COMMENT

Limnol. Oceanogr., 42(5), 1997, 1006–1007

Ratios, regression statistics, and "spurious" correlations

Graphical representation and curve fitting are often needed to interpret the complex relationships between measured quantities in aquatic systems. Although graphical and statistical computer packages have made it simpler to perform analyses, they cannot overcome the statistical problems of certain data manipulations.

For example, one means of expressing relationships between two variables, A and B, is to plot the ratio AB^{-1} against B. Inspection of recent issues of ***Limnology and Oceanography*** reveals the use of this plot in a range of applications, where it has been used to demonstrate that the percentage particulate organic carbon in Amazonian rivers decreases as total suspended solids increase (Hedges et al. 1994), that epilimnetic total N : total P decreases as total P increases in lakes worldwide (Kopácek et al. 1995), that the germanium : silicate ratio declines in glacial meltwater streams as silicate increases (Chillrud et al. 1994), that mass-specific growth rates of cladocera decline with increases in mass (Anderson and Benke 1994), and that weight-specific ammonium release declines with increasing weight in freshwater zooplankton (Haga et al. 1995). No criticism of these studies is implied (in fact, some have been exceptionally careful to avoid the pitfalls that follow); the purpose in citing them is to illustrate the widespread use of this type of plot.

B is involved in both X and Y axes on the plot. Intuitively, if A and B were not at all correlated, we might expect that as B became larger, there would be a tendency for AB^{-1} to become smaller. In fact, this tendency is far stronger than is generally appreciated. To illustrate, a common pseudo-random number generator (a function in Microsoft Excel 5.0) was used to generate two sets of 500 numbers, evenly distributed over the interval 0–10. Linear correlation of A and B shows no significant relationship (Fig. 1a; $r = 0.020$). If, however, AB^{-1} vs. B is plotted, an exponential decline in AB^{-1} with increasing B is seen (Fig. 1b), which appears linear on a log–log plot (Fig. 1c). If these curves are fit to the model

$$AB^{-1} = aB^{b},$$

an apparently satisfying fit is obtained (Fig. 1b, c). The model can be fit with a nonlinear algorithm (Marquardt-Levenberg; SigmaPlot for Windows 2.0), giving $a = 5.15$ and $b = -0.990$, or a linear regression on the log-transformed data can be performed, resulting in similar parameter estimates and $r = 0.67$. The nonlinear fitting package also conveniently provides asymptotic standard errors for both a and b.

The reason for this surprising result is relatively simple. In making the relationship

$$A = aB^{b}$$

B-specific, both sides are divided by B ($=B^{1}$):

$$\frac{A}{B^{1}} = \frac{aB^{b}}{B^{1}}$$

or

$$AB^{-1} = aB^{b-1}.$$

Thus, if there were no relationship initially in a log-transformed linear relationship ($b = 0$), the expectation in a B-specific relationship is an exponent (or slope in the log-transformed case) of -1, exactly what is found.

If the analysis is performed for other distributions (e.g. a normal distribution), other ranges of A and B (e.g. 1,000–1,010), or different ranges for A and B (e.g. 0–1 vs. 1–1,000), the results for b and r remain virtually identical, though a does change.

There are three issues worthy of comment. First, there is obviously a statistical problem with an AB^{-1} vs. B plot. This problem has been pointed out in specific areas of biology, e.g. the case of weight-specific biological scaling (*see* Atchley et al. 1976, Prothero 1986), but it appears it is less widely appreciated in the aquatic sciences in general. The statistical difficulty does *not* invalidate the use of this plot to illustrate a relationship, and in fact for some analyses, it may be argued that the resulting "spurious correlation" is not completely invalid (*see* Prairie and Bird 1989). There is little rationale for using such a plot for statistical analysis, however, because the problems are overcome very simply by performing analyses directly on the variables A and B; the form of the relationship can then be converted mathematically to apply to the B-specific relationship.

A second issue concerns the use of ratios. Ratios are problematic because even when experiments are carefully performed so that the variance of the raw variables A and B conform to statistical requirements (normal distribution, equal variance), the characteristics of the variance about AB^{-1} are unpredictable and even counterintuitive (Atchley et al. 1976), and statistical assumptions about ratios are rarely tested. In general, the use of ratios to scale experimental data is fraught with problems (Atchley et al. 1976, Packard and Boardman 1988).

Finally, whether linear or nonlinear, regressions fitted by least-squares methods require assumptions that are often violated in aquatic science data sets. Some assumptions about distribution can be overcome by using distribution-free (sometimes confusingly referred to as "nonparametric") regression techniques (Cornish-Bowden et al. 1978, Máritz 1981), and the critical assumption that B is estimated without

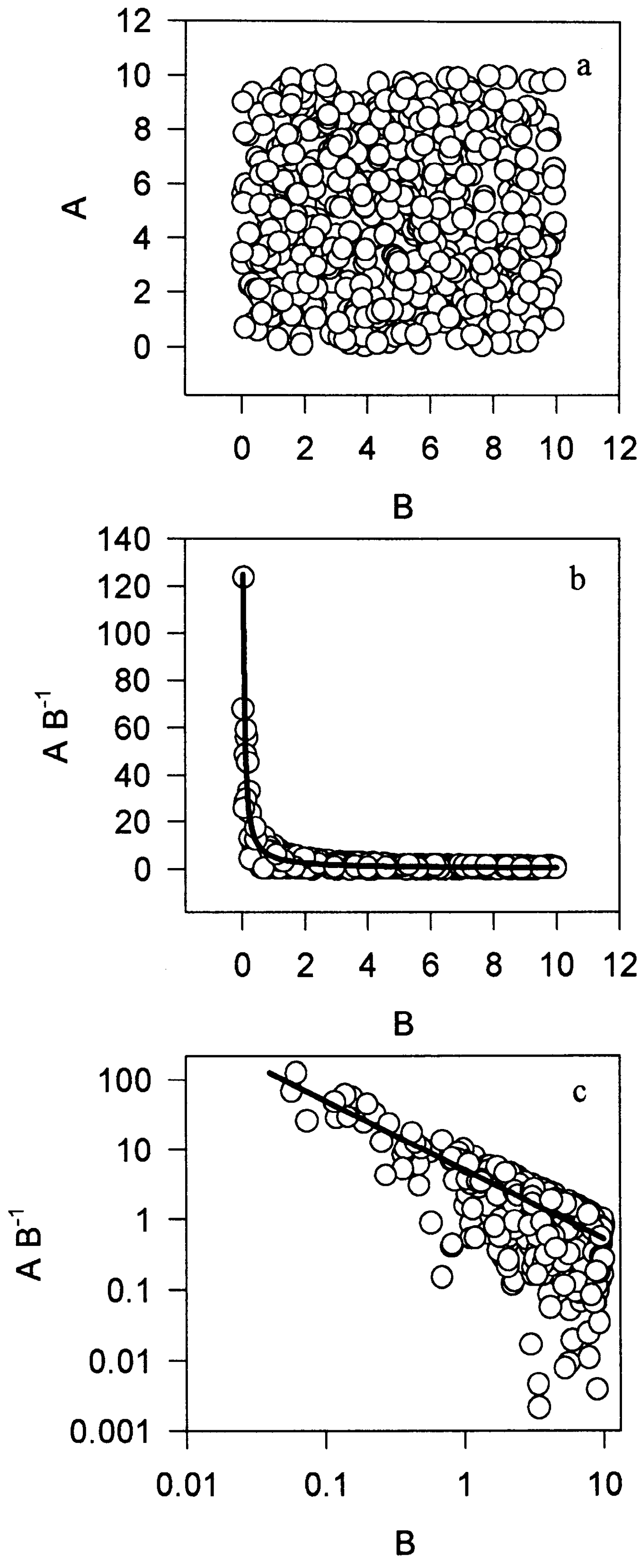

Fig. 1. Plots of the relationship between two independent sets of 500 random numbers (A and B) evenly distributed across the interval 0–10. [a.] Simple plot of A vs. B. [b.] Plot of AB^{-1} vs. B on linear axes and fitted to the nonlinear relationship $AB^{-1} = aB^b$, where $a = 5.15$ and $b = -0.99$. [c.] As in panel b, but plotted on log–log axes. Curve fitting and random number generation are described in the text.

error can be dealt with by geometric methods (Laws and Archie 1981). These techniques are not, however, part of most analysis packages. For nonlinear regression, there is no exact theory for evaluating confidence intervals. Standard errors provided by most plotting packages are asymptotic and, because they tend to underestimate true error, they are usually unsuitable for statistical comparisons; joint confidence intervals should be used in such cases (Johnson 1992). Finally, for relationships that are linear on log–log axes, the coefficient of determination (r^2) can be very misleading; providing the percentage standard error of estimate is one alternative (*see* Smith 1984).

John A. Berges

School of Biology and Biochemistry
The Queen's University of Belfast
Belfast BT9 7BL, Northern Ireland, UK

References

Anderson, D. H., and A. C. Benke. 1994. Growth and reproduction of the cladoceran *Ceriodaphnia dubia* from a forested floodplain swamp. Limnol. Oceanogr. **39:** 1517–1527.

Atchley, W. R., C. T. Gaskins, and D. Anderson. 1976. Statistical properties of ratios. I. Empirical results. Syst. Zool. **25:** 137–148.

Chillrud, S. N., F. L. Pedrozo, P. F. Temporetti, H. F. Planes, and P. N. Froelich. 1994. Chemical weathering of phosphate and germanium in glacial meltwater streams: Effects of subglacial pyrite oxidation. Limnol. Oceanogr. **39:** 1130–1140.

Cornish-Bowden, A., W. R. Porter, and W. F. Trager. 1978. Evaluation of distribution-free confidence limits for enzyme kinetic parameters. J. Theor. Biol. **74:** 163–175.

Haga, H., T. Nagata, and M. Sakamoto. 1995. Size-fractionated NH_4^+ regeneration in the pelagic environments of two mesotrophic lakes. Limnol. Oceanogr. **40:** 1091–1099.

Hedges, J. I., and others. 1994. Origins and processing of organic matter in the Amazon River as indicated by carbohydrates and amino acids. Limnol. Oceanogr. **39:** 743–761.

Johnson, M. L. 1992. Review: Why, when, and how biochemists should use least squares. Anal. Biochem. **206:** 215–225.

Kopácek, J., L. Procházková, E. Stuchlík, and P. Blazka. 1995. The nitrogen-phosphorus relationship in mountain lakes: Influence of atmospheric input, watershed, and pH. Limnol. Oceanogr. **40:** 930–937.

Laws, E. A., and J. W. Archie. 1981. Appropriate use of regression analysis in marine biology. Mar. Biol. **65:** 13–16.

Maritz, J. S. 1981. Distribution-free statistical methods. Chapman and Hall.

Packard, G. C., and T. J. Boardman. 1988. The misuse of ratios, indices and percentages in ecophysiological research. Physiol. Zool. **61:** 1–9.

Prairie, Y. T., and D. F. Bird. 1989. Some misconceptions about the spurious correlation problem in the ecological literature. Oecologia **81:** 285–288.

Prothero, J. 1986. Methodological aspects of scaling in biology. J. Theor. Biol. **118:** 259–286.

Smith, R. J. 1984. Allometric scaling in comparative biology: Problems of concept and methods. Am. J. Physiol. **246:** R152–R160.

←