SOME COMMENTS ON COMPOSITIONAL DATA ANALYSIS IN ARCHAEOMETRY, IN PARTICULAR THE FALLACIES IN TANGRI AND WRIGHT'S DISMISSAL OF LOGRATIO ANALYSIS

J. AITCHISON[1], C. BARCELÓ-VIDAL[2] and V. PAWLOWSKY-GLAHN[2]

1 Department of Statistics, University of Glasgow, Glasgow G12 8QQ

2 Universitat de Girona, Escola Politècnica Superior, Dept. d'Informàtica i Matemàtica Aplicada, Avda. Lluis Santaló, s/n, 17071 Girona, Spain.

This comment exposes the fallacies in the Tangri and Wright (1993) dismissal of the methodology of logratio analysis of compositional data as dangerous surgery. It suggests that compositional data analysts should pay enough attention to the basic nature of compositional data and some elementary principles underlying coherent study in order to avoid meaningless inferences.

KEYWORDS: **COMPOSITIONAL BIPLOT, COMPOSITIONAL PRINCIPLES, PRINCIPAL COMPONENT ANALYSIS, SUBCOMPOSITIONAL COHERENCE**

INTRODUCTION

The purpose of this paper is to advocate the use of what has become known as logratio analysis as a meaningful, interpretable methodology for all problems involving compositional data, and to encourage archaeometricians to take account of underlying and necessary principles of compositional data analysis. As a step towards this and the rejection of the use of 'standard multivariate analysis' we place an emphasis on exposing the substantial fallacies in the Tangri and Wright (1993) attempt to dismiss this logratio methodology.

It is only recently that we have become aware of the interest of archaeologists in compositional data analysis, in particular their analysis of ceramic and glass artefacts. Like any statistical methodology, the analysis of this type of data is founded on solid theoretical developments naturally associated with the algebraic-geometric properties of the sample space. Those developments have led to the realisation that so-called standard multivariate analysis designed for unconstrained multivariate data is entirely inappropriate for the statistical analysis of compositional data. Geometrically this is not difficult to comprehend, since the sample space of compositions is a simplex, a generalisation of a triangle and tetrahedron, a radically different space from real Euclidean space, the space for representing unconstrained vector data. The differences, and therefore the need for different methodologies, has been expressed almost ad nauseam for over a century, for example by Pearson (1897), Chayes (1949,1960,1962, 1971), Sarmanov and Vistelius (1959), Krumbein (1962), Mosimann (1962, 1963), Chayes and Kruskal (1966), Aitchison (1981, 1982, 1983, 1986, 1992, 1994, 1997), Le Maitre (1982), Davis (1986), Pawlowsky (1984), Rock (1988), Woronow (1987), Woronow and Love (1990), Reyment and Savazzi (1999), and variously referred to as the spurious correlation problem, the constant-sum problem, the negative-bias problem, the null-correlation problem, the closure problem, in a variety of disciplines. In particular, these theoretical and practical studies all point to an inevitable truth about compositional data analysis: product-moment correlation of raw components is a meaningless descriptive and analytical tool in the study of compositional variability. What Tangri and Wright (1993) term standard PCA is based on such product moment correlations, and therefore suffers from these criticisms of inappropriate analysis. Attempts to solve this statistical problem up to 1980 were mainly pathological in nature, attempting to analyse what goes wrong when standard multivariate analysis is wrongly applied to compositional data, presumably in the hope that some corrective inference might be made as a result. A similar sequence of applied statistical events took place in the analysis of directional data, but fortunately was quickly resolved by taking account of the special algebraic-geometric nature of the spherical sample space.

OVERVIEW OF THE COMPOSITIONAL PROBLEM

What has come to be known as logratio analysis (Aitchison, 1982, 1983, 1986, 1997) was based on simple intuitive ideas, namely that compositions provide information on relative rather than absolute values of the components of compositions, that relative values are characterised by ratios and that logarithms of ratios are simpler to handle mathematically and interpret statistically than ratios. Since there is a one-to-one correspondence between a composition and a complete set of ratios or logratios, information is neither lost nor gained in the process of transformation. This intuition is strongly supported by a series of logical necessities which any compositional data methodology must satisfy, for example, scale invariance, subcompositional coherence, meaningful groups of operations of change such as perturbation and power, meaningful measures of distance between compositions; see Aitchison (1997, 2001) for a detailed account of these. Later we shall use the compositional principle of subcompositional coherence to illustrate the nature of these necessities, since it has a particular bearing on archaeological compositional analysis and the Tangri and Wright fallacies.

In the light of the preceding comments, let us examine the argument of Tangri and Wright. Put bluntly, it is wonderfully illogical. We have two methods **A** (standard PCA) and **B** (Aitchison's new method). **A** is known by us to be faulty (since we quote Chayes and others who have pointed this out clearly). What about **B**? Let us distort some data sets and see which method seems to change least, according to *our criteria of success*. We find that **A** distorts least, according to *our criteria*. Therefore, despite the fact that we know it to be faulty, **A** must be better, and our analysis of distorted data allows us to state categorically that (Tangri and Wright, 1993, p.104)
'Aitchison's method is dangerous surgery, whatever the demerits of standard PCA.'

The Tangri and Wright argument depends on three assumptions:

(a) In comparing methodologies we can ignore theoretical considerations.

(b) We can judge different methodologies objectively by considering what happens to distorted data sets.

(c) Our criteria of comparison are sensible in the compositional context.

Baxter (1993) has already commented on the Tangri-Wright approach and pointed out a number of doubtful arguments, even fallacies, in their comparison. What appears to us to be missing, however, from most of the archaeometric papers we have seen, is a clear understanding of the basic nature of compositional problems and the logical necessities required by any methodology which purports to be appropriate for the study and interpretation of compositional variability.

A SIMPLE PRINCIPLE OF COMPOSITIONAL DATA ANALYSIS: SUBCOMPOSITIONAL COHERENCE

By assumption (a) Tangri and Wright seem to dismiss theoretical considerations almost as irrelevant. This is in line with the history of compositional data analysis, where the wishful thinking that there is little different about compositional data has led to a century of suspect analysis. Clear thinking about the nature of a compositional problem leads to certain logical necessities, which a meaningful analysis must take into account.

It is clear from the major oxide and trace elements in the data sets BAXTER1 (with parts MgO, Al_2O_3, P_2O_5, K_2O, CaO, MnO, FeO, Cu, Zn, Pb) and BAXTER2 (with parts SiO_2, Al_2O_3, FeO, MgO, CaO, Na_2O, K_2O, TiO_2, P_2O_5, MnO, Sb) that different parts of a composition may be regarded as important for different analytical purposes and by different investigators. In particular the suppression of the ubiquitous SiO_2 from BAXTER1 but not from BAXTER2 indicates that there are occasions when two archaeologists may, for the same data set, be concerned with different parts. This raises the important principle of *subcompositional coherence* in compositional data analysis, most simply explained in terms of a concrete example.

Formally the subcomposition based on parts (1, 2, . . ., C) of a D-part composition $(x_1, ..., x_D)$ is the (1, 2, . . . ,C)-subcomposition $(s_1, . . . , s_C)$ defined by

$$(s_1, ..., s_C) = (x_1, ..., x_C) / (x_1 + ... + x_C). \qquad (1)$$

Now consider two scientists A and B interested in soil samples, which have been divided into aliquots. For each aliquot A records a 4-part composition (animal, vegetable, mineral, water); scientist B first dries each aliquot without recording the water content and arrives at a 3-part composition (animal, vegetable, mineral). Let us further assume for simplicity the ideal situation where the aliquots in each pair are identical and where the two scientists are accurate in their determinations. Then clearly B's 3-part composition (s_1, s_2, s_3) for an aliquot will be a subcomposition of A's 4-part composition (x_1, x_2 , x_3, x_4) for the corresponding aliquot related as in (1) above with $C = 3$, D = 4. It is then obvious that any compositional statements that A and B make about the common parts – animal, vegetable and mineral – must agree. This is the nature of subcompositional coherence.

The ignoring of this principle of subcompositional coherence has been a source of great confusion in compositional data analysis. The literature, even currently, is full of attempts to explain the dependence of components of compositions in terms of product moment correlation of raw components. Consider the simple data set:

Full compositions (x_1, x_2 , x_3, x_4)	Subcompositions (s_1, s_2 , s_3)
(0.1, 0.2, 0.1, 0.6)	(0.25, 0.50, 0.25)
(0.2, 0.1, 0.1, 0.6)	(0.50, 0.25, 0.25)
(0.3, 0.3, 0.2, 0.2)	(0.375, 0.375, 0.25)

Scientist A would report the correlation between animal and vegetable as corr(x_1, x_2) = 0.5, whereas B would report corr(s_1 , s_2) = -1. There is thus incoherence of the product-moment correlation between raw components as a measure of dependence.

Note, however, that the ratio of two components remains unchanged when we move from full composition to subcomposition: $s_i / s_j = x_i / x_j$. Thus as long as we work with scale invariant functions, or equivalently express all our statements about compositions in terms of ratios, we shall be subcompositionally coherent.

In the choice of their distortion technique Tangri and Wright are indeed hoist with their own petard, since they simply place the original compositions in the role of subcompositions of their extended compositions. Their so-called random addition of three parts to the original composition has the effect of distorting only the correlations in the crude PCA analysis. What they are doing therefore is subcompositionally incoherent. The subcompositional coherence of logratio analysis ensures that the submatrix associated with the parts of the original composition of the logratio covariance matrix of the extended composition remains unaltered. These features can be illustrated by a simple example.

Suppose that the original compositional data set is formed by eight samples of 4-part compositions:

Sample	Part 1	Part 2	Part 3	Part 4
1	0.0716	0.2499	0.3702	0.3083
2	0.0045	0.0644	0.7744	0.1567
3	0.1951	0.1861	0.2914	0.3274
4	0.2093	0.4571	0.0667	0.2669
5	0.0092	0.0282	0.8306	0.1320
6	0.2046	0.7376	0.0010	0.0568
7	0.6185	0.1605	0.0488	0.1722
8	0.1395	0.2305	0.3220	0.3080

According to the same partial randomisation procedure proposed by Tangri and Wright (1993), suppose that the selected ‘random parts’ to be added to the original data set are parts 2 and 3. Also suppose that the randomisation of the order of the values of these two added parts is given by

Randomized part 2: 0.1605 [7]; 0.4571 [4]; 0.2305 [8]; 0.0282 [5]; 0.0644 [2]; 0.7376 [6]; 0.1861 [3]; and 0.2499 [1];

Randomized part 3: 0.0010 [6]; 0.7744 [2]; 0.8306 [5]; 0.3702 [1]; 0.0667 [4]; 0.3220 [8]; 0.2914 [3]; and 0.0488 [7],

where the brackets refer to the position in the original data sample set. These randomised parts constitute the new parts 5 and 6 in the extended data set. Thus the Tangri-Wright extended composition is the closure of the data set

Sample	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6
1	0.0716	0.2499	0.3702	0.3083	0.1605	0.0010
2	0.0045	0.0644	0.7744	0.1567	0.4571	0.7744
3	0.1951	0.1861	0.2914	0.3274	0.2305	0.8306
4	0.2093	0.4571	0.0667	0.2669	0.0282	0.3702
5	0.0092	0.0282	0.8306	0.1320	0.0644	0.0667
6	0.2046	0.7376	0.0010	0.0568	0.7376	0.3220
7	0.6185	0.1605	0.0488	0.1722	0.1861	0.2914
8	0.1395	0.2305	0.3220	0.3080	0.2499	0.0488

which is given by

Sample	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6
1	0.0616	0.2152	0.3187	0.2654	0.1382	0.0009
2	0.0020	0.0289	0.3470	0.0702	0.2048	0.3470
3	0.0947	0.0903	0.1414	0.1588	0.1118	0.4030
4	0.1497	0.3269	0.0477	0.1909	0.0202	0.2647
5	0.0081	0.0249	0.7343	0.1167	0.0569	0.0590
6	0.0993	0.3581	0.0005	0.0276	0.3581	0.1563
7	0.4186	0.1086	0.0330	0.1165	0.1260	0.1972
8	0.1074	0.1775	0.2479	0.2372	0.1924	0.0376

The product-moment correlation matrix of the original compositional data set is

	$corr(x_i, x_j)$	j = 1	2	3	4
i	1	1.0000	0.1545	-0.7067	-0.0653
	2	0.1545	1.0000	-0.7408	-0.2623
	3	-0.7067	-0.7408	1.0000	-0.0830
	4	-0.0653	-0.2623	-0.0830	1.0000

whereas the product-moment correlation matrix of the extended data set is j

corr(x_i, x_j)	1	2	3	4	5	6
1	1.0000	0.1243	-0.5970	-0.0161	-0.1070	0.0635
2	0.1241	1.0000	-0.6422	0.0916	0.3119	-0.1908
3	-0.5970	-0.6422	1.0000	0.0887	-0.3228	-0.3732
i 4	-0.0161	0.0916	0.0887	1.0000	-0.4940	-0.4118
5	-0.1070	0.3119	-0.3228	-0.4940	1.0000	-0.0779
6	0.0635	-0.1908	-0.3732	-0.4118	-0.0779	1.0000

with the (1,2,3,4) submatrix substantially different from the original compositional product moment correlations, illustrating the subcompositional incoherence involved in the Tangri-Wright application of what they term standard PCA analysis. In contrast, since ratios and a fortiori logratios are preserved in the formation of subcompositions, the logratio variation matrix of the extended compositions – consisting of the variances of each possible logratio and summarising the logratio covariance structure – is j

$var(\log(x_i/x_j))$	1	2	3	4	5	6
1	0.0000	1.3459	12.1220	2.9296	4.0652	7.3446
2	1.3459	0.0000	9.4185	1.4944	1.7932	6.0357
3	12.1220	9.4185	0.0000	3.5120	7.5541	12.2099
i 4	2.9296	1.4944	3.5120	0.0000	1.9797	6.0410
5	4.0652	1.7932	7.5541	1.9797	0.0000	5.2862
6	7.3446	6.0357	12.2099	6.0410	5.2862	0.0000

with the (1,2,3,4) submatrix identical to the logratio variation matrix of the original compositional data set.

To sum up, the Tangri-Wright investigation of standard PCA and Aitchison's logratio method, in so far as inferences of the original compositions (contained as subcompositions of the extended compositions) are concerned, certainly distorts standard PCA analysis, whereas, since subcompositional ratios and extended compositional ratios are the same, any sensible logratio subcompositional analysis will be unaffected.

THE NATURE OF THE DISTORTED COMPOSITIONAL DATA SETS

In regard to assumption (b) we have already pointed out the fallacy of Tangri and Wright's general argument that comparing the performance of methodologies on distorted data sets on unsubstantiated criteria can reinstate a discredited methodology if compositional principles such as discussed above are ignored. Advances in science usually proceed by a repetitive cycle of observation suggesting theory or hypothesis, which in turn is tested by further observation. We know of no substantial advance that has emerged from creating or distorting data. Certainly there is some merit in simulation of data to illustrate some feature but there must be correct and solid reasoning behind the conclusion which the data analysis illustrates. Because of the relationship (1) the subcomposition (corresponding to the original parts) of the distorted data set will *always* have the invariant ratio and logratio properties illustrated by the example.

THE FALLACIES IN THE TANGRI AND WRIGHT CRITERIA

Before considering assumption (c) and discussing the irrelevance of the criteria of success adopted by Tangri and Wright it is necessary to summarise the effects of distorting the compositional data set of interest on both crude analysis and logratio analysis. First we have seen that a closure operation is required to arrive at the extended composition and that this, as has been known for the last fifty years (Chayes, 1949), alters the product-moment correlations between the original parts. Tangri and Wright claim that this effect is not serious, but from our simple example above we see that it certainly can be sizeable. In contrast, because of the subcompositional coherence property of the logratio method, the logratio covariance structure of the original composition remains unchanged within the relevant subcomposition in the extended data set. But there is another unjustified statement in the Tangri and Wright argument. As Baxter (1993) has already pointed out, they are mistaken in their statement that lack of correlation of a 'variable' with principal components is characterised by coincidence with the centroid: this is simply not true. Principal component plots and the relative variation biplot (Aitchison, 1990) – a useful representation of the complete data set ignored by Tangri and Wright – differ only in their scaling of the axes, but the relative variation diagram provides a more reliable picture of the covariance structure and the relation of the compositions to parts of the composition; see also Aitchison and Greenacre (2001), which provides substantial ideas of how to interpret biplots of compositional data. In biplots correlations between 'variables' are associated with angles between 'rays', not lengths of rays. Lack of correlation between 'variables' in such principal component and relative variation diagrams is associated with orthogonality of rays from the centroid to variable apexes, not closeness to the centroid. Indeed, the lengths of the rays to the new variables, and hence their distance from the centroid, are likely to be very similar to rays associated with the original composition since they have been selected simply as a random ordering of original components.

In short, principal component diagrams and biplots of the distorted compositional data set will attempt to capture an overall picture of the distorted data set. Since in this process the additional parts may well demand as much attention as the original parts and so fail to capture an authentic picture of the original composition, there seems to be no validity in comparing changes between original and distorted as a means of choosing between methodologies.

DISCUSSION AND CONCLUSIONS

Our purpose in writing this short note has been to refute the Tangri-Wright claim that logratio analysis is dangerous surgery, with an implication that archaeometrists may as well continue with the old standard PCA techniques, which are known to be compositionally unsound. In their paper they criticise Aitchison's approach to compositional principal component analysis as emphasising theoretical aspects rather than facing empirical tests. Of course, it is initially theoretical principles of analysis that are important if fundamental errors such as appear in the Tangri-Wright approach are to be avoided; they also seem to overlook the substantial set of post-monograph problems and data analysed with interpretation in Aitchison (1990). There the biplot technique of Gabriel (1971) is developed and in our view gives a much more satisfactory view of compositional variability than principal component diagrams. See also Aitchison and Greenacre (2001) for a more comprehensive set of interpretative tools of such diagrams in the compositional context.

In all statistical inference recognition of the underlying sample space is a first requirement. In compositional data analysis the special nature of the simplex sample space and its algebraic-geometric structure has to be taken into account in the development of any sensible methodology. This can be done – see, for example, Aitchison (2001), Aitchison et al. (2000), Barceló-Vidal et al. (2001), and Pawlowsky-Glahn and Egozcue (2001a, 2001b) - recognising the basic operations of perturbation and powering and the associated metric. The metric is important, as the methodology has to acknowledge the implications of considering whether proportions like 0.1 and 0.2 are as different as 0.5 and 0.6 or not. Standard PCA considers them to be equally distinct, whereas logratio analysis considers 0.2 to be two times 0.1 and 0.6 to be 1.2 times 0.5, recognising that compositional problems are concerned with relative magnitudes. Recognition of the metric vector space structure of the simplex assures the overall coherence of the logratio approach to statistical analysis of compositional data. It is encouraging to see in Buxeda (1999) the first recognition of perturbation as a description of how archaeological compositions change, and we look forward to seeing developments of this approach.

Resistance to the use of logratio analysis takes many forms; see Aitchison (1997) for a detailed account of these. Here we may first pinpoint the leaving-some-out argument, which claims that if the total proportions in the composition sum to less than 100 per cent there is surely no problem. In such a situation the problem still remains compositional and essentially what the investigator is concerned with is subcompositions of some not fully determined compositions. All the principles of compositional data analysis still apply. In an extreme case of this, where only trace elements are involved, the problem is still compositional so that again all principles should be adhered to. The good news for past analyses here is that the common practice of working with logs of the trace elements can be fully justified. Suppose that $x_1, \ldots, x_D$ are trace elements (ppm). These are essentially components of a full composition containing the major oxides, whose amalgamation X say in parts per million will be almost 1 as a ppm ratio. Thus forming logratios, with X as the divisor leads to a logratio vector $(\log (x_1/X), \ldots, \log (x_D /X))$, which because of the fact that X is approximately 1, gives a logratio vector of $(\log x_1, \ldots, \log x_D)$.

A number of investigators choose to work with ratios. This is sound on the basis of scale invariance and subcompositional coherence but does not fully exploit the mathematical and statistical advantages of going further to logratios. For example, there is a substantial difficulty in that the ratio variance-covariance structure is so complex with, for example, no simple, exact relationship between $\mathrm{var}(x_i /x_j)$ and $\mathrm{var}(x_j / x_i)$. The use of logratios simplifies all this with, for example, $\mathrm{var}(\log (x_i /x_j)) = \mathrm{var}(\log(x_j/ x_i))$ and with other operations such as perturbations also much simpler, A further argument for the dismissal of logratio analysis seems to be that numerical results from the use of 'standard unconstrained multivariate analysis (SUMA)' such as 'standard PCA' may turn out to be similar to those from logratio analysis. Such serendipity is rather akin to insistence on a normal distribution assumption and the use of (mean – 2 x standard deviation, mean + 2 x standard deviation) as a 95 per cent prediction interval. On many occasions this will be satisfactory, but if, for example the distribution is skew, disaster may strike with such ridiculous results as suggesting that a urinary excretion rate of a steroid metabolite may be negative with a non-negligible probability. Similarly application of SUMA to compositional data can produce equal disaster; see, for example, Aitchison (1986, 1997, 1999). Furthermore, interpretation of results based on absolute difference between observations rather then relative difference can be quite different. Loosing two pounds of weight might be a reason for celebration for an adult, a reason of concern for a child, and a matter of life and death for a new born. Certainly, in these cases common sense will avoid disasters and will lead to 'reasonable' limits between the different situations, but common sense is not a quantitative measure and a lot of experience is required to gain it. Why not stand on the safe side and use relative differences right from the beginning? Since logratio analysis is no more difficult to compute or interpret than SUMA there seems every reason for adopting this compositionally valid form of statistical analysis.

Another range of compositional problems is where interest is in convex linear mixtures of so-called endmember D-part compositions, say $e_1, \ldots, e_C$, and where a typical mixture x is formed as $\lambda_1 e_1 + \ldots + \lambda_C e_C$, where $\lambda_1, \ldots, \lambda_C$ are the mixing proportions. Although the mixing operation here in its assumption of conservation of matter does not involve logratios, compositional principles still apply to the observed compositional data, namely the observed mixtures, in the search for hypothesized endmembers.

Finally we point out that all that the logratio technique is achieving is to allow the analyst to avoid the unfamiliarity of the simplex and move out to the more familiar Euclidean space to perform and interpret the statistical analysis. Whether we stay in the simplex or logratio to another space does not matter. The inferences will be identical.

REFERENCES

Aitchison, J., 1981, A new approach to null correlations of proportions, *Math. Geology,* **13** (5), 175-89.

Aitchison, J., 1982, The statistical analysis of compositional data (with discussion), *J. R. Statist. Soc*., B, **44** (2), 139-77.

Aitchison, J., 1983, Principal component analysis of compositional data, *Biometrik*a, **70** (1), 57-65.

Aitchison, J., 1986, *The statistical analysis of compositional data*, Chapman and Hall, London.

Aitchison, J., 1990, Relative variation diagrams for describing patterns of variability of compositional data, *Math. Geology*, **22** (4), 487-512.

Aitchison, J., 1992, On criteria for measures of compositional differences, *Math. Geology*, **24** (4), 365-80.

Aitchison, J., 1994, Principles of compositional data analysis, in *Multivariate Analysis and its Applications* (eds. T.W.Anderson, I. Olkin, and K.T.Fang), 73-81, California, Institute of Mathematical Statistics, Hayward, USA.

Aitchison, J., 1997, The one-hour course in compositional data analysis or compositional data analysis is easy, in *Proceedings of the Third Annual Conference of the International Association for Mathematical Geology* (ed. V. Pawlowsky-Glahn), 3-35, CIMNE, Barcelona.

Aitchison, J., 1999, Logratios and natural laws in compositional data analysis, Math. Geology, **31** (5), 563-89.

Aitchison, J., 2001, Simplicial inference, in *Algebraic Methods in Statistics* (eds. M. Viana and D. Richards), Contemporary Series of the American Mathematical Society: in press.

Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J.A., and Pawlowsky-Glahn, V., 2000, Logratio analysis and compositional distance, *Math. Geology*, **32** (3), 271-5.

Aitchison, J., and Greenacre, M., 2001, Biplots for compositional data, *Appl. Statist.*, submitted.

Barceló-Vidal, C., Martín-Fernández, J. A., and Pawlowsky-Glahn, V., 2001, Mathematical Foundations of Compositional Data Analysis, *Proceedings of the Annual Conference of the International Association for Mathematical Geology*, 20 p, Cancún, Mexico.

Baxter, M., 1993, Comment on D. Tangri and R.V.S. Wright, "Multivariate analysis of compositional data …", Archaeometry *,* 35 (1) (1993), *Archaeometry*, **35** (1), 113-5.

Buxeda, J., 1999, Alteration and contamination of archaelogical ceramics: the perturbation problem, *Jour. of Archeological Science,* **26**, 295-313.

Chayes, F., 1949, On ratio correlation in petrography, *Jour. Geology*, **57** (3), 239-54.

Chayes, F., 1960, On correlation between variables of constant sum, *Jour. Geophys.*

Research, **65** (12), 4185-93.

Chayes, F., 1962, Numerical correlation and petrographic variation, *Jour. Geology*, **70** (4), 440-552.

Chayes, F., 1971, *Ratio correlation: a manual for students of petrology and geochemistry*, University of Chicago Press, Chicago.

Chayes, F., and Kruskal, W., 1966, An approximate statistical test for correlation between proportions, *Math. Geology*, **74** (5), 692-702.

Davis, J. C., 1986, *Statistics and data analysis in geology*, Wiley & Sons, New York, 2n ed.

Gabriel, K.R., 1971, The biplot display of matrices with applications to principal component analysis, *Biometrika*, **58**, 455-67.

Krumbein, C., 1962, Open and closed number systems: stratigraphic mapping, *Bull. Amer. Assoc. Petrol. Geologists*, **46**, 322-37.

Le Maitre, R. W., 1982, *Numerical petrography*, Elsevier, Amsterdam.

Mosimann, J. E., 1962, On the compound multinomial distribution, the multivariate β-distribution and correlations among proportions, *Biometrika*, **49** (1), 63-82.

Mosimann, J. E., 1963, On the compound negative binomial distribution and correlations among inversely sampled pollen counts, *Biometrika*, **50** (1), 47-54.

Pawlowsky, V., 1984, On spurious spacial covariance between variables of constant sum, *Science de la Terre*, Ser. Inf. **21**, 107-13.

Pawlowsky-Glahn, V., and Egozcue, J. J., 2001a, About BLU estimators and compositional data, *Math. Geology*: in press.

Pawlowsky-Glahn, V., and Egozcue, J. J., 2001b, Geometric approach to statistical analysis on the simplex*, Stochastic Environmental Research and Risk Assessment*: in press.

Pearson, K., 1897, Mathematical contributions to the theory of evolution: on a form of spurious correlation which may arise when indices are used in the measurements of organs, *Proc. R. Soc*., **60**, 489-98.

Reyment, R. A, and Savazzi, S., 1999, *Aspects of Multivariate Statistical Analysis in Geology*, Elsevier, Amsterdam.

Rock, N. M .S., 1988, *Numerical petrology*, Springer-Verlag, Berlin.

Sarmanov, O. V., and Vistelius, A. B., 1959, On the correlation of percentage values, *Dokl. Akad. Nauk. SSSR*, **126,** 22-5.

Tangri, D., and Wright, R.V.S., 1993, Multivariate analysis of compositional data: applied comparisons favour standard principal components analysis over Aitchison's loglinear contrast method, *Archaeometry*, **35** (1), 103-12.

Woronow, A., 1987, A Book Review: The Statistical Analysis of Compositional Data, by John Aitchison, *Math. Geology*, **19** (5), 579-81.

Woronow, A., and Love, K. M., 1990, Quantifying and testing differences among means of compositional data suites, *Math. Geology*, **22** (7), 837-52.