

Element/Aluminum Ratios in Chemostratigraphy: A Dubious Normalization Resulting in Spurious Correlations

Dr. Grant W. Lowey, P.Geo.
Cordilleran Associates Limited
P.O. Box 21254, Whitehorse, Yukon Y1A 6R2
loweygrant@gmail.com

Introduction

*"The correlation of ratios can lead the user into a great deal of trouble and should normally be avoided."
(Rollinson, 1993)*

Chemostratigraphy, or chemical stratigraphy, is the application of inorganic geochemical data (e.g., major elements, trace elements, rare earth elements, and isotopes) and organic geochemical data (e.g., biomarkers, kerogen type, and total organic carbon) to characterize, subdivide and correlate strata.

Geochemical analyses are reported as parts of a whole (i.e., a fixed total), such as percentages (100) or parts-per-million (1,000,000), and are referred to as 'compositional data'. Compositional data convey only relative information and are not free to vary independently: an increase in one component must cause a decrease in one or more of the other components, resulting in a negative correlation bias (Aitchison, 1986). This is the so called 'constant sum problem' of geochemical data that causes induced 'spurious' correlations (Rollinson, 1993).

However, a more serious form of spurious correlation arises from the use of ratios. Over a century ago, Pearson (1897) warned that spurious correlations result when two ratio variables (i.e., X and Y) have the same common denominator (i.e., the correlation between X/C and Y/C). The spurious correlation produced is caused solely by arithmetic and not due to any natural processes between the variables.

Although the issue of spurious correlation between ratio variables has been raised by numerous authors (i.e., Chayes, 1986; Rollinson, 1993), the use of ratio correlation is still widespread in geochemistry (e.g., so called 'Pearce element ratio diagrams'), and particularly chemostratigraphy. Note that the conversion of geochemical data (i.e., percentages, ppm, etc.) to logarithmic form, molecular proportions, or molar ratios, has no effect on the spurious nature of ratio correlation. However, Aitchison's (1986) log-ratio transformation is a solution to the constant sum problem. In addition, the problem of ratio correlation does not apply to isotopes because the isotopic composition of an element is independent of its concentration or any other process (i.e., it is only affected by radioactive decay).

The aim of this study is to expose the widespread practice in chemostratigraphy of normalizing geochemical data to aluminum, to highlight the problems in using element/aluminum ratios, and to illustrate these problems by re-analyzing published chemostratigraphic results of element/aluminum ratio subdivided and correlated strata from oil wells in Cook Inlet, Alaska.

Dubious Normalization

"The distribution of several metals in East Pacific Rise sediments, when normalized to Al₂O₃, exhibit stronger maxima near the rise crest than when simply plotted on a carbonate-free basis." (Piper, 1973)

It has become common practice in chemostratigraphy to normalize geochemical compositions to aluminum. This method of standardization (i.e., to normalize an element in the numerator to aluminum in the denominator, or E/A, in which aluminum is commonly expressed as Al₂O₃ wt %) is an attempt to compensate for grain-size distribution, variations in mineralogy, dilution or enrichment by other elements, sedimentation rates, and as index of fluvial input (Dunlap et al., 1997; Van der Weijden, 2002).

Implicit in the use of E/A ratios is the assumption that aluminum is derived from detrital minerals that were transported from the continents to the sea. Although fluvial transport was initially proposed as the major source of aluminum to the oceans (i.e., dissolved and suspended loads), it is now known that most aluminum is removed in estuarine and coastal environments, and that eolian dust is the main source of aluminum in the open oceans (Middag et al., 2009). In addition, Murray and Leinen (1996) found a significant source of aluminum scavenged directly from seawater, indicating that the use of E/A ratios as a terrigenous proxy may lead to an overestimate of the actual terrigenous load of marine sediments (and possibly by a factor of two).

Normalization to aluminum also assumes that aluminum has a uniform variability (i.e., a uniform flux from weathering), such that inter-element ratios with Al₂O₃ are constant between samples (Piper and Isaacs, 1995). Furthermore, such ratios are correct only when the element-aluminum variables are collinear (i.e., plot as a straight line) with a zero intercept (Jackson and Somers, 1991). Hence, the premise that relations between normalized values can be treated like a mathematical function describing a linear relation may not be valid (Van der Weijden, 2002).

Perhaps the first use of E/A ratios was by Piper (1973) in a study of metaliferous sediments from the East Pacific Rise. According to Piper (1973), the usual procedure for estimating variations in element concentrations in sediments was to re-calculate the bulk concentrations on a carbonate-free basis. Due to uncertainties in the biogenic component of marine sediments and the imprecision of CaCO₃ analyses at the time, Piper (1973) proposed normalizing element concentrations to Al₂O₃ (which he determined ranged in concentration from 0.16 to 1.15 wt % across the East Pacific Rise), based on the assumption that all of the aluminum was derived from terrestrial detritus and that the rate of accumulation of aluminum was relatively constant. Piper (1973) concluded that normalizing metal concentrations to aluminum resulted in an enhanced maximum near the rise crest, compared to concentrations plotted on a carbonate-free basis: but this is exactly the result that would be expected from a spurious correlation of ratios.

Spurious Correlations

"A ratio is seductively simple." (Curran-Everett, 2013)

The correlation of ratio variables (i.e., Y/Z, and X/W) can be approximated by Pearson's (1897) equation:

$$r_{(Y/Z)(X/W)} = \frac{r_{XY} V_Y V_X - r_{YW} V_Y V_W - r_{XZ} V_X V_Z + r_{ZW} V_Z V_W}{(V_Y^2 + V_Z^2 - 2r_{YZ} V_Y V_Z)^{1/2} (V_X^2 + V_W^2 - 2r_{XW} V_X V_W)^{1/2}}$$

Where r represents the correlation coefficient between two ratio variables and V represents the coefficient of variation (i.e., the standard deviation divided by the mean) of a particular variable.

To illustrate how inter-correlations of ratio variables using aluminum as a common denominator behave, assume that the coefficients of variations of all variables are equal (i.e., $V_X=V_Y=V_Z=V_W$), that variable X equals some particular element (i.e., $X=E_1$) and that variable Y represents some other particular element (i.e., $Y=E_2$), and that variables Z and W represent aluminum (i.e., $Z=W=A$). Pearson's (1897) equation then reduces to (Dunlap et al., 1997):

$$r_{(E_1/A)(E_2/A)} = \frac{1 - r_{E_2A} - r_{E_2A} + r_{E_1E_2}}{2(1 - r_{E_1A})^{1/2} (1 - r_{E_2A})^{1/2}}$$

If the correlations among element 1, element 2 and aluminum are zero (i.e., $r_{E_1A} = r_{E_2A} = r_{E_1E_2} = 0$), the correlation among ratios E_1/A and E_2/A are not zero: in fact, the above equation simplifies to 0.5! Hence, the spurious correlation produced by using aluminum as a common denominator is 0.5 when all variables are uncorrelated. However, aluminum normalization can not only increase the correlation, but it can also decrease the correlation, change the sign of the correlation, or confound the correlation between elements (Van der Weijden, 2002). In addition, the problem of spurious correlation extends to any ratio, including single term ratios, ratios with multiple terms, and ratios with added terms for which there is a common variable (i.e., when one variable is correlated with the ratio of itself to another variable, a common variable in the numerator of the two ratios, or a common variable in the numerator of one ratio and in the denominator of the other ratio)(Kritzer, 1990).

An Example

"Access to the complete data... is a fundamental requirement for the reproducibility of scientific results."
(Goldstein et al., 2009)

A review of the chemostratigraphic literature reveals that most studies using whole rock major and trace element data do not publish results of the geochemical analyses used in the study (e.g., Hildred et al., 2010; Pearce et al., 2010; Ratcliffe, et al., 2010), making it difficult to evaluate the reproducibility of the chemostratigraphic results claimed. However, one study for which geochemical data are available is the chemostratigraphic framework established for gas-prone Oligo-Miocene sequences in Cook Inlet, Alaska (Chemostrat, 2007). Using whole rock geochemistry, this study subdivided the strata into five intervals based mainly on the correlation of $\text{Na}_2\text{O}/\text{Al}_2\text{O}_3$ versus depth profiles, $\text{Rb}/\text{Al}_2\text{O}_3$ versus depth profiles, and binary plots of $\text{Rb}/\text{Al}_2\text{O}_3$ versus $\text{Na}_2\text{O}/\text{Al}_2\text{O}_3$ ratios (Chemostrat, 2007).

Re-analysis of the data shows that a plot of Al_2O_3 versus Na_2O does not have a zero intercept (i.e., there is a positive correlation with 'excess' aluminum), and neither does a plot of Al_2O_3 versus Rb (i.e., no correlation). Furthermore, Na_2O and Rb are negatively correlated (i.e., $r_{\text{Na}_2\text{O Rb}} = -0.71$), but the negative correlation increases when normalized to aluminum (i.e., $r_{\text{Na}_2\text{O}/\text{Al}_2\text{O}_3 \text{ Rb}/\text{Al}_2\text{O}_3} = -0.89$) due to the spurious correlation effect. In addition, re-plotting Rb versus Na_2O shows significantly more scatter than a plot of $\text{Rb}/\text{Al}_2\text{O}_3$ versus $\text{Na}_2\text{O}/\text{Al}_2\text{O}_3$ ratios. Hence, the chemostratigraphic subdivision based on normalization to aluminum may be an artifact of spurious correlation, rather than due to the preferential concentration of certain elements in the strata by natural processes.

Conclusion

"...any interpretation must be tempered by the knowledge that no good method exists for estimating what proportion of the correlation of ratios sharing common elements is spurious and what part is not."

(Dunlop et al., 2001)

The dubious normalization of geochemical data to aluminum (i.e., E/A ratios) is widespread in chemostratigraphy. Although the correlation of such ratios gives apparently reasonable and geologically interpretable results, the correlations are spuriously biased. Warnings against the use of such ratios are legion, but still go largely unheeded. However, it is not possible to distinguish between the effects of spurious correlation caused by ratios and the effects due to geologic processes (Pawlowsky-Glahn and Egozcue, 2006). Therefore, conclusions drawn from E/A ratios are misleading and probably wrong.

Acknowledgements

I am extremely grateful to Dan Hawkins, Professor of Geology at the University of Alaska, for introducing me to geochemistry, and would especially like to thank those folks at the Alberta Energy Regulator for providing me with this opportunity to pursue research in chemostratigraphy.

Selected References

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

Chemostrat, 2007. Chemostratigraphy of Oligo-Miocene sequences in the Cook Inlet, Alaska. Chemostrat Report Ci278, 30 pp. (Also Alaska Materials Center Data Report No. 356.)

Dunlap, W.P., Dietz, J., and Cortina, J.M., 1997. The spurious correlation of ratios that have common variables: a Monte Carlo examination of Pearson's formula. The Journal of General Psychology, 124, 182-193.

Hildred, G.V., Ratcliffe, K.T., Wright, A.M., Zaitlin, B.A., and Wray, D.S., 2010. Chemostratigraphic applications to low-accommodation fluvial incised-valley settings: an example from the lower Mannville Formation of Alberta, Canada. Journal of Sedimentary Research, 80, 1032-1045.

Jackson, D.A., and Somers, K.M., 1991. The spectre of 'spurious' correlations. Oecologia, 86, 147-151.

Murray, R.W., and Leinen, M., 1996. Scavenged excess aluminum and its relationship to bulk titanium in biogenic sediment from the central equatorial Pacific Ocean. Geochimica et Cosmochimica Acta, 60, 3869-3878.

Pearce, T.J., McLean, D., Martin, J.H., Ratcliffe, K., and Wray, D.S., 2010. A whole-rock geochemical approach to the recognition and correlation of "marine bands". In Application of modern stratigraphic techniques: theory and case histories, SEPM Special Publication No. 94, p. 221-238.

Piper, D.Z., 1973. Origin of metaliferous sediments from East Pacific Rise sediments. Earth and Planetary Science Letters, 19, 75-82.

Ratcliffe, K.T., Wright, A.M., Montgomery, P., Palfrey, A., Vonk, A., Vermeulen, J., and Barrett, M., 2010. Application of chemostratigraphy to the Mungaroo Formation, the Gorgon Field, offshore Northwest Australia. APPEA Journal, 50th Anniversary Issue, p. 371-388.

Van der Weijden, C.H., 2002. Pitfalls of normalization of marine geochemical data using a common divisor. Marine Geology, 184, 167-187.