Science of the Total Environment 496 (2014) vi–xii

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Review

Mercury bioaccumulation factors and spurious correlations

Curtis D. Pollman [a], Donald M. Axelrad [b,*]

[a] Aqua Lux Lucis, Inc., 8411 NW 55th Place, Gainesville, FL, USA
[b] Florida A&M University, Institute of Public Health, 1515 South Martin Luther King Blvd, Tallahassee, FL 32307, USA

HIGHLIGHTS

- Spurious correlation is surprisingly common in studies on mercury bioaccumulation.
- Correlations using bioaccumulation factors (BAF), a ratio, have inherent risks.
- BAFs are the ratio of contaminant levels in biota to those in the water column.
- Misleading inferences arise when correlating BAF with its component denominator.
- Direct analysis of component variables is less ambiguous than BAF correlations.

ARTICLE INFO

Article history:
Received 26 June 2014
Received in revised form 8 July 2014
Accepted 8 July 2014
Available online 1 August 2014

Editor: J.P. Bennett

Keywords:
Methylmercury
Aquatic ecosystems
Gambusia
Everglades
Synthetic data

ABSTRACT

While bioaccumulation factors (BAF) – the ratio of biota contaminant concentrations (C_{biota}) to aqueous phase contaminant concentrations (C_w) – are useful in evaluating the accumulation of mercury (Hg) and other contaminants for various trophic levels in aquatic ecosystems, reduction of the underlying relationship between C_{biota} and C_w to a single ratio (BAF) has inherent risks, including spurious correlation. Despite a long and rich history of remonstrations in the literature, several very recent publications evaluating Hg-related BAFs have suffered from false conclusions based on spurious correlation, and thus it seems that periodic reminders of the causes and risks of these errors are required. Herein we cite examples and explanations for unsupported conclusions from publications where authors using BAF-C_w relationships fail to recognize the underlying statistical significance (or lack thereof) of direct relationships between C_w and C_{biota}. This fundamental error leads to other problems, including ascribing mechanistic significance (*e.g.*, mechanisms related to biota contaminant uptake) to "inverse" BAF-C_w relationships that reflect nothing more than regressing the log transformed inverse of C_w (*i.e.*, negative log) against itself (*i.e.*, positive log transformed), and using such regression models of BAF-C_w relationships that appear significant for predictive purposes, but are misleading. Spurious correlation arising in the analysis of BAF relationships can potentially appear in more subtle forms as well, including regressing variables such as dissolved organic carbon (DOC) that are correlated with C_w. We conclude that conducting a direct analysis by examining the relationship between C_{biota} and C_w (or C_{biota} and other variables) rather than evaluating a ratio (BAF) is less ambiguous and subject to error, more easily interpreted, and would lead to more supportable conclusions.

Contents

* Corresponding author. Tel.: +1 850 443 4626.
E-mail addresses: cpollman@aqualuxlucis.org (C.D. Pollman), donald.axelrad@famu.edu, daxe@comcast.net (D.M. Axelrad).

http://dx.doi.org/10.1016/j.scitotenv.2014.07.050

1. Introduction

In his book *The Life of Reason*, the philosopher George Santayana (1906) observed that "[t]hose who cannot remember the past are condemned to repeat it" — an aphorism that certainly appears to be true in aquatic ecology for the issue of spurious correlation. Spurious correlation may occur when, in the process of evaluating the relationship between variables, a common variable appears on both sides of the equation. Typically this occurs when the dependent variable comprises a ratio with the common variable, *x*, in the denominator, and the constructed ratio is then plotted as a function of the common variable by itself (Eq. (1)), or with the common variable appearing in the denominator of a second ratio (Eq. (2)):

$$\frac{y}{x} = f(x) \tag{1}$$

$$\frac{y}{x} = f\left(\frac{b}{x}\right). \tag{2}$$

The difficulty is that the resultant relationship, which can appear to have clear statistical significance, is viewed as something meaningful, when in reality the relationship owes its functional and statistical significance to the common variable appearing on both sides of the equation.

This problem was first identified as a concern in the biological sciences by Pearson (1897), and has long been recognized as a problem that requires attention in aquatic ecology (*e.g.*, Kenney, 1982, Berges, 1997, Jackson and Somers, 1991, Kronmal, 1993; Krambeck, 1995; Brett, 2004; Håkanson and Stenström-Khalili, 2009; Stenström-Khalili and Håkanson, 2009). Despite such a rich history of admonitions, it seems that periodic reminders and expositions of this issue are nonetheless prudent and necessary. As we will demonstrate, we contend that spurious correlation is an issue in studies of mercury (Hg) biogeochemical cycling in aquatic ecosystems where bioaccumulation and bioconcentration factors (BAF and BCF, respectively) often are used as tools for evaluating bioaccumulation of mercury and methylmercury (MeHg) in particular.

A BAF can be defined as the ratio of the concentration of a chemical contaminant in the tissue of an aquatic organism to its concentration in water, in situations where both the organism and its food are exposed to the chemical. In contrast, a BCF is the ratio of the concentration of a chemical in the tissue of an aquatic organism to its concentration in water, in situations where the organism is exposed to the chemical through the water only (US EPA, 2003). As such, Hg BAF values are constructed as the ratio of biota Hg concentrations (C_{biota}) and aqueous phase concentrations of Hg (C_w):

$$\mathrm{BAF} = \frac{C_{biota}}{C_w}. \tag{3}$$

BCFs are constructed in an identical manner, but the ratio is typically used for primary producers because bioconcentration is defined to reflect direct uptake of the contaminant from water only, while BAFs reflect trophic transfer (dietary bioaccumulation) as well. Because methylmercury (MeHg) is the chemical species of mercury that characteristically enters the base of aquatic food webs and is involved in trophic transfer (Seixas et al. 2014), Hg BCF/BAFs typically are calculated using C_w concentrations of MeHg, although total Hg concentrations are used as well. Spurious correlation can thus arise when researchers look for relationships between BAF or BCF and C_w.

2. Illustration of spurious correlation using Everglades survey data

We use data on *Gambusia* (mosquitofish) Hg concentrations and water column MeHg concentrations collected by the USEPA between 1993 and 2005 as part of their Regional Environmental Monitoring and Assessment Program (R-EMAP; Scheidt and Kalla, 2007) to illustrate some fundamental features of spurious correlation as it relates Hg BAFs in aquatic ecosystems. Our analysis has two components — first, we construct plots of the relationship between both observed *Gambusia* Hg concentrations and water column MeHg concentrations, and the resultant BAF values as a function of water column MeHg concentrations. The second component duplicates the analyses conducted for the observed data but uses synthesized *Gambusia* Hg and MeHg concentrations generated randomly with lognormal distributions that match the range of the observed distributions. The synthesized data were constructed with lognormal distributions because the observed data more closely approximate lognormal rather than normal distributions.

Results for both the observed and simulated data are shown in Fig. 1. Because the observed data show only a weak correlation between *Gambusia* Hg and water column MeHg concentrations ($r^2 = 0.11$ (Fig. 1a) for log-transformed concentrations), the observed and simulated results closely mimic each other. For example, because the data are lognormally distributed, the plot of observed and synthesized *Gambusia* Hg concentrations as a function of MeHg concentrations both have few observations where high concentrations of *Gambusia* Hg correspond to high concentrations of MeHg. When BAF values are plotted against MeHg concentrations, both the observed and the synthesized distributions assume a characteristic hyperbolic shape that reflects the fact that the curve is essentially a plot of the inverse of MeHg concentrations – weakly modified by *Gambusia* Hg concentrations – against MeHg concentrations (Fig. 1b and c). This is illustrated in Fig. 2, which superimposes the line representing the inverse of MeHg concentrations scaled by a factor ϕ comprising the geometric mean *Gambusia* Hg concentration (105 ng/g) multiplied by 1000 to make the units equivalent to the BAF values:

$$\varphi = \frac{105}{C_w} \cdot 1000. \tag{4}$$

The asymptotic character of the curve reflects the degree to which the two variables comprising the BAF ratio are correlated. If they are highly correlated, then the resultant BAF values will vary with some error around a value that equals the slope of that relationship and that is invariant with changes in MeHg concentrations (assuming an intercept of zero). This is shown in the following two equations:

$$C_{biota} = \beta \cdot C_w + \varepsilon \tag{5}$$

$$\mathrm{BAF} = \beta' + \varepsilon' \tag{6}$$

where β is the slope of the relationship between C_{biota} and C_w, ε is the residual error in the model relationship, β' is $\beta * 1000$ to ensure dimensional consistency, and ε' is ε divided by $C_w * 1000$. When plotting BAF as a function of C_w, this "scaling" of the residual error by C_w results in greatly magnifying the effect of measurement error on calculated BAF values at the lower end of the distribution of aqueous phase contaminant concentrations. As aqueous phase concentrations increase, the relative magnitude of ε' vis-à-vis β is reduced, and BAF values approach and then approximate β. This is shown in Fig. 3, which assumes that