Mathematical Geology, Vol. 35, No. 8, November 2003 (© 2003)

Compositional Geometry and Mass Conservation[1]

Robert F. Shurtz[2]

A geometrical structure is imposed on compositional data by physical and chemical laws, principally mass conservation. Therefore, statistical or mathematical investigation of possible relations between data values and such laws must be consistent with this structure. This demands that geometrical concepts, such as points that specify both mass and composition in linear space, and lines in projective space that specify composition only, be clearly defined and consistent with mass conservation. Mass thus becomes the norm in composition space in place of the Euclidean norm of ordinary space. Coordinate transformations inconsistent with this geometry are accordingly unnatural and misleading. They are also unnecessary because correlation arising from the constant mass presents no unusual difficulty in the analysis of the underlying quadratic form.

KEY WORDS: mass coordinates, composition ratios, mass norm, compositional statistics, taxicab metric, mass balance, lever rule.

INTRODUCTION

If X, Y, and Z in the triplet (X, Y, Z) represent masses of three components that constitute a batch of system **XYZ**, an increase in any one of them does not necessitate a reduction in the sum of the others: it necessitates only an increase in the total weight of the batch, $M = X + Y + Z$.

But if X, Y, and Z represent the corresponding percentages (ratios, etc.) it is implied that $M = 100$ ($M = 1$, etc.), so that an increase in any one of them must be accompanied by an equal reduction in the sum of the others. For example, an increase of 1% in Z must be accompanied by a decrease of 1% in the sum of X and Y in order to balance the equation

$$Z = 100 - X - Y. \tag{1a}$$

This is the equation of a plane and is indistinguishable from that for a linear statistical regression of Z on X and Y (except for an omitted random term); it

[1]Received 2 May 2003; accepted 11 August 2003.
[2]1661 Pine Street, No. 425, San Francisco, California 94109; e-mail: rfs@infinex.com

0882-8121/03/1100-0927/1

extends naturally to an $(N - 1)$-dimensional hyperplane representing a system of any finite number, N, of components in N-dimensional space

$$X_N = 100 - \sum_{k=1}^{N-1} X_k. \tag{1b}$$

To avoid confusion between this mass balance effect and regression possibly due to other factors, various complicated coordinate transformations have been suggested without apparent attention to their consistency with the law of mass conservation. But, just as spherical coordinates are mandatory in terrestrial cartography, a linear coordinate space equipped with mass coordinates is mandatory in compositional geometry. Such a space underlies every composition, but is most easily understood in the case of the widely used and understood, ternary composition diagram. This is the subject of this paper.

A DETAILED EXAMPLE

The numbers in the triplet (6, 2, 3) may simply be values—in otherwise undefined units—of the coordinates at the point **B** in the three-dimensional orthogonal coordinate system **XYZ** shown in Figure 1 with the **Y**-axis running away from the observer. They may equally well be values of the masses of three identifiable constituents in a recipe for, or a batch of, or an analysis of a compound, combination or mixture of X, Y, and Z, real or conceptual.

In Euclidean geometry the distance from the origin **O** to **B** is 7: the square root of the sum of the squared coordinate differences, $6^2 + 2^2 + 3^2 = 36 + 4 + 9 = 49$. To construct **B**, one starts at **O** and, at least conceptually, lays out the *lengths*, 6, 2, and 3, along the **X**, **Y**, and **Z** axes.

The distance from **O** to **B** in mass coordinates is obviously 11, the ordinary sum of coordinate magnitudes $6 + 2 + 3$. To construct **B**, one starts at **O** and, at least conceptually, lays out *masses*, 6, 2, and 3, of X, Y, and Z along the axes **X**, **Y**, and **Z**.

Technically, distances from the origin to a point are norms (Taylor and Lay, 1980, p. 10), in this case, on the three-dimensional, real arithmetic space whose "points" are triplets of numbers. More generally, they may be multiplets of any finite number N of real positive numbers. The norm of a coordinate difference defines a metric, a distance between each pair of points, on any such space (see any text on functional analysis, e.g., Kreysig, 1978, chap. 2). In Figure 1, the 7-unit distance from **O** to **B** is the Euclidean norm, and the 11-unit distance, the sum of the coordinate magnitudes, may be called the mass norm.

It has no standard mathematical name, but is sometimes called the taxicab norm. In three dimensions it is one of a class of norms

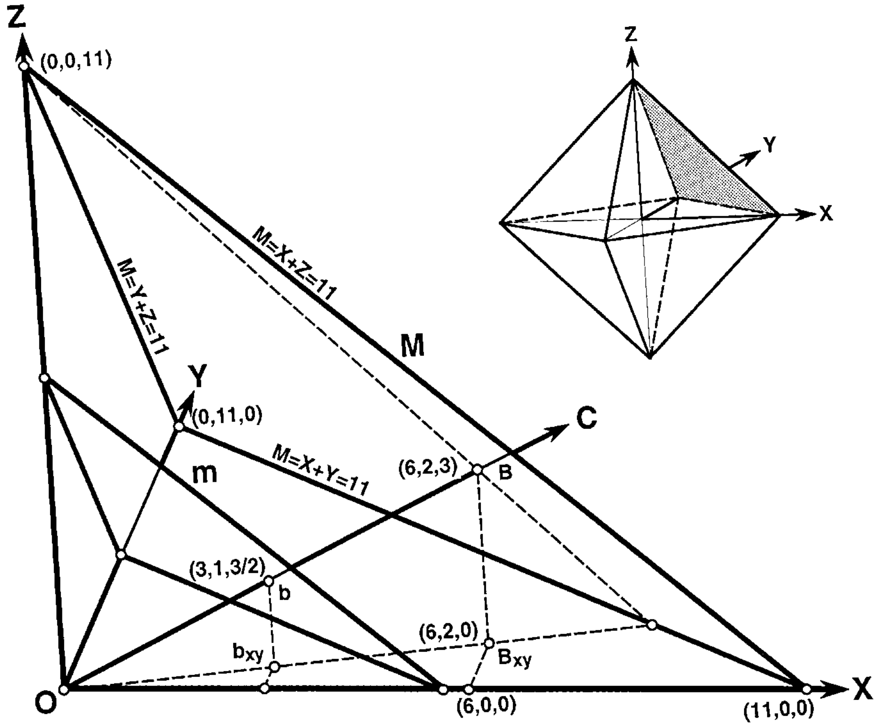

Figure 1. The mass-coordinate space **XYZ** of three-component compositions with the **Y**-axis running away from the observer. Two constant-mass planes, **M** at $M = 11$ and **m** at $m = 3.5$, are shown, along with a "sphere," defined by the mass norm, The point **B** in **M** represents a batch of 11 unit mass, six of **X**, two of **Y**, and three of **Z**. The point **b** in **m** represents a batch half that size and illustrates the fact that all compositions in the ratio $X{:}Y{:}Z = 6{:}2{:}3$ lie on a radial line **C**, at all points of which the mass ratios are the same, $6/11{:}2/11{:}3/11 = 0.55{:}0.18{:}0.27$. For additional explanation see text.

$$\|M\| = (|X|^p + |Y|^p + |Z|^p)^{1/p}, \quad p \geq 1 \tag{2}$$

with obvious extension to N dimensions. When $p = 2$ Equation (2) gives the Euclidean norm, and when $p = 1$ it gives the mass norm. These two norms are topologically equivalent (Kreyszig, 1978, p. 75); but their distinction must be understood and accounted for in compositional geometry, just as the distinction between spherical geometry and plane geometry must be understood in terrestrial cartography.

It is to be expected that effects of this distinction will appear in all calculations involving mass, such as finding a mean, or average composition; and, as shown by the following example, they do. Aitchison (1989, p. 788) calculates the average of 25 percentage compositions in a three-component system and gets the percentages (49, 29, 22). He then transforms the 25 percentage coordinates into logratio coordinates, calculates the average of the 25 logratios, and transforms it back to get the percentage averages (60, 27, 13). Thus, what superficially seems a plausible calculation becomes a physical absurdity: it creates 11 mass percent of component 1, and destroys 2% of component 2 and 9% of component 3, in gross violation of mass conservation. To avoid such bizarre results, compositional geometers must use mass coordinates and the mass norm.

FROM MASS COORDINATES TO PROJECTIVE SPACE

All points lying 11 units of mass distance from the origin in Figure 1 may be "constructed" by combining masses X of **X**, Y of **Y**, and Z of **Z** in consistent units that total to $M = X + Y + Z = 11$. This is the equation of a plane that intersects all three axes at points 11 mass units from the origin and extends without bound in the **XYZ** coordinate system.

In as much as the mass norm is equal at all points in any plane with equal intercepts on the axes, the unit-sphere in mass coordinate space is a unit-octahedron instead, such as that shown in the inset in Figure 1. But, because mass always equals or exceeds zero, only the points in the shaded triangular face of the octahedron, on which all coordinate values equal or exceed zero, appear as compositions. The points in this triangle do not constitute a vector space; but, just as distances between points in a linear vector space are always regarded as positive even though they may be measured in a negative direction, so positive mass values may be measured in a negative direction. Then the entire **XYZ** space becomes mass coordinate space in which the distance between points, generalized to N components, is defined as the sum of the absolute coordinate differences,

$$d(\mathbf{X}, \mathbf{Y}) = \sum\nolimits_1^N |X_i - Y_i|. \tag{3}$$

Mass is continuous—at least down to the atomic scale—except at zero, where it is one–sided continuous upward. So, if two of the three components are absent, their masses may be regarded as having been reduced continuously to zero by infinitesimal changes. There are three such mass coordinate points in Figure 1: (11, 0, 0), (0, 11, 0), and (0, 0, 11). Each of them satisfies the equation of the plane **M**, $X + Y + Z = 11$, at its intercept on the corresponding coordinate axis, all at 11 units of mass distance from **O**.

If only one component is absent, its mass may be presumed to have decreased continuously to zero. All possible mass points then lie on lines of intersection between **M** and the three coordinate planes: (a) for points $(0, Y, Z)$ they all lie on the line $M = Y + Z = 11$; (b) for points $(X, 0, Z)$, they lie on $M = X + Z = 11$; and (c) for points $(X, Y, 0)$, on $M = X + Y = 11$.

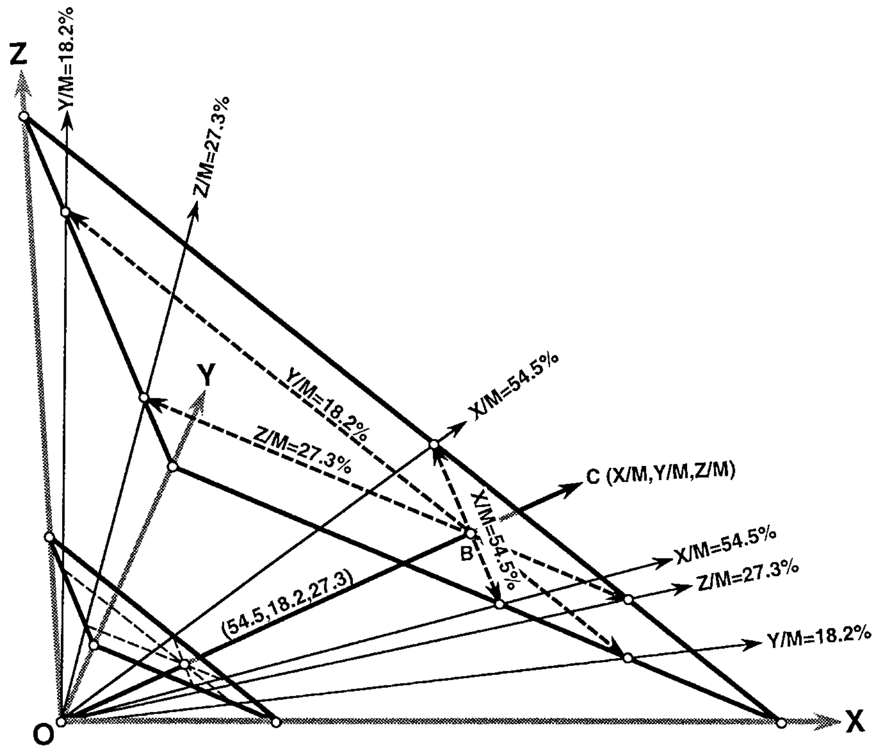

Figure 2. The projective space of three-component compositions corresponding to Figure 1. Radial lines such as **C**, lying at the intersections of radial planes, such as those whose intersections with **M** are shown by dashed lines at $X = 54.5\%$, $Y = 18.2\%$, and $Z = 27.3\%$. Corresponding lines in the constant-mass plane near the origin are also shown. For further explanation see text.

In Figure 2 three radial planes, defined by ratios, $X/M = 6/11 = 54.5\%$, $Y/M = 2/11 = 18.2\%$, and $Z/M = 3/11 = 27.3\%$, are shown. They all intersect in the constant-ratio line **C**, at all points of which the ratios are those of the radial planes, whose intersections with the coordinate planes are shown by arrows, while their intersections with **M** are shown by dashed traces. The sum of the three ratios is one, and of the percentages 100, so any two suffice to define **C**.

It is evident that all radial lines through points in the closed, constant-mass triangle **M** are constant-ratio lines, each identified by the values of (any) two percentages or ratios. A line in space, identified by two such numbers, is thus analogous to a point in a plane, identified by two coordinate values. Accordingly, these radial lines may be regarded as elements of (points in) a two-dimensional space.

In standard mathematics the family of all radial lines in three dimensions is called the projective plane (e.g., Roman, 1992, p. 324; Samuel, 1988, p. 1), so the sheaf, or pencil of radial lines from the origin $\mathbf{O}$ through points in, or on the boundary of, the triangle $\mathbf{M}$ constitutes a closed region in two-dimensional projective space. From a center at the origin, this pencil of lines projects all constant-mass triangles in the entire positive octant of mass coordinate space, such as the small one near the origin in Figure 2, onto any one of them, such as $\mathbf{M}$. Thus, each point in, or on the boundary of, $\mathbf{M}$ is an end view of one of these radial lines looking toward the origin.

The triangle $\mathbf{M}$ thus becomes a ternary composition diagram on which all constant-mass triangles in the positive octant of the system $\mathbf{XYZ}$ have collapsed.

For systems of more than three components, radial lines are still constant-ratio lines and they project as points in hyperplanes of one less dimension than the space itself. When $N = 4$, the three-dimensional hyperplane is a tetrahedron. Each of its points is the projection—the end view—of a radial, constant-ratio line in the four-dimensional mass coordinate space. This argument extends to mass coordinate spaces of any number of components: the end view of a radial line is always a point in a constant-sum, $(N - 1)$-dimensional flat, a hyperplane, in an N-dimensional space.

Any pair of points may be regarded as end views of two radial lines that intersect at the origin of coordinates in a system of *any finite number* of components. Such a pair of lines defines an ordinary two-dimensional radial plane whose edge view is the line connecting them. Any mixture of the two, point compositions must lie somewhere on the line between them at a location which may be found simply and graphically in two dimensions, regardless of the number of components.

CONSTRUCTING A COMPOSITION

Figure 3 shows a shaded triangle lying in a radial plane that is defined by two, constant-ratio radial lines, $\mathbf{C}_1$ and $\mathbf{C}_2$. The third side of the triangle is the line, $\mathbf{m}_0$, on which this plane intersects the constant-mass plane $\mathbf{m}_0$ in which the mass m_0 at all points equals the sum of the masses m_1 and m_2, at all points in the constant-mass planes $\mathbf{m}_1$ and $\mathbf{m}_2$. The system in which the shaded triangle lies may be of any finite dimension (number of components) equal to or greater than two. The phantom three-dimensional mass coordinate system is shown only for orientation with respect to Figures 1 or 2.

Fixing masses m_1 and m_2 defines batch vectors at the points $\mathbf{B}_1$ and $\mathbf{B}_2$ where constant-mass lines $\mathbf{m}_1$ and $\mathbf{m}_2$ intersect constant-ratio lines $\mathbf{C}_1$ and $\mathbf{C}_2$. By conservation of individual component masses, the batch point of the mixture must lie on the constant-mass line $\mathbf{m}_0$ at the point $\mathbf{B}_0 = \mathbf{B}_1 + \mathbf{B}_2$ constructed by the usual vector addition, shown graphically in the shaded triangle.

In mass geometry the mass length of the sum vector must *equal* the sum of the mass lengths of the summed vectors, $|\mathbf{B}_0| = |\mathbf{B}_1| + |\mathbf{B}_2|$, in contrast to the length of the sum vector in Euclidean geometry which may be—and almost always is—*less than* the sum of the lengths of component vectors, in this case, $|\mathbf{B}_0| < |\mathbf{B}_1| + |\mathbf{B}_2|$.

This is the pivotal distinction between mass geometry and Euclidean geometry: Euclidean lengths are measured "as the crow flies," cutting across the coordinate grid, while the mass lengths are measured as the taxi must drive, along the lines, the "streets," of the coordinate grid. In Figure 3, this route in three dimensions is suggested by phantom lines running in order from the origin parallel to the $\mathbf{Z}$, the $\mathbf{X}$, and the $\mathbf{Y}$ axes. These segments do not lie in the shaded plane; and, in higher dimensional cases, they cannot be depicted at all.

The intersection between $\mathbf{m}_1$ and $\mathbf{C}_1$ and that between $\mathbf{m}_2$ and $\mathbf{C}_2$ divide the respective sides of the shaded triangle into two segments. On $\mathbf{C}_1$, the mass lengths of the segments in order, starting from $\mathbf{O}$, are proportional to m_1 by congruence,

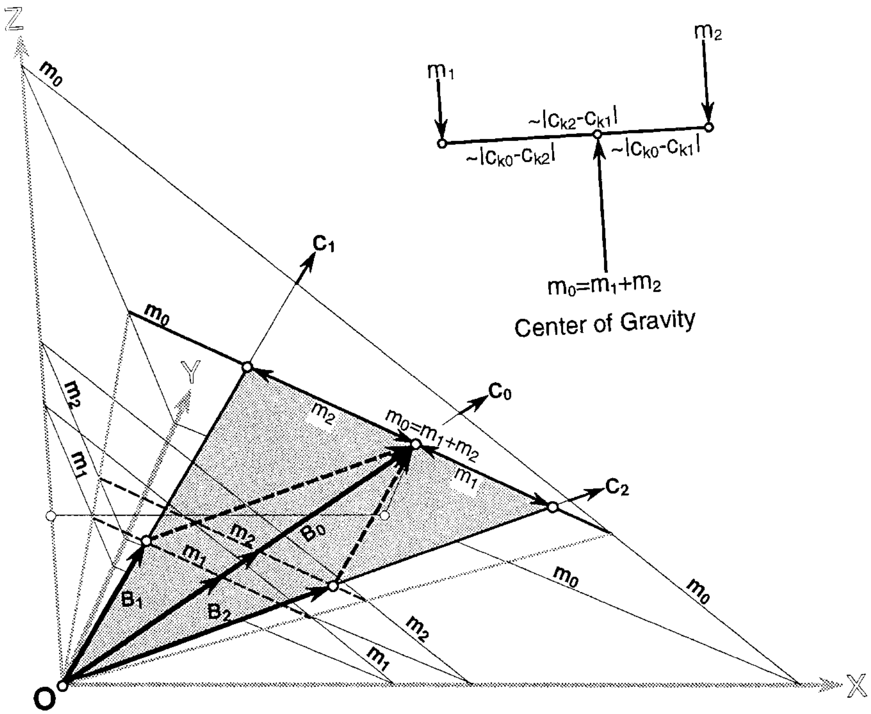

Figure 3. The lever-principle construction of mixed batches in projective space. In the inset, note that the masses appear on the points of the lever that *are not* endpoints of the segments to whose lengths the masses are proportional. For further explanation see text, especially the equations in the Mass Balance section. The phantom three-dimensional coordinates are for general orientation only: only the shaded plane is essential to the construction.

and to m_2 by similarity of triangles. On $\mathbf{C}_2$, the segments are in reverse order, the length of the first is m_2 by congruence, and that of the second is m_1 by similarity of triangles. By similarity of the triangles in the vector construction, $\mathbf{B}_0$ divides the base of the shaded triangle along $\mathbf{m}_0$ into two segments: one of length proportional to m_1 next to its intersection with $\mathbf{C}_2$, and one of length m_2 next to its intersection with $\mathbf{C}_1$: note the inversion. That this graphical construction solves the problem of finding the sum vector $\mathbf{B}_0$ will now be shown analytically.

MASS BALANCE

A process engineer, to whom calculations of mass balance are routine, would see this construction as an awkward and superfluous way of approaching the trivial problem of simultaneously solving two simple algebraic equations: (a) the total mass balance equation, $m_1 + m_2 = m_0$; and (b) a mass balance equation for *any* component common to both $\mathbf{C}_1$ and $\mathbf{C}_2$, no matter how many there may be. For the k-component, equation (b) is $c_{k1}m_1 + c_{k2}m_2 = c_{k0}m_0$, where c_{k1}, c_{k2}, and c_{k0} are the ratios or percentages of the k-th component at all points of the composition lines $\mathbf{C}_1$, $\mathbf{C}_2$, and $\mathbf{C}_0$.

Eliminating m_0, gives $m_1|c_{k1} - c_{k0}| = m_2|c_{k2} - c_{k0}|$. This is the form of solution that illustrates why, in compositional geometry for almost a century, this solution has been called the Lever Rule or the Center of Gravity Principle (Levin and others, 1964, p. 7): it says, simply, that the masses m_1 and m_2 balance on lever arms of length $|c_{k1} - c_{k0}|$ and $|c_{k2} - c_{k0}|$ over a fulcrum at $\mathbf{C}_0$, as shown in the inset in Figure 3. The differences between component concentrations, at the intersections of any constant-mass line with $\mathbf{C}_1$, $\mathbf{C}_2$, and $\mathbf{C}_0$, are proportional to the masses, $|c_{k2} - c_{k0}|{:}|c_{k1} - c_{k0}|{:}|c_{k2} - c_{k1}| = m_1{:}m_2{:}m_0$. The unit of length used to express mass distance between the points is immaterial: it cancels out. This latitude in choice of length units is characteristic of projective geometry whose points are lines between which no linear distance is defined.

COMPOSITIONAL STATISTICS

If a random set of data vectors in three-dimensional space is confined to a plane, instead of spreading randomly over all three dimensions, the rank of the matrix of correlation coefficients is two: accordingly, its determinant must vanish:

$$\begin{vmatrix} 1 & r_{XY} & r_{XZ} \\ r_{XY} & 1 & r_{YZ} \\ r_{XZ} & r_{YZ} & 1 \end{vmatrix} = 0.$$

The equation of the constant-mass plane, Equation (1), imposes two additional constraints on the correlation coefficients: the regression coefficients both equal minus one:

$$a = \begin{vmatrix} r_{XZ} & r_{XY} \\ r_{YZ} & 1 \end{vmatrix} \Bigg/ \begin{vmatrix} 1 & r_{XY} \\ r_{XY} & 1 \end{vmatrix} = -1$$

and

$$b = \begin{vmatrix} 1 & r_{XZ} \\ r_{XY} & r_{YZ} \end{vmatrix} \Bigg/ \begin{vmatrix} 1 & r_{XY} \\ r_{XY} & 1 \end{vmatrix} = -1.$$

These three equations constrain the correlation coefficients when $N = 3$.

But, in systems of more than three components, the vanishing determinant and the equations for the regression coefficients impose fewer constraints, N, than there are correlation coefficients, $(N^2 - N)/2$; accordingly, such systems are not constrained.

The terms of the correlation matrix are sums of products of the form $(x_i - m_i)(x_j - m_j)$ which are measured along specific coordinate axes and so take equal values in both mass and distance units (compare the intercepts of a constant-mass plane). Therefore, the matrix of the quadratic form can be estimated in the usual way in mass coordinate space. Its diagonaliazation will, however, reveal an eigenvalue of zero which means that one of the semiaxes of the characteristic quadric surface (hypersurface) is infinite (Shilov, 1977, p. 291). Diagonalization also yields a new set of random variables, one fewer in number than the original set, but mutually orthogonal and expressed as linear functions of the original variables. These linear functions contain all the information, about relations between variables that is in the data. Other statistical procedures for application in mass coordinate systems must be evaluated on a case-by-case basis.

A constant-mass plane in three dimensions is a two-dimensional surface and it remains two-dimensional under continuous transformation of coordinates. If, however, the transformation is *nonlinear*, the surface remains two-dimensional but is no longer a plane. Accordingly, a distribution of points in the original plane becomes a distribution in whatever curved surface results from transformation. Unless adequate adaptive measures can be found, the usual statistical methods for analzing the data and for testing hypotheses about the original variables become invalid.

DISCUSSION

As already noted, correlation "built-in" by the implied constant-mass of compositions, has motivated efforts to mitigate its effect by transforming rectilinear coordinate systems to curvilinear coordinates. These are inconsistent with rectilinear mass coordinate geometry just as, vice versa, the rectilinear coordinates of plane maps are inconsistent with the spherical surface of the earth.

Distances between points in curvilinear coordinates, such as the logratio coordinates mentioned in the Introduction, must be measured along curves, not along straight lines, for the same reason that the distance between San Francisco and Paris must be measured on a great circle on the surface of the earth, not on a straight line through the earth.

Further, noninvertible coordinate transformations are unacceptable because the original data cannot be recovered from the converted data. This means that clusters or trends observed in the converted data cannot be attributed to possible natural causes because they may be only mathematical artifacts. An example is the closure operation,

$$C(z_i) = cz_i/(z_1 + z_2 + z_3 \ldots + z_d) \tag{4}$$

which appears in a transformation proposed by Pawlowsky-Glahn and Egozcue (2002, p. 261): it restricts the sum of the transformed coordinates to a chosen c, such as $c = 100\%$. The operation in Equation (4) is not invertible because the sum in the denominator is lost and cannot be recovered from the values of the $C(z_i)$, once the division is performed.

More importantly, the closure in Equation (4) imposes a constant-sum constraint on the transformed data that is no different from that on the original data, and which is the purported justification for transforming that data in the first place. Moreover, this new constraint is in addition to the transform of the original constant-sum plane which is still there, although no longer a plane. In the logratio transformations, it is a surface which curves away to infinity.

Values of the logratio transform approach positive infinity as the denominator of the ratio approaches zero, and they approach negative infinity as the numerator approaches zero. All such boundary points lie in flats—the points, lines, planes, and hyperplanes—that bound the positive region of mass coordinate space, on all of which one or more of the masses is zero.

In efforts to avoid the embarrassment of infinite values, all these boundary points have been relegated to a multitude of flats of the necessary dimensionality, and called subcompositions. The idea is clearly stated in Atchison and others (2000, p. 274): ". . . a composition with one of its parts absent may be chemically, physically, or biologically completely different from compositions with all parts positive." This grossly contradicts everyday experience: to add a microgram of sugar to a liter of pure water does not make any such difference in its properties; and to add a molecule of sugar renders this statement preposterous.

Finally, there is no practical difference between values that increase to infinity and infinite values themselves: they become far too large to make sense long before reaching those boundaries. This is illustrated in the simple series of logratio compositions discussed by this writer (Shurtz 2000, Fig. 1). As the two compositions there discussed approach boundary points on the sides of a ternary diagram from its interior, the back transform of their logratio average approaches the corner of the diagram at which both the originally predominant component masses have vanished. The gross violation of mass conservation in this sequence is abundantly evident long before the corner is reached.

CONCLUSION

In conclusion, it is evident that these efforts to transform away the constant-sum constraint are unnecessary. The usual diagonalization of the fundamental quadratic form in mass coordinate space reveals that the data lies in a space of fewer than N dimensions. In addition, it reveals a linear coordinate transformation in which the new coordinates are expressed as linear functions of the original mass coordinates, thus ensuring that original masses can be recovered. The quadratic form contains all the information about underlying geological processes that can be extracted from second- and lower-order moments of the data.

So it seems to this commentator that the best way to avoid the many problems posed by physically impossible coordinate transformations is to "... be careful not to befuddle ourselves with transformations so exotic that we lose sight of the original nature of the geologic properties we are attempting to understand" (Davis, 1986, p. 92).

REFERENCES

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

Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J. A., and Pawlowsky-Glahn, V., 2000, Logratio analysis and compositional distance: Math. Geol. v. 32, no. 7, p. 271–275.

Davis, J. C., 1986, Statistics and data analysis in geology: Wiley, New York, 646 p.

Levin, E. M., Robbins, C. R., and Mc Murdie, H. F., 1964–1991, Phase Diagrams for Ceramists: v. I through v. VIII, compiled by the National Institute of Standards and Technology (formerly National Bureau of Standards) for The American Ceramic Society, Inc., Westerville, O. H., 8180 Diagrams.

Kreyszig, E., 1978, Introductory functional analysis with applications: Wiley, New York, 688 p.

Pawlowsky-Glahn, V., and Egozcue, J. J., 2002, BLU estimators and compositional data: Math. Geol., v. 34, no. 3, p. 259–274.

Roman, S., 1992, Advanced linear algebra: Springer-Verlag, New York, 363 p.

Samuel, P., 1988, Projective geometry: Springer-Verlag, New York, 156 p.

Shilov, G. E., 1977, Linear algebra, (translated and edited): R. A. Silverman, Dover, New York, 387 p.

Shurtz, R. F., 2000, Comment on: "Logratios and natural laws in compositional data analysis" by J. Atchison: Math. Geol., v. 32, no. 5, p. 645–647.

Taylor, A. E., and Lay, D. C., 1980, Introduction to functional analysis: Krieger, Malabar, FL, 467 p.