Mathematical Geology, Vol. 38, No. 2, February 2006 (© 2006)
DOI: 10.1007/s11004-006-9026-z

Short Note

The Graph Median–A Stable Alternative Measure of Central Tendency for Compositional Data Sets[1]

W. E. Sharp[2]

KEY WORDS: geometric mean, half-taxi, log-ratio, ternary diagram.

INTRODUCTION

In a letter to the editor some 16 years ago, Aitchison noted that the arithmetic average for compositions such as illustrated by the hongite subcomposition ABC (Fig. 1) is "clearly useless as a measure of location because it falls outside of the array of compositions and is indeed very atypical of the data set" (Aitchison, 1989, p. 789). He then states that the geometric mean "serves equally well for curved data sets and for more linear and elliptical data sets" (Aitchison, 1989, p. 789). This property of the geometric mean went unchallenged until 11 years later when, in another letter to the editor, Shurtz (2000) pointed out that the position of the geometric mean is not conserved and varies greatly with small changes in marginal values. To this reader, Aitchison had made an important point in that there may be instances when one wants a measure of location which consistantly belongs to the data set. However, the results of Shurtz suggested that the geometric mean might not always belong to the data set. The purpose of this note was to check on the geometric mean as a consistent measure of location and should it fail, to then propose a suitable alternative.

VARIATIONS ON THE HONGITE ABC SUBCOMPOSITIONS

To see if the geometric mean would remain with the data set and be a consistent measure of location, a series of variations on the ABC hongite subcomposition

[1]Received 22 May 2005; accepted 29 August 2005; Published online: 16 May 2006.
[2]Department of Geological Sciences, University of South Caolina, Columbia, South Carolina 29208, USA; e-mail: sharp@sc.edu

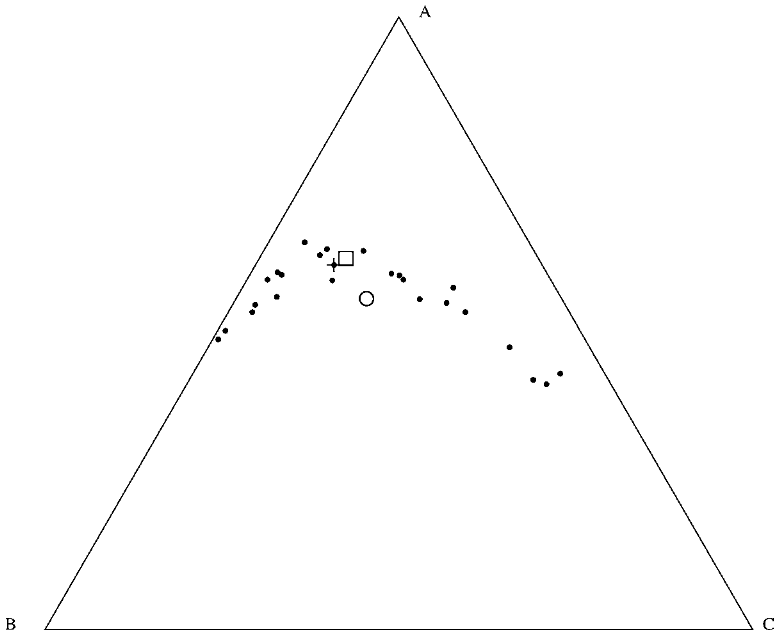

Figure 1. Ternary plot of the original Hongite ABC subcomposition. Mean (◯), geometric mean (□), graph median (+).

(Aitchison, 1986, p. 9) were constructed. For all variations, the arithmetic average of the subcompositional data set was held constant. For hongite 2 (Table 1), the data for hongite (Table 1) subcomposition were projected onto a line parallel to the end members BC and passing through the average (Fig. 2). For hongite 3 (Table 1), the data was reflected across a line parallel to the line BC passing through the average composition (Fig. 3). For hongite 4 (Table 1), the variance of the data about a line perpendicular to BC end members and passing through the average was shrunk by a factor of three (Fig. 4) while for hongite 5 (Table 1), the variance of the data about this same line was expanded by a factor of two (Fig. 5).

STABILITY OF GEOMETRIC MEAN

If (x_a, x_b, x_c) represent the compositions in the ternary diagram A, B, C, then the logratio values will be given by $y_a = \log(x_a/x_c)$ and $y_b = \log(x_b/x_c)$ and their mean by $\langle y_a \rangle$ and $\langle y_b \rangle$. The geometric mean location can then be inverted from: $(\langle x_a \rangle, \langle x_b \rangle, \langle x_c \rangle) = (\exp\langle y_a \rangle, \exp\langle y_b \rangle, 1)/(\exp\langle y_a \rangle + \exp\langle y_b \rangle + 1)$ (Aitchison, 1989, p. 788).

Table 1. Variations on the ABC Components Subcomposition of Hongite

	Hongite Subcomposition ABC[a]			Hongite 2[b]			Hongite 3[c]			Hongite 4[d]			Hongite 5[e]		
Sample	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
H1	57.9	37.6	4.5	54.00	40.44	5.56	50.09	43.15	6.76	57.90	30.01	12.09	57.90	41.47	0.63
H2	59.5	29.4	11.1	54.00	32.09	13.92	48.49	34.48	17.03	59.50	25.46	15.04	59.50	35.19	5.31
H3	46.1	11.3	42.6	54.00	8.63	37.37	61.89	6.42	31.69	46.10	23.73	30.17	46.10	3.36	50.54
H4	62.1	29.1	8.8	54.00	34.14	11.86	45.89	38.77	15.34	62.10	24.47	13.43	62.10	34.55	3.35
H5	51.8	44.8	3.4	54.00	42.31	3.69	56.19	39.85	3.96	51.80	35.80	12.40	51.80	47.91	0.29
H6	63.2	31.7	5.1	54.00	38.83	7.18	44.79	45.61	9.60	63.20	25.47	11.33	63.20	35.82	0.98
H7	54.3	40.1	5.6	54.00	39.64	6.36	53.69	39.15	7.16	54.30	32.09	13.61	54.30	44.78	0.92
H8	41.8	6.3	51.9	54.00	4.40	41.61	66.19	2.87	30.94	41.80	21.97	36.23	41.80	0.80	57.40
H9	51.8	14.7	33.5	54.00	12.72	33.29	56.19	11.00	32.81	51.80	23.23	24.97	51.80	7.41	40.79
H10	47.4	51.8	0.8	54.00	45.20	0.80	60.59	38.63	0.78	47.40	43.69	8.91	47.40	52.59	0.01
H11	61.8	24.1	14.1	54.00	27.51	18.49	46.19	30.48	23.33	61.80	22.69	15.51	61.80	28.04	10.16
H12	53.0	43.8	3.2	54.00	42.44	3.56	54.99	41.08	3.93	53.00	35.02	11.98	53.00	46.74	0.26
H13	40.8	10.6	48.6	54.00	7.34	38.66	67.19	4.69	28.12	40.80	25.12	34.08	40.80	2.54	56.66
H14	53.3	16.6	30.1	54.00	14.92	31.08	54.69	13.43	31.87	53.30	23.39	23.31	53.30	10.42	36.28
H15	58.1	22.0	19.9	54.00	22.56	23.44	49.89	22.95	27.16	58.10	23.40	18.50	58.10	22.45	19.45
H16	40.1	9.1	50.8	54.00	6.21	39.80	67.89	3.87	28.24	40.10	24.46	35.44	40.10	1.76	58.14
H17	48.8	50.1	1.1	54.00	44.87	1.13	59.19	39.67	1.14	48.80	41.66	9.54	48.80	51.17	0.03
H18	58.3	38.0	3.7	54.00	41.38	4.63	49.69	44.62	5.69	58.30	30.30	11.40	58.30	41.29	0.41
H19	53.9	20.1	26.0	54.00	18.50	27.50	54.09	17.05	28.86	53.90	24.39	21.71	53.90	16.63	29.47
H20	55.8	14.4	29.8	54.00	13.62	32.38	52.19	12.89	34.92	55.80	21.65	22.55	55.80	7.98	36.22
H21	57.8	21.0	21.2	54.00	21.30	24.70	50.19	21.46	28.35	57.80	23.19	19.01	57.80	20.29	21.91

Table 1. Continued

	Hongite Subcomposition ABC[a]			Hongite 2[b]			Hongite 3[c]			Hongite 4[d]			Hongite 5[e]		
Sample	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
H22	61.1	30.6	8.3	54.00	35.07	10.93	46.89	39.20	13.91	61.10	25.44	13.46	61.10	36.09	2.81
H23	57.1	40.0	2.9	54.00	42.47	3.54	50.89	44.85	4.26	57.10	31.99	10.91	57.10	42.66	0.24
H24	57.1	20.8	22.1	54.00	20.72	25.29	50.89	20.54	28.56	57.10	23.39	19.51	57.10	19.54	23.36
H25	57.0	30.9	12.1	54.00	31.73	14.27	50.99	32.40	16.60	57.00	26.91	16.09	57.00	36.99	6.01
MN	54.00	27.56	18.44	54.00	27.56	18.44	54.00	27.56	18.44	54.00	27.56	18.44	54.00	27.54	18.46
GM	60.57	27.19	12.24	61.75	25.87	13.38	61.01	24.53	14.45	54.95	24.70	17.35	70.73	24.13	5.14
MD	59.5	29.4	11.1	54.00	31.73	14.27	46.19	30.48	23.33	58.10	23.40	18.50	61.80	28.04	10.16

Note. MN, GM, and MD are mean, geometric mean, and graph median, respectively.

[a] Original hongite subcomposition.

[b] Compositions projected onto line parallel to BC.

[c] Compositions reflected across a line parallel to BC.

[d] Compositions contracted about a line perpendicular to BC.

[e] Compositions expanded about a line perpendicular to BC.

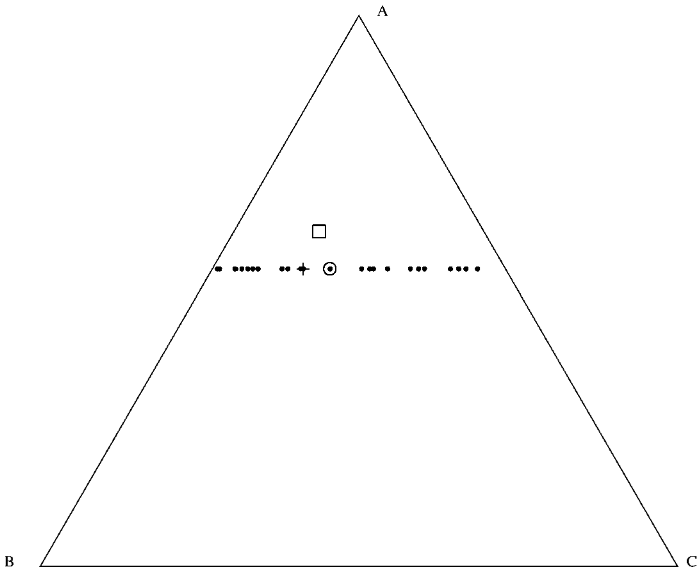

Figure 2. Ternary plot of Hongite 2. Mean (◯), geometric mean (□), graph median (+).

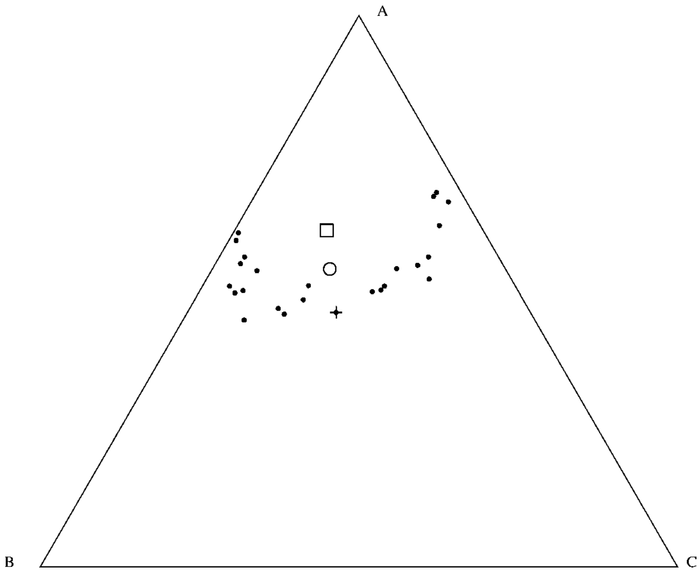

Figure 3. Ternary plot of Hongite 3. Mean (◯), geometric mean (□), graph median (+).

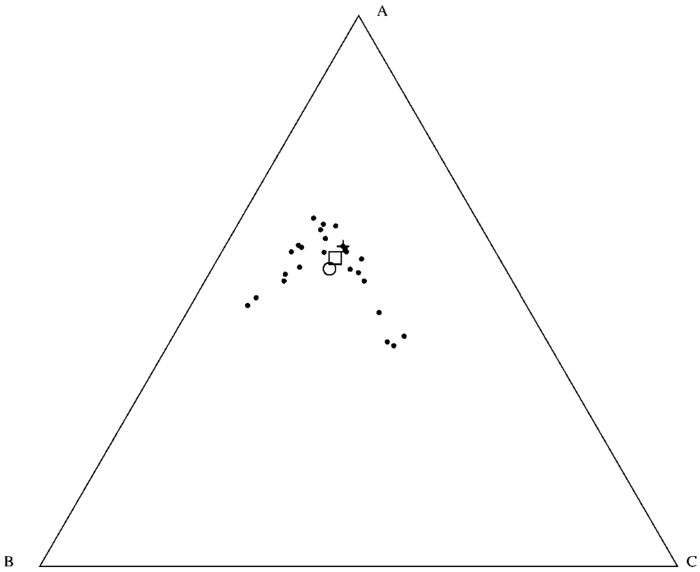

Figure 4. Ternary plot of Hongite 4. Mean (◯), geometric mean (□), graph median (+).

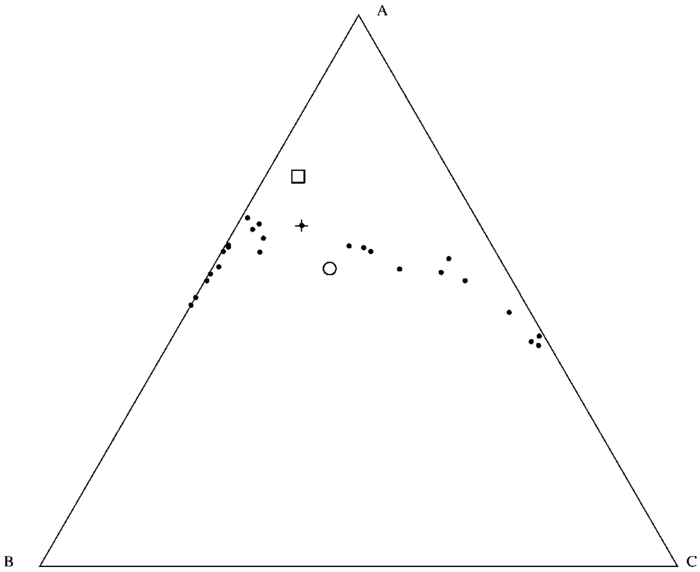

Figure 5. Ternary plot of Hongite 5. Mean (◯), geometric mean (□), graph median (+).

A plot of the hongite subcomposition ABC (Fig. 1) and the location of its geometric mean shown as a square corresponds with the published figure of Aitchison, 1989, p. 788). However the arithmetic average shown here with a circle is closer to the data set than that shown as a circle in the original figure. This is because there seems to have been a computational error in the published value (cf. Table 1 and Aitchison, 1989, p. 789).

A plot of hongite 2 (Fig. 2) shows the geometric mean has virtually the same location as it was in Figure 1 however its position lies well away from the line containing all compositions. In this case the ordinary average now falls within the data set and the geometric mean falls outside.

A plot of hongite 3 (Fig. 3) clearly shows the geometric mean having nearly same location as its position in Figures 1 and 2, but notice that it is now farther removed from the actual data set than is the ordinary average.

For hongite 4, the data was shrunk away from the margins of the diagram while maintaining the same set of values of the component A. As illustrated in a plot of hongite 4 (Fig. 4), this shrinkage causes the geometric mean to now approach the location of the arithmetic average.

For hongite 5, the data was spread out towards the margins of the diagram while maintaining the same set of values of component A. This expansion now causes the geometric mean to migrate toward the apex of component A and well away from the data set. The results for hongite 4 and 5, are consistent with the perturbation effects described in detail by Shurtz (2000).

GRAPH MEDIAN

The graph median (Small, 1997) and the spatial median (Brown, 1983) are extensions of the ordinary median to multidimensional data. A preliminary evaluation of these measures suggested that the spatial median would follow the arithmetic average, but that the graph median should always remain a part of the data set. The graph median is the centroid of the data and as in the one-dimensional case is unique if the multidimensional set of data points are connected to one another by single lines in the form of a tree and the tree is formed by the one set of connecting lines whose sum is a minimum. This graph is referred to as a minimum spanning tree.

For compositional data, a direct measure of distances between compositions can be obtained by using the half-taxi metric (Miller, 2003) and from the set of distances among all compositions, the minimal spanning tree can then be constructed using the method of Prim (1957). The graph median is then obtained by stripping away the outermost branches of the tree (leaves) in a series of successive stages until only one point or a pair of points remain. This point or the mid-point between the pair of points is then the graph median. It is unique, it is a part of the data set and it lies at the center.

The determination of the graph median is considerably more complicated than the computation of the arithmetic average or even the geometric mean, however, it always remains a part of the data set.

The location of the graph median is shown in all of the compositional plots of hongite variations as a plus (+) sign. Note that even though the graph median may shift from one data point to another as the data set is stretched or compressed (cf. Figs. 1–5), the graph median always remains a part of the data set.

CONCLUSIONS

An evaluation of the geometric mean composition for variations on hongite subcompositions ABC shows that the geometric mean in general does not belong to the data set. Worst of all, when all compositions lie on a single line, the geometric mean lies well off the line, an unacceptable compositional inconsistency. In hongite 2 and 5 not only is the geometric mean not a part of the data set, it lies well outside the confines of the cloud of points. That is to say the geometric mean which may have ratio consistency, has no intuitive graphical relationship with compositions in an ordinary ternary diagram.

This is in marked contrast with the arithmetic average (mean) which in all cases lies within the confines and at the center of the cloud of points even when the arithmetic average is not a member of the data.

In all cases the graph median is not only a consistent measure of location it is always a member of the data set even where the arithmetic average is not. While it is computationally much more difficult to obtain, it can serve as an alternative measure of location in those cases, such as principal component analysis, in which the preferred measure of location should be a member of the data set.

REFERENCES

Aitchison, J., 1986, The statistical analysis of compositional data: Chapman and Hall, London, 416 p.

Aitchison, J., 1989, Measures of location of compositional data sets: Math. Geol., v. 21, no. 7, p. 787–790.

Aitchison, J., 1992, On criteria for measures of compositional differences: Math. Geol., v. 24, no. 4, p. 365–379.

Brown, B.M., 1983, Statistical uses of the spatial median: J. R. Stat. Soc. Ser. B Met., v. 45, no. 1, p. 25–30.

Miller, W.E., 2003, Revisting the geometry of a ternary diagram with the half-taxi metric: Math. Geol., v. 34, no. 3, p. 275–290.

Prim, R.C., 1957, Shortest connection networks and some generalizations: Bell Syst. Tech. J., v. 36, p. 1389–1401.
Shurtz, R.F., 2000, Comment on "Logratios and natural laws in compositional data analysis" by J. Aitchison: Math. Geol., v. 32, no. 5, p. 645–647.
Small, C.G., 1997, Multidimensional medians arising from geodesics on graphs: Ann. Stat., v. 25, no. 2, p. 478–494.