_2, Fe2O_3, SO_3, Cr_2O_3, CaO. Al_2O_3, TiO_2, P_2O_5, MgO, K_2O, MnO_2, BaO.

The chemostratigraphical profile

SiO_2 was selected as component divisor for the application of equation (2). The 16 vector-lengths were computed by means of equation (3) and the resulting values plotted against stratigraphical location (text-fig. 1). The sequence of points discloses several interesting features. Firstly the isolated position occupied by the single sample from the Furao Formation., indicating the likelihood of a distinctly different palaeoecological background for that material. The published section for Furao Formation indicates that there is a high content of organic carbonaceous material in the sediments (Hirano et al. 2003, p. 19). All older points tend to oscillate around a fairly stable value with the exception of the first observation recorded for the Taipinglinchang Formation.

The points for the Yong'ancun Formation form a group distinguishable by having lower vectorial lengths on the average than the points deriving from the Taipinglinchang Formation. There does therefore on the basis of this admittedly restricted sample evidence for slightly different palaeoecological backgrounds for the three sequences considered. In any event, the profile established here should be of value for subsequent comparative studies in the area.

The effect of removing half of the parts was tested. For this analysis, only the oxides of the major elements Si, Fe, Ca, Al Mg and K were retained. Text-figure 2 reiterates the isolated placement of the sample from the Furao Formation. Levels 6 and 8 have new positions: Both of those levels contain calcareous nodules with substantial values for Al and Mg (dolomitic?). The points for the Yong'ancun Formation occupy approximately the same relative positions as in text-figure 1.

There is a difference for a part of the sequence which implies that valid comparisons between profiles must be based on the same set of identical parts.

REFERENCES

AITCHISON, J., 1986. *The statistical analysis of compositional data.* Chapman and Hall, London, xv + 416 pp.

CHEN, P. 1996. Freshwater biota, stratigraphic correlation of Late Cretaceous of china. *Memoirs Geological Society of India*, 37: 35-62.

HIRANO, H., SAJA, Y., CHEN, P., ZHAO, W., MITSUKI, T. 2003. Stratigraphy of the Upper cretaceous along the Heilongjiang River, China. *Series of Biology and earth sciences, school of Education, Waseda University*, 51: 1-14.

REYMENT, R. A. 1980. *Morphometrical methods in biostratigraphy.* Academic Press, London, 175 pp.

REYMENT, R. A. and JÖRESKOG, K. G. 1993. *Applied factor analysis in the natural sciences.* Cambridge University Press, xii+371 pp.

Manuscript received January 2, 2004
Manuscript accepted August 3, 2004